JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

Foundations of Quaternion Quantum Mechanics

DAVID FINKELSTEIN
Yeshiva University, New York, New York

JOSEF M. JAUCH
University of Geneva and CERN, Geneva, Switzerland

SAMUEL SCHIMINOVICH
Yeshiva University, New York, New York

AND

DAVID SPEISER
University of Geneva, Geneva, Switzerland

(Received October 2, 1961)

A new kind of quantum mechanics using inner products, matrix elements, and coefficients assuming values that are quaternionic (and thus noncommutative) instead of complex is developed. This is the most general kind of quantum mechanics possessing the same kind of calculus of assertions as conventional quantum mechanics. The role played by the new imaginaries is studied. The principal conceptual difficulty concerns the theory of composite systems where the ordinary tensor product fails due to noncommutativity. It is shown that the natural resolution of this difficulty introduces new degrees of freedom similar to isospin and hypercharge. The problem of the Schrödinger equation, "which i should appear?" is studied and a generalization of Stone's theorem is used to resolve this problem.

1. WHY QUATERNION QUANTUM MECHANICS?

IN conventional quantum mechanics propositions (the result of measurements) are represented as subspaces in an infinite-dimensional Hilbert space $\mathfrak{H}$. In particular the pure states are the one-dimensional subspaces (rays) of $\mathfrak{H}$. The relation between measurements of different kinds are expressed in the lattice structure of these subspaces. One of the most important features in this structure is the absence of the distributive law which characterizes the propositional calculus of classical systems. One might say, therefore, that the lattice structure of subspaces incorporates an essential ingredient of quantum systems, viz., that measurements of different kinds may interfere leading to the well-known uncertainty relations and complementary properties.

The representation of the propositional calculus of atomic systems as a lattice structure was given as long ago as 1936 by Birkhoff and von Neumann.[1] There it is shown that a propositional calculus exists that we can call *general quantum mechanics* (as distinguished from complex quantum mechanics) in as much as no number system or vector space at all is assumed in its formulation.[2] The relevance for the present work is found in

[1] G. Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936). The first suggestion of quaternion quantum mechanics appears in a footnote of this paper. C. N. Yang has also pointed out the interest of this possibility [*Proceedings of the Seventh Rochester Conference on High-Energy Nuclear Physics 1957* (Interscience Publishers, Inc., New York, 1957), p. IX-26].

[2] We can present the propositional calculus of general quantum mechanics as follows, if we consider finite-dimensional Hilbert spaces only, thus excluding systems with continuous variables except as limiting cases. The elements $A, B, C, \cdots$ (which may indifferently be regarded as representative ensembles, propositions about a physical system, or "operational" rules for testing the truth of statements) are subject to the basic operation of *negation* $A \rightarrow \sim A$ and the basic relation of *implication* $A \subset B$. In addition, unlike the classical propositional calculus, the propositions make up (are the points of) a topological space. The axioms are: 1. *The axioms for a complemented lattice.* Implication is reflexive, transitive, and antisymmetric (reversible only for equals). Any A, B, possess both a g.l.b. $A \cap B$ (A and B) and a l.u.b. $A \cup B$ (A or B) with respect to implication. There exists an over-all g.l.b. O and l.u.b. I. Negation is an involutory[3] anti-automorphism of the lattice. 2. *Axioms of cardinality.* To each proposition A may be associated a non-negative integer $|A|$ such that if $A \subset B$ then $|A| < |B|$; if $A \subset B$ and $|A| = |B|$ then $A = B$; $|O| = 0$; and

the following remarkable result:

It is always possible to represent the pure states of a system of "general quantum mechanics" by rays in a vector space in a one-to-one manner, and for this it is necessary and sufficient to employ orthogonal vector spaces (Hilbert spaces) over the following numl er systems:

$\mathfrak{R}$, the real numbers,
$\mathfrak{C}$, the complex numbers, and
$\mathfrak{Q}$, the quaternions.[3]

Moreover it is always possible to represent these pure states by rays in a vector space over $\mathfrak{Q}$ (not every ray may be needed for this), but not so over $\mathfrak{R}$ or $\mathfrak{C}$.

This result suggests two things. First, inasmuch as the propositional calculus reflects empirical information, some of this information appears in the number system which is used for the construction of the Hilbert space. Secondly, it is not necessary to go beyond the three possibilities $\mathfrak{R}$, $\mathfrak{C}$, and $\mathfrak{Q}$ for the representation of general quantum mechanics.

We can thus formulate the following precise problem: Which of the three possibilities for the representation of general quantum mechanics is the one most suitable for the description of the actual physical world?

A preliminary analysis of this problem might start with the question: Why has conventional quantum mechanics been developed with complex numbers instead of reals? The relation between complex and real quantum mechanics has been extensively investigated by Stueckelberg and collaborators and is now completely understood. Briefly stated the situation is as follows: Complex quantum mechanics is completely equivalent to real quantum mechanics plus a superselection rule: All observables in real quantum mechanics must commute with a fixed linear operator J, which is antisymmetrical ($J^T=-J$) and satisfies $J^2=-I$. The operator J is intimately related to the symmetry of time reversal. Indeed, the time-reversal transformations are precisely those symmetry transformations which anticommute with J. This result shows that there exists a connection between supersymmetries and the choice of the number field.

While thus the difference between complex and real quantum mechanics is relatively simple, quaternion quantum mechanics has many new features which make it a much richer theory. It is perhaps surprising that such a promising possibility has not yet been more fully developed. There appear to be reasons for this. For example, the problem of how to write a Schrödinger equation is not a trivial one in $\mathfrak{Q}$ quantum mechanics because of the appearance of a square root of minus one in the ordinary Schrödinger equation; and the description of interacting systems by a direct product is made difficult by the noncommutativity of $\mathfrak{Q}$-valued wave functions.

In this paper we shall present the general features of a quaternion quantum mechanics. In a subsequent paper we shall show how these features can be utilized

---

$|{\sim}A|=|I|-|A|$. We assume, without loss of generality, that the least of the positive values assumed by this integer is normalized to be 1; otherwise any positive integral multiple of A would also satisfy these requirements. 3. *Axiom of superposition.* If $|A|=|B|=1$ then there exists a C with $|C|=1$ such that $A\cup B=B\cup C=C\cup A$. 4. *Axiom of continuity.* $|A|$ is a continuous function of A.

Evidently Axioms 1 and 2 are valid for the propositional calculus of a classical system with a finite number of states. Axiom 3 is the very essence of quantum logic; the C whose existence it asserts is a superposition of A and B in the quantum sense. Were the classical distributive law of logic $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$ adjoined to Axiom 1, Axiom 3 would be inconsistent. Even without continuity (Axiom 4) it would follow from Axioms 1–3 that[1] the propositions correspond one-to-one to substances of some vector space over a skew field. Since *every* subspace of the vector space is utilized in this one-to-one realization, there is no room for superselection principles in what we have called general quantum mechanics; but if Axiom 3 is simply dropped we find a "supersum" (direct sum with superselection rules between addends) of systems for each of which Axiom 3 is satisfied. Thus it is not necessary to go beyond the quaternions until Axioms 1, 2, or 4 are weakened.

We have consigned this matter to inferior print and omitted much mathematical beauty; it concerns mostly how one arrives at $\mathfrak{Q}$ quantum mechanics, and in the final analysis it is more important to know where a theoretical path leads than how one fell upon it.

[3] We give here the algebra of quaternions. Every quaternion can be written in the form

$$q=q_0+q_1i_1+q_2i_2+q_3i_3,$$

where the four coefficients q_k are real. The multiplication of quaternions is associative, distributive, and obeys

$$i_k^2=-1 \quad k=1,2,3;$$
$$i_1i_2i_3=-1.$$

In the last equation, the anticyclic order of factors might have been taken. Every quaternion q possesses an inverse q^{-1}.

In Hamilton's notation a quaternion is regarded as the sum of a "scalar" (*real*) part and a "vector" (*imaginary*) part: $q=q_0+\mathbf{q}\cdot\mathbf{i}$. The quaternions that commute with all other quaternions are just the reals. The quaternions that commute with a given nonreal quaternion form a subset isomorphic to the complex numbers. There exists an operation $q\rightarrow q^Q$ on the quaternions that is *involutory* ($q^{QQ}=q$), Hermitian definite (q^Qq is real, and vanishes only when $q=0$), and *anti-automorphic* ($p^Qq^Q=(qp)^Q$), and it is called the *quaternion conjugate* ($\mathfrak{Q}$ conjugate): $i_k^Q=-i_k$. On the other hand, the *automorphisms* of the quaternions are all of the form $q\rightarrow aqa^{-1}$. (The quaternion a associated with a particular automorphism is not uniquely defined by this equation; by requiring that the *norm* of a, meaning a^Qa, be unity, the ambiguity is reduced to an extremely important matter of sign.)

It is sometimes convenient to represent quaternions by pairs of complex numbers (c^0,c^1) according to

$$q=c^0+i_2c^1,$$

where c^0, c^1 commute with i_3, and are therefore essentially complex numbers. Treating these pairs as vectors in a two-dimensional complex vector space C^2, we find that every linear transformation of $\mathfrak{Q}$ is represented by a linear transformation of C^2, that is by a 2×2 complex matrix. In particular the *left multiplication* $q\rightarrow aq$, by a fixed quaternion a, is represented by a matrix a_{ij}, the *symplectic* representation of a. The symplectic representations of left multiplication by i_1, i_2, i_3 are just the Pauli spin operators (times i). But the symplectic representation of right multiplications by quaternions are sums of linear and antilinear operators.

Computation yields that $q^Q=c^{0*}-i_2c^1$ where Q denotes the quaternion, and the star the ordinary complex, conjugate. The "scalar product" of two quaternions p,q, $(p,q)=p^Qq$, then becomes (with $p=b^0+i_2b^1$) $p^Qq=(b^{0*}c^0+b^{1*}c^1)+i_2(b^0c^1-b^1c^0)$. We separated the quaternion with respect to i_2 and identified i_3 with the complex i. But, of course, we could have used any pair of anticommuting units as well.

for the description of the multiplicities of the strongly interacting particles, for a fundamental theory of electromagnetism, and for a possible unification of the theory of electromagnetic and Fermi interactions.

2. GENERAL QUANTUM MECHANICS

By a suitable choice of development it is possible to present three kinds of quantum mechanics to be called $\mathfrak{R}$, $\mathfrak{C}$, and $\mathfrak{Q}$ all at once. Therefore let $\mathfrak{F}$ be one of these number systems (fields) in which there are defined the notions of continuity, addition, multiplication, inverse, and conjugate (*). We term the ensuing schema $\mathfrak{F}$ quantum mechanics. We need the concept of a Hilbert space over $\mathfrak{F}$, $\mathfrak{H}(\mathfrak{F})$ and may define it by taking as fundamental the algebraic operations of vector addition, multiplication by $\mathfrak{F}$ numbers, and conjugate (†). The vector addition requires no discussion. Because $\mathfrak{F}$ may be noncommutative, multiplication of a vector Ψ or $|\cdot\rangle$ by a number a must be set up carefully.[1] We write Ψa, or $|\cdot\rangle a$, and require the associative law

$$(\Psi a)b=\Psi(ab) \tag{1}$$

and the distributive law. The conjugate $\Psi\dagger$ or $\langle\cdot|$ of a vector Ψ or $|\cdot\rangle$ is in the dual Hilbert space $\mathfrak{H}\dagger(\mathfrak{F})$, which means that it is a linear function on $\mathfrak{H}(\mathfrak{F})$:

$$\Psi\dagger(\Phi a)=(\Psi\dagger\Phi)a, \quad \Psi\dagger(\Phi+\Phi')=\Psi\dagger\Phi+\Psi\dagger\Phi'.$$

The conjugate operation is to be Hermitian, $\Psi\dagger\Phi=(\Phi\dagger\Psi)^*$; antilinear, $(\Psi a)\dagger=a^*\Psi\dagger$, $(\Psi+\Phi)\dagger=\Psi\dagger+\Phi\dagger$; and definite,

$$\Psi\dagger\Psi=0 \quad \text{if and only if } \Psi=0.$$

Thus $\Psi\dagger\Phi$ has the properties of a scalar product.[4] We use the distance $|\Psi-\Phi|=[(\Psi-\Phi)\dagger(\Psi-\Phi)]^{\frac{1}{2}}$ to define the topology of $\mathfrak{H}(\mathfrak{F})$.

A linear manifold of $\mathfrak{H}(\mathfrak{F})$ is a set of vectors closed under vector addition and multiplication by $\mathfrak{F}$ numbers. A subspace is a closed linear manifold. The subspaces play a fundamental role in the interpretation. We assume that they are in one-to-one correspondence with the propositions about the physical system[2] in the way that subsets of classical phase space are in one-to-one correspondence with propositions about the classical physical system. Momentarily we are excluding the important cases of quantum mechanics with superselection principles by this assumption. The notion of quantum mechanics defined in this manner, using such "unobservable" concepts as the Hilbert space vectors and their scalar products, coincides in the finite-dimensional case with the general quantum mechanics defined in terms of "observable" non-numerical concepts such as implication and negation.[1,2]

3. THE SCHRÖDINGER EQUATION

How do we link this structure to reality? Basic physical quantities such as energy and momentum are recognized by their relation to such symmetry properties of the physical system as time- and space-translational symmetry. We therefore consider groups of transformations acting on the system.[5]

For this end define a mapping T on $\mathfrak{H}(\mathfrak{F})$ into itself that has the property

$$T(\Psi a+\Phi b)=(T\Psi)a'+(T\Phi)b' \tag{1}$$

to be *colinear*. Here a' is to be obtained from a, and b' from b, by an automorphism of $\mathfrak{F}$ associated with T, independent of Ψ and Φ, and called the automorphism belonging to T; for $\mathfrak{F}=\mathfrak{Q}$,

$$a'=qaq^{-1},$$
$$b'=qbq^{-1},$$

q independent of Φ, Ψ. The name *colinear* is suggested by the name used for the closely related concept of a *collineation* in projective geometry. Evidently colinear transformations (and indeed only such mappings of vectors) carry subspaces into subspaces. A colinear transformation for which the associated automorphism of $\mathfrak{F}$ is the identity is termined *linear*.

The elements of groups of transformations in $\mathfrak{F}QM$ are going to be *co-unitary* transformations, which are defined to be colinear transformations U on $\mathfrak{H}(\mathfrak{F})$ enjoying the additional property

$$(U\Phi)\dagger(U\Psi)=(\Phi\dagger\Psi)'=q(\Phi\dagger\Psi)q^{-1}, \tag{2}$$

where the prime indicates that the automorphism of $\mathfrak{F}$ belonging to the colinear transformation U is applied to the $\mathfrak{F}$ number $\Phi\dagger$, Ψ. If linear, a co-unitary transformation is called *unitary*.

Let us now consider the passage of time. We shall suppose it is represented by a one-parameter group of unitary transformations U_t on $\mathfrak{H}(\mathfrak{F})$. Requiring U_t to be unitary, amounts to nothing more or less than requiring the logical relations between propositions concerning the system to be independent of the time origin.[6] For

[4] Usually it is the scalar product that is taken as fundamental, but except with the Dirac notation this leads to a doubling of symbols, and to an ambiguity about which factor is the linear one, which the antilinear. We shall find the Dirac notation extremely convenient for $\mathfrak{Q}$ quantum mechanics, since it manages automatically certain rules of order that are not important in $\mathfrak{C}$ quantum mechanics. Equivalent definitions of $\mathfrak{H}(\mathfrak{Q})$ are given by von Neumann and Birkhoff, (reference 3); E. H. Moore, *General Analysis* (American Philosophical Society, Philadelphia, Pennsylvania, 1935); and O. Teichmuller, Z. Math. 174, 73 (1935).

[5] Again two approaches present themselves, the "synthetic" and the "analytic"; just as the definition of "general quantum mechanics" in footnote 2 is the "synthetic" version of the "analytic" one given in the text of this section for $\mathfrak{F}QM$. Again we relegate the "synthetic" formulation to a footnote: An automorphism U on the propositional calculus of $\mathfrak{F}QM$ is a mapping of *propositions* to *propositions*, $U:A\rightarrow A'=A^U$, that possesses an inverse and preserves the operation of negation and the relation of implication. It is then a theorem[3] that every such mapping is effected by a mapping of vectors of the kind to be called co-unitary above. Likewise any mapping that preserves implication is represented by a colinear vector transformation.

[6] Why unitary and not simply co-unitary? Since after all the essential requirements from the point of view of the propositional calculus are that implication and negation of propositions (linear dependence and orthogonality of vectors) be preserved by the passage of time, and this is a property of the co-unitary operators.

example, suppose Ψ is a vector representing the state prepared by some definite process performed on the system by an external apparatus. For the process to be a definite one, we mean in particular that the times of operation of the parts of the apparatus are specified. Then $U_t\Psi$ represents the state that would be produced by the same apparatus operating at times retarded by the amount t.

We now wish to make the passage to a Schrödinger equation. In $\mathcal{C}$ quantum mechanics Stone's theorem is the bridge from the unitary group U_t to the Schrödinger equation for $\Psi(t)=U_{-t}\Psi(0)$,

$$d\Psi/dt=-iH\Psi. \tag{3}$$

For $\mathcal{R}$ and $\mathcal{Q}$ quantum mechanics, this equation makes no sense; quite obviously so for $\mathcal{R}$ quantum mechanics where there is no symbol i at all, but equally so for $\mathcal{Q}$ quantum mechanics as well, since the various i's that appear there are not linear operators at all and their appearance on the left of a Ψ is undefined. Therefore we require the following modification of Stone's theorem:

Theorem: Every one-parameter unitary group U_t on $\mathcal{H}(\mathcal{F})$ is generated by an equation of the form

$$d\Psi/dt=-\eta H\Psi, \tag{4}$$

where H is Hermitian non-negative, η is anti-Hermitian unitary, and both commute strongly with U_t. H is unique, and η is unique except on the null space of H within which it acts as an arbitrary anti-Hermitian unitary operator.

Proof: See Appendix C.

Since this decomposition results in a non-negative generator H, it is an appropriate one for the time translation but not for the other one-parameter groups, where reflection may be a physically possible process. This theorem gives us a way to construct the energy from the time translation group. How shall we construct the momentum from the space translations, etc.? Let us write A for the infinitesimal anti-Hermitian generator of time translation, so that

$$d\Psi/dt=A\Psi.$$

Then the absolute value of A gives the non-negative energy operator of the theory and the phase of A gives a unitary square root of -1:

$$H=|A|\equiv(A\dagger A)^{\frac{1}{2}},$$
$$\eta=|A|A^{-1}.$$

For ordinary quantum mechanics $\mathcal{F}=\mathcal{C}$ and $\eta=i$. The theorem shows exactly in what directions this case can be generalized. For instance we can assume that all the observables commute with the operator η. If we do this, we obtain the following possibilities:

$\mathcal{F}=\mathcal{R}$: This case has been studied by Stueckelberg and it was shown by him that it is equivalent to $\mathcal{F}=\mathcal{C}$, $\eta=i$.

$\mathcal{F}=\mathcal{C}$: If $\eta=i$ we obtain ordinary quantum mechanics without superselection rules. If $\eta\neq i$, we obtain a more general case with superselection rules.

$\mathcal{F}=\mathcal{Q}$: This is the case we are especially concerned with in this paper. If we drop the assumption that all the observables commute with η we obtain more general cases about which little is known.

Only the case $\mathcal{F}=\mathcal{R}$, which is equivalent to $\mathcal{F}=\mathcal{C}$ and antilinear observables, has been investigated in some detail by Stueckelberg, Guenin, and Piron (unpublished).

At some points in the paper we shall thus make the following assumption: The operator η associated with the time-translation operator U_t by the above theorem is a superselection operator: All the observables commute with η. But it should be pointed out that in that case we are not very far from complex quantum mechanics. We state without proof the following reduction theorem.

Theorem: The propositions of $\mathcal{Q}$ quantum mechanics that commute with a fixed anti-Hermitian unitary operator are isomorphic (with respect to the logical operations of intersection, span, and orthocomplement) to the propositions of $\mathcal{C}$ quantum mechanics.

Of course there is more to quantum mechanics than its propositional calculus. Even if the η superselection principle is imposed, eliminating the new kinds of complementarity peculiar to $\mathcal{Q}$ quantum mechanics, there will still remain new dynamical variables and symmetries.

It will be observed that the points at which the operator η appears in the formulation of each particular quantum theory and the points at which Planck's constant enters are the same. The operator η can best be thought of as taking the place of the combination $i/\hbar$ in ordinary quantum mechanics. Therefore the

---

The reason the unitary operators are sufficient varies slightly for the three cases $\mathcal{F}=\mathcal{R}$, $\mathcal{C}$, $\mathcal{Q}$:

For $\mathcal{F}=\mathcal{R}$, the only automorphism of $\mathcal{F}$ is the identity $a\rightarrow a$; all colinear operators are linear, and all co-unitary operators are unitary.

For $\mathcal{F}=\mathcal{C}$, the automorphisms of $\mathcal{F}$ are the identity I and the complex conjugate $C:a\rightarrow a^*$; all colinear operators are either linear, or if not, are called *antilinear*, the two classes being disconnected. Since a one-parameter group U_t is connected and $U_0=1$ is linear, all U_t are necessarily linear.

For $\mathcal{F}=\mathcal{Q}$, the automorphisms of $\mathcal{F}$ are the conjugations $a\rightarrow a^q=qaq^{-1}$; any colinear operator T can be expressed in terms of an associated (nonunique) linear operator L and a quaternion q according to

$$T\Psi=L\Psi q. \tag{*}$$

Now we see that the linear operators (q real) are continuously connected to the other colinear operators (q not real). Thus the continuity argument does not work here. On the other hand the colinear T and the associated linear operator L of (*) define the same correspondence of propositions to propositions (subspaces to subspaces). Therefore, for $\mathcal{Q}$ quantum mechanics every such correspondence, being representable by a colinear operator, is representable by a linear operator. By choosing the q in (*) to be of unit norm, it is readily seen, L is determined up to sign and is unitary if T is co-unitary. We thus obtain a unitary function of time U_t obeying

$$U_tU_{t'}=\pm U_{t+t'}.$$

By continuity, it is always possible to redefine U_t so that the upper sign is chosen.

introduction of operators that fail to commute with η, as when η is a quaternion, can be regarded as taking Planck's "constant" to be a dynamical variable of the theory. It will be interesting in particular to examine the consequences of quantum fluctuations in the quantum of action. The superselection principle on the other hand makes η a more classical variable by excluding interference between its different values. In principle this makes it consistent to "freeze" the value of η and therefore to suppress these new possibilities.

4. COMPOSITE SYSTEMS

The other extension of the Birkhoff-von Neumann $\mathfrak{Q}$ quantum mechanics that is needed concerns the description of composite systems. In $\mathfrak{C}$ quantum mechanics we are accustomed to multiplying the wave functions of subsystems to get wave functions of composite systems. This multiplication is called the *tensor product*. Even in the quantum theory of fields, where other techniques are used to describe a system of several particles of the same kind, this process of multiplying wave functions is needed to set up a theory for the interaction of particles of different kinds (quanta of different fields).

This multiplication procedure is acceptable in $\mathfrak{C}$ quantum mechanics but not in $\mathfrak{Q}$ quantum mechanics. Suppose $|1\rangle$ and $|2\rangle$ are vectors describing states of distinct systems 1 and 2. Introducing bases in the corresponding Hilbert spaces $\mathfrak{K}_1$ and $\mathfrak{K}_2$, we are led to wave functions:

$$\psi(x_1)=\langle x_1|1\rangle, \quad \psi(x_2)=\langle x_2|2\rangle, \tag{1}$$

where x_1, x_2 enumerate the basis vectors for $\mathfrak{K}_1$ and $\mathfrak{K}_2$, respectively. In order to describe the state of the composite system 1×2, we introduce a tensor product space $\mathfrak{K}_1\times\mathfrak{K}_2$ by giving it a symbolic basis $\{|x_1,x_2\rangle\}$. Then the composite state is supposed to be represented by the wave function

$$\psi(x_1,x_2)=\psi(x_1)\psi(x_2).$$

In order for this coordinate-dependent method to be acceptable, it must be shown to be essentially coordinate independent. The intrinsic relations (intersection, span, orthogonality) between statements (subspaces) of the product space must depend only upon the intrinsic relations between the factors of which they are made.

This coordinate independence is proven when one shows that a change of basis in either of the factor spaces (i.e., a co-unitary transformation in $\mathfrak{K}_1$ or $\mathfrak{K}_2$) results merely in a change of basis in the product space (i.e., a co-unitary transformation in $\mathfrak{K}_1\times\mathfrak{K}_2$). The important formulas for this are the following:

$$\begin{aligned}\psi'(y_1,x_2)&=\langle y_1|1\rangle\langle x_2|2\rangle=\langle y_1|x_1\rangle(\langle x_1|1\rangle\langle x_2|2\rangle)\\&=\langle y_1|x_1\rangle\psi(x_1,x_2),\\ \psi'(x_1,y_2)&=\langle x_1|1\rangle\langle y_2|2\rangle=\langle y_2|x_2\rangle(\langle x_1|1\rangle\langle x_2|2\rangle)\\&=\langle y_2|x_2\rangle\psi(x_1,x_2).\end{aligned} \tag{2}$$

The symbol $\langle x_1|y_1\rangle$ represents the matrix elements between two bases $\{|x_1\rangle\}$ and $\{|y_1\rangle\}$ of $\mathfrak{K}_1$. A summation convention is used. Obviously these formulas are valid for $\mathfrak{R}$ and $\mathfrak{C}$ Hilbert spaces, where the matrix elements, being $\mathfrak{R}$ or $\mathfrak{C}$ numbers, commute with one another. Obviously the second formula is not valid in $\mathfrak{Q}$ quantum mechanics, where the matrix elements are quaternions and do not commute.

Indeed there seems to be no satisfactory definition of the tensor product of vectors in $\mathfrak{Q}$ Hilbert spaces as an operation that is unique, commutative, and invariant under the entire unitary groups of the spaces being multiplied.

As a result, the gap between $\mathfrak{Q}$ quantum mechanics and classical physics is greater than the gap between $\mathfrak{C}$ quantum mechanics and classical physics. In classical physics there are no phase relations to be considered when systems are imbedded as subsystems in a larger system, either by adding or multiplying their phase spaces. In $\mathfrak{C}$ quantum mechanics there are phase relations between states that are important when sums are formed but not when products are formed. In $\mathfrak{Q}$ quantum mechanics the phase relations are important when states are either added or multiplied.

This novel feature of $\mathfrak{Q}$ quantum mechanics is expressed in another way in terms of complementarity. In classical physics there are no complementarity relations. In $\mathfrak{R}$ and $\mathfrak{C}$ quantum mechanics complementarity relations exist between physical properties of one physical system, but not between properties of different systems; the momentum of an electron and the position of a neutron, for example. In $\mathfrak{Q}$ quantum mechanics, there exists a complementarity between some properties of any two systems. There is no reasonable way of forming a composite system such that all the observables associated with one of the systems commute with all the observables associated with the other.

What are we to make of this peculiar unitary nature of $\mathfrak{Q}$ quantum mechanics, that prevents one at the very start from speaking of absolutely independent systems? One point of view is that in nature, after all, there are no truly independent systems. In a unitary field theory of all the elementary particles, the problem of describing composite systems is to be solved without ever introducing such an artificial concept as a tensor product. This possibility exists in $\mathfrak{Q}$ quantum mechanics as well as in $\mathfrak{C}$ quantum mechanics. Yet we know that on occasion systems can be treated as if they were very nearly independent. Therefore we must approximate such independence in $\mathfrak{Q}$ quantum mechanics.

5. NEW DEGREES OF FREEDOM

In order to express composite systems in $\mathfrak{F}$ quantum mechanics (in particular in $\mathfrak{Q}QM$ it is) necessary to understand the new degrees of freedom, that fail to

commute even when they refer to distinct and independent systems. Their physical meaning we will study later. Now we express them as follows.

There exist continuous transformations acting on the $\mathfrak{F}$ numbers that leave invariant the intrinsic relations $+, \times, *$. These are the *automorphisms* of $\mathfrak{F}$. For $\mathfrak{F}=\mathfrak{R}$, $\mathfrak{C}$, $\mathfrak{Q}$ they are enumerated in footnote 6. Covariance of a law under these automorphisms is intended to mean that the form of the law does not depend on the choice of a realization of the number system $\mathfrak{F}$. In the case of $\mathfrak{C}QM$, it expresses equivalence of i with $-i$. In the case of $\mathfrak{Q}QM$ it expresses the equivalence of i_1, i_2, i_3 with any other like-handed set of anticommuting units. We shall make an assumption that simultaneously introduces new degrees of freedom associated with these automorphisms of the number system and guarantees a conservation law associated with these new degrees of freedom.

For every automorphism $A: a \rightarrow a^A$ of the number field $\mathfrak{F}$, there is given a co-unitary transformation $\Psi \rightarrow \Psi^A$ of the Hilbert space such that

$$(\Psi a)^A = \Psi^A a^A. \tag{1}$$

All the laws of the system are covariant under this transformation. This is the principle of covariance under automorphism of $\mathfrak{F}$, or of $\mathfrak{F}$ covariance for short. It permits the automorphisms of $\mathfrak{F}$ to act on $\mathfrak{H}_F$ and requires covariance under them. Let us develop this somewhat elliptic expression before proceeding to deduce its consequences.

There is a subset of the number system $\mathfrak{F}$ that is invariant under all the automorphisms A, namely the reals. In parallel, let us call a state vector real if it is invariant under all the automorphisms A. It is readily shown there exist complete sets of real vectors. It is now customary to define the action of A on any operator X by decreeing that the two relations

$$X^A\Psi^A = \Phi^A \quad \text{and} \quad X\Psi = \Phi$$

are to be equivalent, and we can speak of any operator X as real for which $X^A = X$. These real operators are just those that are invariant under the automorphisms of $\mathfrak{F}$.

Now it not to be inferred from the principle of $\mathfrak{F}$ covariance that all physical quantities are invariant under all automorphisms, i.e., real. We remarked in the introduction that the most interesting new features of $\mathfrak{Q}QM$ involve nonreal Hamiltonians, for example. Rather, it is the form of the laws relating them that is to be real. We suppose there is given a fundamental list of operators Ω in terms of which all observable physical statements (projections) can be expressed. By the laws of the system we mean a collection of algebraic relations $F(\Omega)=0$ among the operators Ω from which the intrinsic relations among the observable physical properties are to be deduced. Then $\mathfrak{F}$ covariance of the physical laws is the requirement that the relations

$$F(\Omega)=0 \quad \text{and} \quad F(\Omega^A)=0$$

are equivalent.

Let us examine the consequences for the three cases $\mathfrak{F}=\mathfrak{R}$, $\mathfrak{C}$, $\mathfrak{Q}$.

The principle of $\mathfrak{R}$ covariance is null; there are no automorphisms of $\mathfrak{R}$ other than the identity I. As an example of $\mathfrak{R}QM$, take the following formulation of the linear harmonic oscillator, in which position, momentum, and energy are taken as fundamental and η is the superselection operator discussed in Sec. 3:

$$\begin{aligned} px - xp &= -\eta \\ \tfrac{1}{2}(p^2+x^2) &= H \\ \eta^2 &= -1. \end{aligned} \tag{2}$$

It is obvious that this theory is real in name only as long as η is taken to be a superselection operator. That is, its observables are isomorphic to the observables of the conventional theory of the oscillator. (Were operators that failed to commute with η accepted as observables, this would correspond to accepting antilinear operators into the usual theory as observables and would be a significant change. We shall not consider this possibility.) A representation of the relations (2) is given by the usual differential operators for x and H, a 2×2 matrix for η,

$$\eta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3}$$

and a matric-differential operator for p,

$$p = -\eta(\partial/\partial x). \tag{4}$$

The principle of $\mathfrak{C}$ covariance is not null. It requires the existence of an antilinear operation $\psi \rightarrow \psi^C$ corresponding to the automorphism $a \rightarrow a^C \equiv a^*$ of the complex numbers. It further requires invariance of the dynamical relations under C. Thus, consider a set of relations defining a harmonic oscillator. Taking position, momentum, and energy as basic, they are usually

$$\begin{aligned} \tfrac{1}{2}(p^2+x^2) &= H, \\ px - xp &= -i, \\ \eta &= i. \end{aligned} \tag{5}$$

Evidently, these are not real relations; they are not satisfied by the complex conjugate quantities. They contain the symbol i explicitly. We may construct a $\mathfrak{C}$ covariant harmonic oscillator by adjoining a real anti-Hermitian η commuting with the other operators of the theory and taking (2) instead of (5) for the basic relations. Now i no longer appears explicitly; but not x is not a complete set of commuting observables by itself. There is the additional Hermitian operator

$$\epsilon = i\eta$$

that commutes with x and has the eigenvalues ± 1.

This operator is a superselection operator since η is one.

To put it differently, the relations (2) establish a correlation between the sign of i and the sense of time, while $\mathfrak{C}$ covariance implies that the sign of i has no physical correlate. If what we have called the $\mathfrak{R}QM$ theory of the oscillator coincides with the conventional theory (because of the superselection rule), what we have called the $\mathfrak{C}QM$ theory of the oscillator corresponds to one with an internal charge degree of freedom (because of the $\mathfrak{C}$ covariance rule). This new degree of freedom is not made effective in the free oscillator, the levels with $\epsilon=+1$ being degenerate with the levels with $\epsilon=-1$, but will be effective when the oscillator is coupled to other systems with forces that depend on ϵ.

Now let us examine the consequences of $\mathfrak{Q}$ covariance. The colinear transformation $\Psi\rightarrow\Psi^A$ is not as convenient to work with as a linear transformation, especially for defining observables. However if A is the automorphism $p\rightarrow qpq^{-1}$, $\Psi^A q$ depends linearly on Ψ since by (1)

$$(\Psi a)^A q=\Psi^A a^A q=\Psi^A(qaq^{-1})q=(\Psi^A q)a.$$

Let us designate this combination by $q\Psi$:

$$q\Psi\equiv\Psi^A q \quad \text{or} \quad \Psi^A=q\Psi q^{-1}\equiv\Psi^q.$$

Thus, $\mathfrak{Q}$ covariance permits us to identify the quaternions with linear operators on $\mathfrak{K}(\mathfrak{Q})$. Henceforth they may act on vectors from the left, a license that is taken for granted for complex numbers. In particular the units i_1, i_2, i_3 are now defined linear operators when they are written on the left of a state vector.[7] A quaternion acting in this capacity will be called a quaternion operator. It is easy to prove the existence of a basis for $\mathfrak{K}(\mathfrak{Q})$ in which the quaternion operator q is represented by a diagonal matrix whose diagonal elements are q. This is the same thing as proving the existence of a real basis. By using a real basis it is easy to prove that every operator X can be uniquely expressed in terms of the operators i_α in the form

$$X=\sum_0^3 X_\alpha i_\alpha,$$

where the coefficients X_α are real linear operators.

In complex quantum mechanics we are familiar with the decomposition of a general operator X into Hermitian and anti-Hermitian parts H, A:

$$X=H+A$$
$$H=\tfrac{1}{2}(X+X\dagger)$$
$$A=\tfrac{1}{2}(X-X\dagger).$$

Of course the same decomposition exists in $\mathfrak{Q}QM$. We wish to mention here that the structure we have introduced in $\mathfrak{Q}QM$ also makes possible a decomposition of the most general operator into 8 parts.

[7] Many of the properties of a system of three anticommuting, anti-Hermitian unitary operators on $\mathfrak{K}(\mathfrak{Q})$ like i_1, i_2, i_3 have been discussed by Teichmüller, reference 4.

The general X can be expanded, we have pointed out, in the form

$$X=\sum_0^3 X_\alpha i_\alpha,$$

where the coefficients X_α commute with the quaternions i_α (i.e., are real). It is natural to define the *quaternion conjugate* of an operator X by changing the sign of its imaginary terms:

$$X^q=\sum_0^3 X_\alpha i_\alpha^q.$$

Further we introduce the transpose as an intrinsic operation according to

$$X^T=X\dagger^q=X^q\dagger.$$

Then any X is the sum of symmetric and skew-symmetric operators S, K according to

$$X=S+K$$
$$S=\tfrac{1}{2}(X+X^T)=S^T$$
$$K=\tfrac{1}{2}(X-X^T)=-K^T.$$

Finally the symmetric and skew-symmetric operators can be decomposed into four parts relative to the quaternion basis i:

$$X=\sum_0^3 S_\alpha i_\alpha+\sum_0^3 K_\alpha i_\alpha.$$

If we merely seek a theory of the linear harmonic oscillator in $\mathfrak{Q}QM$, it is sufficient to replace the symbol i in the complex theory (5) by i_3, say. If we seek a $\mathfrak{Q}$ covariant theory, this is insufficient. However the relations (2) are $\mathfrak{Q}$ covariant.

To define the theory we must somehow specify the nature of the operator η.

6. TENSOR PRODUCTS

The additional structure given the quaternion Hilbert space by the principle of $\mathfrak{Q}$ covariance also makes possible a unique definition of tensor product. The group of the geometry, which is now to leave the quaternion operators fixed as well as the other Hilbert space concepts, is thereby reduced to a real unitary group. The problem of invariance is naturally easier when the group of the geometry is reduced.

We choose real bases x_1, x_2 in the two spaces $\mathfrak{K}_1$, $\mathfrak{K}_2$ to be multiplied. We take as a formal basis of $\mathfrak{K}_1\times\mathfrak{K}_2$ the symbolic products $|x_1\rangle|x_2\rangle=|x_2\rangle|x_1\rangle=|x_1, x_2\rangle$. We define the tensor products of two vectors $|1\rangle$, $|2\rangle$ in either order by their matrix elements in the $|x_1, x_2\rangle$ basis:

$$\langle x_1, x_2|1\rangle|2\rangle=\langle x_1|1\rangle\langle x_2|2\rangle$$
$$\langle x_1, x_2|2\rangle|1\rangle=\langle x_2|2\rangle\langle x_1|1\rangle.$$

Briefly, we multiply vectors by multiplying their matrix elements in real bases. Likewise we will form the tensor products of operators by multiplying their matrix elements in real bases:

$$\langle x_1, x_2|A_1\times A_2|x_1', x_2'\rangle=\langle x_1|A_1|x_1'\rangle\langle x_2|A_2|x_2'\rangle,$$
$$\langle x_1, x_2|A_2\times A_1|x_1', x_2'\rangle=\langle x_2|A_2|x_2'\rangle\langle x_1|A_1|x_1'\rangle.$$

Since the matrix elements do not commute, the tensor product does not commute in general. The projection in $\mathfrak{C}_1 \times \mathfrak{C}_2$ corresponding to P_1 in $\mathfrak{C}_1$ is taken to be $P_1 \times 1_2$ where 1_2 is the identity in $\mathfrak{C}_2$. Likewise

$$P_2 \rightarrow P_2 \times 1_1.$$

As the result of the noncommutativity we have the unusual circumstance that the projection in $\mathfrak{C}_1 \times \mathfrak{C}_2$ representing the logical conjunction of two propositions P_1 and P_2 may vanish when P_1 and P_2 do not. This conjunction is to be found as the intersection of the subspaces on which $P_1 \times 1_2$ and $P_2 \times 1_1$ project, and it is easy to construct examples $[\psi_1(x_1) = \exp(i_1 x_1), \psi_2(x_2) = \exp(i_2 x_2)]$ in which this intersection is zero although P_1 and P_2 are both one-dimensional projections. This anomaly does not occur for real projections.

In ordinary quantum mechanics there are three different ways in which corresponding operators on different systems combine when the systems are composed. Unitary transformations are examples of operators that compose *multiplicatively*: $U = U_1 \times U_2$. Infinitesimal generators therefore compose *additively*, as do many physical quantities. But the underlying field elements, which are also operators, are composed by *identification*; for example, the imaginary unit i has the property $i(|1\rangle|2\rangle) = (i|1\rangle)|2\rangle = |1\rangle(i|2\rangle)$. We have come as close as possible to this situation in our formulation of the tensor product in quaternion quantum mechanics. (Namely, the effect of multiplying a tensor product by a quaternion (from either side) may be computed by letting the quaternion multiply the factor in the tensor product it abuts.) This raises a certain conceptual problem of some importance. What can it mean physically to perform such an identification?

The elements of the number field represent logical relations between possible states of the physical system under consideration. (Indeed they may be represented by ordered triples of pure states.) It is evidently possible to say when two such relations are the "same" even for states of distinct physical systems without going beyond the concepts of pure logic. However this only defines the numbers up to automorphisms, evidently. For real quantum mechanics, this is sufficient to define them completely. In complex quantum mechanics an ambiguity remains: the i of one physical system may be identified either with i or $-i$ of another by this method. (It is possible to multiply the states of one system by the time-reversed states of the other and obtain a tensor product which meets all requirements of mere logic.) In quaternion quantum mechanics the corresponding ambiguity is infinitely greater. Yet it must be resolved in order to discuss composite systems. This requires us to introduce further elements of structure into the theory.

The operator η also presents a problem in this regard. The anti-Hermitian infinitesimal generator of time is $A = -\eta H$ and is composed additively for noninteracting systems because of its meaning. The Hamiltonian H should also be composed additively for noninteracting systems. These two requirements are enough to require that the operator η be composed by identification, as is multiplication by i in complex quantum mechanics. Moreover it then follows that the relation between η and the quaternion operators must be a universal one, not involving specific dynamical features of the system under consideration. *We shall pursue the hypothesis that the operator η actually corresponds to a member of the underlying number field* as in ordinary quantum mechanics. This reduces the problem to the one previously stated.

This appears to destroy the $\mathfrak{Q}$ covariance of the theory by singling out one imaginary. To preserve $\mathfrak{Q}$ covariance we will have to formulate a dynamical theory of the operator η so that its value is determined by an initial condition. This will be considered in a subsequent paper. In the mean time we will speak of $\mathfrak{Q}$ covariance "modulo η": agreeing to transform η as a quaternion when examining a theory for covariance.

In general, we see that the quantum theories will possess less symmetry than their classical limits. *The very process of quantization singles out an axis in the imaginary space.*

7. QUANTUM OSCILLATORS IN $\mathfrak{Q}$QM

Any linear field may be regarded as an aggregate of harmonic oscillators. In the previous section we sketched the $\mathfrak{R}$, $\mathfrak{C}$, and $\mathfrak{Q}$ quantum mechanics of a one-dimensional oscillator. This is the kind of oscillator a real scalar field gives. In the decomposition of the field in terms of some appropriate complete set of real orthogonal functions, or modes, real expansion coefficients appear as amplitudes and become the dynamical variables. Each of these amplitudes can be thought of as the displacement or coordinate of an oscillator. The principle quantum number n of one oscillator then counts the number of quanta in one mode, and is called an occupation number. Such oscillators we will call *quantum oscillators*. According to the statistics we have boson or fermion oscillators.

Sometimes it is convenient to express the field as an aggregate of two- or three-dimensional oscillators instead of a one-dimensional oscillator. If a complex field is expanded in terms of real orthogonal functions the amplitudes are complex numbers. Each amplitude can be regarded as the coordinate of an oscillator in the complex plane instead of as the coordinates of two real oscillators. This is useful when there is symmetry with respect to rotations in the complex plane. Instead of two separate occupation numbers n_1, n_2 the energy levels of the complex oscillator are most aptly described by a principal quantum number n and an angular momentum m about the origin of the complex plane. The number n is the number of quanta of that type and

the angular momentum m is the total charge that they carry:

$$n=n_+ + n_-,$$
$$m=n_+ - n_-,$$

where n_+ and n_- count positive and negative quanta.

In decomposing an isovector field in terms of a real family of (isoscalar) functions, we get oscillator amplitudes which are themselves triples (isovectors) of real numbers. Then it is convenient to work with three-dimensional oscillators. They possess a principal quantum number n, an orbital angular momentum l, and a magnetic quantum number m. The number n is again the number of quanta present of that type. The angular momentum $l \geq 0$ measures the charge multiplicity of the level, which is $2l+1$. The number m according to isospin theory gives the total charge of that member of the charge multiplet:

$$n=n_+ + n_0 + n_-,$$
$$m=n_+ - n_-,$$

where n_+, n_0, n_- count positive, neutral, and negative quanta.

Boson Oscillators

In this section we will study the oscillators that will make up a field that is quaternionic instead of real or complex. We shall call the system we are describing the quaternionic oscillator ($\mathfrak{Q}$ oscillator), having already touched on $\mathfrak{R}$ and $\mathfrak{C}$ oscillators. (It is quite possible to consider an oscillator in the general field $\mathfrak{F}$, developing all three cases in parallel. We shall not.)

The quaternionic oscillators move in the four-dimensional space $\mathfrak{Q}$. The invariance their Hamiltonians possess is called $\mathfrak{Q}$ covariance in Sec. 5. If we call the coordinate of the oscillator q,

$$q=\sum_0^3 q_\alpha i_\alpha, \quad i_0=1 \tag{1}$$

this means invariance under

$$q \longrightarrow q^a = aqa^{-1}$$

or under

$$q_0 \longrightarrow q_0$$
$$q_n \longrightarrow \sum_n r_{mn} q_n \quad (m, n=1, 2, 3). \tag{2}$$

Here we are taking note that the automorphism $q \longrightarrow q^a$ leaves the real part of q invariant and subjects the imaginary part of q to a Euclidean rotation represented by a matrix r_{mn}. The range of the variable q must be invariant under (2). There are just three non-null linear subspaces of $\mathfrak{Q}$ that are invariant under these automorphisms. The real axis is invariant: the three-dimensional space of pure imaginaries is invariant; and of course the entire space $\mathfrak{Q}$ is also. There are therefore three kinds of $\mathfrak{Q}$-covariant linear oscillators in the space $\mathfrak{Q}$; a real oscillator, an imaginary one, and a four-dimensional system we will call the full $\mathfrak{Q}$ oscillator. The real oscillator has already been treated in Sec. 5 and we will study the imaginary quaternionic oscillator now. (For a $\mathfrak{C}$ oscillator the automorphism invariance, which is there $\mathfrak{C}$ covariance, leads again to three kinds of oscillators, but the real axis and the imaginary axis are both one dimensional, and are not essentially different.)

As a guide for the development we write the classical Lagrangian in the form

$$L=(\dot{q}^*\dot{q}-\omega^2 q^* q)/2 \tag{2}$$

and take as classical coordinates the three real coefficients q_m of the expansion

$$q=\sum i_m q_m, \quad m=1, 2, 3. \tag{4}$$

This leads to a classical Hamiltonian and Poisson brackets

$$H=(p_m p_m+\omega^2 q_m q_m)/2$$
$$[q_m, q_n]_P=[p_m, p_n]_P=0 \tag{5}$$
$$[p_m, q_n]_P=\delta_{mn}$$

where $p_m=\dot{q}_m$. We replace these Poisson brackets by η times the quantum commutator. Since η is actually a particular quaternion in disguise this temporarily destroys the $\mathfrak{Q}$ covariance of the theory as a whole, as already mentioned. From the commutation relations

$$[q_m, q_n]=[p_m, p_n]=0$$
$$[p_m, q_n]=-\eta\delta_{mn}, \tag{6}$$

it follows by use of Jacobi's identity that η commutes with q_m and p_m. The operators p_m, q_m are not completely defined yet. Their commutation relations with the quaternion units perpendicular to η are undetermined. For example we can suppose that the q_m are represented by real symmetric operators. In that case the order of factors in the expansion (4) is irrelevant. The p_m are then represented by imaginary operators.

It is readily seen that the operators H and $\mathbf{Q}=\mathbf{q}\times\mathbf{p}$ are constants of the motion. The first is the energy or total occupation number of the three oscillators, the second the "angular momentum." Moreover the operators H, Q^2, Q_3 form a complete commuting set of observables, in spite of the "extra degrees of freedom" associated with quaternions ζ, $\eta\zeta$ that anticommute with η. ζ may be used to couple the oscillator to another system, but not to construct any Hermitian operator commuting with the three enumerated.

The ground state is unique (nondegenerate) and is given by the usual real wave function of the three-dimensional isotropic oscillator in the q representation. The possibility exists of multiplying this state by the quaternionic phase $\exp \mathbf{i}\cdot\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is composed of three arbitrary real numbers, without changing the energy, but this does not change the ray in Hilbert space $\mathfrak{H}(\mathfrak{Q})$.

We notice that the group of the theory associated with its $\mathfrak{Q}$ covariance contains the transformation $\eta \longrightarrow -\eta$ usually called time reversal.

Now consider the full $\mathcal{Q}$ oscillator. It is obvious that it is a composite of a real and an imaginary $\mathcal{Q}$ oscillator, so that in principle no new behavior is to be expected. Nevertheless it possesses an additional symmetry (if all the frequencies are the same) that deserves special note:

There is a definite sense in which neither the real nor imaginary $\mathcal{Q}$ oscillators are linear. To illustrate this sense, consider the Lagrangian and equation of motion of the complex oscillator, a classical mechanical system:

$$L=\tfrac{1}{2}(\dot{z}\dagger\dot{z}-\omega^2 z\dagger z)\quad \ddot{z}=-\omega^2 z.$$

The equations of motion are invariant under the substitution

$$z\rightarrow zc$$

where c is any complex constant. It is the $\mathcal{Q}$ analog of this linear transformation of the dynamical variables (not of the quantum state vector) that we now have in mind when we say that the real and imaginary oscillator are *nonlinear*: they will submit to the transformation

$$q\rightarrow qk,$$

not for any constant quaternion k, but only for real k. The full $\mathcal{Q}$ oscillator can be a genuinely linear system in the sense of quaternions. It is therefore interesting; but not therefore fundamental. (The basic requirement we impose on all our mechanical systems is $\mathcal{Q}$ covariance, not $\mathcal{Q}$ linearity. All actual systems are at least a little nonlinear.)

With this forward we present the $\mathcal{Q}$ linear oscillator. The classical Lagrangian is

$$L=\tfrac{1}{2}(\dot{q}\dagger\dot{q}-\omega^2 q\dagger q),\quad q=\textstyle\sum_0^3 q_\alpha i_\alpha. \tag{7}$$

The quantum Hamiltonian and commutators are then

$$\begin{aligned} H&=\tfrac{1}{2}\textstyle\sum_0^3 (p_\alpha^2+\omega^2 q_\alpha^2) \\ [p_\alpha,q_\beta]&=-\eta\delta_{\alpha\beta} \\ [p_\alpha,p_\beta]&=0 \\ [q_\alpha,q_\beta]&=0. \end{aligned} \tag{8}$$

η is again one of the unit imaginary quaternions. For simplicity we suppose here that the q_α are symmetric operators, as well as Hermitian. Let us be more explicit about the general definition of the generator designated in a special case by $\mathbf{Q}$. The automorphism $q\rightarrow aqa^{-1}$, $p\rightarrow apa^{-1}$ $(a\in\mathcal{Q})$ acting on the quaternion units induces a transformation of the real components q_α, p_α that leaves invariant their commutation relations (8). Accordingly defining q_m' by

$$aqa^{-1}=\sum_{m=0}^{3} q_m' i_m \tag{9}$$

and defining p_m' similarly, we demand $Q(a)$ such that

$$\begin{aligned} q_m'&=Q(a)q_m Q(a)^{-1}, \\ p_m'&=Q(a)p_m Q(a)^{-1}, \\ \eta'&=Q(a)\eta Q(a)^{-1}=\eta. \end{aligned} \tag{10}$$

The infinitesimal form of this relation deals with the three infinitesimal i_1, i_2, i_3 instead of the finite a, and seeks corresponding Hermitian operators Q_1, Q_2, Q_3. The infinitesimal i_m rotation $a=1+i_m\delta\theta/2$ generates variations

$$\delta_m q=aqa^{-1}=[i_m,q]\delta\theta/2\equiv\textstyle\sum \delta_m q_n i_n \tag{11}$$

and analogously for $\delta_m p$ and $\delta_m p_n$. We require Q_m to be a Hermitian operator satisfying

$$\delta_m q_n=\eta[Q_m,q_n] \tag{12}$$

and analogously for $\delta_m p_n$. We also require

$$[Q_m,\eta]=0.$$

From this follows

$$\mathbf{Q}=\mathbf{q}\times\mathbf{p}. \tag{13}$$

For the $\mathcal{Q}$ linear oscillator, infinitesimal left multiplication by a unit quaternion is also an invariant transformation. We shall call its generator $\mathbf{T}$, requiring that for infinitesimal left multiplications

$$\delta_m q\equiv i_m\delta\theta q/2\equiv\textstyle\sum_\alpha \delta_m q_\alpha i_\alpha$$

we have

$$\delta_m q_\alpha=\eta[T_m,q_\alpha]\delta\theta. \tag{14}$$

This leads to an essentially unique real $\mathbf{T}$ which like $\mathbf{Q}$ can be expressed in terms of q and p:

$$\begin{aligned} \mathbf{T}&=\textstyle\sum \mathbf{t}_{\alpha\beta}q_\alpha p_\beta, \\ \mathbf{T}&=(T_m), \\ \mathbf{t}&=(t^m). \end{aligned} \tag{15}$$

Substituting (15) into (14) yields the following unique forms for the coefficient matrices t:

$$\begin{aligned} {t_{01}}^1&=-\tfrac{1}{2}, \\ {t_{23}}^1&=-\tfrac{1}{2}. \end{aligned} \tag{15'}$$

The coefficients $\mathbf{t}$ are unchanged by cyclic permutations of (123) and are skew symmetric in the lower index pair. The coefficients whose values do not thereby follow from (15') are zero. The three real 4×4 matrices $\mathbf{t}_{\alpha\beta}$ generate certain orthogonal transformations in the space of the q.

Infinitesimal right multiplication is also an invariant transformation of this oscillator. We call its real generator $\mathbf{Y}$, requiring that for $\delta_m q=qi_m\delta\theta/2=\sum \delta_m q_\alpha i_\alpha$

$$\begin{aligned} \delta_m q_\alpha&=[y_m,q_\alpha]\delta\theta \\ 0&=[y_m,\eta]. \end{aligned} \tag{16}$$

This leads to an essentially unique real $\mathbf{Y}$, which can be expressed in the form

$$\mathbf{Y}=\mathbf{y}_{\alpha\beta}q_\alpha p_\beta. \tag{17}$$

Substituting (17) into (16) yields the following unique forms for the coefficient matrices $\mathbf{y}$:

$$\begin{aligned} {y_{01}}^1&=-\tfrac{1}{2}, \\ {y_{23}}^1&=\tfrac{1}{2}. \end{aligned} \tag{18}$$

The coefficients $\mathbf{y}$ are unchanged by cyclic permutations of (123) and are skew symmetric in the lower index pair. The coefficients whose values do not thereby follow from (18) are zero. The three real 4×4 matrices $y_{mn}{}^{p}$ generate certain orthogonal transformations in the space of the q_α. Obviously

$$\mathbf{Q}=\mathbf{T}+\mathbf{Y}. \tag{19}$$

We will call $\mathbf{T}$ the *isospin* and $\mathbf{Y}$ the *hypercharge* vectors of this quantum oscillator. However they do not have the universal significance of the charge vector $\mathbf{Q}$, which always exists. $\mathbf{T}$ and $\mathbf{Y}$ are strictly definable as conserved quantities only for $\mathfrak{Q}$ linear systems. The decomposition (19) is closely related to the well-known decomposition of the orthogonal group O_4 (of which $\mathbf{Q}$ generates a subgroup) into two groups O_3. It is easy to verify the following commutation relations from the definitions (13), (14), and (16), without computations:

$$\begin{aligned}
&[Q_m,Q_n]=\eta\epsilon_{mnp}Q_p\\
&[T_m,T_n]=\eta\epsilon_{mnp}T_p\\
&[Y_m,Y_n]=\eta\epsilon_{mnp}Y_p\\
&[T_m,Y_n]=0.
\end{aligned} \tag{20}$$

For example the last relation holds because $\mathbf{T}$ generates a left multiplication and $\mathbf{Y}$ generates a right multiplication. Left and right multiplications always commute:

$$a(qb)=(aq)b.$$

The vector operators $\mathbf{T}$ and $\mathbf{Y}$ both have the algebraic properties of an angular momentum and commute with each other. Therefore it is possible to form commutative scalars Q, T, and Y in the familiar manner.

$$\begin{aligned}
&\mathbf{Q}^2=Q(Q+1),\\
&\mathbf{T}^2=T(T+1),\\
&\mathbf{Y}^2=Y(Y+1).
\end{aligned} \tag{21}$$

It is convenient to introduce the operators

$$\begin{aligned}
&a_\alpha=(p_\alpha+\eta\omega q_\alpha)/(2\omega)^{\frac{1}{2}},\\
&a_\alpha\dagger=(p_\alpha-\eta\omega q_\alpha)/(2\omega)^{\frac{1}{2}},
\end{aligned} \tag{22}$$

obeying

$$[a_\alpha,a_\beta]=0,\quad [a_\alpha\dagger,a_\beta]=\delta_{\alpha\beta}.$$

Except for a zero-point energy we have

$$H=\sum\omega a_\alpha\dagger a_\alpha. \tag{23}$$

Because of the relation

$$Ha_\alpha-a_\alpha H=-\omega a_\alpha, \tag{24}$$

the operators a_α reduce the number of quanta, and the operators $a_\alpha\dagger$ increase the number of quanta. All the one-quantum states are of the form

$$\Psi_1=\textstyle\sum_\alpha a_\alpha\dagger\cdot c_\alpha\Psi_0, \tag{25}$$

where the c_m are $\mathfrak{Q}$ constants and Ψ_0 is the real ground state.

There exists a frame that makes the a_α, $a_\alpha\dagger$ real instead of the q_α. This reflects the arbitrariness mentioned in connection with the commutation relations (6).

Fermion Oscillators

For the fermion oscillator we again start from the classical Poisson brackets (5), but now we replace them with anticommutators instead of commutators. We may just as well write the relations defining the full (linear) $\mathfrak{Q}$ oscillator, since the real or imaginary oscillator can be extracted easily. They have been given essentially the following form by Gursey[8]:

$$\begin{aligned}
&\{p_\alpha,q_\beta\}=\delta_{\alpha\beta}=1,\,2,\,3\\
&\{p_\alpha,p_\beta\}=0\\
&\{q_\alpha,q_\beta\}=0\\
&\{p_\alpha,i_\beta]=0\\
&[q_\alpha,i_\beta]=0\\
&H=\tfrac{1}{2}\eta\omega\textstyle\sum_\alpha[p_\alpha,q_\alpha].
\end{aligned}$$

The $\mathfrak{Q}$ covariance of this system is clear, modulo η, and in addition it possesses $\mathbf{T}$ and $\mathbf{Y}$ invariance, which are defined as in (14,16) in terms of left and right multiplication by quaternions. Where the boson oscillator possessed an infinity of distinct energy eigenvalues the fermion oscillator has but five in the linear case (two in the real case; four in the imaginary). These correspond to $N=0,\,1,\,2,\,3,\,4$.

There is an essential difference between the fermion and boson symmetry properties. For the boson the Hamiltonian is $\mathfrak{Q}$ covariant but the anti-Hermitian time generator involves η. For the fermion the anti-Hermitian time generator is $\mathfrak{Q}$ covariant but the Hamiltonian involves η.

ACKNOWLEDGMENTS

We are grateful to E. C. G. Stueckelberg for enlightening discussions on the relation between imposing superselection principles and augmenting the underlying number field of the quantum theory. T. Kannelopoulos helped us with many quaternionic problems. T. D. Lee's advice helped to shape much of the attitude adopted in this work. A remark by C. N. Yang that the neutron and proton might be distinguished by their behavior under the automorphism aqa^{-1} of the quaternions determined our approach to the problem of electric charge, and his severe criticism has influenced the form in which this work is presented as well. We are indebted to F. Gürsey, who has long studied quaternionic formulations of elementary particle theories, for valuable suggestions. For hospitality and support we are severally indebted to CERN, the Institute for Advanced Studies, Stevens Institute of Technology, the University of Colorado, Oak Ridge National Laboratory, the University of Buenos Aires, and the Swiss Commission of Atomic Science.

[8] F. Gürsey (private communication).

APPENDIX A. $\mathfrak{Q}$ HILBERT SPACE

Most of the statements in the theory of a complex Hilbert space can be taken over for a $\mathfrak{Q}$ Hilbert space. We will develop the elementary theory in a way that makes the analogy between the two obvious. The principle results of this Appendix are the Spectral theorem and our modification of Stone's theorem.

A Hilbert space $\mathfrak{H}$ over the field of the quaternions is a set of vectors $\varphi, \psi, \cdots$ which have the following property.

1. H is a linear space:

If $\varphi, \psi \in \mathfrak{H}$ then $\varphi+\psi \in \mathfrak{H}$.
If $\varphi \in \mathfrak{H}$ then $\varphi q \in H$, $q \in \mathfrak{Q}$.

$$(\varphi+\psi)q = \varphi q + \psi q$$
$$\varphi(p+q) = \varphi p + \varphi q$$
$$(\varphi p)q = \varphi(pq).$$

2. In $\mathfrak{H}$ a duality and therefore a scalar product are defined: To *any* vector φ is associated a dual vector $\varphi\dagger$ in the dual space, such that the *scalar product* $\varphi\dagger\psi$ is Hermitian symmetric, linear in the second factor, and antilinear in the first factor:

$$\varphi\dagger\psi = (\psi\dagger\varphi)^q$$
$$\varphi\dagger(\psi_1+\psi_2) = \varphi\dagger\psi_1 + \varphi\dagger\psi_2$$
$$\varphi\dagger(\psi q) = (\varphi\dagger\psi)q.$$

We remark that

$$(\varphi q)\dagger\psi = [\psi\dagger(\varphi q)]^q = [(\psi\dagger\varphi)q]^q = q^q(\psi\dagger\varphi)^q = q^q(\varphi\dagger\psi).$$

Finally, $\psi\dagger\psi$ is definite:

$$\psi\dagger\psi \geqq 0, \quad \text{and} \quad \psi\dagger\psi = 0 \quad \text{implies} \quad \psi = 0.$$

This is consistent because $q\dagger q \geq 0$.

3. $\mathfrak{H}$ is complete.

Schwarz's Inequality

Since in the theory of Hilbert space nearly every theorem depends on Schwarz's inequality, we will derive this relation as an example exhibiting the minor departure from complex Hilbert space. In fact, a careful handling of the scalars is all one needs.

Consider

$$0 \leq (\varphi p - \psi q)\dagger(\varphi p - \psi q)$$
$$= p^q(\varphi\dagger\varphi)p - p^q(\varphi\dagger\psi)q - q^q(\psi\dagger\varphi)p + q^q(\psi\dagger\psi)q.$$

Putting $p = (\psi\dagger\psi)$, $q = (\psi\dagger\varphi)$, we have:

$$0 \leq \psi\dagger\psi\{(\psi\dagger\psi)(\varphi\dagger\varphi) - (\psi\dagger\varphi)(\varphi\dagger\psi)\}$$

or since $(\psi\dagger\psi) \geq 0$, $(\varphi\dagger\varphi)(\psi\dagger\psi) \geq (\varphi\dagger\psi)(\psi\dagger\varphi)$. In the usual way one then derives the triangle inequality, and the statements about continuity, continuous sequences, and convergence.

Operators

A linear operator L is a mapping of $\mathfrak{H}$ into itself with the following properties

$$L(\varphi+\psi) = L\varphi + L\psi; \quad \varphi, \psi, L\varphi, L\psi \in \mathfrak{H}$$
$$L(\varphi q) = (L\varphi)q.$$

(This definition is adequate for operators in finite-dimensional spaces or for bounded operators. For unbounded operators a more precise specification of the domain of definition is necessary.) According to this definition of a linear operator, multiplication of a vector by a quaternion (from the right) is not a linear operator, for if we compute the effect of the transformation defined by $\psi \rightarrow \psi' = \psi q$ upon a linear combination

$$\psi = \varphi_1 p_1 + \varphi_2 p_2,$$

we find

$$\psi \rightarrow \psi' = \psi q = \varphi_1 p_1 q + \varphi_2 p_2 q \neq \varphi_1' p_1 + \varphi_2' p_2.$$

Instead $\psi' = \varphi_1'(q^{-1}p_1 q) + \varphi_2'(q^{-1}p_2 q)$. We therefore define a *colinear* operator Λ as a correspondence of H into H such that

$$\Lambda(\varphi+\psi) = \Lambda\varphi + \Lambda\psi,$$
$$\Lambda(\psi q) = \Lambda\psi q',$$

where q' is a quaternion associated with q and Λ. This just expresses the fact that Λ carries subspaces into subspaces. It may be shown that q' is a quaternion of the same class as q, and it will be written q^Λ.

Hermitian, anti-Hermitian, unitary, and normal operators are defined, respectively, through the equations

$$\varphi\dagger H\psi = (H\varphi)\dagger\psi, \quad \varphi\dagger A\psi = -(A\varphi)\dagger\psi$$
$$(U\varphi)\dagger(U\psi) = \varphi\dagger\psi, \quad (N\varphi)\dagger(N\psi) = (N\dagger\varphi)\dagger(N\dagger\psi)$$

for every $\varphi, \psi \in \mathfrak{H}$, or

$$H = H\dagger, \quad A = -A\dagger, \quad U^{-1} = U\dagger, \quad NN\dagger = N\dagger N.$$

We note at this point an important difference in the relation between Hermitian and anti-Hermitian operators in the complex and the quaternionic case. In the former the imaginary unit i provides a unique transition from the Hermitian to the anti-Hermitian and vice versa: $H = iA$, $A = -iH$. (Moreover, in that case,

$$U = \frac{1-iH}{1+iH}.)$$

That an analogous relation in a $\mathfrak{Q}$ space cannot hold is easily seen from the fact that in an n-dimensional space H has $2n^2-n$, but A has $2n^2+n$, independent $\mathfrak{Q}$ elements.

However,

$$U = (1-A)/(1+A)$$

still holds. (In this case the numerator and the denominator commute.)

As in a complex space, a unitary operator does not provide us with the most general norm-preserving transformation (isometry).

This is given through a colinear operator Y with

$$(Y\varphi)\dagger(Y\psi)=(\varphi\dagger\psi)^{Y}$$

$(\varphi\dagger\psi)^{Y}$ being a quaternion of the same class as $(\varphi\dagger\psi)$. Such an operator Y is termed co-unitary.

Symplectic Representation

Finally we mention (following Chevalley) a special representation of a $\mathfrak{Q}$ operator in a n-dimensional $\mathfrak{Q}$ space $\mathfrak{H}(\mathfrak{Q})$ through a $\mathfrak{C}$ operator in a $2n$-dimensional $\mathfrak{C}$ space $\mathfrak{H}(\mathfrak{C})$. This is nothing but a straightforward generalization of the decomposition of a quaternion in an ordered pair of complex numbers.

We introduce in both spaces an orthornormal system, and associate to every vector ψ with components $(\psi_1\cdots\psi_n)$ the vector ψ_s with components $(c_1\cdots c_n, d_1\cdots d_n)$, where $\psi_m=c_m+i_2d_m$, and $[c_m,i_3]=[d_m,i_3]=0$. For each operator L on $\mathfrak{H}(\mathfrak{Q})$ we define L_s on $\mathfrak{H}(\mathfrak{C})$ by $L_s\psi_s=(L\psi)_s$. One can easily show that the correspondence $\psi\leftrightarrow\psi_s$ is bi-unique, and that to every operator L in $\mathfrak{H}(\mathfrak{Q})$ corresponds one and only one L_s in $\mathfrak{H}(\mathfrak{C})$, the *symplectic representative* of L. In particular, to an isometry in $\mathfrak{H}(\mathfrak{Q})$ corresponds an isometry in the $\mathfrak{H}(\mathfrak{C})$. We designate the isometry $\psi\rightarrow\psi i_2$ by K and its symplectic representative by K_s. K_s is anti-unitary.

Let the operator J be defined so that $J\psi$ has components $(i_3\psi_1, i_3\psi_2, \cdots, i_3\psi_n)$. The condition that an operator on $\mathfrak{H}(\mathfrak{Q})$ be unitary, $U\dagger=U^{-1}$, is expressed for the corresponding operator on $\mathfrak{H}(\mathfrak{C})$ through the two equations $U_s\dagger=U_s^{-1}$ and $U_s{}^{T}J_sU_s=J_s$, where $U_s{}^{T}$ means the transposed operator and $J_s{}^{T}=-J_s$, $J_s{}^{2}=-1$. This shows that the group of unitary transformations in a $\mathfrak{Q}$ space is isomorphic to the so-called unitary-restricted symplectic group $Sp(n)$ and therefore is simple. (The group of unitary transformations in a $\mathfrak{C}$ space, as is well known, is not.)

In order than an operator L' on $\mathfrak{H}(\mathfrak{C})$ be the symplectic representative of some L on $\mathfrak{H}(\mathfrak{Q})$, $L'=L_s$, it is necessary and sufficient that L' commute with the operator K_s:

$$L'K_s=K_sL'. \tag{1}$$

APPENDIX B. THE EIGENVALUE PROBLEM

Defining an eigenvalue as a root of the secular determinant is not very convenient here, even in the finite dimensional case, because of noncommutativity. We will treat the diagonalization problems in the finite-dimensional case with the same means as the infinite dimensional, that is to say with spectral resolution techniques. The goal of this and the next chapter will be the theorem that every normal operator admits a spectral resolution.

We give here only a short survey, concentrating on those points where the statements or the methods differ essentially from what one is used to in the complex case and referring the reader for the rest to the literature. We state the spectral theorems for bounded Hermitian, anti-Hermitian, and, finally, normal operators. These theorems then will easily be extended to unbounded operators.

We begin with some general theorems.

We state without proof: The eigenvalues of a Hermitian, anti-Hermitian, or unitary operator are real, pure imaginary, or have norm 1, respectively. Whereas in a $\mathfrak{C}$ space, eigenvectors belonging to different eigenvalues are orthogonal, here we can assert only the following theorem: eigenvectors of normal operators belonging to different *eigenclasses* are orthogonal.

Proof: If $N\varphi=\varphi n$ and $N\varphi'=\varphi'n'$ (n, n' quaternions) and N is normal then $N\dagger\varphi=\varphi n^*$ and $N\dagger\varphi'=\varphi'n^*$. Thus

$$n^*\varphi\dagger(\varphi'n')=(\varphi n)\dagger(\varphi'n')=\varphi\dagger N\dagger N\varphi'=\varphi\dagger NN\dagger\varphi'$$
$$=n\varphi\dagger\varphi'n'.$$

Thus $n^*qn'=nqn'^*$ where $q\equiv\varphi\dagger\varphi'$. It follows that if $q\neq0$, then n and n' (which can only vanish together then) belong to the same class. And on the other hand, from

$$L\varphi=\varphi\alpha \quad (L \text{ any linear operator})$$

follows:

$$L\varphi q=\varphi\alpha q=(\varphi q)q^{-1}\alpha q.$$

Together with φ, φr (where r is a real number) is an eigenvector belonging to the eigenvalue α; but φq belongs to the eigenvalue $q^{-1}\alpha q$. In contrast to the complex case in general $q^{-1}\alpha q\neq\alpha$, but belongs to the same class. One expects, therefore, that only the class (i.e., the norm and the real part) of an eigenvalue might have an invariant meaning. We term the class of an eigenvalue an eigenclass. Every number in an eigenclass is an eigenvalue.

If the eigenvectors belonging to the eigenclass $\{\alpha\}$ span a subspace $\mathfrak{M}$ of dimension n, we call n the degeneracy of the eigenclass. It is always possible to span $\mathfrak{M}$ with an orthogonal system of eigenvectors *belonging to the same eigenvalue*. This can be done with a construction analogous to the Schmidt orthogonalization process. The choice of the orthogonal eigenvectors can even be made in such a way that their common eigenvalue involves only one arbitrarily selected imaginary quaternion, say i_3:

$$L\varphi_r=\varphi_r\alpha_r$$
$$(\varphi_r\dagger\varphi_s)=\delta_{rs}, \quad \alpha_r=a_r+b_ri_3$$

a_r and b_r real. b_r can even be made ≥0.

We now are able to state the needed spectral theorems, and to prove them where new features might appear.

Hermitian Operators

Theorem: If H is a bounded Hermitian operator, [that is, $\varphi\dagger H\psi=(H\varphi)\dagger\psi$, $||H\psi||\leq M||\psi||$] then

there exists a unique spectral family E_λ such that

$$H=\int_{-\infty}^{+\infty}\lambda dE_\lambda.$$

This theorem is generalized to the quaternion case without any change in wording whatsoever. Even the proof can be literally transferred to the quaternion case. The numbers λ occurring in the spectral resolution are real and they commute with all quaternions. The essential noncommutative property of the quaternions has therefore no chance to make itself felt in this case. This is why the analogy to the case of complex spaces is complete.

Anti-Hermitian Operators

An anti-Hermitian operator $A=-A\dagger$ admits a unique spectral resolution:

$$A=\int_0^\infty J\lambda dE_\lambda$$

where E_λ is a (unique) spectral family over the interval $0\leq\lambda$, and J a unique operator with the properties:

$$J^+=-J,\quad J^2=-E,\quad [J,E_\lambda]=[J,A]=0.$$

In contrast to the complex case, the operator J, which is the appropriate generalization of i, appears and the integral is extended only from 0 to $+\infty$.

Unitary Operators

Every unitary transformation U in the quaternion space admits a unique spectral resolution

$$U=\int_0^\pi e^{J\theta}dE(\theta)$$

where J is a uniquely determined linear operator on $E(\pi)[1-E(+0)]V$ with the properties $J\dagger=-J$, $[J,E(\theta)]=0$, and $J^2=-1$ on its domain, and $E(\theta)$ is a spectral family over the interval $0\leq\theta\leq\pi$.

APPENDIX C. STONE'S THEOREM IN A $\mathfrak{Q}$ HILBERT SPACE

A one-parameter group of unitary $\mathfrak{Q}$ operators $U(t)$ can be written in the form:

$$U(t)=e^{At}\quad\text{where}\quad A\equiv\lim_{t\to0}[U(t)-E]/t=-A\dagger.$$

Furthermore (and here it is possible to go a step further than in a $\mathfrak{C}$ Hilbert space):

$$A=\eta H\quad\text{where}\quad[\eta,H]=0,\quad\eta^2=-1,\quad\eta\dagger=-\eta,$$
$$H\dagger=H,\quad H\geq0.$$

This theorem can be proved with help of the symplectic representation. To every $U(t)$ we can associate a symplectic representative $U_s(t)$ acting on $\mathfrak{K}(\mathfrak{C})$, such that all group-relations are preserved. Stone's theorem (proved for $\mathfrak{C}$ Hilbert spaces) then asserts the existence of

$$A_s=\lim_{t\to0}[U_s(t)-E]/t=-A_s\dagger,$$

such that

$$U_s(t)=e^{A_st}.$$

But from

$$[U_s(t),K_s]=0$$

follows

$$[A_s,K_s]=0.$$

Therefore by (1) of Appendix A, there is an A of which A_s is the symplectic representative, and

$$A=\lim_{t\to0}[U(t)-E]/t.$$

Now:

$$A\equiv\int_0^\infty\eta\lambda dE_\lambda=\eta\int\lambda dE_\lambda\equiv\eta H,\quad H\geq0.$$

(Of course it is desirable to have a direct proof avoiding the complex.)

Analyticity Properties of the Momentum-Space Vertex Function*†

W. S. BROWN‡
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received September 13, 1961)

The problem of finding the maximal domain of analyticity of the momentum-space vertex function implied solely by Lorentz covariance, local commutativity, and mass spectrum with thresholds not all zero, is restated as a holomorphy envelope problem. If only one of the threshold masses is nonzero, the problem can be divided into two simpler problems. The triangle diagram suggests the holomorphy envelope of one, but only gives an upper bound on the holomorphy envelope of the other. It is shown that the boundary of the latter most probably consists of the three cuts and a single quasi-analytic hypersurface with certain specified properties. If two of the threshold masses are equal and the third vanishes, an appealing conjecture suggested by the triangle diagram is shown to contradict a generalization of Jost's example. In the general case some upper and lower bounds are obtained. The Källén-Toll representation and conjecture are briefly discussed. The relation of the first three terms in the representation of the position-space vertex function to the Mercedes diagram in perturbation theory is displayed, and it is shown that there is no analogous relation for the fourth term. In the single threshold case this fourth term must account for the singularities of the momentum-space vertex function on the quasi-analytic hypersurface which bounds the holomorphy envelope. The motivation for studying the analyticity properties of vertex diagrams is discussed, and the simplest totally symmetric ones are investigated.

INTRODUCTION

THIS paper is a summary of the results of an earlier unpublished report, which we shall refer to as I.[1] The problem under investigation is to find the enlargement of the analyticity domain of the momentum-space vertex function which is implied by the imposition of mass spectral conditions. The maximal domain which is implied solely by Lorentz covariance, local commutativity, and positiveness of energy was first found by Källén and Wightman[2] in a paper which we shall refer to as KW. A simplified proof (due to Ruelle)[3] of their result is given in I.

This paper is divided into four major sections as follows: Section 1 is devoted to general theory including a precise formulation of the problem to be investigated. In Sec. 2 an upper bound on the enlarged domain, which can be deduced with the aid of the triangle diagram, is presented. The single-threshold problem is discussed in Sec. 3. It can be divided into two simpler problems, one of which is completely solved. For the other, the qualitative form of the solution and several upper and lower bounds are given. Section 4 consists of five subsections on "Other Topics"; namely, the Källén-Toll representation, some perturbation theory examples, a generalization of Jost's example, the double-threshold problem, and the outlook for further progress. The subsection on perturbation theory examples was written in collaboration with Dr. D. B. Fairlie.

1. GENERAL THEORY

In order to establish our notation, we shall begin by defining the position-space and momentum-space vertex functions corresponding to a set of three local scalar fields $A_i(x)$, $i=1, 2, 3$. Let (ijk) denote any permutation of the indices (123). The threefold vacuum expectation values are defined by

$$W_{ijk}(x_1,x_2,x_3)=\langle 0|A_i(x_i)A_j(x_j)A_k(x_k)|0\rangle. \quad (1.1)$$

By translational invariance, the Fourier transform involves a δ function, which we factor off for convenience:

$$(2\pi)^{12}\delta(p_1+p_2+p_3)\tilde{W}_{ijk}(p_1,p_2,p_3)=\int dx_1dx_2dx_3$$
$$\times\exp[i(p_1x_1+p_2x_2+p_3x_3)]W_{ijk}(x_1,x_2,x_3). \quad (1.2)$$

If we substitute (1.1) into the right side and insert a sum over intermediate states, we find the support property

$$\tilde{W}_{ijk}(p_1,p_2,p_3)=0;\ \text{unless}\ p_i\in V_+^{\,i}\ \text{and}\ p_k\in V_-^{\,k} \quad (1.3)$$

where the $V_\pm^{\,i}$ are defined by

$$V_\pm^{\,i}=\{p:\ p^2\geq M_i^2,\ p_0\gtrless\}, \quad (1.4)$$

with mass thresholds M_i defined by

$$M_i^2=\min(p^2)\ \text{such that}$$
$$\langle 0|A_i(0)|p\rangle\langle p|A_j(\xi)A_k(0)|0\rangle\neq 0 \quad (1.5)$$

for some momentum eigenstate $|p\rangle$ and some four-vector ξ. It follows that $W_{ijk}(x_1,x_2,x_3)$ is the boundary

* This work was supported in part by the National Science Foundation and in part by the U. S. Air Force, Air Office of Scientific Research.

† Based on a dissertation submitted to Princeton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

‡ Present address: Bell Telephone Laboratories, Murray Hill, New Jersey.

[1] W. S. Brown, "On the Analytic Properties of the Vertex Function with Mass Spectral Conditions," Princeton University Report (1961). (This report contains a chapter on perturbation theory examples written in collaboration with Dr. D. B. Fairlie.) Copies may be obtained, while the supply lasts, from the secretary of the Physics Department, Princeton University, Princeton, New Jersey.

[2] G. Källén and A. Wightman, Mat. fys. Skrifter Danske Videnskab. Selsk. 1, No. 6 (1958).

[3] D. Ruelle, Thesis (Bruxelles, 1959).

value of the analytic function

$$W_{ijk}(\zeta_1,\zeta_2,\zeta_3)=\int dp_1dp_2dp_3\delta(p_1+p_2+p_3)$$
$$\times\tilde{W}_{ijk}(p_1,p_2,p_3)\exp[-i(p_1\zeta_1+p_2\zeta_2+p_3\zeta_3)] \quad (1.6)$$

from the tube

$$T_{ijk}: \begin{cases} y_i-y_j\in V_- \\ y_j-y_k\in V_-, \end{cases} \quad (1.7)$$

where $\zeta_i=x_i+iy_i$ and $V_\pm$ is the open cone $p^2>0$, $p_0\gtrless 0$. By complex Lorentz invariance, W_{ijk} can be continued throughout the extended tube T_{ijk}', which is defined as the set of all ζ such that $\Lambda\zeta$ is in T_{ijk} for some complex Lorentz transformation Λ. It now follows from local commutativity that all six of the analytic functions $W_{ijk}(\zeta)$ are the same and continue each other. Thus there is a single position-space analytic function $W(\zeta_1,\zeta_2,\zeta_3)$ analytic in the union of the extended tubes T_{ijk}'. However, the six vacuum expectation values $W_{ijk}(x)$ are not the same, since they are boundary values of $W(\zeta)$ from different tubes. Because of its complex Lorentz invariance $W(\zeta)$ depends only on the scalar products[4]

$$z_i=(\zeta_j-\zeta_k)^2; \quad i=1,2,3, \quad (1.8)$$

so we define

$$W(\zeta_1,\zeta_2,\zeta_3)\equiv W(z_1,z_2,z_3). \quad (1.9)$$

Clearly $W(z)$ is analytic in the domain U in z space which is obtained from the union of the extended tubes T_{ijk}' in ζ space by the mapping (1.9). This domain is not a holomorphy domain.[5] Its holomorphy envelope was first found by Källén and Wightman in KW, so we shall call it the KW domain or simply D_{KW}.

Before investigating D_{KW} in more detail we want to prove that it is also relevant for the momentum-space vertex function $H(z)$. We begin by introducing the retarded functions[6]

$$r_i(x_1,x_2,x_3)=\theta(x_i-x_j)\theta(x_j-x_k)$$
$$\times\langle 0|[[A_i(x_i),A_j(x_j)],A_k(x_k)]|0\rangle$$
$$+\theta(x_i-x_k)\theta(x_k-x_j)$$
$$\times\langle 0|[[A_i(x_i),A_k(x_k)],A_j(x_j)]|0\rangle, \quad (1.10)$$

where $\theta(x)$ is the step function

$$\theta(x)=\begin{cases} 1; & x_0>0 \\ 0; & x_0<0. \end{cases} \quad (1.11)$$

It is easily shown that $r_i(x)$ is invariant under the inhomogeneous Lorentz group and has the support property

$$r_i(x_1,x_2,x_3)=0, \text{ unless } x_i-x_j\in V_+ \text{ and } x_i-x_k\in V_+. \quad (1.12)$$

Its Fourier transform is defined by

$$(2\pi)^{12}\delta(p_1+p_2+p_3)\tilde{r}_i(p_1,p_2,p_3)=\int dx_1dx_2dx_3$$
$$\times\exp[i(p_1x_1+p_2x_2+p_3x_3)]r_i(x_1,x_2,x_3). \quad (1.13)$$

By translation invariance $r_i(x)$ depends only on the difference vectors

$$\begin{aligned} \xi_j&=x_i-x_j \\ \xi_k&=x_i-x_k. \end{aligned} \quad (1.14)$$

Therefore (1.13) can be rewritten

$$(2\pi)^8\tilde{r}_i(p_1,p_2,p_3)=\int d\xi_jd\xi_k$$
$$\times\exp[-i(p_j\xi_j+p_k\xi_k)]r_i(x_1,x_2,x_3). \quad (1.15)$$

Using (1.12) we see that this is the boundary value of the analytic function

$$(2\pi)^8H_{i+}(s_1,s_2,s_3)=\int d\xi_jd\xi_k$$
$$\times\exp[-i(s_j\xi_j+s_k\xi_k)]r_i(x_1,x_2,x_3) \quad (1.16)$$

from the tube

$$T_{i+}: \begin{cases} q_j\in V_- \\ q_k\in V_-, \end{cases} \quad (1.17)$$

where $s_i=p_i+iq_i$ and of course $s_1+s_2+s_3=0$. Similarly, we introduce the advanced functions

$$a_i(x_1,x_2,x_3)=\theta_-(x_i-x_j)\theta_-(x_j-x_k)$$
$$\times\langle 0|[[A_i(x_i),A_j(x_j)],A_k(x_k)]$$
$$+\theta_-(x_i-x_k)\theta_-(x_k-x_j)$$
$$\times\langle 0|[[A_i(x_i),A_k(x_k)],A_j(x_j)]|0\rangle \quad (1.18)$$

where $\theta_-(x)=\theta(-x)$. By the same reasoning we find that $\tilde{a}_i(p_1,p_2,p_3)$ is the boundary value of an analytic function $H_{i-}(s_1,s_2,s_3)$ from the tube

$$T_{i-}: \begin{cases} q_j\in V_+ \\ q_k\in V_+. \end{cases} \quad (1.19)$$

[4] This follows from the Hall-Wightman theorem. See D. Hall and A. Wightman, Kgl. Danske Videnskab. Selsk. Mat.-fys. Medd **31**, No. 5 (1957).

[5] In other words *every* function analytic in it can be continued into a larger domain called the holomorphy envelope. For an introduction to the theory of functions of several complex variables see any of the following: A. S. Wightman, in *Relations des dispersions et particules élémentaires*, edited by C. De Witt and R. Oehme (Hermann & Cie, Paris, France, 1960). S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton University Press, Princeton, New Jersey, 1948). H. Behnke and P. Thullen, *Théorie der Funktionen Mehrerer Komplexer Veränderlichen* (Chelsea Publishing Company, New York, 1934). For a "physicists summary" see I (Chapter 2) or KW (pages 23–25).

[6] We use the notation of Burgoyne. See N. Burgoyne, Thesis (Princeton, 1961) and H. Araki and N. Burgoyne, Nuovo cimento **18**, 342 (1960). This is related to the more familiar notation of Lehmann, Symanzik, and Zimmermann by the formula

$$r_i(x_1,\cdots,x_n)=i^{n-1}r(x_i;x_1,\cdots,x_{i-1},x_{i+1},\cdots,x_n).$$

It follows directly from the definitions that

$$r_i(x_1,x_2,x_3)-a_j(x_1,x_2,x_3)=\theta(x_i-x_j)K_{ijk}(x_1,x_2,x_3), \quad (1.20)$$

where $K_{ijk}(x)$ is the double commutator

$$\begin{aligned}K_{ijk}(x_1,x_2,x_3)&\\ &=\langle 0|[[[A_i(x_i),A_j(x_j)],A_k(x_k)]|0\rangle\\ &=W_{ijk}(x)-W_{jik}(x)-W_{kij}(x)+W_{kji}(x). \quad (1.21)\end{aligned}$$

By (1.3), its Fourier transform has the support property

$$\begin{aligned}\tilde{K}_{ijk}(p_1,p_2,p_3)=0;\quad &\text{unless}\quad p_k\in V_{\pm}{}^k\\ &\text{and}\quad [p_i\in V_{\mp}{}^i\quad\text{or}\quad p_j\in V_{\mp}{}^j]. \quad (1.22)\end{aligned}$$

Taking the Fourier transform of (1.20) (with $i=1$ and $j=2$), we find

$$\begin{aligned}\tilde{r}_1(p)-\tilde{a}_2(p)&\\ &=(2\pi i)^{-1}\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}dk_0\frac{\tilde{K}_{123}(p_1+k,\,p_2-k,\,p_3)}{k_0-i\epsilon} \quad (1.23)\end{aligned}$$

where $k=(k_0;0,0,0)$, and using (1.22) we conclude that

$$\tilde{r}_i(p)-\tilde{a}_j(p)=0;\quad \text{unless}\quad p_k{}^2\geq M_k{}^2. \quad (1.24)$$

By complex Lorentz invariance, each of the functions $H_{i\pm}$ can be continued throughout its extended tube $T_{i\pm}'$, and (1.24) implies that all six are the same and continue each other. Thus there is a single momentum-space analytic function $H(s_1,s_2,s_3)$ analytic in the union of the extended tubes $T_{i\pm}'$. However, the Fourier transformed retarded and advanced functions are not the same, since they are boundary values of $H(s)$ from different tubes. Because of its complex Lorentz invariance, $H(s)$ depends only on the scalar products

$$z_i=s_i{}^2;\quad i=1,\,2,\,3 \quad (1.25)$$

so we define

$$H(s_1,s_2,s_3)\equiv H(z_1,z_2,z_3). \quad (1.26)$$

Clearly $H(z)$ is analytic in U and hence also in D_{KW}.

Manifolds and Hypersurfaces in C^n

Next let us introduce a few definitions which are helpful for the description of domains in C^n, the space of n complex variables.

An analytic variety in a domain D in C^n is a relatively closed set in D which is locally definable by the zeros of a finite number k of analytic functions. A point P in an analytic variety V is a regular point if the Jacobian of these functions with respect to some subset of k of the z's is nonzero at P. The complex dimension of an analytic variety V in the neighborhood of a regular point P is $n-k$. An analytic manifold in D is an analytic variety in D, every point of which is a regular point.

A hypersurface in a domain D in C^n is a relatively closed set H in D, such that for each point P in H there is a neighborhood N in which a coordinate system can be defined so that $H\bigcap N$ is a $(2n-1)$-real-dimensional coordinate plane. An analytic hypersurface of type t (analytic type) in D is a hypersurface H in D which is locally definable in the form

$$H=\{z:\ f(z,\rho)=0\ \text{ for some real }\rho\}, \quad (1.27)$$

where f is analytic in z, and t-times continuously differentiable (analytic) in ρ, and $\partial f/\partial z_i$ is nonzero for some i. A quasi-analytic hypersurface of rank r and type t (analytic type) in D, is a hypersurface H in D which is locally definable in the form

$$\begin{aligned}H=\{z:\ f_i(z,\rho)=0,\quad &i=1,\cdots,n-r,\\ &\text{for some}\quad \rho\in R^{2n-1-2r}\}, \quad (1.28)\end{aligned}$$

where f_i is analytic in z, and t times continuously differentiable in ρ, and where the Jacobian of the f's with respect to some subset of $n-r$ of the z's is nonzero.

The KW Domain

We are now prepared to describe the KW domain, in which both $W(z)$ and $H(z)$ are analytic. It is bounded by pieces of the following seven analytic hypersurfaces:

$$\begin{aligned}C_i:&\quad z_i=\rho;\quad i=1,\,2,\,3\\ F_{jk}':&\quad \rho^2+\rho(z_i-z_j-z_k)+z_jz_k=0;\quad i=1,\,2,\,3\\ \mathfrak{F}:&\quad \rho^2-\rho(z_1+z_2+z_3)+(z_1z_2+z_1z_3+z_2z_3)=0\\ &\quad \rho\geq 0. \quad (1.29)\end{aligned}$$

Let z_1 be chosen anywhere in its cut plane. Then divide the z_2 plane into three sectors as shown in Fig. 1. If z_2 is chosen in Sector III, then D_{KW} is the entire cut z_3 plane. For z_2 in Sector I or II, D_{KW} is illustrated in Fig. 2 and Fig. 3, respectively. Clearly, the exterior of D_{KW} is the union of the three cuts C_i and a six-real-dimensional region, bounded by pieces of the F_{ij}' and $\mathfrak{F}$, which we shall call the singularity solid. It will be convenient to subdivide the singularity solid into regions

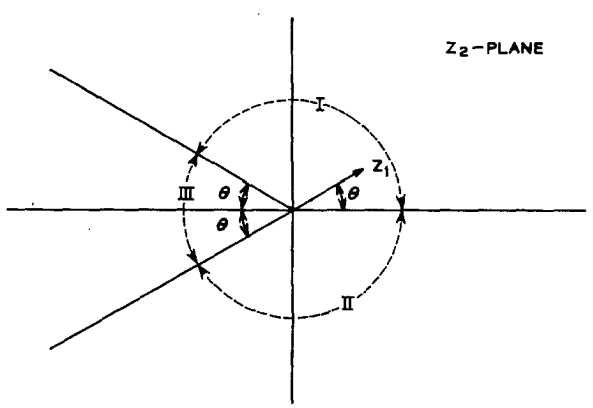


FIG. 1. Three sectors in the z_2-plane for fixed z_1. If z_2 is chosen in Sector III, then D_{KW} is the entire cut z_3 plane. For z_2 in Sector I or II, D_{KW} is illustrated in Figs. 2 and 3, respectively.

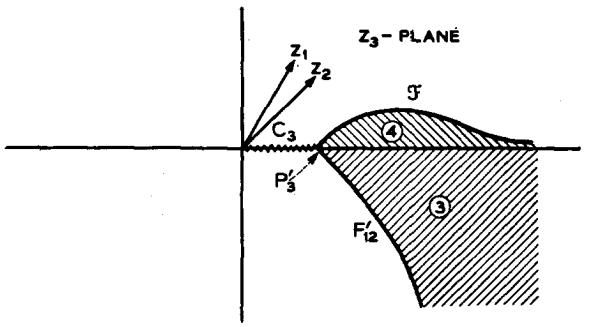


FIG. 2. This figure illustrates the KW domain (exterior shaded) for z_2 in Sector I of Fig. 1. (The wiggly line is to be understood as shading.)

in the following way:

$$\begin{aligned} &\text{Region } (i): \quad y_iy_j<0, \quad y_iy_k<0 \\ &\text{Region } (4): \quad y_1y_2>0, \quad y_1y_3>0 \\ &\qquad z_i=x_i+iy_i \\ &\qquad i=1,\,2,\,3. \end{aligned} \tag{1.30}$$

The relevant portions of F_{jk}' and $\mathfrak{F}$ are exactly those portions in Region (i) and Region (4), respectively.

Spectral Conditions

If the threshold masses [see (1.5)] are all zero, then the only restrictions[7] on $W(z)$ and $H(z)$ are analyticity in the KW domain and certain boundedness requirements[8] of polynomial type. If one or more of the threshold masses is nonzero, then $W(z)$ is further restricted by (1.3), which implies that $H(z)$ is restricted by (1.24).[9]

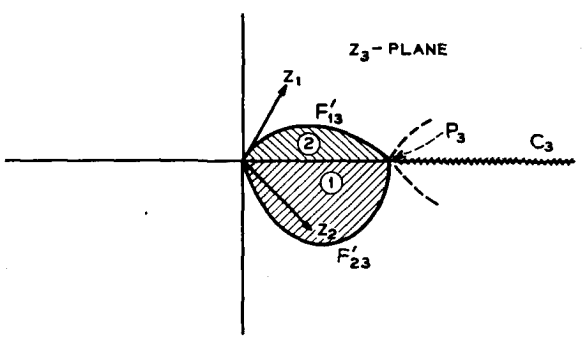


FIG. 3. This figure illustrates the KW domain (exterior shaded) for z_2 in Sector II of Fig. 1.

Using the edge-of-the-wedge theorem[10] and the strong continuity theorem,[11] it can be shown that (1.24) holds if and only if $H(z)$ is analytic at all points satisfying[12–14]

$$\begin{aligned} &z_1 \text{ not positive real,} \\ &z_2 \text{ not positive real,} \\ &0\leq z_k<M_k^2 \text{ (real),} \\ &(z_1,z_2,z_3)\in\bar{D}_{\mathrm{KW}}. \end{aligned} \tag{1.31}$$

This result is illustrated in Fig. 4. $H(z)$ is analytic not only in D_{KW} but also at the points indicated by the dashed lines.

Our problem is to find the holomorphy envelope of the enlarged domain which is obtained by adding the points (1.31), for $k=1,\,2,\,3$, to D_{KW}. The exterior of this domain is the union of the cuts C_i (now defined as $z_i\geq M_i^2$) and the singularity solid. Unfortunately there is no known practicable procedure for finding holomorphy envelopes, but once found or guessed their verification is relatively straightforward (see KW and I).

2. IMPLICATIONS OF THE TRIANGLE DIAGRAM

In this section we shall deduce an upper bound[15] on the holomorphy envelope by studying the analytic function which is associated with the momentum-space triangle diagram[16] (see Fig. 5). This function, which depends on the squared internal masses a_i $(i=1,\,2,\,3)$, is given by

$$H(a,z)=\int_0^1\int_0^1\int_0^1 \frac{d\alpha_1d\alpha_2d\alpha_3\delta(1-\sum_{i=1}^{3}\alpha_i)}{\sum_{i=1}^{3}[z_i\alpha_j\alpha_k-a_i\alpha_i]}, \tag{2.1}$$

where the integration variables α_i are the usual one-dimensional Feynman parameters. By the general theory, H is represented by this formula in the domain U, and has an analytic continuation throughout D_{KW}. Actually it is easily shown that H, considered as a function of six complex variables, is analytic everywhere

[7] The unitarity of the S matrix and the positive definiteness of the metric in Hilbert space both imply restrictions which mix the vacuum expectation values of different orders. It is conceivable that they imply some restriction on the three point functions alone, but this does not seem likely.

[8] These are equivalent to the temperateness of the distribution boundary values. See L. Schwartz, *Transformations de Laplace des distributions*, Seminar Math. Lund. (Sweden), 196 (1952), and J. McKenna, Thesis (Princeton, 1960).

[9] It is easily shown that (1.3) is equivalent to the four-term identity

$\tilde{r}_i+\tilde{a}_i-\tilde{r}_j-\tilde{a}_j=0$; unless $p_k\in V_\pm{}^k$ and $[p_i\in V_\mp{}^i$ or $p_j\in V_\mp{}^j]$.

If the triangle inequalities, $M_i\leq M_j+M_k$, $i=1,\,2,\,3$, are satisfied, this is equivalent to (1.24). Otherwise this contains additional information which does not imply any further enlargement of the domain of analyticity of $H(z)$.

[10] For a simple proof and references to earlier work see F. J. Dyson, Phys. Rev. **110**, 580 (1958).

[11] This theorem is a generalization to n variables (see I) of a theorem of Bremermann which is proved in Math. Ann. **127**, 406 (1954).

[12] $\bar{D}$ means the closure of D.

[13] This result was first stated by R. Jost, Helv. Phys. Acta **31**, 263 (1958). A rigorous proof is given in I.

[14] It is shown by Källén and Toll, Helv. Phys. Acta **33**, 753 (1960), that (1.3) does not imply any enlargement of the domain of analyticity of $W(z)$.

[15] Our upper bound E' will be schlicht, but it is conceivable that the holomorphy envelope E may not be. If E is schlicht, it is a subdomain of E'. Otherwise E lies over a subdomain of E'. The meaning of this is explained in the review by A. S. Wightman cited in reference 5.

[16] It is worth noting that H can also be obtained as the *exact* momentum-space vertex function corresponding to the three local scalar fields, $A_i=\varphi_j\varphi_k$, where φ_i is a free field of squared mass a_i.

except possibly on the manifolds

$$z_i=[(a_j)^{\frac{1}{2}}\pm(a_k)^{\frac{1}{2}}]^2;\quad i=1,\ 2,\ 3$$
$$\Phi(a,z)=0 \tag{2.2}$$
$$a_i=0;\quad i=1,\ 2,\ 3,$$

where

$$\Phi(a,z)=\sum_{i=1}^{3}[a_i^2z_i+a_ja_k(z_i-z_j-z_k)+a_iz_i(z_i-z_j-z_k)]+z_1z_2z_3. \tag{2.3}$$

If the α integrations were carried out explicitly, they would lead to Spence functions, which have singularities of logarithmic type. Therefore we must introduce branch cuts if we want H to be single valued. These must be located in such a way that H is represented[17] by (2.1) throughout the domain U. If z is in Region (4) [see (1.30)], then the denominator of (2.1) never vanishes, so the integral itself continues H from U to all points of Region (4). If z_i is real while z_j and z_k are both in the same half-plane, then the denominator can only vanish if $\alpha_i=0$ and

$$z_i\geq[(a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}}]^2. \tag{2.4}$$

We conclude that the singularity $z_i=[(a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}}]^2$ is a branch point whose branch cut must be given by (2.4)

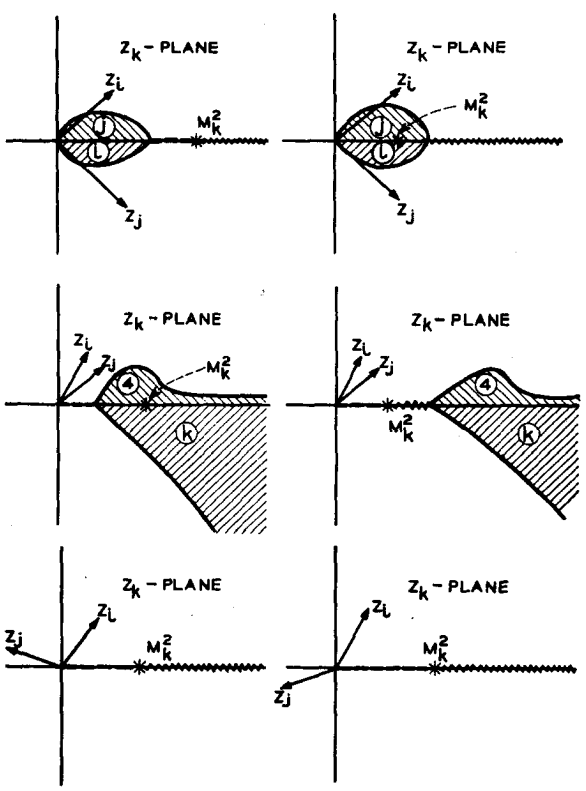


FIG. 4. In this figure the dashed lines indicate the points (1.31) to which $H(z)$ can be continued when $M_k^2\neq 0$.

[17] From now on we consider H as a function of three complex variables z_i with fixed positive real a_i.

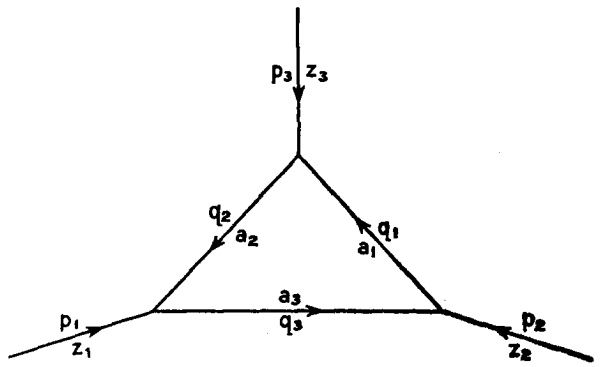


FIG. 5. The triangle diagram.

if it is to be independent of z_j and z_k and outside of D_{KW}. It can be interpreted as the "threshold" for the creation of a pair of particles with masses $(a_j)^{\frac{1}{2}}$ and $(a_k)^{\frac{1}{2}}$, respectively. We also conclude that the singularity $z_i=[(a_j)^{\frac{1}{2}}-(a_k)^{\frac{1}{2}}]^2$ does not appear on the physical sheet.

Let us now turn our attention to the manifold, $\Phi(a,z)=0$, which we shall call the Φ manifold. Solving for z_i we find

$$(\Phi=0):\quad z_i=a_j+a_k-(1/2a_i)\times[(z_j-a_i-a_k)(z_k-a_i-a_j)\pm(R_j)^{\frac{1}{2}}(R_k)^{\frac{1}{2}}], \tag{2.5}$$

where

$$R_i=\lambda(z_i,a_j,a_k), \tag{2.6}$$

with

$$\lambda(a,b,c)=a^2+b^2+c^2-2ab-2ac-2bc. \tag{2.7}$$

On the physical sheet the Φ manifold is never relevant in Region (4), and it is relevant in Region (i) if and only if (2.5) holds with the (+) sign and with each of the square roots in the upper half-plane.

In summary, $H(a,z)$ is analytic in the product of the cut planes, with cuts given by (2.4), except for a logarithmic singularity on the portion of the Φ manifold which we have just described.[18] Furthermore, it satisfies the spectral conditions for thresholds

$$M_i=(a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}};\quad i=1,\ 2,\ 3, \tag{2.8}$$

as is apparent from Fig. 4.[19]

For a given set of threshold masses, it follows that the function

$$\hat{H}[\varphi(a),z]=\int_0^\infty d^3a\,\varphi(a)H(a,z) \tag{2.9}$$

is analytic in the KW domain and satisfies the spectral conditions provided φ is chosen so that

$$\varphi(a_1,a_2,a_3)=0;\ \text{unless}\ (a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}}\geq M_i,\ i=1,\ 2,\ 3 \tag{2.10}$$

[18] The branch cut which connects this portion of the Φ manifold to infinity may be located wherever is most convenient provided it does not pass into D_{KW} or Region (4).

[19] The relevant portion of the Φ manifold is contained in the singularity solid.

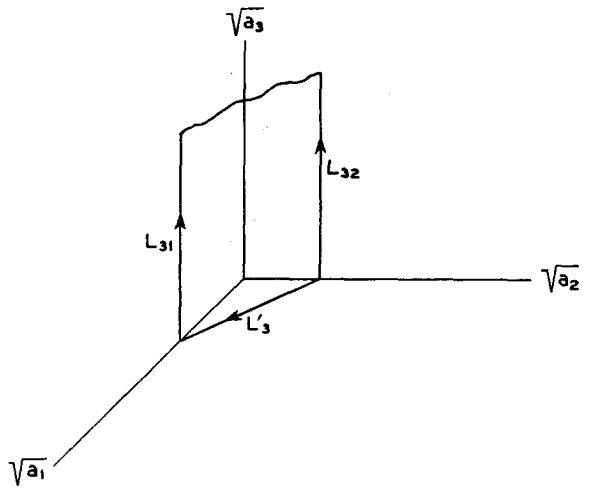


FIG. 6. The support of $\varphi(a)$ in the single-threshold case, $M_3 \neq 0$.

and so that the integral is sufficiently convergent. Clearly this is not a representation of the most general function analytic in the KW domain and satisfying the spectral conditions, since it has no singularities in Region (4) even in the case of zero threshold masses. In this case, however, $\hat{H}$ can have singularities anywhere outside of the KW domain except in Region (4), and it is this fact which enabled Källén and Wightman to guess (and then verify) that F_{jk}' is the boundary of D_{KW} in Region (i).

In the general case, (2.10) restricts the region of possible singularities and provides an upper bound on the holomorphy envelopes. In the single-threshold case $M_i \neq 0, M_j = M_k = 0$, this upper bound coincides with the holomorphy envelope in Regions (j) and (k), but not in Region (i). In all cases with two or more nonzero threshold masses there is no coincidence.

The Single-Threshold Case

Let us consider the single-threshold case, $M_3 \equiv M > 0$, $M_1 = M_2 = 0$. Then the support region of $\varphi(a)$ is given by

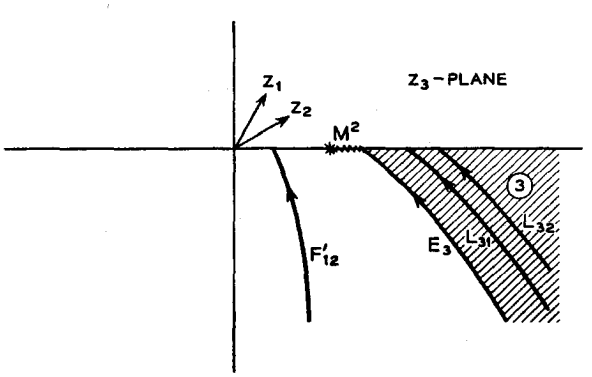


FIG. 7. The shading indicates the possible singularities of $\hat{H}$ in Region (3) in the single-threshold case, $M_3 \neq 0$.

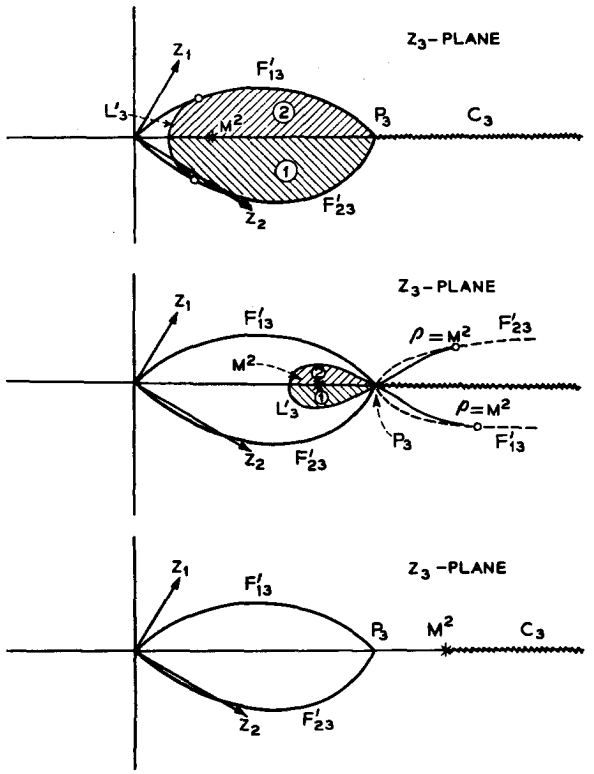


FIG. 8. The shading indicates the possible singularities of $\hat{H}$ in Regions (1) and (2) in the single threshold case, $M_3 \neq 0$, for three different values of $M_3 \equiv M$. It is shown in Sec. 3 that the holomorphy envelope of D_{3+-} coincides with the unshaded region.

$$a_i \geq 0; \quad i = 1, 2, 3$$
$$(a_1)^{\frac{1}{2}} + (a_2)^{\frac{1}{2}} \geq M \tag{2.11}$$

and is illustrated in Fig. 6.

First let us determine the possible singularities of $\hat{H}$ in Region (3). Let z_1 and z_2 be fixed in the upper half-plane. Then (2.5) [with $i=3$ and with the (+) sign] defines a many-one mapping from a space into the z_3 plane. If the image point, corresponding to a given point (a_1, a_2, a_3) in the support region of φ, is in the

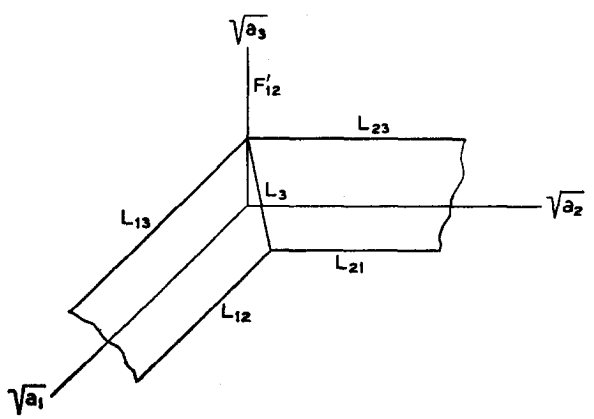


FIG. 9. The support of $\varphi(a)$ in the double-threshold case $M_1 = M_2 \neq 0$.

lower half-plane, then it is in Region (3) and is a possible singularity of $\hat{H}$. If it is in the upper half-plane, then it is in Region (4) and is irrelevant. We want to find the image of the support region [Eq. (2.11)]. In the limit as M approaches zero, the plane $(a_1)^{\frac{1}{2}}+(a_2)^{\frac{1}{2}}=M$ approaches the a_3 axis whose image is F_{12}'. For $M>0$ the boundary of the image region is an envelope E_3, which is the image of some curve in the plane $(a_1)^{\frac{1}{2}}+(a_2)^{\frac{1}{2}}=M$, whose equation depends on z_1 and z_2. It can be shown that E_3 crosses the real axis above threshold as illustrated in Fig. 7, or else lies entirely in the upper half-plane. It is fairly easy to display E_3 as a quasi-analytic hypersurface of rank one and of analytic type. Furthermore, it can be shown that the rank is no more than one; i.e., that E_3 is not an analytic hypersurface.

To find the possible singularities of $\hat{H}$ in Regions (1) and (2) we carry out similar procedures in the z_1 plane and in the z_2 plane. The portion of F_{23}' in Region (1) with $0\leq\rho<M^2$ and the portion of F_{13}' in Region (2) with $0\leq\rho<M^2$ are replaced by the analytic hypersurface L_3' as illustrated in Fig. 8. The pre-image of L_3' is indicated in Fig. 6. Thus L_3' is given by

$$L_3': \begin{cases} \Phi(a,z)=0 \\ a_1=m^2 \\ a_2=(M-m)^2 \\ a_3=0 \\ 0\leq m\leq M. \end{cases} \tag{2.12}$$

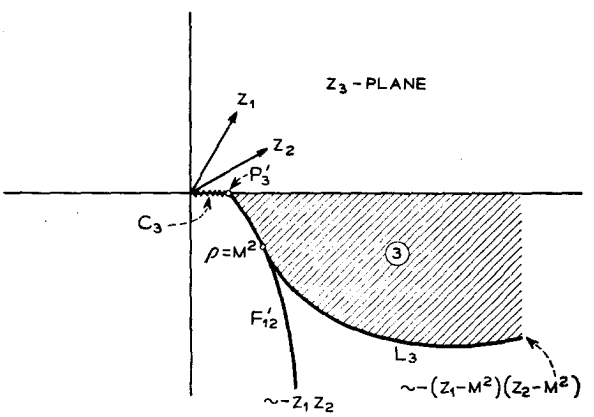


FIG. 10. The shading indicates the possible singularities of $\hat{H}$ in Region (3) in the double-threshold case, $M_1=M_2\neq 0$.

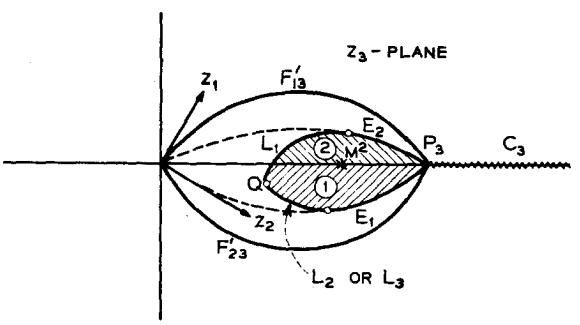


FIG. 11. The shading indicates the possible singularities of $\hat{H}$ in Regions (1) and (2) in the double-threshold case, $M_1=M_2\neq 0$.

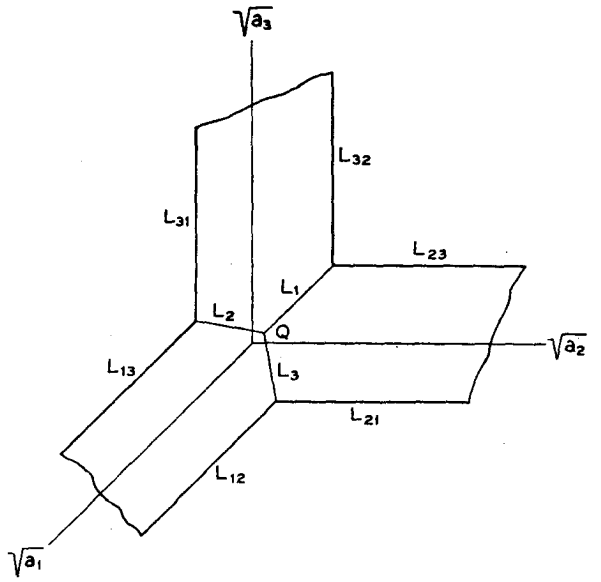


FIG. 12. The support of $\varphi(a)$ in the triple-threshold case, $M_1=M_2=M_3$.

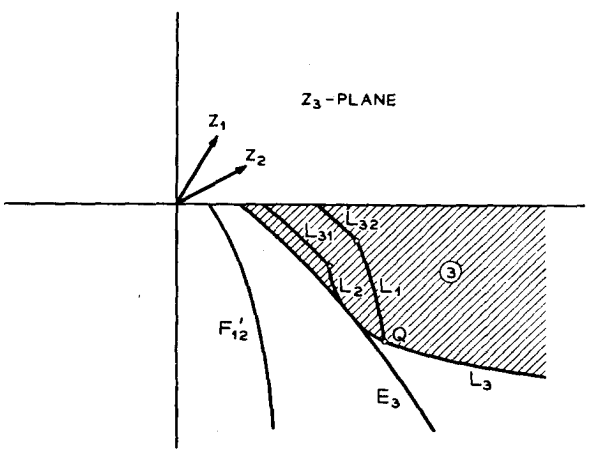


FIG. 13. The shading indicates the possible singularities of $\hat{H}$ in Region (3) in the triple-threshold case, $M_1=M_2=M_3$.

As we shall see in the next section, this is the boundary of the holomorphy envelope in Regions (1) and (2), but E_3 is not the boundary in Region (3).

Other Cases

In the "double-threshold case" $M_1=M_2\equiv M>0$, $M_3=0$, the support region of φ is illustrated in Fig. 9, and the region of possible singularities is illustrated in Figs. 10 and 11.

In the "triple-threshold case" $M_1=M_2=M_3\equiv M>0$, the support region of φ is illustrated in Fig. 12 and the region of possible singularities is illustrated in Figs. 13 and 14.

In every case the most complicated parts of the boundary of the singular region are the envelopes. Now

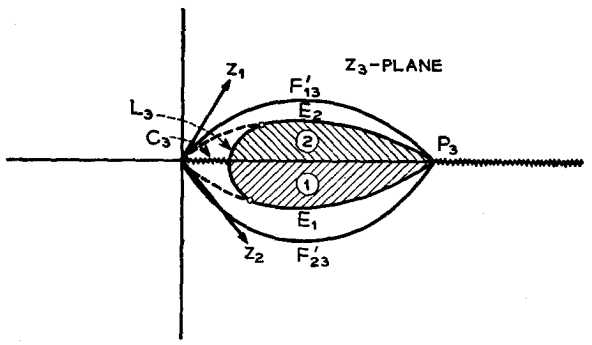


FIG. 14. The shading indicates the possible singularities of $\hat{H}$ in Regions (1) and (2) in the triple threshold case, $M_1=M_2=M_3$.

E_i appears already in the single-threshold case, and it depends on M_i alone. We therefore conjecture that whatever hypersurface replaces F_{jk}' in the single-threshold case $M_i\neq 0$ will be involved in the boundary of the holomorphy envelope in the general case, except that a portion of it near $\rho=0$ will be irrelevant.

Nambu's Theorem

For real z_1 and z_2 with $0\leq z_1\leq M_1^2$ and $0\leq z_2\leq M_2^2$, the function $H(a,z)$ is analytic in the cut z_3 plane. The cut may begin at the normal threshold, $z_3=M_3^2$ or at an "anomalous threshold" $0<z_3<M^2$, depending on z_1 and z_2. This can be seen by studying the limiting form of Fig. 10 (or its analog with $M_1\neq M_2$) as z_1 and z_2 approach the real axis. Nambu has shown[20] that the momentum-space vertex function has this analyticity property to all orders in perturbation theory, but Jost's example (see Sec. 4) shows that it does not have it in general. The failure may be due to the special interactions which are assumed in perturbation theory, or perhaps to nonconvergence of the perturbation series.

3. THE SINGLE-THRESHOLD PROBLEM ($M_3\neq 0$)

In this section we shall discuss the single-threshold problem, $M_3\equiv M>0$, $M_1=M_2=0$. The problem is to find the holomorphy envelope of the domain D_3 whose *exterior* is the union of the singularity solid and the three cuts

$$\begin{aligned} C_1&: \ z_1\geq 0\\ C_2&: \ z_2\geq 0\\ C_3&: \ z_3\geq M^2. \end{aligned} \tag{3.1}$$

Consider the subdomains

$$\begin{aligned} D_{3++}&=D_3\cap(y_1>0)\cap(y_2>0)\\ D_{3+-}&=D_3\cap(y_1>0)\cap(y_2<0), \end{aligned} \tag{3.2}$$

where $z_i=x_i+y_i$. It is shown in I that the holomorphy envelope of D_3 is the union[21] of the holomorphy envelopes of D_{3++} and D_{3+-} and their conjugate domains D_{3--} and D_{3-+}, together with the negative extensions of C_1 and C_2.[22] It is sufficient to find the holomorphy envelopes of D_{3+-} and D_{3++}.[23] For reasons which will become apparent we shall refer to these as the P_3 problem and the P_3' problem, respectively.

The P_3 Problem

First let us consider the P_3 problem. The domain D_{3+-} is bounded by the analytic hypersurfaces

$$\begin{aligned} \tilde{C}_1&: \ y_1=0\\ \tilde{C}_2&: \ y_2=0\\ C_3&: \ z_3\geq M^2\\ F_{13}'&: \ \rho^2+\rho(z_2-z_1-z_3)+z_1z_3=0\\ F_{23}'&: \ \rho^2+\rho(z_1-z_2-z_3)+z_2z_3=0\\ &\rho\geq 0, \end{aligned} \tag{3.3}$$

and is illustrated in Fig. 15. It is locally a holomorphy domain at every boundary point except along the portion of the edge P_3 which lies below threshold.[24] By the Kantensatz[25] a completion[26] through this "indented edge" is possible. In fact by an almost trivial application of the completion technique used in KW and in I, the "bulb" in the z_3 plane can be "removed" whenever the point P_3 (which depends on z_1 and z_2) lies below threshold as in Fig. 15(b). This partial completion yields an enlarged domain whose boundary involves a new analytic hypersurface B_3 given by

$$B_3: \ (M^2-\rho)z_1+\rho z_2=\rho(M^2-\rho); \quad \rho\geq 0, \tag{3.4}$$

the indented portion of the edge P_3 has now been replaced by two new indented edges, $F_{13}'\cap B_3$ and $F_{23}'\cap B_3$, through which further completion is possible.

In the preceding section we found an upper bound on the holomorphy envelope involving a new analytic hypersurface L_3' [see Eq. (2.12) and Fig. 8], and we conjectured that this upper bound would in fact coincide with the holomorphy envelope. An appropriate change of variables and a further application of the completion technique[27] verifies this conjecture. We find that the

[20] Y. Nambu, Nuovo cimento **6**, 1064 (1957); see also K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).

[21] In the event that one or more of the latter holomorphy envelopes is nonschlicht, the definition of this union requires some discussion, which is given in detail in I.

[22] By the negative extension of a cut, $z\geq 0$, we mean the hypersurface $z<0$.

[23] The holomorphy envelopes of D_{3-+} and D_{3--} are the complex conjugates of the holomorphy envelopes of D_{3+-} and D_{3++} respectively. This follows from the remark that for every function $f(z)$ analytic in a given domain D the function $f(z^*)^*$ is analytic in D^*.

[24] See Fig. 15(b). By definition P_3 is the four-real-dimensional intersection $F_{13}'\cap F_{23}'$. The intersection of P_3 with the z_3 plane for fixed z_1 and z_2 is always a single point on the real axis.

[25] This theorem was first proved by H. Kneser in Math. Ann. **106**, 656 (1932). An exposition of it is given in the review by A. S. Wightman cited in reference 5.

[26] We use the terminology of Bochner and Martin (see reference 5). Analytic continuation refers to functions and analytic completion to domains.

[27] The method is elementary, purely geometrical, and completely rigorous. The proof in this case (see Chapter 9 of I) reflects the complexity of the geometry.

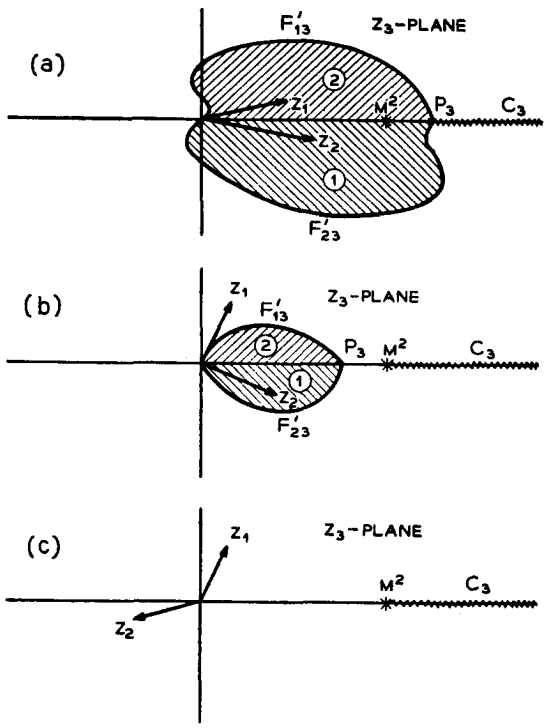


FIG. 15. This figure illustrates the domain D_{3+-} (exterior shaded) in the z_3 plane for several choices of z_1 and z_2.

holomorphy envelope of D_{3+-} is the domain bounded by the analytic hypersurfaces

$$\begin{aligned} \tilde{C}_1&: \ y_1=0 \\ \tilde{C}_2&: \ y_2=0 \\ C_3&: \ z_3=\rho \\ F_{13}'&: \ \rho^2+\rho(z_2-z_1-z_3)+z_1z_3=0 \\ F_{23}'&: \ \rho^2+\rho(z_1-z_2-z_3)+z_2z_3=0 \\ &\rho\geq M^2 \end{aligned} \tag{3.5}$$

and

$$L_3'\colon \begin{cases} \Phi(a,z)=0 \\ a_1=m^2 \\ a_2=(M-m)^2 \\ a_3=0 \\ 0\leq m\leq M, \end{cases} \tag{3.6}$$

and illustrated in Fig. 8.

Clearly the holomorphy envelope of D_3 in Regions (1) and (2) is also bounded by the hypersurfaces (3.5) and (3.6) and illustrated in Fig. 8. This follows from the theorem stated below (3.2). Alternatively, since D_{3+-} is contained in D_3, its holomorphy envelope gives a lower bound on the holomorphy envelope of D_3, which coincides [in Regions (1) and (2)] with the upper bound obtained in Sec. 2.

The formula for L_3' can be solved for z_3 with the result

$$L_3'\colon\ z_3=-\frac{[m^2z_1-(M-m)^2z_2][m^2-(M-m)^2+z_1-z_2]}{[m^2-z_2][(M-m)^2-z_1]}; \qquad 0\leq m\leq M. \tag{3.7}$$

The following theorems provide some insight into the geometry of Fig. 8.

Theorem 3.1. The manifolds $m=0$ of L_3' and $\rho=M^2$ of F_{13}' are identical.

Theorem 3.2. The manifolds $m=M$ of L_3' and $\rho=M^2$ of F_{23}' are identical.

Theorem 3.3. In the z_3 plane, L_3' is tangent to F_{13}' at the point $\rho=M^2$ of F_{13}'.

Theorem 3.4. In the z_3 plane, L_3' is tangent to F_{23}' at the point $\rho=M^2$ of F_{23}'.

Theorem 3.5. In the z_3 plane, if P_3 is above threshold, L_3' passes through it twice. If P_3 is below threshold, L_3' does not intersect it.

Theorem 3.6. L_3' does not intersect F_{13}' in the closure of Region (2) or F_{23}' in the closure of Region (1) except at the points cited in the preceding theorems.

Theorem 3.7. L_3' never intersects $\tilde{C}_3$ outside of the interval $0<z_3<M^2$ except at P_3.

Theorem 3.8. Let

$$\begin{aligned} z_1&=x_1+i\epsilon_1 \\ z_2&=x_2-i\epsilon_2 \\ z_3&=z_3^{(0)}, \end{aligned} \tag{3.8}$$

where x_1, x_2, ϵ_1, ϵ_2 are positive and $z_3^{(0)}$ is in the cut plane (with cut $z_3\geq M^2$). If $(x_1)^{\frac{1}{2}}+(x_2)^{\frac{1}{2}}<M$, then the point (z_1,z_2,z_3) is in the holomorphy envelope for all sufficiently small ϵ_1 and ϵ_2. If $(x_1)^{\frac{1}{2}}+(x_2)^{\frac{1}{2}}>M$, then the point (z_1,z_2,z_3) is outside of the holomorphy envelope for all sufficiently small ϵ_1 and ϵ_2.

The Intersection $C_1\bigcap C_2\bigcap L_3'$

It is clear from Theorem 3.8 that the intersection $C_1\bigcap C_2\bigcap L_3'$ is given by

$$\begin{aligned} z_1&\geq 0 \\ z_2&\geq 0 \\ (z_1)^{\frac{1}{2}}+(z_2)^{\frac{1}{2}}&=M, \end{aligned} \tag{3.9}$$

which can be rewritten in the form

$$\begin{aligned} z_1&=(M-m)^2 \\ z_2&=m^2 \\ 0&\leq m\leq M. \end{aligned} \tag{3.10}$$

By definition this intersection belongs to the distinguished boundary of the holomorphy envelope of D_3, but only its "boundary points" [z_3 real and $\lambda(z_1,z_2,z_3)\geq 0$] belong to the boundary of D_3. This is somewhat surprising, since ordinarily the distinguished boundary of a domain is the same as the distinguished boundary of its

holomorphy envelope.[28] However, it is the Silov boundary that is *always* conserved, and since the intersection $C_1 \cap C_2 \cap L_3'$ is a one-parameter family of one-complex-dimensional analytic manifolds on which the maximum principle holds, its "interior points" cannot be in the Silov boundary of the holomorphy envelope even though they are in its distinguished boundary.[29]

The P_3' Problem

Next let us consider the P_3' problem. This problem has not been solved, but several interesting results have been obtained. The problem is to find the holomorphy envelope of the domain D_{3++} which is bounded by the analytic hypersurfaces

$$\begin{aligned} \bar{C}_1&: \ y_1=0 \\ \bar{C}_2&: \ y_2=0 \\ C_3&: \ z_3 \geq M^2 \\ F_{12}'&: \ \rho^2+\rho(z_3-z_1-z_2)+z_1z_2=0 \\ \mathfrak{F}&: \ \rho^2-\rho(z_1+z_2+z_3)+(z_1z_2+z_1z_3+z_2z_3)=0 \\ &\qquad \rho \geq 0 \end{aligned} \tag{3.11}$$

and is illustrated in Fig. 16. It is locally a holomorphy domain at every boundary point except along the portion of the edge P_3' which lies below threshold.[30] By the Kantensatz a completion through this "indented edge" is possible. By the strong continuity theorem (see footnote 11) it can be shown that every point of F_{12}' in Region (3) and every point of $\mathfrak{F}$ in Region (4), as well as every point of this indented edge, is in the holomorphy envelope of D_{3++}.

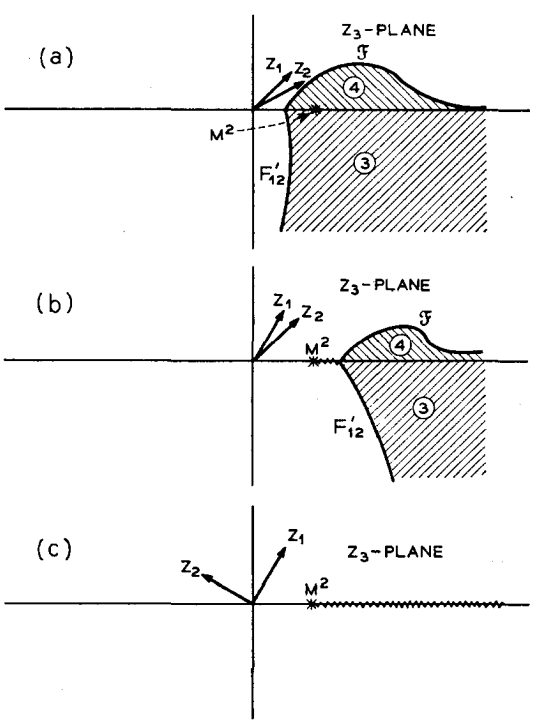


FIG. 16. This figure illustrates the domain D_{3++} (exterior shaded) in the z_3 plane for several choices of z_1 and z_2.

We have already found an upper bound on the holomorphy envelope in Region (3); namely, the quasi-analytic hypersurface E_3 (see Figs. 6 and 7). A number of comparable upper bounds can be obtained by purely geometrical arguments. The simplest of these in Region (3) is the analytic hypersurface $F_{12}'(M^2)$ defined by

$$F_{12}'(M^2): \ \rho^2+\rho(z_3-M^2-z_1-z_2)+z_1z_2=0, \ \rho \geq 0, \tag{3.12}$$

which is obtained by translating F_{12}' to the right through the distance M^2 in the z_3 plane. This is illustrated in Fig. 17. In the special case $z_1=z_2 \equiv z$, E_3 can be obtained from $F_{12}'(M^2)$ by the substitution $z \rightarrow z-\frac{1}{4}M^2$. Therefore E_3 is a weaker upper bound than $F_{12}'(M^2)$ in this case[31] (see Fig. 18). This means that E_3 is not the boundary of the holomorphy envelope in Region (3). In the limit as z_1 and z_2 approach the real axis in Fig. 17 the entire lower half-plane is outside of the upper bound $F_{12}'(M^2)$. It follows with the aid of a rather complicated Silov boundary argument that the entire upper half-plane is also outside of the holomorphy envelope in this limit. In other words every point of $C_1 \cap C_2$, except for the "physical real points" (see footnote 46) has a full complex neighborhood in the *exterior* of the holomorphy envelope of D_{3++}.

There is a simple partial completion which can be performed with the aid of Ruelle's lemma.[3] This leads to two new analytic hypersurfaces, G_1 and G_2, given by

$$G_1: \begin{cases} \rho^2[M^4+M^2(z_1-z_2-z_3)+z_2z_3] \\ -\rho z_1[2M^2+(z_1-z_2-z_3)]+z_1^2=0 \\ 0 \leq \rho \leq 1 \end{cases} \tag{3.13}$$

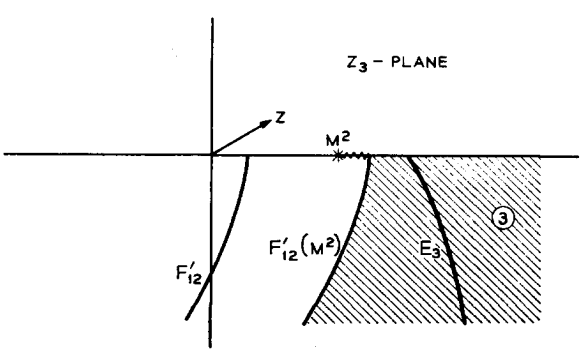


FIG. 18. This figure illustrates the upper bounds $F_{12}'(M^2)$ and E_3 in the special case $z_1=z_2 \equiv z$.

[28] The author is indebted to Dr. H. Rossi for several helpful discussions on this and related topics.

[29] For further discussion and references see I.

[30] See Fig. 16(a). By definition P_3' is the four-real-dimensional intersection $F_{12}' \cap \mathfrak{F}$. The intersection of P_3' with the z_3 plane for fixed z_1 and z_2 is always a single point on the real axis.

[31] In some other special cases the opposite is true.

and

$$G_2: \begin{cases} \rho^2[M^4+M^2(z_2-z_1-z_3)+z_1z_3] \\ -\rho z_2[2M^2+(z_2-z_1-z_3)]+z_2^2=0 \\ 0\leq\rho\leq 1, \end{cases} \quad (3.14)$$

respectively. The reader will note that both of these hypersurfaces end (when $\rho=1$) on the manifold $\rho=M^2$ of $\mathfrak{F}$. Clearly they provide a lower bound on the holomorphy envelope, but it is a very weak one.

The hypersurfaces (3.11) which bound D_{3++} intersect in the following edges, among others:

$$\begin{aligned} P_1&=C_1\cap F_{12}' \\ P_2&=C_2\cap F_{12}' \\ P_1'&=C_1\cap\mathfrak{F} \\ P_2'&=C_2\cap\mathfrak{F} \\ P_3'&=C_3\cap F_{12}'\cap\mathfrak{F}. \end{aligned} \quad (3.15)$$

We have already seen that the portion of P_3' which lies below threshold is in the holomorphy envelope. Using an upper bound constructed especially for the purpose it can be shown that the portion of P_3' which lies above threshold is an edge in the boundary of the holomorphy envelopes. In fact, when z_1 and z_2 are in the same half-plane, every point in the singularity solid and sufficiently close to C_3 above threshold is in the exterior of the holomorphy envelope. This result is illustrated in Figs. 19 and 20. Similarly it can be shown that certain portions of the edges, P_1, P_2, P_1', P_2', are also edges in the boundary of the holomorphy envelope.

The hypersurface (hypersurfaces) which replaces (replace) F_{12}' and $\mathfrak{F}$ is (are) severely restricted by the requirement that it (they) contain the required portions of these edges. If it is assumed that the holomorphy envelope is bounded piecewise by analytic hypersurfaces of analytic type,[32] then at least three are needed to replace F_{12}' and at least three more to replace $\mathfrak{F}$. The idea of such a complicated boundary is extremely uninviting. Furthermore it is difficult to imagine the Silov boundary of such a domain being the same as the Silov boundary of D_{3++}.

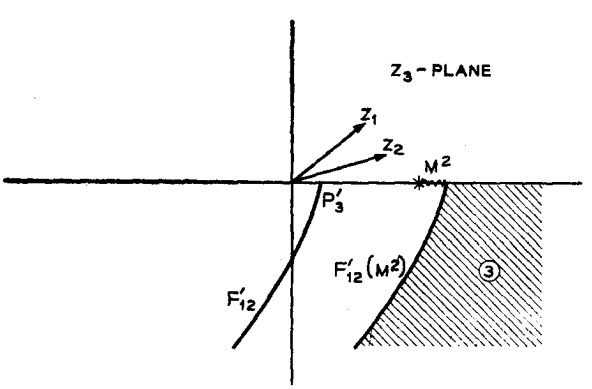


FIG. 17. This figure illustrates the upper bound $F_{12}'(M^2)$, which is obtained by translating F_{12}' to the right through the distance M^2.

[32] The hypotheses of Theorem 12.5 in I are slightly stronger than this clause would require but the differences do not appear to be significant.

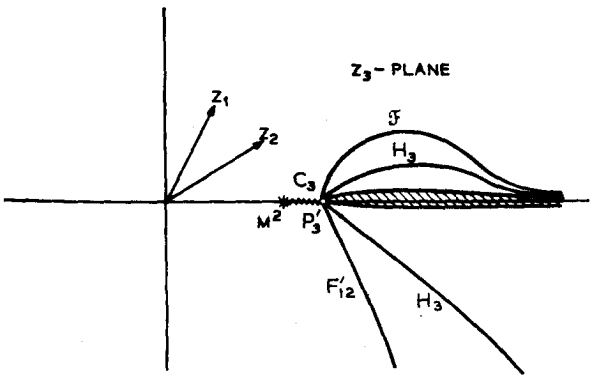


FIG. 19. When P_3' is above threshold, the boundary, H_3, of the holomorphy envelope of D_{3++} must pass through it as illustrated in this figure. The shaded strip indicates known exterior points.

If the boundary of the holomorphy envelope is assumed to be piecewise four-times continously differentiable, then it must be piecewise analytic or quasi-analytic. Since it is probably not piecewise analytic, the simplest remaining possibility is that F_{12}' and $\mathfrak{F}$ are replaced by a *single*[33] quasi-analytic hypersurface H_3, which reduces to the product of F_{12}' and $\mathfrak{F}$ when M is set equal to zero. The qualitative form of H_3 is illustrated in Figs. 19 and 20. In the borderline case where P_3 and M^2 coincide, the ρ values on the two branches of H_3 at P_3' must be the same. For slightly larger M^2, they must turn complex, perhaps because of a change of sign in some square root.[34]

It is reasonable to assume that H_3 can be globally

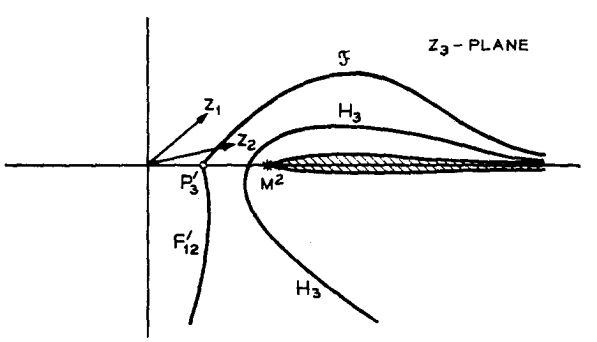


FIG. 20. When P_3' is below threshold, the boundary, H_3, of the holomorphy envelope of D_{3++} must cross the real axis between P_3' and the threshold (by the Kantensatz and Jost's example) as illustrated in this figure. The shaded strip indicates known exterior points.

[33] If there were two, one to replace F_{12}' and one to replace $\mathfrak{F}$, then the intersection of the two in Fig. 20 would almost certainly violate the Kantensatz. For further discussion see Chapter 12 in I.

[34] In the P_3 problem precisely such a phenomenon occurs. As a result L_3' (or its extension obtained by increasing the range of the parameter) contains all of the portion of P_3 with $z_3\geq M^2$ but none of the portion with $z_3<M^2$.

defined in the form

$$H_3: \begin{cases} f(z,\rho,M^2)=0 \\ g(z,\rho,M^2)=0 \\ \qquad\qquad \rho\in R^3 \\ \qquad\qquad z\in C^3, \end{cases} \tag{3.16}$$

where the analytic functions f and g are as yet unknown.

The problem of finding f and g, given that H_3 contains the required portions of the edges (3.15), might be considered as a boundary value problem, but no serious attempt has yet been made to explore this approach. A more helpful approach, suggested by Källén and Toll, is discussed briefly in the following section.

4. OTHER TOPICS

In this section we shall discuss the Källén-Toll representation, some perturbation theory examples, a generalization of Jost's example, the double threshold problem, and the outlook for further progress.

The Källén-Toll Representation

Using the Bergmann-Weil formula, Källén and Toll have shown[35] that every function $W(z)$ analytic in the KW domain [and sufficiently bounded at infinity] can be represented in the form

$$W(z)=\sum_{\mu=1}^{4}\int_0^\infty\int_0^\infty\int_0^\infty da_1da_2da_3\Big[\prod_{i=1}^{3}\Delta^{(+)}(a_i,z_i)\Big] \times\Big[\int_0^\infty d\rho\psi_\mu(a,\rho)f_\mu(z,\rho)\Big], \tag{4.1}$$

where the ψ_μ are weight functions and the f_μ are given by

$$f_i(z,\rho)=\frac{\ln(-\rho z_i)-\ln(\rho-z_j)-\ln(\rho-z_k)}{\rho z_i+(\rho-z_j)(\rho-z_k)}; \quad i=1,\,2,\,3 \tag{4.2}$$

$$f_4(z,\rho)=\frac{\sum_{i=1}^{3}[\ln(\rho-z_i)-\ln(-z_i)]}{\rho^2-\rho(z_1+z_2+z_3)+(z_1z_2+z_1z_3+z_2z_3)}$$

with

$$-\pi\leq\mathrm{Im}\,\ln\leq\pi. \tag{4.3}$$

The denominators in (4.2) are the defining polynomials of F_{jk}' and $\mathfrak{F}$, respectively. The numerators are zero on the irrelevant parts of these manifolds, but nonzero on the relevant parts.

Let the bracketed expressions in (4.1) be called A and B, respectively. Both are analytic in the KW domain. Replacing z_i by $(\zeta_j-\zeta_k)^2$, where $\zeta_i=x_i+iy_i$ as in Eqs. (1.6)–(1.8), we see that B is analytic in all six of the tubes T_{ijk} defined in (1.7). Therefore B is the Fourier-Laplace transform of a set of temperature distributions $\tilde{B}_{ijk}$ having the support property (1.3) with zero mass thresholds. An identical argument holds for A except that the $\tilde{A}_{ijk}$ satisfy (1.3) with mass thresholds

$$M_i=(a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}}; \quad i=1,\,2,\,3. \tag{4.4}$$

By the convolution theorem, AB is the Fourier-Laplace transform of the set of convolution products $\tilde{A}_{ijk}{}^{*}\tilde{B}_{ijk}$, and these have the support property (1.3) with mass thresholds (4.4). Therefore if the weights ψ_μ in (4.1) have the support property

$$\psi_\mu=(a_1,a_2,a_3;\rho)=0, \quad \text{unless} \quad (a_j)^{\frac{1}{2}}+(a_k)^{\frac{1}{2}}\geq M_i \quad \text{for} \quad i=1,\,2,\,3, \tag{4.5}$$

then the temperature distributions $\tilde{W}_{ijk}$ satisfy (1.3).

Källén and Toll have conjectured that the converse is true; in other words that (1.3) implies (4.5).[36] If the conjecture is true, then our problem can be solved in full generality by finding the enlargement of the domain of analyticity of H which is implied by the restriction (4.5) on the weight functions ψ_μ in the Källén-Toll representation of W.

Let W_μ be the μth term in the representation (4.1), and let H_μ be the corresponding momentum-space vertex function.[37] It is shown by Källén and Toll that

$$W_i(z)=\int_0^\infty\int_0^\infty d^3ad^3b\,\varphi_i(a,b)M(a,b,z); \quad i=1,\,2,\,3 \tag{4.6}$$

where φ_i is a new weight function

$$\varphi_i(a,b)=k\delta(b_j)\delta(b_k)\int_0^\infty d\rho\psi_i(a,\rho)(\partial/\partial\rho)\bar{\Delta}(b_i,\rho), \tag{4.7}$$

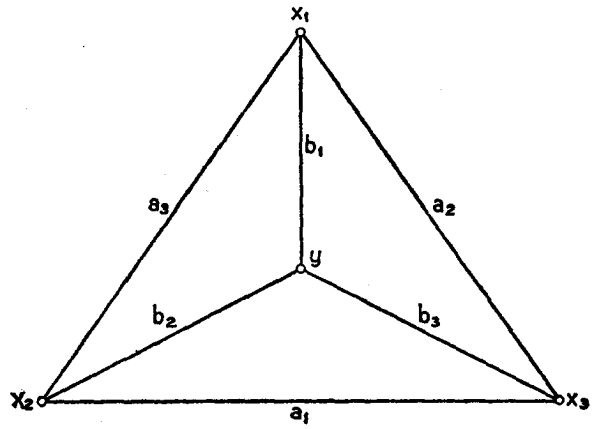


FIG. 21. The position-space Mercedes diagram. The corresponding analytic function is called $M(a,b,z)$.

[35] G. Källén and J. Toll, Helv. Phys. Acta 33, 753 (1960).

[36] For a given W the weights ψ_μ are not unique. The conjecture is that at least one set of weights satisfying (4.5) can be found if the spectral conditions are satisfied.

[37] Clearly $H=\sum_{\mu=1}^{4}H_\mu$, but H_μ is not necessarily the μth term in the Källén-Toll representation of H.

k being a numerical constant, and where M is the analytic function corresponding to the position-space Mercedes diagram, which is illustrated in Fig. 21. It follows that

$$H_i(z)=\int_0^\infty\int_0^\infty d^3a d^3b\,\varphi_i(a,b)M(a,b,z);\quad i=1,2,3 \quad (4.8)$$

where M is the analytic function corresponding to the momentum-space Mercedes diagram (which is illustrated in Fig. 23).

It can be shown[38] that the analytic function corresponding to any momentum-space vertex diagram is analytic throughout Region (4). Therefore[39] $H_i(z)$ is analytic throughout Region (4), for $i=1, 2, 3$, and $H_4(z)$ must account for the singularities of $H(z)$ in Region (4). It follows that $H_4(z)$ cannot be represented as an integral over perturbation theory examples, analogous to (4.8). Finally it can be shown that $H_4(z)$ must account for the singularities of $H(z)$ on the hypersurface H_3 in the P_3' problem. It seems hopeful that the hypersurface H_3 could be found by computing the Fourier transform of the time-ordered boundary value of $W_4(z)$ and investigating the analytic properties of the resulting momentum-space function, $H_4(z)$.

Many physicists feel that the singularities in Region (4) are unphysical because they do not occur in perturbation theory. If this feeling could be supported (for example by an argument from unitarity), then it might turn out that all the singularities allowed by the general theory could be related to perturbation theory diagrams. This hope has led the author to a study, in collaboration with Fairlie, of the analyticity properties of several vertex diagrams, which we shall now discuss.

Some Examples from Perturbation Theory

(By W. S. Brown and D. B. Fairlie)

The three simplest totally symmetric vertex diagrams are the triangle (Fig. 5), the triskelion (Fig. 22), and the Mercedes (Fig. 23). We have already emphasized the importance of the triangle diagram in Sec. 2. Our object here is to determine the singular manifolds for the triskelion and the Mercedes, which are analogous to the Φ manifold for the triangle [see (2.2) and (2.3)]. In applying the Landau conditions[40–42] we shall not consider the "end-point singularities," which lead to the expected threshold manifolds and some other complex manifolds, since these can also be obtained from the lower order diagrams which are "contractions" of the given diagrams.

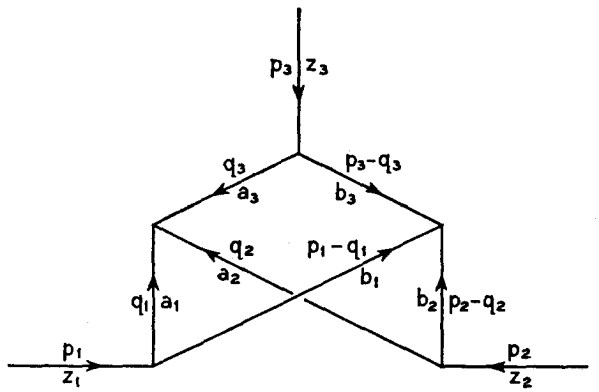


FIG. 22. The triskelion diagram.

For the triskelion diagram the desired singular manifold is given by

$$\Psi(a,b,z)\equiv 4D^2(4D+A^2)+4AB^2(9D+2A^2)-27B^4=0 \quad (4.9)$$

where

$$A=\tfrac{1}{8}\lambda(z+a+b)-\tfrac{1}{4}[\lambda(z)+\lambda(a)+\lambda(b)]$$

$$B=-\tfrac{1}{4}\det(z,a,b)$$

$$D=-\tfrac{1}{8}\sum_{i=1}^{3}\{z_i^2(a_jb_k+a_kb_j)+z_jz_k(a_iB_i+b_iA_i)+z_i[2a_ib_jb_k+a_jb_kB_k+a_kb_jB_j+2b_ia_ja_k+b_ja_kA_k+b_ka_jA_j]\}, \quad (4.10)$$

with

$$A_i=a_i-a_j-a_k,\qquad B_i=b_i-b_j-b_k \quad (4.11)$$

and with $\lambda(z)\equiv\lambda(z_1,z_2,z_3)$ defined by (2.7). It can be shown that Ψ is a homogeneous twelfth-degree polynomial in its nine arguments, and is irreducible over the rationals. Furthermore it is invariant under permutations of the indices, 1, 2, 3, permutation of the vectors, z, a, b, and transposition of the matrix of these vectors.

For the Mercedes diagram we have not succeeded in writing down an explicit formula for the desired singular

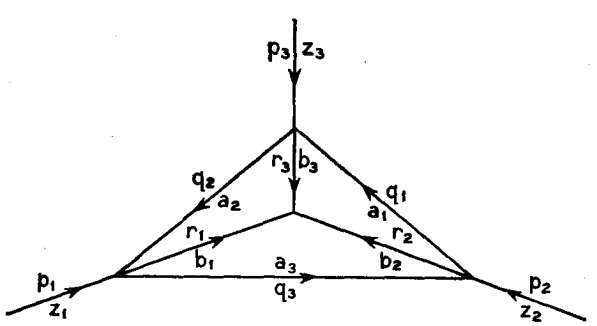


FIG. 23. The Mercedes diagram.

[38] This follows easily from the formula for the analytic function corresponding to an arbitrary diagram, which is given by Symanzik in Progr. Theor. Phys. (Kyoto) **20**, 690 (1958).

[39] Here we proceed formally, assuming that the analyticity is determined by the kernel and not by the integrations. However it is entirely possible that the opposite is the case. For further discussion see I.

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

[41] J. Tarski, J. Math. Phys. **1**, 154 (1960).

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

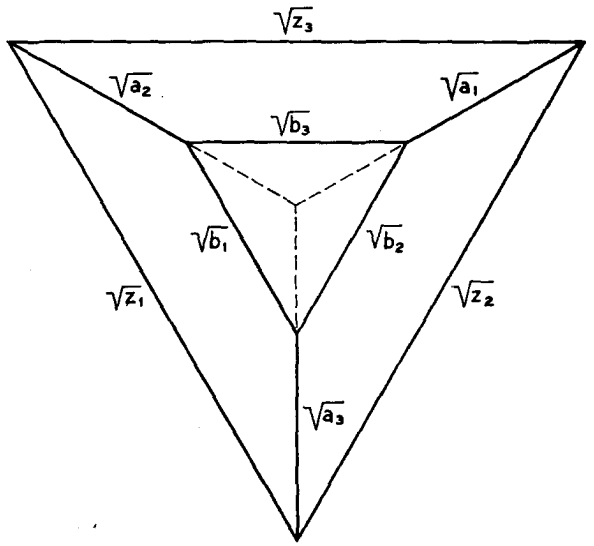


FIG. 24. The dual of the Mercedes diagram.

manifold. However, the equations which determine it can be given a very simple geometrical interpretation by means of the dual diagram[43] (see Fig. 24), which in this case is two dimensional. Given eight of the nine sides, the ninth is determined, and the polynomial equation, $P(a,b,z)=0$, which determines it is the equation of the singular manifold.

The Generalized Jost Example

Jost's example[44] was originally introduced in order to prove that the dispersion relation for the pion-nucleon vertex does *not* follow for the observed masses from the general requirements of Lorentz covariance, local commutativity, and mass spectrum. It can also be used to disprove the conjecture[45] that $H(z)$ must be analytic in the cut z_3 plane for fixed real $z_1<M_1^2$ and $z_2<M_2^2$. We shall now discuss a generalization of Jost's example, which will permit us to disprove a very appealing conjecture for the double threshold problem, and which has a simple interpretation.

In Sec. 1 it is shown that every momentum-space vertex function is analytic in the enlarged KW domain, and that the general requirements mentioned above imply no further restrictions except those mentioned in footnotes 7–9. Consider the function

$$H(C,N,z)=[(N_1^2-z_1)^{\frac{1}{2}}+(N_2^2-z_2)^{\frac{1}{2}}+(N_3^2-z_3)^{\frac{1}{2}}-C]^{-1}, \quad (4.12)$$

where the N's are positive real numbers such that

$$\begin{aligned} N_i &\geq M_i \\ N_j+N_k &> N_i, \end{aligned} \quad (4.13)$$

where C is any complex number in the triangle

$$\begin{aligned} 0 &\leq \mathrm{Re}C \leq N \\ 0 &\leq |\mathrm{Im}C| \leq N-\mathrm{Re}C \end{aligned} \quad (4.14)$$

with

$$N=\min(N_i, N_j+N_k-N_i)>0 \quad (4.15)$$

and where the square-root function is defined so that

$$\mathrm{Re}\sqrt{\ }\geq 0. \quad (4.16)$$

It can be shown that this function has the required analyticity, and is therefore a possible momentum-space vertex function. The region of possible singularities of H as C varies over the triangle (4.14) is discussed in detail in I. In particular, H may have singularities anywhere in the polycylinder

$$|z_i-N_i^2| \leq N^2/18; \quad i=1, 2, 3. \quad (4.17)$$

In other words H may have singularities anywhere in a full complex neighborhood of any unphysical real point[46] on the intersection of the three cuts, $z_i \geq M_i^2$.

The Double-Threshold Problem ($M_1=M_2$)

We shall now discuss an appealing conjecture for the double-threshold problem ($M_1=M_2\equiv M$, $M_3=0$), and we shall prove that the conjecture is contradicted by the generalized Jost example. The problem is to find the holomorphy envelope of the domain D_{12} whose

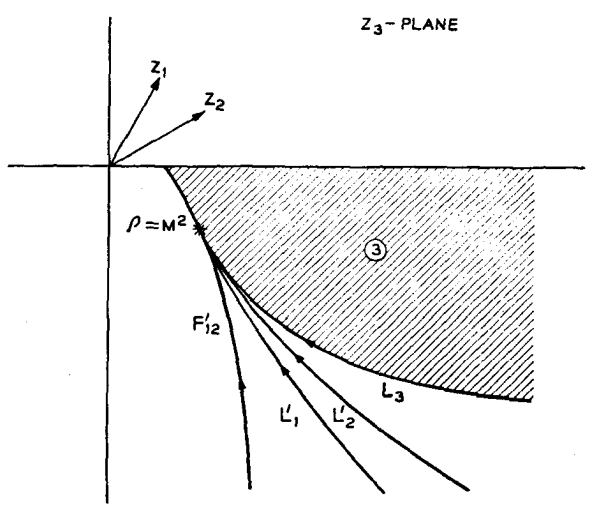


FIG. 25. In Region (3) the domain D_{12} can be completed at least up to L_1' and at least up to L_2', but not beyond L_3.

[43] A general discussion of dual diagrams is given by Landau in the work cited in reference 40. The dual diagram of the Mercedes diagram is discussed by L. B. Okun and A. P. Rudik in Nuclear Phys. **14**, 271 (1960).

[44] R. Jost, Helv. Phys. Acta. **31**, 263 (1958).

[45] See the discussion of Nambu's Theorem at the end of Sec. 2.

[46] A real point (z_1,z_2,z_3) is called *physical* if there exist real four-vectors (p_1,p_2,p_3) such that $p_i^2=z_i$ for $i=1, 2, 3$ and $p_1+p_2+p_3=0$. Otherwise it is called *unphysical*. A real point (z_1,z_2,z_3) is unphysical if and only if $z_i>0$ for $i=1, 2, 3$ and $\lambda(z)<0$. Now the inequalities $N_j+N_k>N_i$ imply that $\lambda(N^2)<0$, so the point (N_1^2,N_2^2,N_3^2) is an unphysical real point.

exterior is the union of the cuts

$$z_1 \geq M^2$$
$$z_2 \geq M^2 \qquad (4.18)$$
$$z_3 \geq 0$$

and the singularity solid. This domain has four indented edges; P_1, P_2, P_1', P_2'. From our study of the P_i problem (see Sec. 3) we know that D_{12} can be completed through the P_1 edge up to L_1' in Regions (2) and (3) and through the P_2 edge up to L_2' in Regions (1) and (3). Even if we knew the solution of the P_i' problem, we would not thereby have any further information in Region (3). The hypersurfaces L_1' and L_2' are illustrated in Fig. 25. In Region (3), where both are relevant, they intersect in an indented edge (for example when $z_1=z_2$) through which further completion is possible.

In Sec. 2 we obtained a simple upper bound on the completion into Region (3); namely, the analytic hypersurface L_3 (see Figs. 9, 10, and 25). It is not unreasonable to conjecture that L_3 is in fact the boundary of the holomorphy envelope of D_{12} in Region (3). In the special case $z_1=z_2\equiv z$, L_3 dips into the lower half-plane if and only if $\mathrm{Re}z>\frac{1}{2}M^2$. In the same special case Oehme has shown[47] that $H(z)$ must be analytic in the cut z_3 plane if $\mathrm{Re}z\leq\frac{1}{2}M^2$. These remarks make our conjecture especially appealing. However, in the subcase $z=M^2-\epsilon M+i\epsilon^2$, where ϵ is sufficiently small compared to M, L_3 lies entirely outside of the upper bound provided by the generalized Jost example (see Fig. 26), so our conjecture is false.

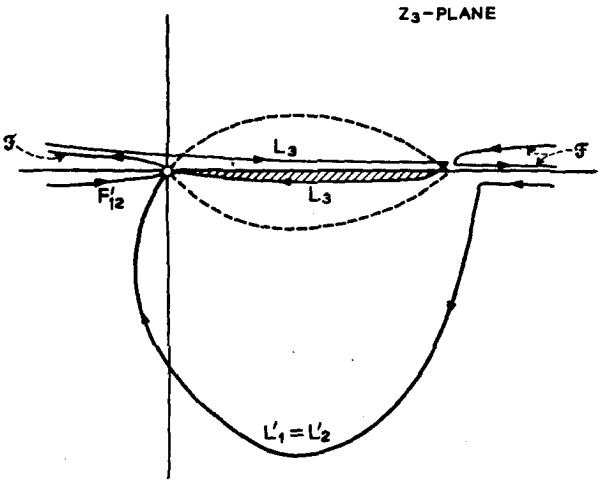


FIG. 26. In the special case $z_1=z_2=M^2-\epsilon M+i\epsilon^2$, L_3 lies entirely outside of the upper bound (dashed lines) provided by the generalized Jost Example.

Outlook

In the single threshold case $M_i\neq 0$, we have seen that for z_j and z_k in opposite half-planes the holomorphy envelope is bounded by portions of F_{ij}', F_{ik}', C_i and a new analytic hypersurface L_i'; while for z_j and z_k in the same half-plane it is most probably bounded by portions of C_i and a single quasi-analytic hypersurface H_i, whose explicit formula is not known.

There is reason to hope that portions of all three of the H_i are relevant when all three threshold masses are nonzero, and this hope adds interest to the problem of finding the formula for H_i. This problem might be treated as a boundary value problem (as suggested near the end of Sec. 3) or by way of the Källén-Toll conjecture (which is discussed in this section).

Aside from these ideas the outlook is not very bright. The conventional method[48] for finding holomorphy envelopes is to guess the boundary by studying examples of functions analytic in the given domain, and then to change to a new set of variables in which the completion method of KW and I can be applied. For our problem the only known source of examples is perturbation theory, and as we have seen in our study of the Källén-Toll representation we cannot expect to find H_i by studying the singular manifolds from perturbation theory diagrams. A study of these diagrams might, however, lead to the solution of the problem in which Region (4) is added to the enlarged KW domain, and this would be of considerable interest. The triangle diagram does not solve this problem, and the Mercedes and the triskelion seem to be too complicated to deal with. However, there is no reason to consider only totally symmetric diagrams, and the author feels that it would be worthwhile to undertake a systematic study of the simplest vertex diagrams in which one of the three indices is singled out.

Another question which deserves investigation is whether the positive definiteness inequalities and/or the unitarity of the S matrix imply any enlargement of the analyticity domain of $H(z)$ and/or $W(z)$. Although it seems unlikely, it is not inconceivable.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor A. S. Wightman for suggesting the problem and for his invaluable guidance and unfailing encouragement throughout the course of this work. He also wishes to thank R. Gunning, H. Rossi, H. Epstein, D. Fairlie, and N. Burgoyne for helpful discussions.

[47] R. Oehme, Phys. Rev. **117**, 1151 (1960).

[48] Another common procedure is to construct integral representations and show that they can be continued. This is the method used by Oehme in the work cited in reference 47.

Analytic Regularization in Relativistic Quantum Field Theory*

JAMES MCKENNA†
Department of Mathematics, Princeton University, Princeton, New Jersey
(Received October 6, 1961)

The terms of a perturbation expansion of the vacuum expectation value of the time ordered product of three scalar fields are examined in x space. It is shown that these terms can be considered as the boundary values of certain analytic functions, where the variables of these functions include complex convergence parameters. Several different integral representations for these analytic functions are obtained which exhibit part of the domain of analyticity. The singular behavior of these functions as the complex convergence parameters tend to zero is discussed, and it is conjectured that the removal of the singular part of these functions is equivalent to renormalization.

I. INTRODUCTION

THIS paper is a report on an investigation of certain properties of the vacuum expectation value of the time-ordered product of three scalar fields. The purpose of the paper is to show that we can consider the terms of a perturbation expansion of this expectation value as boundary values of analytic functions if we introduce certain complex regularizing parameters into the divergent integrals defining the terms of the expansion.

Our reasons for wishing to do this are twofold. On the one hand we hope to be able to show that by performing certain analytic or limiting operations on these analytic functions, the result will be the renormalized series in the sense of Dyson and Salam. Recent work of Caianiello gives considerable hope that this can be done.[1] Since the regularized terms of the perturbation expansion are in a considerably more tractable condition than are the corresponding terms when renormalized by the techniques of Dyson and Salam, we feel that the regularized form of the expansion may more easily yield information about the possible convergence of the expansion, and, especially since we work exclusively in x space, about analyticity properties of the individual terms.

On the other hand this work shows promise as a way of fitting some of the results of covariant perturbation theory into the framework of the axiomatic approach to relativistic quantum field theory. Wightman[2] has shown that the content of the theory of a single scalar field $\Phi(x)$, which satisfies certain basic axioms, is completely contained in a knowledge of all the vacuum expectation values:

$$F_n(x_1,\cdots,x_n)=(\Psi_0,\Phi(x_1)\Phi(x_2)\cdots\Phi(x_n)\Psi_0), \qquad n=1,\,2,\,\cdots. \quad (1.1)$$

Furthermore, $F^n(x_1,\cdots,x_n)$ is completely determined as the boundary value of an analytic function. This theory must be Lorentz covariant; there can be no states of negative energy, and there must exist a vacuum state Ψ_0 whose energy is zero; and finally local commutativity must be satisfied. In view of these results of Wightman, any knowledge of the analytic functions whose boundary values are the vacuum expectation values is most useful. In particular, since at present perturbation theory is the most practical method for studying realistic field theories, we feel that the study of the analytic functions arising from the perturbation expansion may provide useful links between perturbation theory and questions of existence studied by the axiomatic method.

It should be remarked at this point that Bogoliubov and Parasiuk[3] have used a somewhat similar regularization technique in their elegant version of renormalization theory, and in certain respects our approach is similar to theirs. However, there are substantial differences in the two approaches; mainly involving our use of analytic functions of complex variables and a new representatation of the terms of the perturbation series. Also, our emphasis on the nature of the singularities of the regularized integrals differs considerably in outlook from their approach to the problem.

We have chosen to use three local scalar fields, which we represent in the Heisenberg picture by $\Phi_1(x)$, $\Phi_2(x)$, and $\Phi_3(x)$. The fields describe particles of masses m_1, m_2, and m_3, respectively. We demand that $m_j>0$, $1\leq j\leq 3$, but there are no other restrictions on the masses. The interaction between the particles is described by the interaction Lagrangian density

$$L_I=-g\Phi_1(x)\Phi_2(x)\Phi_3(x). \quad (1.2)$$

We have chosen this model because it seems to be the simplest theory which at the same time has an interaction consisting of the product of three different fields. We feel that once the basic technique has been worked out, it can be applied to more realistic theories involving spinor fields.

The particular vacuum expectation value studied is

$$T(x_1,x_2,x_3)=(\Psi_0,T\{\Phi_1(x_1)\Phi_2(x_2)\Phi_3(x_3)\}\Psi_0). \quad (1.3)$$

* This paper is based on the author's thesis, submitted to the faculty of Princeton University in partial fulfillment of the requirements of the Ph.D. degree.

† Now at the Bell Telephone Laboratories, Murray Hill, New Jersey.

[1] E. R. Caianiello, Nuovo cimento **13**, 637 (1959); **14**, 185 (1959).

[2] A. S. Wightman, Phys. Rev. **101**, 860 (1956).

[3] N. N. Bogoliubov and O. S. Parasiuk, Doklady Akad. Nauk. S. S. S. R. **100**, 25 (1955); Acta Math. **97**, 227 (1957); N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959).

With the aid of well-known transformations[4] this can be written as

$$T(x_1,x_2,x_3)=\frac{(\psi_0,T\{S\varphi_1(x_1)\varphi_2(x_2)\varphi_3(x_3)\}\psi_0)}{(\psi_0,S\psi_0)}. \tag{1.4}$$

In (1.4), all quantities are written in the interaction picture, and S is the S matrix. The numerator of (1.4) can be expanded, using standard techniques, to give

$$(\psi_0,T\{S\varphi_1(x_1)\varphi_2(x_2)\varphi_3(x_3)\}\psi_0)=\sum_{n=0}^{\infty}\frac{(ig)^{2n+1}}{(2n+1)!2^{3n+3}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}d^4u_1\cdots d^4u_{2n+1}[u_1,\cdots,u_{2n+1},x_1]_1 \times[u_1,\cdots,u_{2n+1},x_2]_2\times[u_1,\cdots,u_{2n+1},x_3]_3. \tag{1.5}$$

In (1.5),

$$[u_1,\cdots,u_{2n+1},x_k]_k=\sum_{j_1\cdots j_{2n+1}}\{\prod_{p=1}^{n}\Delta_{Fk}(u_{j_{2p-1}}-u_{j_{2p}})\}\Delta_{Fk}(u_{j_{2p-1}}-x_k). \tag{1.6}$$

In (1.6), the sum is over all permutations $j_1,\cdots,j_{2n+1}$ of the integers $1,\cdots,2n+1$, which satisfy the restrictions

$$j_{2p-1}<j_{2p},\quad p=1,\cdots,2n,\quad j_1<j_3<\cdots<j_{2n-1}$$

and $\Delta_{Fk}(x_k)$ is the standard Feynman propagator, where the subscript k reminds us that it is formed with mass m_k.

The denominator in (1.4) can be expanded in a similar fashion, and it is well known that the result of formally dividing the series for the numerator in (1.4) by the series for the denominator has the effect of removing all the vacuum-vacuum terms.[4] In terms of Feynman graphs, this means that after this formal division, only integrals corresponding to connected Feynman graphs remain.

It must be emphasized, of course, that all these manipulations are completely formal, because most of the integrals comprising the terms of (1.5) are divergent, even after the formal division by the denominator. This paper is concerned with those individual integrals in the terms of (1.5) which remain after the formal division by the denominator, that is, with those integrals which have connected Feynman graphs. We will show that each such integral in the term of order $2n+1$ of series (1.4) can be made convergent by introducing $3n$ complex convergence parameters, w_j, $j=1,2,\cdots,3n$. From this point on, all calculations can be done in a completely rigorous fashion. However, since most of the calculations and proofs are long and involved, and furthermore since all the proofs and calculations are given in complete detail in the author's thesis which is available on microfilm, for the most part only the results will be given here, with an occasional outline of the calculations involved.

In Sec. II, by introducing complex convergence parameters into a suitable integral representation for $\Delta_F(x)$, we give a well-defined mathematical meaning to the individual integrals which remain after the formal division in (1.4) is carried out. From this point on, the emphasis is on mathematical rigor. In the last part of this section an extension of the domain of analyticity is demonstrated.

Section III is devoted to discussing the existence of a modified Schwinger representation for each of these regularized integrals and drawing several important conclusions from this representation. In the first place, this representation demonstrates a considerable enlargement of the domain of analyticity of the analytic functions being studied. In the second place, the Schwinger representation enables us to derive a very useful integral representation for these functions.

Section IV is devoted to a discussion of the singular behavior of these regularized integrals as the complex regularizing parameters tend to zero. The chief tool used in this study is the integral representation derived in Sec. III. This study is not yet complete, and it is on the completion of this investigation of the singularities of the regularized integrals that our version of the renormalization program depends.

SECTION II

A. Definition of Integrals to be Regularized and Their Feynman Diagrams

Our starting point will be a formal integral of the type

$$F_n'=\left\{\frac{(ig)^{2n+1}}{2^{3n+3}(2n+1)!}\right\}\int_{-\infty}^{\infty}d^4u_1\cdots d^4u_{2n+1} \times\{\Delta_{F1}(u_{j_1}-u_{j_2})\Delta_{F1}(u_{j_3}-u_{j_4})\cdots\Delta_{F1}(u_{j_{2n+1}}-x_1)\} \times\{\Delta_{F2}(u_{k_1}-u_{k_2})\Delta_{F2}(u_{k_3}-u_{k_4})\cdots\Delta_{F2}(u_{k_{2n+1}}-x_2)\} \times\{\Delta_{F3}(u_{l_1}-u_{l_2})\Delta_{F3}(u_{l_3}-u_{l_4})\cdots\Delta_{F3}(u_{l_{2n+1}}-x_3)\}. \tag{2.1}$$

$\{j_1,\cdots,j_{2n+1}\}$, $\{k_1,\cdots,k_{2n+1}\}$, and $\{l_1,\cdots,l_{2n+1}\}$ are three permutations of the integers $1,\cdots,2n+1$.

Since we have assumed that the indicated division in (1.4) has been carried out, (2.1) has a connected Feynman diagram. The Feynman diagram corresponding to (2.1) is constructed in the usual fashion. The points

[4] S. S. Schweber, H. A. Bethe, and F. de Hoffmann, *Mesons and Fields* (Row, Peterson & Company, Evanston, Illinois, 1956), Vol. 1, p. 384.

corresponding to the vectors $u_1, \cdots, u_{2n+1}$ are called internal points and the three points corresponding to x_1, x_2, and x_3 are called external points. Likewise, lines corresponding to propagators connecting internal points are called internal lines, while the three remaining lines connected to x_1, x_2, and x_3, respectively, are called external lines. Each internal line carries the label 1, 2, or 3, according to whether it corresponds to the Feynman propagator constructed with mass m_1, m_2, or m_3. The external line into the external point x_j necessarily carries the label j. Two points are called linked if they are connected by a line in the diagram in which there are no intervening points. Two points are said to be connected by an arc if there is a sequence of points starting with the first of the given points and ending with the second point, and such that any two consecutive points in the sequence are linked. The order of a diagram is just the number of its internal points. Any two points in the diagram corresponding to (2.1) are connected by an arc. This plays an important role in the proofs of many of the properties of the regularized integrals. Since the single integral corresponding to the diagram of order one is well defined as it stands, only integrals corresponding to diagrams of order greater than one will be considered here.

The next step in giving a well-defined meaning to (2.1) is to introduce an integral representation of the Feynman propagator[5]:

$$\Delta_{Fk}(x)=-\frac{i}{2\pi^2}\int_0^\infty \exp\left[-i\left\{\frac{m_k^2}{4t}+tx^2\right\}\right]dt. \quad (2.2)$$

This integral does not converge in the ordinary sense, but is rather to be considered as a distribution in the sense of Schwartz. If this expression for Δ_{Fk} were to be substituted into (2.1), the result would still be a divergent integral. Therefore, we start by regularizing (2.2); that is, we consider instead the expression

$$\Delta_{Fk}'(x)=-\frac{i}{2\pi^2}\times\int_0^\infty \exp\left[-i\left\{\frac{m_k^2-i\mu}{4t}+t(x^2+w)\right\}\right]dt. \quad (2.3)$$

In (2.3) we assume $\mu>0$ and $\mathrm{Im}w<0$. Integral (2.3) is absolutely convergent and is easily handled. Also, this is a Lorentz invariant method of regularization. Unfortunately, even after replacing each Δ_F in (2.1) by the regularized integral (2.3), the resulting integral still does not converge absolutely. This is because of the integrations over the four vectors u_j. To get around this difficulty we insert the factor

$$d(u,\delta)=\exp\left(-\delta\sum_{j=1}^{2n+1}\sum_{k=0}^{3}(u_{j,k})^2\right), \quad \delta>0 \quad (2.4)$$

under the integral sign in (2.1), perform the u integrations, and then let $\delta\to 0+$. In (2.4), $u_{j,k}$ is the kth component of the four-vector u_j. This operation, which is essentially the same as symmetric integration,[6] turns out to be Lorentz invariant. Therefore, this method of regularizing integral (2.1) is Lorentz invariant.

B. Mathematical Conventions

Before going on to a more detailed examination of the regularized integral, it will be helpful to list a few mathematical conventions. We first discuss the labeling of the Feynman parameters appearing in the integrals for $\Delta_{Fk}'(x)$. In a diagram having $2n+1$ internal points there are $3n+3$ lines. Corresponding to each such line there is an integral (2.3). Index the integration variables in the following way: The variable corresponding to the line going into x_j will be λ_j, $j=1, 2, 3$. The variables t_j, $j=1, \cdots, n$, will be assigned to the lines labeled 1, the variables t_j, $j=n+1, \cdots, 2n$, will be assigned to the lines labeled 2, and the variables t_j, $j=2n+1, \cdots, 3n$, will be assigned to the lines labeled 3. The remaining parameters in the integrals with the variables λ_j will be $m_j-i\epsilon_j$ and ζ_j, $j=1, 2, 3$. In the integral which has the variable of integration t_j, the corresponding mass parameter will be $M_j-i\mu_j$. This means that $M_j=m_1$, $j=1, \cdots, n$, $M_j=m_2$, $j=n+1, \cdots, 2n$, and $M_j=m_3$, $j=2n+1, \cdots, 3n$. The complex convergence parameter for each internal line will be w_j, $j=1, \cdots, 3n$.

Since we will be dealing extensively with multiple integrals in the sequel, it will be useful to adopt a convenient notation for n-tuple integrals and vectors in C^n. In general the notation $x=(x_1,\cdots,x_n)$ will be used, and the corresponding column vector is x^T. Define

$$x\cdot y=\sum_{j=1}^{n}x_jy_j.$$

Unless express mention is made to the contrary it will always be assumed that the above Euclidian metric is being used. As is usually the case, $x+y$ is the vector with components x_j+y_j, and if α is a complex number, αx is the vector with components αx_j. In general, those vectors determined by the three lines of a diagram which end in the three external points are of dimension three, while those determined by the remaining lines will be of dimension $3n$. Statements made about vectors, such as $x\geq 0$, $\mathrm{Im}y<0$, etc., mean that the statement holds for each component of the vector. Thus in the product space of n complex variables z_j, the direct product of the closed lower half planes is conveniently described as the region $\mathrm{Im}z\leq 0$. There is another notation which will be frequently used. If $t=(t_1,\cdots,t_n)$, then the vector $1/t$ has components $(1/t_1,\cdots,1/t_n)$. As long as no confusion results, we will write scalar valued functions of vector arguments, $f(x_1,\cdots,x_n)$, as $f(x)$. Also, when there is no

[5] J. Schwinger, Phys. Rev. 75, 651 (1949).

[6] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1955), p. 455.

risk of confusion, multiple integrals in which the range of integration for each variable is the same will be written as single integrals, e.g.,

$$\int_0^\infty dx_1 \int_0^\infty dx_2 \cdots \int_0^\infty dx_n f(x_1,\cdots,x_n) \equiv \int_0^\infty dx f(x).$$

Integrals of functions $g(u_1,\cdots,u_n)$ have already appeared where each variable u_j stands for a four vector, $(u_{j0},u_{j1},u_{j2},u_{j3})$, and the range of integration is from $-\infty$ to ∞ for each component, and the standard notation $d^4u_j = du_{j0}du_{j1}du_{j2}du_{j3}$ has been used. Again when there is no risk of confusion such integrals will be written

$$\int_{-\infty}^\infty d^4u g(u).$$

There is one last convention we want to mention. The amplitude of any complex variable will be assumed to be between 0 and 2π. Thus the stament $\mathrm{Im}z<0$ means $\pi<\mathrm{amp}z<2\pi$, and it does not mean $3\pi<\mathrm{amp}z<4\pi$.

C. The Definition of $F_n(z,w;\mathfrak{y})$

If we now replace Δ_{Fk} by Δ_{Fk}' in (2.1), insert (2.4) under the integral sign, and take the limit as $\delta \to 0+$, we get

$$F_n = \lim_{\delta\to 0+}\left\{\frac{(ig)^{2n+1}}{(2n+1)!}\left(-\frac{i}{2\pi^2}\right)^{3n+3}\frac{1}{2^{3n+3}} \times\int_{-\infty}^\infty d^4u d(u,\delta)\int_0^\infty d\lambda dt M_n(\lambda,t) \times\exp(-iK_n(\mu,\lambda,t))H_n(\lambda,t)\right\}. \quad (2.5)$$

In this equation we have set

$$M_n(\lambda,t) = \exp\left(-i\left\{\sum_{j=1}^3 \frac{m_j^2-i\epsilon_j}{4\lambda_j}+\sum_{j=1}^{3n}\frac{M_j^2-i\mu_j}{4t_j}\right\}\right) = \exp[-\tfrac{1}{4}i\{(m^2-i\epsilon)\cdot 1/\lambda + (M^2-i\mu)\cdot 1/t\}], \quad (2.6)$$

$$H_n(\lambda,t) = \exp\left(-i\left\{\sum_{j=1}^3 \lambda_j\zeta_j+\sum_{j=1}^{3n} t_j w_j\right\}\right) = \exp[-i(\lambda\cdot\zeta+t\cdot w)], \quad (2.7)$$

$$K_n(u,\lambda,t) = \sum_{p=1}^n t_p(u_{j_{2p}}-u_{j_{2p-1}})^2+\lambda_1(u_{j_{2n+1}}-x_1)^2 + \sum_{p=1}^n t_{n+p}(u_{k_{2p}}-u_{k_{2p-1}})^2+\lambda_2(u_{k_{2n+1}}-x_2)^2 + \sum_{p=1}^n t_{2n+p}(u_{l_{2p}}-u_{l_{2p-1}})^2+\lambda_3(u_{l_{2n+1}}-x_3)^2. \quad (2.8)$$

In (2.8) the Lorentz metric is used, and the integral (2.5) converges absolutely as long as $\epsilon>0$, $\mu>0$, $\delta>0$, $\mathrm{Im}\zeta<0$, and $\mathrm{Im}w<0$.

Because of our choice of integration variables λ and t, the factors M_n and H_n are the same for all diagrams of the same order, and it is the factor K_n which changes from diagram to diagram and dictates the changes in the structure of the integral. K_n can be expanded in the following fashion. Let g^{ik}, $j,k=0,1,2,3$, be the Lorentz metric tensor, i.e., $g^{00}=-g^{11}=-g^{22}=-g^{33}=1$, $g^{ik}=0$, $j\neq k$. Then (2.8) can be expanded to give

$$K_n = \sum_{l=0}^3 g^{ll}\{\sum_{j,k=1}^{2n+1} a_{jk}u_{j,l}u_{k,l} - 2\sum_{j=1}^{2n+1}\sum_{k=1}^3 b_{jk}u_{j,l}x_{k,l} + \sum_{j=1}^3 C_{jj}(x_{j,l})^2\}. \quad (2.9)$$

We can define the matrices

$$A=(a_{ij}), \quad B=(b_{ij}), \quad C=(\delta_{ij}C_{jj}). \quad (2.10)$$

C is a 3×3 matrix, whose elements are $C_{jj}=\lambda_j$, $j=1,2,3$. B is a $(2n+1)\times 3$ matrix which has only 3 elements not identically zero. These are $b_{p1}=\lambda_1$, $b_{q2}=\lambda_2$, and $b_{r3}=\lambda_3$. A is a $(2n+1)\times(2n+1)$ symmetric matrix of the form

$$A = \begin{pmatrix} \sum_{j=2}^{2n+1}\alpha_{1j}+\epsilon_1 & -\alpha_{12} & \cdots & -\alpha_{1(2n+1)} \\ -\alpha_{21} & \sum_{\substack{j\neq 2\\ j=1}}^{2n+1}\alpha_{2j}+\epsilon_2 & \cdots & -\alpha_{2(2n+1)} \\ \vdots & \vdots & \vdots & \vdots \\ -\alpha_{(2n+1)1} & -\alpha_{(2n+1)2} & \cdots & \sum_{j=1}^{2n}\alpha_{(2n+1)j}+\epsilon_{(2n+1)} \end{pmatrix}. \quad (2.11)$$

The elements α_{jk} are either identically zero or are of the form t_l or t_l+t_m. The quantities ϵ_j are equal to the sum of the elements in the corresponding row of B. Thus at most, three of the ϵ_j do not vanish identically. If there were only one nonvanishing element ϵ_j, it would imply that the corresponding Feynman diagram was not connected, so this possibility is excluded. A has the further properties that if $t>0$, $\lambda>0$, then $\det A>0$, and every cofactor of $\det A$ is positive. Consequently A^{-1} exists and every element of A^{-1} is positive. These statements depend on the fact that the corresponding Feynman diagram is connected.

If we define the vectors

$$x_l=(x_{1,l},x_{2,l},x_{3,l}), \quad l=0, 1, 2, 3, \tag{2.12}$$

then the result of performing the u integrations and carrying out the limiting process $\delta\to 0+$ in (2.5) is

$$F_n=G_n\int_0^\infty dtd\lambda D^{-2}M_n(\lambda,t)H_n(\lambda,t) \times\exp(-i\sum_{l=0}^{3} g^{ll}x_l(C-B^TA^{-1}B)x_l{}^T). \tag{2.13}$$

In (2.13) we have set

$$D=\det A, \quad G_n=\frac{(ig)^{2n+1}}{(2n+1)!}\left(-\frac{i}{2\pi^2}\right)^{3n+3}\frac{(\pi^2)^{2n+1}}{2^{3n+3}}. \tag{2.14}$$

In order to show explicitly that (2.13) is a Lorentz invariant expression, the argument of the exponential in the last factor of (2.13) can be expanded to give

$$\sum_{l=0}^{3} g^{ll}x_l(C-B^TA^{-1}B)x_l{}^T$$

$$=D^{-1}\{e_1(x_1-x_2)^2+e_2(x_2-x_3)^2+e_3(x_3-x_1)^2\}, \tag{2.15}$$

$$e_1=\lambda_1\lambda_2D_{(qp)}, \quad e_2=\lambda_2\lambda_3D_{(rq)}, \quad e_3=\lambda_3\lambda_1D_{(pr)}, \tag{2.16}$$

where $D_{(ts)}$ is the co-factor of the (ts) element in $D=\det A$. $(x_i-x_j)^2$ is the Lorentz square of the difference of the two four-vectors x_j and x_k.

If we write

$$z_1=(x_1-x_2)^2, \quad z_2=(x_2-x_3)^2, \quad z_3=(x_3-x_1)^2, \tag{2.17}$$

then the modified integral F_n can be written in a form explicitly showing its Lorentz invariance

$$F_n=G_n\int_0^\infty dtd\lambda D^{-2}M_n(\lambda,t)H_n(\lambda,t) \times\exp(-iD^{-1}e\cdot z). \tag{2.18}$$

At this point it should be remarked that Symanzik[7] and Nambu[8] have previously obtained the result (2.18) in a slightly different form. The aims of both Symanzik and Nambu in their papers were quite different from ours, and all our subsequent results bear no resemblance to their results.

As we remarked earlier, the elements of A^{-1} are all non-negative when $t\geq 0$, $\lambda\geq 0$, so that $e\geq 0$ over the range of integration. As a consequence, the integral (2.18) still converges if z_1, z_2, and z_3 are complex numbers with negative imaginary parts. At this stage, the integral (2.18) is over regularized. It is readily seen that as long as $\mathrm{Im}z<0$, the integral F_n still converges if $\epsilon=0$, $\zeta=0$. We assume from now on that $\epsilon=0$, $\zeta=0$, and call the resulting function $F_n(z,w;\mu)$. It will shortly be shown that the regularizing parameter μ is also unessential and can be set equal to zero. The resulting function, $F_n(z,w)$, will be called the regularized integral. The original integral F_n' can now be considered as the limit of $F_n(z,w)$ as $\mathrm{Im}z\to 0-$ and $w\to 0$. In general, $F_n(z,w)$ is highly singular at $w=0$, and it is believed that the standard renormalization of F_n' is, in some yet undetermined fashion, equivalent to removing the singular portion of the function $F_n(z,w)$ and then setting $w=0$.

D. Definition of $F_n(z,w)$

In order to show the possibility of setting $\mu=0$, and to determine a portion of the domain of analyticity of $F_n(z,w)$, we make a change of variables in the integral defining $F_n(z,w;\mu)$:

$$F_n(z,w;\mu)=G_n\int_0^\infty d\lambda dtD^{-2} \times\exp[-i\{\tfrac{1}{4}(M^2-i\mu)\cdot(1/t)+w\cdot t +\tfrac{1}{4}m^2\cdot(1/\lambda)+D^{-1}e\cdot z\}]. \tag{2.19}$$

Make the change of variables

$$\lambda_j=r\eta_j; \quad j=1, 2, 3; \quad t_j=rs_j; \qquad j=1, 2, \cdots, 3n-1; \quad t_{3n}=r(1-\sum_{j=1}^{3n-1} s_j). \tag{2.20}$$

The resulting integral over r can be performed, and introducing the redundant variable

$$s_{3n}=1-\sum_{j=1}^{3n-1} s_j,$$

the final result is

$$F_n(z,w;\mu) =2^{n-1}\pi e^{-(n-2)\frac{1}{2}\pi i}G_n\int_0^\infty dsd\eta\delta(1-\sum_{j=1}^{3n} s_j) \times\tilde{D}^{-2}B^{\frac{1}{2}(n-1)}a^{-\frac{1}{2}(n-1)}H_{-n+1}{}^{(1)}(e^{-\pi i}\sqrt{a}\sqrt{B}). \tag{2.21}$$

In (2.21)

$$a=(M^2-i\mu)\cdot(1/s)+m^2\cdot(1/\eta), \qquad B=s\cdot w+\tilde{D}^{-1}(\tilde{e}\cdot z), \tag{2.22}$$

$\tilde{D}$ and $\tilde{e}$ are the functions D and e with λ and t replaced

[7] K. Symanzik, Prog. Theoret. Phys. (Kyoto), **20**, 690 (1958).
[8] Y. Nambu, Nuovo cimento **6**, 1064 (1957).

by η and s and $H_n^{(1)}$ denotes a Hankel function of the first kind.

Using standard properties of the Hankel function, it follows that integral (2.21) converges when $\mu=0$ and defines an analytic function of z and w in a domain including the region $\mathrm{Im}z<0$, $\mathrm{Im}w<0$. A portion of this domain can be described as follows: Let l be any straight line through the origin of the complex plane. Then l divides the complex plane into two open sets, of which at least one, say L, does not contain the positive real axis. Then $F_n(z,w)$ is analytic for all z and w in L.

There is one case of particular interest. Let $\mathrm{Im}w=0$, $\mathrm{Re}w<0$. Then considered as a function of z alone, $F_n(z,w)$ is analytic as long as z lies on the opposite side from the positive real axis of any straight line through the origin. This result is independent of where w lies on the negative real axis. Since the eventual hope is to separate $F_n(z,w)$ into a part which is analytic at $w=0$ and a part which is singular at $w=0$, this last result suggests that after setting $w=0$ in the regular part, the resulting function of z will be analytic at least in the region just described.

Summary of Sec. II

In Sec. II it was shown that the integrals which make up the individual terms of the perturbation series can be regularized in a Lorentz invariant fashion. The original integrals are the boundary values of a function $F_n(z,w)$ of the $3n$ complex convergence parameters w_j, and the three complex variables z_j. This boundary value is taken as $w\to 0$ and $\mathrm{Im}\ z\to 0-$. A new integral representation of $F_n(z,w)$ was exhibited from which an enlargement of the original domain of analyticity could be obtained.

SECTION III

A. New Representation for $F_n(z,w;\mu)$

In this section a new representation of $F_n(z,w)$ will be discussed. It will be restricted to values of z lying on the negative real axis and satisfying another condition to be specified later. This representation will exhibit a further enlargement of the domain of analyticity, and most important, it will lend itself to the derivation of a third representation of $F_n(z,w)$ which will be very useful in the study of the singularities of $F_n(z,w)$ near $w=0$.

The initial representation we shall exhibit is based on an idea of Schwinger's.[9] It is well known that the singular function $\Delta^+(m^2,x^2)$ is the boundary value of an analytic function which we write as $\Delta^+(m^2,z)$.[10] In fact

$$\Delta^+(m^2,z)=-\frac{1}{(2\pi)^2}\int_0^\infty dt\exp\left(-i\left\{\frac{m^2}{4t}+tz\right\}\right). \quad (3.1)$$

It is well known that (3.1) can also be written as

$$\Delta^+(m^2,z)=-\left(\frac{m^2}{8\pi}\right)\frac{H_1^{(1)}(m\sqrt{z})}{m\sqrt{z}}. \quad (3.2)$$

Equation (3.2) shows that $\Delta^+(m^2,z)$ is a single-valued analytic function of z in the plane cut from 0 to ∞ along the positive real axis.

We now propose to write

$$F_n(z,w;\mu)=\int_0^\infty d\beta\Phi(z,\beta)\prod_{j=1}^{3n}\Delta^+(\beta_j-i\mu_j,w_j), \quad (3.3)$$

and we must show that such a representation does exist. It will turn out that Φ is really a temperature distribution in the sense of Schwartz.[11] In order to prove the existence of Φ, strong restrictions must be placed on z. Once this is done, we can set $\mu=0$, giving a new representation for $F_n(z,w)$, which for fixed z will be a function of w, analytic in the direct product of the w planes, each cut from 0 to ∞ along the positive real axis.

B. Discussion of the Proof of the Existence of Φ

We now give an outline of a proof that representation (3.3) for $F_n(z,w;\mu)$ does exist. First replace $\Delta^+(\beta_j-i\mu_j,w_j)$ in (3.3) by its integral expression (3.1). Formally changing orders of integration, this gives

$$F_n(z,w;\mu)$$
$$=\left(-\frac{1}{4\pi^2}\right)^{3n}\int_0^\infty dt\exp(-i\{-\tfrac{1}{4}i\mu\cdot(1/t)+w\cdot t\})$$
$$\times\int_0^\infty d\beta\Phi(z,\beta)\exp(-i\tfrac{1}{4}\beta\cdot 1/t). \quad (3.4)$$

If we compare this with expression (2.19) for $F_n(z,w;\mu)$, we see that both look like Laplace transforms when w is taken to be the variable of the transform. We formally take the inverse transforms of both sides and equate the results to get

$$G_n\exp(-i\tfrac{1}{4}M^2\cdot 1/t)$$
$$\times\int_0^\infty d\lambda D^{-2}\exp(-i\tfrac{1}{4}m^2\cdot(1/\lambda)-iD^{-1}e\cdot z)$$
$$=\left(-\frac{1}{4\pi^2}\right)^{3n}\int_0^\infty d\beta\Phi(z,\beta)\exp(-\tfrac{1}{4}i\beta\cdot 1/t). \quad (3.5)$$

Again, if we set $p_j=1/4t_j$, $j=1,\cdots,3n$, the right-hand side looks just like the Laplace transform of Φ. The problem is then to show that the left-hand side of (3.5) has just the necessary analyticity and boundedness

[9] J. Schwinger, *Proceedings of the Seventh Conference on High-Energy Nuclear Physics, 1957* (Interscience Publishers, Inc., 1957).

[10] W. E. Thirring, *Principles of Quantum Electrodynamics* (Academic Press Inc., New York, 1958).

[11] L. Schwartz, *Theorie des distributions* (Herman et Cie, Paris, France, 1951), Vols. 1 and 2.

properties to be the Laplace transform of a temperate distribution.

At this point we quote several results from the theory of distributions. The first theorem is due to Schwartz.[12]

Theorem: A distribution T is temperate ($T\in S'$) if and only if T can be written as the derivative of a continuous function of slow growth. More explicitly, if and only if

$$T=D^s[\{\prod_{j=1}^{n}(1+x_j^2)^{r_j}\}f(x)]=D^s g(x). \tag{3.6}$$

In (3.6) $f(x)$ is a continuous, bounded function; s_j, $j=1, \cdots, n$, are non-negative integers, $D^s=\partial^s/\partial x_1^{s_1}\partial x_2^{s_2}\cdots\partial x_n^{s_n}$; and $s_1+\cdots+s_n=s$. The derivatives are taken in the sense of distributions, of course.

The second result is theorem due to Gärding.[13]

Theorem: Let $L(p)$ be a function of the n complex variables $p_j=\xi_j+i\eta_j$, $j=1, 2, \cdots, n$. Then in order that $L(p)$ be the Laplace transform of a temperate distribution, where $e^{\eta\cdot x}T$ is also a temperate distribution for all $\eta<0$, it is necessary and sufficient that

(a) $L(p)$ is analytic in the direct product of the lower half-planes, i.e., in the region $\eta<0$.

(b) In any strip $-\infty<b_j<\eta_j<a_j<0$, $j=1, 2, \cdots, n$,

$$|L(p_1,\cdots,p_n)|\leq P(|\xi_1|,\cdots,|\xi_n|), \tag{3.7}$$

where P is a polynomial which may depend on the strip.

(c) In any strip $-\infty<b_j<\eta_j<0$, $j=1, 2, \cdots, n$ there exist non-negative integers m_j and r_j, and positive constants A_j and B_j, $j=1, 2, \cdots, n$ (which are all independent of ξ, but which may depend on the strip) such that the following inequality holds

$$|L(p_1,\cdots,p_n)|\leq\prod_{j=1}^{n}\{|\xi_j+i\eta_j|^{m_j}(A_j+B_j|\eta_j|^{-r_j})\}. \tag{3.8}$$

Furthermore, a function of slow growth whose derivative gives T [cf. Eq. (3.6)] is given in terms of $L(p)$ by

$$(i)^s g(x)=(2\pi)^{-n}\int_{-\infty}^{\infty}d\xi e^{i(\xi+i\eta)\cdot x}L(\xi+i\eta)\prod_{j=1}^{n}(\xi_j+i\eta_j)^{-s_j}. \tag{3.9}$$

The final theorem which we quote is due to Lions and it gives information about the support of a distribution from a knowledge of its Laplace transform.[14]

Theorem: Let T be a distribution such that $e^{\eta\cdot x}T$ is a temperate distribution for all $\eta<0$. Then in order that the support of T be contained in the semi-space $x\cdot\lambda\geq A$, it is necessary and sufficient that the Laplace transform of T, $L(\xi+i\eta)$, be such that for every $B<A$, and for $\eta<0$,

$$\exp(tB)|L[\xi+i(\eta-t\lambda)]|\leq P(|\xi+i\eta|), \tag{3.10}$$

for all positive values of the scalar parameter t, where P is a polynomial which can depend on η.

From Gärding's theorem we see that a temperate distribution $\Phi(z,\beta)$ satisfying (3.4) will exist if the left-hand side of (3.5) is analytic in the direct product of the lower half p_j planes, and if in that region it satisfies inequalities (3.7) and (3.8).

In order to prove the necessary analyticity, essentially it is necessary only to show that $D(\lambda,1/p)$ has no zeros in the region $\mathrm{Im}p<0$, $\lambda>0$. This is true, and its proof depends strongly on the fact that D corresponds to a connected Feynman diagram.

In order to prove that the left-hand side of (3.5) satisfies the necessary boundedness conditions, a somewhat circuitous argument is used. First simple contour integration is used to show that λ can be replaced by $i\lambda$ in (3.5). Then the assumption is made that the complex numbers z_j satisfy condition R

$$z_j<0, \quad j=1, 2, 3; \quad (-z_j)^{\frac{1}{2}}<(-z_k)^{\frac{1}{2}}+(-z_l)^{\frac{1}{2}}, \quad (R)$$

(where $(j,k,l)=(1,2,3)$ in cyclic order. When z satisfies condition R it is possible to find three real, Euclidian, four-vectors a, b, and c such that

$$-z_1=(a-b)^2, \quad -z_2=(b-c)^2, \quad -z_3=(c-a)^2. \tag{3.11}$$

Let $f_j=(a_j,b_j,c_j)$, $j=1, 2, 3, 4$ where a_j is the jth component of the four-vector a, etc. Then with the aid of a well-known integral relationship we can write

$$\begin{aligned}D^{-2}&\exp[iD^{-1}e\cdot(-z)]\\&=D^{-2}\exp[i\sum_{j=1}^{4}f_j(C-B^TA^{-1}B)f_j^T]\\&=(-\pi^2)^{-(2n+1)}\int_{-\infty}^{\infty}d^4v\\&\times\exp(i\sum_{j=1}^{4}\{u_jAu_j^T-2u_jBf_j^T+f_jCf_j^T\}).\end{aligned} \tag{3.12}$$

In (3.12), $u_j=(v_{1,j},\cdots,v_{2n+1,j})$ where $v_{l,j}$ is the jth component of the real, *Euclidian*, four-vector v_l, f_j has just been defined above; and A, B, and C are the matrices defined in (2.10) except that λ has been replaced by $i\lambda$ and t by $1/4p$, where $\mathrm{Im}p<0$. The integral on the right of (3.12) is readily seen to converge absolutely, and it is important to notice that we deal with a Euclidian metric and not with a Lorentz metric. If the expression (3.12) for $D^{-2}\exp(-iD^{-1}e\cdot z)$ is substituted into the left-hand side of (3.5), it becomes a straightforward matter to show that the desired boundedness conditions are indeed satisfied.

Using the form of the left-hand side of (3.5) just obtained, it can readily be shown with the aid of Lion's theorem that the support of Φ lies in the region $\beta_j\geq M_j^2$. This completes the outline of a proof that a temperate distribution $\Phi(z,\beta)$ exists such that representation (3.3) holds as long as z satisfies condition R, $\mathrm{Im}w<0$, $\mu<0$.

[12] Reference 11, Vol. 2, p. 95.
[13] L. Garding (unpublished).
[14] J. L. Lions, J. D'Analyse Math., 2, 369 (1952).

Furthermore, the support of Φ lies in the region $\beta_j \geq M_j^2$, $j=1, \cdots, 3n$.

C. A New Representation for $F_n(z,w)$

If we now employ the canonical form for a temperate distribution given in Schwartz's theorem, we can write

$$F_n(z,w;\mu)=(-1)^\chi \int_{M_1^2}^{\infty} d\beta_1 \cdots \int_{M_{3n}^2}^{\infty} d\beta_{3n} g_n(z,\beta) \times \prod_{j=1}^{3n} \frac{\partial^{s_j}}{\partial \beta_j^{s_j}} \Delta^+(\beta_j - i\mu_j, w_j), \quad (3.13)$$

where $\chi=\sum_{j=1}^{3n} s_j$. In this expression, neither the function $g_n(z,\beta)$ nor the integers s_j are unique, but the final integral (3.13) is unique, as it must be. This will soon be seen explicitly. The first important use to be made of (3.13) is to show that we can set $\mu=0$ and that the resulting function, the regularized integral $F_n(z,w)$, is for fixed z satisfying condition R, a function of w analytic in the direct product of the w planes, each cut from zero to infinity along the positive real axis. These statements can be proved in a straightforward fashion from (3.13), making use of the properties of $g_n(z,\beta)$ given in Schwartz's theorem, and the well-known properties of $\Delta^+(\beta,w)$.

This new representation for $F_n(z,w)$ has given us new and important information about the domain of analyticity. The restriction on z is severe, and was imposed in order to prove the validity of representation (3.3). While this restriction is sufficient to insure the validity of (3.3), it is unknown whether it is necessary. It should also be remarked that no new information about the domain of analyticity of $F_n(z,w)$ can be obtained from this representation. Because of well-known properties of the Hankel function, if any w_j is put on the positive real axis, $\Delta^+(\beta_j,w_j)$ no longer damps exponentially as $\beta_j \to +\infty$, and our integral need no longer converge. If we let w_j wind once around the origin, say, then $\Delta^+(\beta_j,w_j)$ *grows* exponentially as $\beta_j \to \infty$ and the integral then certainly diverges. This of course does not prove that $F_n(z,w)$ is not analytic in a larger domain, in fact it is a very attractive conjecture that $w=0$ is a branch point of a complicated Riemann surface on which $F_n(z,w)$ is analytic. The results of Sec. IV will lend further credence to this conjecture.

We now use the relationship (3.9) between the generating function $g_n(z,\beta)$ and the corresponding Laplace transform $L(z,p)$, given in the last part of Gärding's theorem, to derive our final form of the regularized integral. For the purposes of this paper, this will be the most useful expression for $F_n(z,w)$. The quantity $L(z,p)$ to be used in computing $g_n(z,\beta)$ is just $(-4\pi^2)^{3n}$ times the left-hand side of (3.5), namely,

$$L(z,p)=-i(-\pi^2)^{3n} G_n \exp(-iM^2\cdot p)\int_0^\infty d\lambda D^{-2} \times \exp[-\tfrac{1}{4}M^2\cdot(1/\lambda)-iD^{-1}e\cdot z]. \quad (3.14)$$

In (3.14) D and e are as defined in (2.14) and (2.16) except that λ has been replaced by $i\lambda$ and t by $1/4p$, $\mathrm{Im}\, p<0$. Now the expression for $D^{-2}\exp(-iD^{-1}e\cdot z)$ which is given in (3.12) can be inserted into (3.14). If this result is substituted into (3.9), the integrations over λ and ξ can be performed explicitly, leaving an integration over the $2n+1$ Euclidian four-vectors v_j.

If this expression for $g_n(z,\beta)$ is substituted into integral (3.13), (with $\mu=0$), and if we express the derivatives of the Δ^+ functions in terms of Hankel functions, then the integrations over β can be performed explicitly, and the result is

$$F_n(z,w)=Q_n\int_{-\infty}^{\infty} d^4v \frac{K_1[m_1\{(v_{j_{2n+1}}-a)^2\}^{\frac{1}{2}}]}{\{(v_{j_{2n+1}}-a)^2\}^{\frac{1}{2}}} \frac{K_1[m_2\{(v_{k_{2n+1}}-b)^2\}^{\frac{1}{2}}]}{\{(v_{k_{2n+1}}-b)^2\}^{\frac{1}{2}}} \frac{K_1[m_3\{(v_{l_{2n+1}}-c)^2\}^{\frac{1}{2}}]}{\{(v_{l_{2n+1}}-c)^2\}^{\frac{1}{2}}} \times \prod_{p=1}^{n} \left[\frac{H_1^{(1)}[m_1\{w_p-(v_{j_{2p-1}}-v_{j_{2p}})^2\}^{\frac{1}{2}}]}{\{w_p-(v_{j_{2p-1}}-v_{j_{2p}})^2\}^{\frac{1}{2}}} \frac{H_1^{(1)}[m_2\{w_{p+n}-(v_{k_{2p-1}}-v_{k_{2p}})^2\}^{\frac{1}{2}}]}{\{w_{p+n}-(v_{k_{2p-1}}-v_{k_{2p}})^2\}^{\frac{1}{2}}} \times \frac{H_1^{(1)}[m_3\{w_{p+2n}-(v_{l_{2p-1}}-v_{l_{2p}})^2\}^{\frac{1}{2}}]}{\{w_{p+2n}-(v_{l_{2p-1}}-v_{l_{2p}})^2\}^{\frac{1}{2}}} \right]. \quad (3.15)$$

$$Q_n=i\left(\frac{m_1}{2\pi}\right)^n\left(\frac{m_2}{2\pi}\right)^n\left(\frac{m_3}{2\pi}\right)^n m_1m_2m_3(\pi^2)^{n-1}G_n.$$

In (3.15) K_1 is a modified Bessel kind, $H_1^{(1)}$ is a Hankel function, and the integration four vectors are real and Euclidian. It is most important to note that all squares such as $(v-a)^2$ or $(v_j-v_k)^2$ are taken with respect to a Euclidian metric.

Although integral (3.15) looks quite formidable, it is a straightforward matter to write it down if the Feynman diagram to which it corresponds is given. First assign the integration four vectors $v_1, \cdots, v_{2n+1}$ to the internal points, and the vectors a, b, and c to the external points. Then assign to each line its mass m, and to each internal line its convergence parameter w. There is first

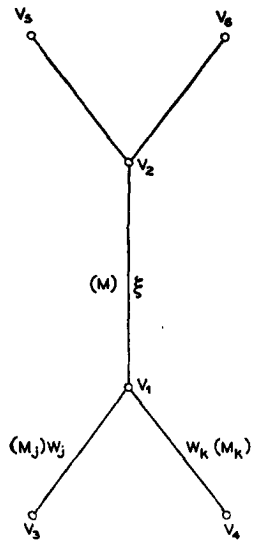


FIG. 1. Situation of internal line with convergence parameter ξ.

the constant $i(\pi^2)^{n-1}G_n$ which depends only on the order of the diagram. Second, for each of the three external lines write down a factor of the form

$$\frac{mK_1[m\{(v-a)^2\}^{\frac{1}{2}}]}{\{(v-a)^2\}^{\frac{1}{2}}}. \tag{3.16}$$

Third, for each of the remaining lines there is a term of the form

$$\frac{mH_1^{(1)}[m\{w-(v_j-v_k)^2\}^{\frac{1}{2}}]}{2\pi\{w-(v_j-v_k)^2\}^{\frac{1}{2}}}. \tag{3.17}$$

Finally, integrate over all four vectors.

D. The Euclidian p-Space Representation of $F_n(z,w)$

We conclude Sec. III by deriving the Euclidian p-space representation of $F_n(z,w)$ from (3.15). To do this we write the expressions (3.16) and (3.17) in terms of their Laplace transforms, substitute the result into (3.15), and carry out the v integrations. The result is

$$F_n(z,w)=i(-1)^{3n}(2\pi)^{-n-2}(\pi^2)^{n-1}G_n\int_{-\infty}^{\infty}d^4p\{\prod_{s=1}^{3n+1}\delta^4(p_{j_s}\pm p_{k_s}\pm p_{l_s})$$

$$\times(p_{3n+1}{}^2+m_1^2)^{-1}(p_{3n+2}{}^2+m_2^2)^{-1}(p_{3n+3}{}^2+m_3^2)^{-1}\exp[-i(p_{3n+1}\cdot a+p_{3n+2}\cdot b+p_{3n+3}\cdot c)]$$

$$\times\prod_{j=1}^{n}\Bigg[\frac{(w_j)^{\frac{1}{2}}H_1^{(1)}[\{w_j(p_j{}^2+m_1^2)\}^{\frac{1}{2}}]}{(p_j{}^2+m_1^2)^{\frac{1}{2}}}\,\frac{(w_{j+n})^{\frac{1}{2}}H_1^{(1)}[\{w_{j+n}(p_{j+n}{}^2+m_2^2)\}^{\frac{1}{2}}]}{(p_{j+n}{}^2+m_2^2)^{\frac{1}{2}}}$$

$$\times\frac{(w_{j+2n})^{2}H_1^{(1)}[\{w_{j+2n}(p_{j+2n}{}^2+m_3^2)\}^{\frac{1}{2}}]}{(p_{j+2n}{}^2+m_3^2)^{\frac{1}{2}}}\Bigg]. \tag{3.18}$$

In writing down (3.18), the variables p have been labeled in the following fashion. The variables p_{3n+1}, p_{3n+2}, and p_{3n+3} have been assigned to the external lines ending in x_1, x_2, and x_3, respectively. The variable p_j has been assigned to the line with convergence parameter w_j. There is one delta function at each internal point, and the signs before the p's are chosen to conserve momentum.

Using the properties of Hankel functions, it is easy to show that

$$\lim_{w\to 0} w^{\frac{1}{2}}H_1^{(1)}[\{w(p^2+m^2)\}^{\frac{1}{2}}]=-(2i/\pi)\{p^2+m^2\}^{-\frac{1}{2}}. \tag{3.19}$$

Now formally let $w_j\to 0$, $j=1, 2, \cdots, 3n$, in (3.18) and make use of (3.19). The result, which is a highly divergent integral, is, up to a factor of $(-1)^{3n+3}(i)^{n+2}$, what one would get by writing down integral (2.1), expressing each Feynman propagator in terms of its Laplace transform, and then performing the u integrations. It is the same, that is, with the important exception that in integral (3.18) the propagator appears as $(p^2+m^2)^{-1}$ where p^2 is the Euclidian square, while in (2.1) the propagator would appear as $(p^2-m^2)^{-1}$, and here p^2 is the Lorentz square. In other words, in the integral (3.18) Feynman's method[15] of rotating the path of the p_0 integration through 90° has already been performed. This can be regarded as a justification of one of the previously slightly dubious manipulations of renormalization theory.

This concludes Sec. III. In this section we have indicated the existence of a Schwinger type representation for the regularized integral $F_n(z,w)$ for any given diagram when the variables z are properly restricted, and $F_n(z,w)$ is considered as a function of w. This representation is exhibited in (3.13) (after μ has been set equal to zero). With the aid of this representation, a further important representation of $F_n(z,w)$ was exhibited in (3.15). Finally, (3.15) was used to examine the form of $F_n(z,w)$ as a p-space integral.

[15] R. P. Feynman, Phys. Rev. 76, 769 (1949).

SECTION IV

A. The Singularity Corresponding to a Single Internal Line

In Sec. IV we will indicate how (3.15) can be used to determine the behavior of $F_n(z,w)$ as certain groups of convergence parameters w_j vary in neighborhoods of the origin, while the remaining parameters are held fixed. It is assumed throughout Sec. IV that z satisfies condition R.

We first discuss in some detail the case where the convergence parameter of a single internal line varies, while the remaining convergence parameters are held fixed in the cut planes. We call the convergence parameter under consideration ξ, to distinguish it from the remaining parameters w_j, and we write $F_n(z,w)=F_n(\xi)$. We assume that this line is not one half of a self-energy loop, where by a self-energy loop we mean two lines which have the same end points. We assume further that at least one endpoint of this line is not also the endpoint of an external line. This portion of the Feynman diagram under consideration, labeled with the proper variables, is shown in Fig. 1. It is assumed that neither v_3 nor v_4 are external points.

Now in integral (3.15) make the change of variables

$$u_1=v_1-v_2, \quad u_j=v_j, \quad j=2, 3, \cdots, 2n+1. \tag{4.1}$$

After this change of variables, the three terms in the integral containing u_1 are

$$\frac{H_1^{(1)}(m\{\xi-u_1^2\}^{\frac{1}{2}})}{\{\xi-u_1^2\}^{\frac{1}{2}}}\frac{H_1^{(1)}[M_j\{w_j-(u_1+u_2-u_3)^2\}^{\frac{1}{2}}]}{\{w_j-(u_1+u_2-u_3)^2\}^{\frac{1}{2}}} \times\frac{H_1^{(1)}[M_k\{w_k-(u_1+u_2-u_4)^2\}^{\frac{1}{2}}]}{\{w_k-(u_1+u_2-u_4)^2\}^{\frac{1}{2}}}. \tag{4.2}$$

We next transform u_1 to polar coordinates

$$\begin{aligned} u_{11}&=r\sin\theta_1\sin\theta_2\sin\theta_3\\ u_{12}&=r\sin\theta_1\sin\theta_2\cos\theta_3\\ u_{13}&=r\sin\theta_1\cos\theta_2\\ u_{14}&=r\cos\theta_1\\ d^4u_1&=r^3\sin^2\theta_1\sin\theta_2 dr d\theta_1 d\theta_2 d\theta_3=r^3 dr d\omega. \end{aligned} \tag{4.3}$$

Let

$$\varphi(r)=\int_{-\infty}^{\infty} d^4u_2\cdots d^4u_{2n+1}G(u_2,\cdots,u_{2n+1}) \times\int d\omega\frac{H_1^{(1)}[M_j\{w_j-(u_1+u_2-u_3)^2\}^{\frac{1}{2}}]}{\{w_j-(u_1+u_2-u_3)^2\}^{\frac{1}{2}}}\frac{H_1^{(1)}[M_k\{w_k-(u_1+u_2-u_4)^2\}^{\frac{1}{2}}]}{\{w_k-(u_1+u_2-u_4)^2\}^{\frac{1}{2}}}. \tag{4.4}$$

In (4.4) $G(u_2,\cdots,u_{2n+1})$ is the product of all those terms in integral (3.15) representing $F_n(\xi)$ which do not have u_1 as an argument. The function $\varphi(r)$ is, of course, also a function of z and the remaining w_l, which remain fixed throughout our discussion. It can also be shown that there is an $\epsilon>0$ such that $\varphi(r)$ is an analytic function of r when $|r|<\epsilon$, and furthermore, only even powers of r appear in the Taylor series expansion of $\varphi(r)$. The analyticity of $\varphi(r)$ depends on the fact that w_j and w_k are held fixed in the cut plane.

We can then write

$$F_n(\xi)=\int_0^{\infty} dr r^3\varphi(r)\frac{H_1^{(1)}(M\{\xi-r^2\}^{\frac{1}{2}})}{(\xi-r^2)^{\frac{1}{2}}}. \tag{4.5}$$

If we pick a so that $0<a<\frac{1}{2}\epsilon^2$, the integral appearing in (4.5) can be split into two integrals,

$$\int_0^{\infty}=\int_0^{\sqrt{a}}+\int_{\sqrt{a}}^{\infty}. \tag{4.6}$$

The function of ξ corresponding to the integral from $\sqrt{a}$ to ∞ is readily seen to be a single-valued analytic function of ξ for all ξ satisfying $|\xi|<\frac{1}{2}a$. To determine the nature of the function defined by the integral from 0 to $\sqrt{a}$, we can replace $\varphi(r)$ by its Taylor series expansion, and the Hankel function can be replaced by its well-known expression in terms of infinite series and logarithms.[16] Upon interchanging summation and integration, the resulting elementary integrals can be performed. The net result of these calculations is that we can write

$$F_n(\xi)=\xi\ln\xi S(\xi)+R(\xi), \tag{4.7}$$

where $S(\xi)$ and $R(\xi)$ are both single-valued analytic functions of ξ when $|\xi|<\frac{1}{2}a<\frac{1}{4}\epsilon^2$. From (4.7) it is readily seen that $\xi S(\xi)$ is just $-1/2\pi i$ times the "jump" of $F_n(\xi)$ across the positive real axis. That is, if $\xi>0$, then

$$\xi S(\xi)=-(1/2\pi i)[F_n(\xi+0i)-F_n(\xi-0i)]. \tag{4.8}$$

We see that at $\xi=0$, $F_n(\xi)$ has a branch point of infinite order. Furthermore, since $\xi\ln\xi\to 0$ as $\xi\to 0$, there is no "infinity" in the singular part of $F_n(\xi)$ at $\xi=0$.

It should be remarked, that if the condition does not hold that at least one endpoint of the line under consideration is not also the endpoint of an external line, it can still be shown that $\xi=0$ is a branch point of

[16] W. Magnus and F. Oberhetinger, *Formulas and Theorems for the Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1954), p. 17.

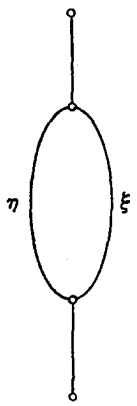


FIG. 2. A self-energy loop shown with corresponding parameters.

infinite order and that there is no "infinity" in the singular part of $F_n(\xi)$ at $\xi=0$. This is done by integrating several times by parts.

B. The Singularity Corresponding to a Self-Energy Loop

The same techniques which were outlined in part A of Sec. IV can be used to study the singularity corresponding to a self-energy loop. The convergence parameters corresponding to the two sides of the loop [shown in Fig. 2] will be called ξ and η to distinguish them from the remaining w_j, which of course will be held fixed in the cut plane. We will write $F_n(z,w)\equiv F_n(\xi,\eta)$. The results in this case are more complex in that the functions exhibting the multiple valuedness of $F_n(\xi,\eta)$ are not elementary functions, but are Spence functions or dilogarithms.[17] The dilogarithm is defined as

$$Li_2(z)=-\int_0^z \frac{\ln(1-t)}{t}dt. \tag{4.9}$$

$Li_2(z)$ has branch points at $z=1$ and $z=\infty$, $Li_2(0)=0$, and it satisfies the two identities:

$$Li_2(z)+Li_2(1/z)=-\tfrac{1}{6}\pi^2-\tfrac{1}{2}[\ln(-z)]^2, \tag{4.10}$$

$$Li_2(z)+Li_2(1-z)=\tfrac{1}{6}\pi^2-\ln z\ln(1-z). \tag{4.11}$$

Now proceeding as in the case of a single internal line, we get

$$\begin{aligned}F_n(\xi,\eta)=&\xi\left\{\tfrac{1}{2}\ln^2\xi+Li_2\left(\frac{\xi-\eta}{\xi}\right)\right\}S_1(\xi,\eta)\\&+\eta\left\{\tfrac{1}{2}\ln^2\eta+Li_2\left(\frac{\eta-\xi}{\eta}\right)\right\}S_2(\xi,\eta)\\&+\xi\ln\xi S_3(\xi,\eta)+\eta\ln\eta S_4(\eta,\xi)\\&+\frac{\xi\ln\xi S_5(\xi)-\eta\ln\eta S_5(\eta)}{\xi-\eta}+R(\xi,\eta).\end{aligned} \tag{4.12}$$

[17] L. Lewin, *Dilogarithms and Associated Functions* (Macdonald, London, England, 1958).

In (4.12), the functions $S_j(\xi,\eta)$, $j=1, 2, 3, 4$, are single valued and analytic in a neighborhood of the origin, and in general $S_j(0,0)\neq 0$. $S_5(z)$ is a single-valued analytic function of z in a neighborhood of the origin, and $S_5(0)\neq 0$, and $R(\xi,\eta)$ is also single valued and analytic in a neighborhood of the origin. The behavior of the two terms involving dilogarithms can be determined with the aid of identities (4.10) and (4.11). The behavior of $F_n(\xi,\eta)$ near the origin is seen to be quite complex. It is multiple valued, and has a logarithmic "infinity" at the origin. This infinity arises from the term

$$\frac{\xi\ln\xi S_5(\xi)-\eta\ln\eta S_5(\eta)}{\xi-\eta}. \tag{4.13}$$

This can be seen quite readily by setting $\eta=0$, and then letting $\xi\to 0$. The rather mild infinity encountered here undoubtedly reflects the fact that the model we are using is a super-renormalizable one. In fact, if we treated this three-field theory with standard techniques, we would find that there is only one primitively divergent integral. This is the p-space integral corresponding to the self-energy loop, and it diverges logarithmically.

C. Behavior of $F_n(z,w)$ as a Function of the Convergence Parameters of More Complex Groups of Internal Lines

We will now briefly indicate the conditions under which the techniques used in discussing the case of a single internal line and a self-energy loop can be extended to the case of larger groups of internal lines and self-energy loops. Let us consider a collection of g internal lines, which we denote by G. We assume that G consists of s single lines and t self-energy loops, so that $s+2t=g$. We want to know the behavior of the singular portion of $F_n(z,w)$ as the convergence parameters corresponding to the lines of G vary in a neighborhood of the origin.

There are two main features of our procedure. The first is that it is necessary to be able to perform a change of variables [as in (4.1)] so that the argument of each Hankel function corresponding to a line of G is of the form w_j-r^2. That is, it depends only on the radius of a single integration variable, and is independent of angles. The second feature is that the integral over all the terms not corresponding to lines of G is an analytic function of the integration variables r^2, corresponding to lines of G.

Let P be the collection of end points of the lines in G. Then any collection of internal lines for which the above two features hold can be exactly characterized by the following two conditions.

Condition I: Of all the lines in the Feynman diagram which are not in G but which have one end point in P, only the lines connected to at most one point of P can be external lines, and the rest must be internal lines.

Condition II: There can be no closed loops in G aside from self-energy loops.

Given any collection of lines G satisfying conditions I and II, the form of the singular part of $F_n(z,w)$ corresponding to the convergence parameters of these lines can be determined. The result is that each single interior line will contribute singularities of the form $w \ln w$, while each self-energy loop will additionally contribute singularities of the form

$$w_j\left\{\tfrac{1}{2}\ln^2 w_j + Li_2\left(\frac{w_j-w_k}{w_j}\right)\right\}$$

and

$$\frac{w_j \ln w_j S(w_j) - w_k \ln w_k S(w_k)}{w_j - w_k}.$$

These singularities will appear both by themselves and multiplied together in various combinations. In fact, the general form in which the singularities will appear can be obtained by writing down a factor (4.7) for each of the s single-internal lines, and a factor (4.12) for each of the t self-energy loops, and then by multiplying them all together. This will not give the coefficients correctly, but it will show the correct form in which the singular functions occur. It should be noticed that no new types of singularities have appeared so far. The worst infinities so far encountered are logarithmic.

In Sec. IV we have been able to exploit (3.15) in order to determine a good deal of information about the singular behavior of $F_n(z,w)$ near $w=0$. Although it is clear that we have not completely determined this singular behavior of $F_n(z,w)$, we are hopeful that the main features have been mapped out, and in particular we believe that no infinities worse than logarithmic will appear.

V. CONCLUSION

In Secs. I through IV we have indicated that by introducing the complex regularizing parameter w, each term in the formal perturbation series for $T(x_1,x_2,x_3)$ can be regarded as the boundary value of an analytic function. We have also examined a number of the analytic properties of these regularized integrals $F_n(z,w)$.

We believe that one possible way of relating our regularized integrals $F_n(z,w)$ to the finally desired renormalized integrals $F_n(z)$ lies in the recent work of Caianiello.[1]

Caianiello has shown that any scheme for taking the finite part of integrals such as (2.1), provided that it satisfies certain conditions, is equivalent to renormalization. A final finite renormalization may be necessary to get complete equivalence with the Dyson-Salam method. Once we have completely determined the singular behavior of $F_n(z,w)$ at $w=0$, the natural procedure will be to remove the singular part, and then set $w=0$ in the remainder. With the aid of Caianiello's criteria, it may be possible to prove that this is a renormalization.

From the remarks concluding part D of Sec. II, we saw that if the variable w were set on the negative real axis, there was a domain in z space in which each $F_n(z,w)$, regarded as a function of z, was analytic, regardless of the position of the w_j on the negative real axis. This domain in z space included the product of the negative real axes. Therefore, even though in our study of the singularities of $F_n(z,w)$ we have required that z satisfy condition R, it seems reasonable to expect that this restriction will not turn out to be a serious hindrance.

We feel that this road to renormalization is a natural one in the light of the remarks made in the introduction.

ACKNOWLEDGMENT

I want to express my deepest thanks to Professor A. S. Wightman for suggesting the research on which this paper is based, and for giving continued advice, encouragement, and enthusiastic support during the course of the research.

Postulates of Quantum Field Theory

R. HAAG AND B. SCHROER*
Department of Physics, University of Illinois, Urbana, Illinois†
(Received August 11, 1961)

We give a survey of the postulates which provide the framework of quantum field theory. Particular attention is given to the lesser known ones, namely, completeness requirements and "primitive causality." Three field theoretical models are analyzed. These examples show that "primitive causality" is not a consequence of the commonly used postulates (which include the "causal commutation relations"). "Asymptotic completeness" is not assured by any of the other postulates listed. In the formulation and analysis the concept of operator rings associated with space-time regions is used extensively.

I. INTRODUCTION

A LARGE part of recent work in quantum field theory has been devoted to a scrutiny of a few basic postulates, sometimes called "axioms." Three of these postulates, namely Lorentz covariance, local (causal) commutation relations, and statements concerning the structure of the energy-momentum spectrum are adopted by practically all authors. Beyond this it has been recognized that the theory should possess a number of other properties in order to be physically reasonable. Among these are completeness requirements, the "asymptotic condition," and a property which we shall call "primitive causality." We shall show here by a discussion of explicit models that the completeness requirements and "primitive causality" are independent of the earlier mentioned postulates. The asymptotic condition on the other hand combines two distinct statements, one of which is a consequence of the three commonly used postulates and the other one is a completeness requirement.

In Sec. II we list all the relevant postulates, give some brief comment on their significance, and describe the main concepts which will be used in the analysis: the polynomial algebra and the von Neumann ring associated with a space-time region. Section III gives a formulation of the completeness requirement in terms of Wightman functions and shows that models recently discussed by Sudarshan and Bardakci[1] violate this postulate. In Sec. IV we analyze Greenberg's "generalized free field."[2] It is found to satisfy the completeness requirement but to violate the "primitive causality." It shows therefore that the latter property is independent of Postulates 1 through 7. In Sec. V another model, which was probably invented by Wightman, is analyzed. It satisfies all the postulates except the one of asymptotic completeness.

* Supported in part by the joint program of the Office of Naval Research and the U. S. Atomic Energy Commission.

† Part of this work was done while the authors were guests at the Summer Institute for Theoretical Physics in Madison, Wisconsin in July, 1961.

[1] E. C. G. Sudarshan and K. Bardakci, J. Math. Phys. 2, 767 (1961).

[2] O. W. Greenberg, "Generalized free fields and models of local field theory" (to be published). See also, A. L. Licht and J. S. Toll, Bull. Am. Phys. Soc. 6, 59 (1961). A similar model has been proposed by M. Wellner (unpublished preprint, 1959).

II. THE POSTULATES

For the sake of completeness we list all the postulates of a sufficiently reasonable nature which have been proposed by various authors.[3–8] In the case of Postulates 2, 5, and 8 two versions are given. In each case the statement labeled (a) is weaker than the one labeled (b). In the formulation of the postulates and in later discussions we aim to maintain a certain amount of mathematical conciseness, but a completely rigorous treatment is beyond our ability and beyond the scope of this paper. For simplicity, we consider only theories of a single, scalar field. The generalization to the case of several fields (of scalar, spinor, vector type, etc.) is probably obvious. The reader who is unfamiliar with some of the mathematical concepts used may find some help in the Appendix where the less usual terms are defined.

The Manifold of States

Postulate 1. The states of the system correspond to the vectors of a (separable) Hilbert space $\mathfrak{H}$ with positive definite metric.[9]

Definition of the Field

The "field" is introduced in the following way:

Postulate 2(a). For every point x in space time we have a bilinear form $A(x)$ in the Hilbert space.

This means that if Ψ and Φ are two vectors of $\mathfrak{H}$ which lie in a certain dense domain D_0, then the matrix element

[3] Postulates 1 through 5 and Postulate 7 have been most thoroughly discussed in references 4 and 5. See also references 6 and 7. The completeness postulate 6 is mentioned in references 6 and 7 but has so far never been used. Primitive causality was postulated in reference 8. The first correct description of the asymptotic condition is given in reference 7.

[4] A. S. Wightman, Phys. Rev. 101, 860 (1956).

[5] A. S. Wightman, in *Colloque international sur les problemes mathematiques de la theorie quantique des champs*, Lille, 1957 (Centre National de la Recherche Scientifique, Paris, 1959).

[6] R. Haag, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. 29, No. 12 (1955).

[7] H. Lehmann, K. Symansik, and W. Zimmermann, Nuovo cimento 1, 205 (1955).

[8] R. Haag in reference 5.

[9] The case of a Hilbert space with indefinite metric is of some interest (Gupta-Bleuler formulation of quantum electrodynamics, Heisenberg's approach to a theory of elementary particles). We have no comments on that case in this paper.

$\langle\Phi|A(x)|\Psi\rangle$ should be a finite number and it shall, of course, depend linearly on Ψ, antilinearly on Φ.

Postulate 2(b). If $f(x)$ is any real function (in space time) which belongs to class $\mathcal{S}$, i.e., which is infinitely often differentiable and vanishes stronger than any power for large $|x_\mu|$, then

$$A(f)=\int A(x)f(x)d^4x \tag{1}$$

shall be an (unbounded) self-adjoint operator in $\mathfrak{H}$. The same shall be true of the "polynomials" in the field

$$\int f(x)A(x)d^4x+\int f^{(2)}(x_1x_2)A(x_1)A(x_2)d^4x_1d^4x_2+\cdots$$

$$\times\int f^{(n)}(x_1\cdots x_n)A(x_1)\cdots A(x_n)d^4x_1\cdots d^4x_n, \tag{2}$$

provided that the coefficient functions $f^{(n)}$ are functions of class $\mathcal{S}$ in the $4n$-dimensional configuration space and that they possess the reality property

$$f^{(n)*}(x_1\cdots x_n)=f^{(n)}(x_n\cdots x_1). \tag{3}$$

There shall be a common dense domain D_1 in $\mathfrak{H}$ on which all these polynomials are applicable. If we do not insist on self-adjoint operators we can use polynomials whose coefficient functions $f^{(n)}$ are not restricted by (3).

. . .

It is convenient, at this point, to introduce two algebraic objects which may be associated with any space-time region $\mathfrak{G}$. Consider the subclass of complex functions $f^{(n)}(x_1\cdots x_n)$ in $\mathcal{S}$, which vanish if any argument x_i lies outside the region $\mathfrak{G}$. The collection of all polynomials of the field with such restricted coefficient functions, forms an algebra. It will be called the polynomial algebra of the region $\mathfrak{G}$ and denoted by $\mathcal{P}_\mathfrak{G}$.[10] Its elements are unbounded operators in $\mathfrak{H}$. Since the mathematical theory of algebras of unbounded operators is barely existent, it is useful to consider along with $\mathcal{P}_\mathfrak{G}$ the algebra of bounded opertors (von Neumann ring) which is associated with $\mathcal{P}_\mathfrak{G}$.[11] This ring will be denoted by $R_\mathfrak{G}$. Practically speaking, $R_\mathfrak{G}$ consists of all bounded operators which are functions of the operators of $\mathcal{P}_\mathfrak{G}$. If the region $\mathfrak{G}$ is the whole of space-time, then we write $\mathcal{P}_\infty$, R_∞ for the corresponding algebras.

One remark of caution may be added regarding Postulate 2(b). Although the functions of class $\mathcal{S}$ are very decent indeed (namely, infinitely often differentiable) there are several known field theoretic models in which the functions are not decent enough to make $A(f)$ an operator. In these models one has to restrict oneself to test functions $f(x)$ whose Fourier transform differs from zero only in a finite region of momentum space (integer analytic functions in x space). In the Lagrangian approach these models would be classified as "unrenormalizable." This suggests that one might regard 2(b) as a postulate which excludes "nonrenormalizable" theories. Of course also the possibility of defining $\mathcal{P}_\mathfrak{G}$ and $R_\mathfrak{G}$ (physically speaking, of defining observables which relate to exactly specified space-time regions) depends crucially on 2(b).

Lorentz Invariance

Postulate 3. To every inhomogeneous Lorentz transformation[12] (a,Λ), there is a unitary operator $U(a,\Lambda)$ in $\mathfrak{H}$. These operators form a representation of the inhomogeneous Lorentz group in $\mathfrak{H}$.

Covariance of the Field

Postulate 4.

$$U(a,\Lambda)A(x)U^{-1}(a,\Lambda)=A(\Lambda x+a). \tag{4}$$

Structure of Energy-Momentum Spectrum

The energy-momentum operators $P\mu$ are the infinitesimal generators of the translation operators $U(a)$:

$$U(a)=e^{-iPa}. \quad Pa=\mathbf{Pa}-P_0a_0. \tag{5}$$

The mass operator is

$$M=(-P^2)^{\frac{1}{2}}; \quad P^2=\mathbf{P}^2-P_0^2. \tag{6}$$

We state the spectral postulate first in its weakest form:

Postulate 5(a). The four-vector formed by the simultaneous eigenvalues of the four components of $P\mu$ never lies outside of the forward light cone (i.e., the energy is always non-negative, the mass is real).

. . .

A much stronger statement can be made which relates the representation of the inhomogeneous Lorentz group uniquely (up to equivalence) to the mass and spin values of the stable particles which are described by the theory. The details of this connection are irrelevant to our present purpose. (See references 5 and 6.) The following slightly stronger version of the spectral postulate will, however, be useful.

Postulate 5(b). There is exactly one state in $\mathfrak{H}$ which has zero energy momentum (and is invariant under all Lorentz transformations). This is the vacuum state $|0\rangle$. There are one or several discrete eigenvalues $m_1, m_2, \cdots$ of the mass operator, corresponding to the states of a single stable particle and there is a continuum of mass values above $2m_1$.

[10] If X and Y belong to $P_\mathfrak{G}$ and c is a complex number, then clearly also XY, $X+Y$, cX, and the adjoint X^+ belong to $P_\mathfrak{G}$. The coefficient functions of X^+ are obtained from those of X by forming the complex conjugate and reversing the order of the arguments $x_1\cdots x_n$.

[11] An exact definition and a description of a few basic properties of von Neumann rings is given in the Appendix. For the mathematical theory of von Neumann rings see e.g., M. A. Naimark, *Normed Rings*; translated from 1st Russian edition by L. F. Boron (P. Noordhoff, Groningen, 1959). We say that an unbounded self-adjoint operator T is "associated" with the von Neumann ring R if all the spectral projectors of T belong to R.

[12] Here Λ stands for a homogeneous Lorentz transformation; a is a translation 4-vector.

Completeness (Irreducibility)

Let $R(\mathfrak{H})$ be the von Neumann ring which consists of *all* bounded operators in $\mathfrak{H}$. We call a collection of operators *complete* if the von Neumann ring associated with them is identical with $R(\mathfrak{H})$. Another equivalent definition of completeness of a set $\mathfrak{A}$ of operators is that there exists no bounded operator (apart from multiples of the identity) which commutes with all the operators in $\mathfrak{A}$ and $\mathfrak{A}^\dagger$. The Hilbert space is then irreducible with respect to the algebra generated by $\mathfrak{A}$, $\mathfrak{A}^\dagger$, i.e., it does not contain any invariant subspace. The two terms "completeness of a collection of operators" and "irreducibility of $\mathfrak{H}$" with respect to that collection are used synonymously.

Postulate 6. (See note added in proof). The set of operators consisting of all the $A(f)$ with $f\in\mathfrak{S}$ shall be complete:

$$R_\infty = R(\mathfrak{H}). \tag{7}$$

Causality Postulates

The usual causality postulate which is adopted by practically all authors may be expressed in the following way:

Postulate 7. (Einstein causality) If $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are two space-time regions which are completely space-like with respect to each other then $R_{\mathfrak{G}_1}$ and $R_{\mathfrak{G}_2}$ commute.

. . .

This is an expression of the principle that no physical effect can travel faster than the velocity of light and hence measurements in the two regions $\mathfrak{G}_1$ and $\mathfrak{G}_2$ cannot disturb each other. To distinguish this principle from another causality condition (Postulate 8) we shall call it "Einstein causality."

Obviously there is no remnant of principle 7 in a nonrelativistic theory. Yet the concept of causality is also relevant there. We formulate therefore another causality principle which should hold both in relativistic and nonrelativistic theories.

Postulate 8(a). ("Primitive causality") Let $\mathcal{T}$ be a "time slice," i.e., a region in space-time, infinitely extended in space but restricted in the time coordinate to an interval

$$t-\tau<x_0<t+\tau; \quad |\mathbf{x}| \text{ arbitrary}. \tag{8}$$

Then we require

$$R_\tau = R_\infty \text{ for any } \tau. \tag{9}$$

. . .

This, together with Postulate 6, means that we can uniquely determine a state of the system by observations made in an arbitrarily small time interval. If Postulate 6 holds but 8 is violated, then we need observations at all times to determine a state, and hence, statistical predictions from observations at one time to those at another time would not in general be possible.

One can sharpen Postulate 8(a) by including the special features of relativistic dynamics. Let us call a region $\mathfrak{G}_2$ causally dependent on region $\mathfrak{G}_1$ if every ray in the backward cone (or the forward light cone) originating from any point in $\mathfrak{G}_2$ passes through $\mathfrak{G}_1$. Then we want

Postulate 8(b). $R_{\mathfrak{G}_2}$ is contained in $R_{\mathfrak{G}_1}(R_{\mathfrak{G}_2}\subset R_{\mathfrak{G}_1})$ if $\mathfrak{G}_2$ is causally dependent on $\mathfrak{G}_1$.

. . .

Loosely speaking, Postulate 8(a) stipulates that there should be a field equation which determines the field at an arbitrary time from that in a time slice. Postulate 8(b) then says that this field equation should have a hyperbolic propagation character. We shall show that Postulate 8(a) is independent of Postulate 7 and all the previously mentioned postulates. We can derive 8(b), however, from 8(a) and 7 if we add one technical assumption about the nature of the rings $R_{\mathfrak{G}}$. (See Sec. VI.)

Asymptotic Condition

Originally the asymptotic condition served a dual purpose. On the one hand, it was a definition of how the mathematical objects discussed above should be related to quantities of physical interest, namely the cross sections for collision processes of particles. On the other hand, it imposed further restrictions on the mathematical scheme which have to be met in order to make a particle interpretation possible at all. In recent years it was recognized that a large fraction of the statements which had been combined into the "asymptotic condition" was, in fact, already implied by the earlier mentioned postulates and their physical interpretation. While talking about the causality postulates we had tacitly assumed that the "observables" (=Hermitian operators) in $R_{\mathfrak{G}}$ correspond to physical measurements which can be performed within the region $\mathfrak{G}$. Otherwise neither Postulate 7 nor 8 would make sense. This tacit assumption provides, however, a sufficient amount of physical interpretation of the theory to uniquely determine which states in $\mathfrak{H}$ (if any) correspond to a given configuration of incoming or outgoing particles. Thus the asymptotic condition is not necessary for the purpose of physical interpretation.[13,14] Also it was found that one part of it (Postulate 9) is a consequence of the earlier mentioned postulates. There remains, however, a part (Postulate 10) which does impose a further restriction on the theory. For clarity we separate the asymptotic condition into two distinct statements.

Postulate 9.[15] Let $g(x)$ be a normalizable solution of the Klein-Gordon equation[16] to the mass m, and Ψ, Φ a

[13] R. Haag, Phys. Rev. 112, 669 (1958).

[14] In reference 13 the statements in Postulate 9 were proved starting from certain assumptions about the behavior of vacuum expectation values $\langle 0|A(x_1)\cdots A(x_n)|0\rangle$ when all times are equal and the space distances large. Recently D. Ruelle (to be published) succeeded in proving these assumptions from the Postulates 1 through 5 and 7, provided that there is no particle with vanishing rest mass in the theory.

[15] As mentioned above, this is not an independent postulate.

[16] i.e., $(g,g)=i\int[g^*(x)(\partial g/\partial x_0)-(\partial g^*/\partial x_0)g(x)]d^3x<\infty$.

pair of states from a certain dense domain of $\mathfrak{H}$. Then

$$\lim_{t\to\pm\infty}\left\langle\Phi\left|i\int_{x_0=t}\left[A(x)\frac{\partial g}{\partial x_0}-\frac{\partial A}{\partial x_0}g(x)\right]d^3x\right|\Psi\right\rangle \quad (10)$$

exists. We have adopted the convention $\langle 0|A(x)|0\rangle=0$ which can always be achieved by subtracting a suitable constant from A. The limit is zero if m is not one of the masses of the stable particles which are described by the theory or if the matrix elements of $A(x)$ between the vacuum and the states of a single particle of that particular mass happen to vanish. To avoid burdening the discussion by unessential detail let us assume that the latter case does not arise, i.e.,

$$\langle p,i|A(0)|0\rangle=c_i\neq 0 \quad (11)$$

for all i.[17–19] Here $|p,i\rangle$ denotes a state of a single particle of type i and with momentum p. If the particle has no spin then c_i is, of course, independent of p due to Lorentz invariance.

If we make the simplifying assumption (11) then the limits (10) differ from zero exactly if m is chosen to be one of the masses m_i. For each m_i and each g the limit of (10) for $t\to\infty$ defines an operator (since we can let Φ and Ψ vary through a dense domain). These limit operators will be denoted by $c_iA_{i,g}{}^{\text{in}\dagger}$. Similarly the limit $t\to+\infty$ defines the operators $c_iA_{i,g}{}^{\text{out}\dagger}$.

The operators $A_{i,g}{}^{\text{in}\dagger}$, $A_{i,g}{}^{\text{out}\dagger}$ are interpreted as the creation operators of incoming (outgoing) particles of type i and wave function g. Let $\mathfrak{H}^{(\text{in})}$ be the subspace of $\mathfrak{H}$ (hopefully the whole space) which is generated by repeated application of the $A_{i,g}{}^{\text{in}\dagger}$ on the vacuum. Then within $\mathfrak{H}^{(\text{in})}$ the $A_{i,g}{}^{\text{in}\dagger}$ have all the properties of a system of creation operators of noninteracting particles, in particular

$$[A_{i,g}{}^{\text{in}},A_{k,g'}{}^{\text{in}\dagger}]=\delta_{ik}(g,g'). \quad (12)$$

Here $A_g{}^{\text{in}}$ means the adjoint of $A_g{}^{\text{in}\dagger}$ and (g,g') is the scalar product of Klein-Gordon wave functions as defined in footnote 16.

As mentioned above, all the statements contained in Postulate 9 may be derived from the earlier postulates. In order to have a reasonable particle theory we need, however, still one more property:

Postulate 10. (Asymptotic completeness)

$$\mathfrak{H}^{\text{in}}=\mathfrak{H}=\mathfrak{H}^{\text{out}}. \quad (13)$$

We shall see that this requirement does not follow from the other postulates. If it holds, then the statements 9 imply that the S matrix is unitary. In fact the slightly weaker requirement $\mathfrak{H}^{\text{in}}=\mathfrak{H}^{\text{out}}$ would suffice for that purpose. If it does not hold then there are states in the theory which cannot be described in terms of asymptotic particle configurations.

[17] The treatment of the case where some c_i vanish is given in references 13, 18, and 19.

[18] W. Zimmermann, Nuovo cimento 4, 597 (1958).

[19] K. Nishijima, Phys. Rev. 111, 995 (1958).

III. COMPLETENESS

Wightman[4] has shown that if the set of functions

$$W^{(n)}(x_1\cdots x_n)=\langle 0|A(x_1)\cdots A(x_n)|0\rangle \quad (14)$$

is known, then the theory is completely determined. The Hilbert space $\mathfrak{H}$ and the operators $A(f)$ and $U(a,\Lambda)$ in it, can be easily constructed with the help of the $W^{(n)}$. The restrictions which the Postulates 1–5 and 7 impose on the functions $W^{(n)}$ have been given in reference 4.

Let us formulate the completeness postulate 6 in terms of the Wightman functions $W^{(n)}$. For this purpose it is useful to realize that the construction of the field theory from the vacuum expectation values reduces to a standard technique in mathematics if one uses a slightly more abstract terminology: The $W^{(n)}$ may be regarded as defining a positive linear form over the polynomial algebra $\mathcal{P}_\infty$. This means the following. If the $W^{(n)}$ are known, then we can, by virtue of (14), assign to every element Y of $\mathcal{P}_\infty$ [i.e., every operator of the form (2)] a number $F(Y)$ which shall be the vacuum expectation value of Y:

$$F(Y)=\langle 0|Y|0\rangle. \quad (15)$$

The dependence on Y is linear, i.e.,

$$F(aY_1+bY_2)=aF(Y_1)+bF(Y_2). \quad (16)$$

Furthermore, for positive operators this number must be positive (or rather non-negative).

$$F(Y^\dagger Y)\geqslant 0 \text{ for any } Y \text{ in } \mathcal{P}_\infty. \quad (17)$$

If a positive linear form over an algebra is given, then we also have a representation of that algebra by operators in a Hilbert space. What one does essentially is to consider the elements of the algebra as the vectors of the Hilbert space. Let Ψ_Y denote the vector which corresponds to the element Y of the algebra. The scalar product is then defined by

$$\langle\Psi_Y,\Psi_X\rangle=F(Y^\dagger X). \quad (18)$$

The operator corresponding to Y in the Hilbert space is defined by

$$Y\Psi_X=\Psi_{YX}. \quad (19)$$

This method will be called the Gelfand construction.[20] The crucial point is, of course, the positive definiteness condition (17) which is easily expressed in terms of the $W^{(n)}$ [see reference 4, Eq. (21)], but not as easily satisfied. A similar situation pertains with respect to the completeness of the operator algebra $\mathcal{P}_\infty$ in the representation space. According to the general theory,[20] $\mathcal{P}_\infty$ is complete if and only if the positive linear form F cannot be decomposed into a sum of two other distinct positive linear forms. In a formula, if

$$W^{(n)}=W_1{}^{(n)}+W_2{}^{(n)}, \quad (20)$$

[20] See, e.g., Naimark.[11] Actually the construction there is carried out for normed algebras but if one is not afraid of unbounded operators, the norm is not needed in this specific context.

where $W_1^{(n)} \neq \lambda W^{(n)}$ and both W_1 and W_2 satisfy the positivity condition, then $\mathcal{P}_\infty$ is *not* complete in the representation space constructed with the help of W. Therefore Postulate 6 requires that W be indecomposable in the sense of Eq. (20). Recently Sudarshan and Bardakci[1] discussed field theoretical models which are obtained just by taking Wightman functions of the type (20) where W_1 and W_2 are the Wightman functions of two other theories. From what has been said it is clear that these models necessarily violate the completeness postulate. (See note added in proof).

IV. PRIMITIVE CAUSALITY

Another recently described model is the "generalized free field" of Greenberg.[2] It satisfied Postulates 1–5 as well as 7 in an obvious manner. We shall show that it also satisfies the completeness condition 6, but not the primitive causality 8. It serves therefore to show that the primitive causality is not a consequence of the earlier postulates.

The model is defined by prescribing the commutation relations

$$[A(x),A(y)]=i\int_0^\infty \rho(\kappa^2)\Delta(x-y,\kappa)d\kappa^2 \qquad (21)$$

where $\rho(\kappa^2)$ is a non-negative weight function and $\Delta(x-y,\kappa)$ the usual commutator function of a free Klein-Gordon field with mass κ. We shall take

$$\rho(\kappa^2)=\delta(\kappa^2+m^2)+\sigma(\kappa^2), \qquad (22)$$

where $\sigma(\kappa^2)$ vanishes below $\kappa=2m$ and is a smooth, non-vanishing function above this threshold. In addition we assume a unique vacuum state $|0\rangle$ from which the Hilbert space is generated by application of polynomials of A and we adopt the convention $\langle 0|A(x)|0\rangle=0$. Then it follows immediately that, as in the case of a free field, A has no Fourier components for space-like momenta. Therefore it can be uniquely decomposed into a creation part (negative frequency part) $A^{(-)}$ and a destruction part $A^{(+)}$. The Hilbert space may then be constructed in exactly the same manner which is used for a free field. One easily sees that $\mathfrak{H}$ is separable.

We first prove the completeness of R_∞ in this model using the lemma:

Lemma: If the representation space $\mathfrak{H}$ of the ring R contains a cyclic vector[21] Φ, and if R contains also the projection operator on that vector, then R is complete.

Proof of the lemma: Let Ψ be an arbitrary vector in $\mathfrak{H}$, Φ the cyclic vector, and P_ϕ the projection operator on Φ. In Dirac's notation $P_\phi=|\Phi\rangle\langle\Phi|$. There exists an operator $T\in R$ such that

$$\langle\Phi,T\Psi\rangle\neq 0.$$

Otherwise $\langle R^\dagger\Phi,\Psi\rangle=\langle R\Phi,\Psi\rangle=0$ which would imply $\Psi=0$ due to the cyclicity of Φ. Therefore, given any two vectors Ψ and Ψ' we can find operators T and T' in R such that $T'P_\phi T\Psi$ is arbitrarily close to Ψ'. Here again the cyclicity of Φ is used. If $P_\phi\in R$ then $T'P_\phi T\in R$ and hence there can be no invariant subspace.

Application to our case: It is easily seen that the vacuum is a cyclic vector with respect to R_∞. In the first place, by the very construction of $\mathfrak{H}$, $\mathcal{P}_\infty|0\rangle$ and even $(\mathcal{P}_\infty)_H|0\rangle$ is a dense set.[22] On the other hand, if C is an Hermitian operator in $\mathcal{P}_\infty$, E_λ its spectral projection corresponding to eigenvalues less than λ, then $CE_\lambda\in R_\infty$. The state $\Psi=C|0\rangle$ is arbitrarily well approximated by $\Psi_\lambda=CE_\lambda|0\rangle$ as $\lambda\to\infty$.

We have to show now that the projector on the vacuum (denoted by P_0) is contained in R_∞. For this purpose we construct a countable collection of Hermitean operators in $\mathcal{P}_\infty$ with the common property of having the vacuum as an eigenstate to the eigenvalue 0. If one such operator is N_k, then the projector $P_0^{(k)}$ which projects into the subspace belonging to the eigenvalue 0 of N_k is an element of R_∞. If, besides the vacuum, there is no other simultaneous eigenstate of all the N_k to the eigenvalue zero, then,

$$P_0=\prod_k P_0^{(k)}. \qquad (23)$$

The infinite product on the right-hand side still belongs to R_∞ since it is the strong limit of a sequence of operators in R_∞. The collection N_k is formed in complete analogy to the occupation number operators of free field theory. Consider a set of functions $f_k(p)$ which vanish for momenta outside the forward light cone, satisfy the orthogonality conditions

$$\int f_k^*(p)f_l(p)\rho(-p^2)d^4p=\delta_{kl} \qquad (24)$$

and are a complete basis in the Hilbert space formed by the square integrable functions $\varphi(p)$ with respect to the measure $d\mu(p)=\rho(-p^2)d^4p$. Take

$$a_k=\int f_k(p)e^{ipx}A(x)d^4pd^4x \qquad (25)$$

$$N_k=a_k^+a_k. \qquad (26)$$

One shows exactly, as for a free field, that the N_k have the desired properties.

We have seen thus that the completeness postulate 6 is satisfied. Now we show that the "primitive causality" (Postulate 8) is violated. Consider the time slice $\mathcal{T}$ consisting of all points with the time coordinate $|x_0|<\tau$ and let $\mathcal{P}_\tau$ and R_τ be the corresponding polynomial and von Neumann algebras. We construct an Hermitian operator

$$C=\int c(x)A(x)d^4x;\quad C\in\mathcal{P}_\infty \qquad (27)$$

[21] Φ is called cyclic if $R\Phi$ is a dense set of vectors in the whole space.

[22] $(\mathcal{P}_\infty)_H$ means the collection of Hermitian operators in $\mathfrak{H}$.

which commutes will all elements of $\mathcal{P}_\tau$ (in the common dense domain). It will then also commute with R_τ and since C is not a multiple of the identity, R_τ is not complete. The formal expression for the commutator between C as given by (27) and $A(x)$ is the c-number function

$$f(x)=[A(x),C]=i\int\Delta(x-x',\kappa)\rho(\kappa^2)d\kappa^2 c(x')d^4x'. \quad (28)$$

Its Fourier transform

$$\tilde{f}(p)=\epsilon(p)\rho(-p^2)\tilde{c}(p) \quad (29)$$

vanishes where ρ vanishes and has the symmetry property

$$\tilde{f}^*(-p)=-\tilde{f}(p) \quad (30)$$

if C is Hermitian. If we can find a function $f(x)$ in class $\mathcal{S}$ with the properties

$$\begin{aligned} f(x)&=0 \quad \text{for} \quad |x_0|<\tau \\ \tilde{f}(p)&=0 \quad \text{for} \quad -p^2<4m^2 \end{aligned} \quad (31)$$

and the symmetry (30) then we can define

$$\tilde{c}(p)=\begin{cases} 0 & \text{for} \quad -p^2<4m^2 \\ \dfrac{\tilde{f}(p)}{\epsilon(p)\rho(-p^2)} & \text{for} \quad -p^2>4m^2. \end{cases} \quad (32)$$

In this manner we get an Hermitian operator C which commutes with all $A(x)$ for $|x_0|<\tau$. The construction of such a function $f(x)$ is easily done with the help of the Jost-Lehmann-Dyson method.[23] Consider a solution of the 5-dimensional wave equation

$$\left(\Delta-\frac{\partial^2}{\partial x_0^2}+\frac{\partial^2}{\partial\xi^2}-4m^2\right)\varphi(x,\xi)=0 \quad (33)$$

where ξ is a single 5th variable and put

$$f(x)=\lim_{\xi\to 0}\varphi(x,\xi). \quad (34)$$

This function automatically possesses the support properties demanded for $\tilde{f}(p)$ in momentum space. The function $\varphi(x,\xi)$ is determined by its values and the values of its time derivative in the 4-dimensional plane $x_0=0$. Due to the hyperbolic propagtion character, the values of φ at a point (x,ξ) are influenced only by the initial values in the region

$$|\mathbf{x}-\mathbf{x}'|^2+|\xi-\xi'|^2<x_0'^2 \quad (35)$$

In particular, if we chose

$$\varphi\Big|_{x_0=0}=0; \quad \frac{\partial\varphi}{\partial x_0}\Big|_{x_0=0}=0 \quad \text{for} \quad \xi<\tau \quad (36)$$

then

$$f(x)=\varphi(x,\xi)\Big|_{\xi=0}=0 \quad \text{for} \quad |x_0|<\tau \quad (37)$$

so that (31) is satisfied. Condition (30) just means that $f(x)$ shall be chosen real.

We see therefore that the Greenberg model violates the "primitive causality" (Postulate 8) but satisfies Postulates 1–7. Greenberg[2] has proved that the model also violates asymptotic completeness. Postulate 9 is satisfied in a trivial fashion.

V. ASYMPTOTIC COMPLETENESS

Next we give an example which violates the asymptotic completeness while satisfying all other mentioned postulates. The model in question is also an old friend. The field is defined by

$$A(x)=:A_0^2(x):. \quad (38)$$

The double dots denote the Wick product. A_0 is a free Klein-Gordon field. The Hilbert space is that subspace of the free-field Hilbert space which results if one applies only polynomials of A (instead of A_0) on the vacuum. Obviously $\mathfrak{H}$ contains then only states with an even number of particles. There is no discrete eigenstate of the mass operator (apart from the vacuum state) in $\mathfrak{H}$. Postulates 1–5 and Einstein causality (Postulate 7) are satisfied.[24] The only nontrivial problem in this connection is to check Postulate 2(b) which has been done by Wightman.[25] The first part of the asymptotic condition (Postulate 9) is trivially satisfied; there are no incoming fields, the limits of (10) vanish. Hence asymptotic completeness is flagrantly violated. We shall show now that the completeness postulate 6 and the "primitive causality" (Postulate 8) are satisfied in this model.

The validity of Postulate 6 is demonstrated in much the same way as in the case of the Greenberg model. Instead of the N_k, we use here operators

$$A(f)=\int A(x)f(x)d^4x \quad (39)$$

with functions f whose Fourier transforms are nonzero only for space-like momenta. These operators all have the vacuum as an eigenstate to eigenvalue zero and one shows that there is no other vector in $\mathfrak{H}$ which is annihilated by all these $A(f)$. We omit the details of the calculation.

To demonstrate that primitive causality holds in this model we construct the Hamiltonian operator using only the field A and its derivatives in a small time interval τ. The Hamiltonian density expressed in terms of A_0 is

$$h=\tfrac{1}{2}\{:(\mathrm{grad}A_0)^2:+:\dot{A}_0^2:+m^2:A_0^2:\}. \quad (40)$$

On the other hand we have

$$B(x)\equiv\tfrac{1}{2}\Box A(x)=:(\mathrm{grad}A_0^2):-:\dot{A}_0^2:+m^2:A_0^2: \quad (41)$$

[23] R. Jost and H. Lehmann, Nuovo cimento **5**, 1460 (1957); F. J. Dyson, Phys. Rev. **110**, 1460 (1958).

[24] The strong form of Postulate 5 is not satisfied since there are no discrete eigenvalues (besides 0) of the mass operator in the model. This default can, however, be easily amended.

[25] A. S. Wightman, Paris lectures, 1956 (unpublished).

and we shall show that

$$C(x)=\tfrac{1}{3}:(\mathrm{grad}A_0)^2:+:\dot{A}_0^2: \tag{42}$$

is also a local function[26] of A. Clearly then

$$h(x)=\tfrac{1}{4}[m^2A(x)+B(x)+3C(x)]. \tag{43}$$

To obtain C consider the following expression

$$C_\lambda(f,g)=\int f(t)g\left[\frac{t'-t}{\lambda}\right]$$
$$\times[B(\mathbf{x},t)B(\mathbf{x},t')-\langle 0|B(\mathbf{x},t)B(\mathbf{x}',t')|0\rangle]dtdt' \tag{44}$$

where $B=\frac{1}{2}\Box A$ and where $f(t)$, $g(t)$ are smooth test functions whose support is inside the interval $-(\tau-\delta)<t<(\tau-\delta)$. The averaging over the times t, t' is sufficient to make $C_\lambda(f,g)$ a good operator. For this reason we have not introduced any averaging over the space coordinates $\mathbf{x}$ in (44). If λ is sufficiently small then $C_\lambda(f,g)$ is associated with the ring R_τ of the time slice. We want to look at the behavior of C_λ for $\lambda\to 0$. The Wick ordering of the bracket in (44) gives

$$\begin{aligned}B(x)B(y)-\langle 0|B(x)B(y)|0\rangle&\\ =:B(x)B(y):-4i:[\partial_\mu A_0(x)][\partial_\nu A_0(y)]:&\\ \times\partial^\mu\partial^\nu\Delta^{(+)}(x-y)+4im^2:[\partial_\mu A_0(x)]A_0(y):&\\ \times\partial^\mu\Delta^{(+)}(x-y)-4im^2:A_0(x)[\partial_\mu A_0(y)]:&\\ \times\partial^\mu\Delta^{(+)}(x-y)+4im^4:A_0(x)A_0(y):\Delta^{(+)}(x-y).&\end{aligned} \tag{45}$$

Imagine this inserted in (44) and then $C_\lambda(f,g)$ applied to a typical m-particle state Ψ with a momentum space wave function which decreases rapidly for large momenta. As λ gets smaller the contributions from the different terms in (45) to $\|C_\lambda(f,g)\psi\|$ will behave quite differently. The first one vanishes proportional to λ, the others become singular at $\lambda=0$, the strength of the singularity depending on the number of derivatives of the $\Delta^{(+)}$ function. Let us evaluate the term with the strongest singularity (i.e., the one which contains two differentiations of the $\Delta^{(+)}$ function). For simplicity we put $\mathbf{x}=0$. The contribution is then, apart from irrelevant numerical constants

$$\lambda\int p_\mu^{(1)}p_\nu^{(2)}q^\mu q^\nu:\tilde{A}_0(p^{(1)})\tilde{A}_0(p^{(2)}):\tilde{f}(p_0^{(1)}+p_0^{(2)})$$
$$\times\tilde{g}[\lambda(q_0-p_0^{(2)})]\delta(q^2+m^2)\theta(q)d^4qd^4p^{(1)}d^4p^{(2)}\Psi. \tag{46}$$

Here, $\tilde{f}$, $\tilde{g}$, $\tilde{A}_0$ are the Fourier transforms of f, g, A_0, respectively. The integration over $p^{(1)}$ and $p^{(2)}$ gives no trouble. If $p_0^{(i)}$ is positive, i.e., if $\tilde{A}(p^{(i)})$ represents a destruction operator, then a cutoff is provided by the wave function of the state Ψ. If one or both of the $p_0^{(i)}$ are negative then $\tilde{f}$ provides the cutoff for one or for both those momenta. Thus the integration volume with respect to $p^{(1)}$, $p^{(2)}$ may be considered as essentially a finite one. The q integration, on the other hand, diverges for $\lambda=0$. For small λ the cutoff introduced by $\tilde{g}$ becomes effective only at very high momenta. Therefore we can neglect $p_0^{(2)}$ in the argument of $\tilde{g}$ if $\lambda\to 0$ and write (46) as

$$\lambda\int q^\mu q^\nu\delta(q^2+m^2)\theta(q)\tilde{g}(\lambda q_0)d^4q\int\tilde{f}(p_0^{(1)}+p_0^{(2)})p_\mu^{(1)}p_\nu^{(2)}$$
$$:\tilde{A}_0(p^{(1)})\tilde{A}_0(p^{(2)}):d^4p^{(1)}d^4p^{(2)}\Psi$$
$$=\lambda\cdot I^{\mu\nu}\int_{|\mathbf{x}|=0}f(t):\partial_\mu A_0(x)\partial_\nu A_0(x):dt\Psi \tag{47}$$

where $I^{\mu\nu}$ stands for the first integral. We have

$$I^{\mu\nu}=0 \text{ if } \mu\neq\nu \text{ for symmetry reasons.} \tag{48}$$

$$\begin{aligned}I^{ii}&=\frac{1}{3}\int\mathbf{q}^2\delta(q^2+m^2)\theta(q)\tilde{g}(\lambda q_0)d^4q\\ &=\tfrac{1}{3}I^{00}-\tfrac{1}{3}m^2\int\delta(q^2+m^2)\theta(q)\tilde{g}(\lambda q_0)d^4q.\end{aligned} \tag{49}$$

Since the last term in (49) is less singular than I^{00} we have in the limit $\lambda\to 0$

$$I^{ii}=\tfrac{1}{3}I^{00}. \tag{50}$$

Finally

$$I^{00}(\lambda)=4\pi\int\frac{q_0^2}{2q_0}\tilde{g}(\lambda q_0)(q_0^2-m^2)^{\frac{1}{2}}q_0dq_0$$
$$\approx 2\pi\int q_0^3\tilde{g}(\lambda q_0)dq_0. \tag{51}$$
$$I^{00}(\lambda)=2\pi\lambda^{-4}\int n^3\tilde{g}(n)dn.$$

Thus we get

$$\lim_{\lambda\to 0}[\lambda I^{00}(\lambda)]^{-1}C_\lambda(f,g)\Psi$$
$$=\int f(t)[\tfrac{1}{3}:(\mathrm{grad}A_0)^2+:\dot{A}_0^2:]dt\Psi. \tag{52}$$

We see that in a dense domain of states the Hermitian operator $C(f)=\int C(x)f(x)d^4x$ with C defined by (42) and f vanishing outside a region $\mathfrak{G}$, is obtainable as the strong limit of a sequence of Hermitian operators from $\mathcal{P}_{\mathfrak{G}}$.[27] To prove that C is a local function of A we still have to go over from the unbounded operators to the von Neumann rings with which they are associated. This may be done by means of the following theorem[28]:

[26] The statement "C is a local function of A" shall mean that the von Neumann ring generated by the field C in any region $\mathfrak{G}$ is contained in the von Neumann ring generated by A in the same region.

[27] Actually, we have used $\mathcal{P}_{\mathfrak{G}'}$ with $\mathfrak{G}'$ slightly larger than $\mathfrak{G}$ in our construction, but this could be avoided by a more refined argument.

[28] F. Riesz and B. Sz.-Nagy, *Vorlesungen über Funktionalanalysis* (Deutscher Verlag der Wissen., Berlin, Germany, 1956), Chap. IX.

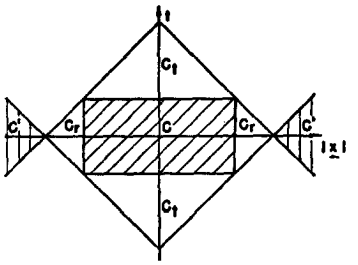


FIG. 1.

Let T_n be a sequence of unbounded self adjoint operators and let

$$\lim_{n\to\infty} T_n\Psi = T\Psi \tag{53}$$

for a dense domain of states Ψ; the limit T shall again be a self-adjoint operator. Then the spectral projections of T_n converge strongly toward those of T.

In our case we consider the sequence

$$C_\lambda \equiv [\lambda I^{00}(\lambda)]^{-1} C_\lambda(f,g)$$

as $\lambda \to 0$. The spectral projections of the C_λ are in $R_{\mathcal{O}}$ and their strong limits (as $\lambda \to 0$) are still in $R_{\mathcal{O}}$ since a von Neumann ring is strongly closed. Therefore the spectral projections of the limit operator $C(f)$ are in $R_{\mathcal{O}}$, i.e., the operator $C(f)$ is also associated with $R_{\mathcal{O}}$.

VI. HYPERBOLIC PROPAGATION CHARACTER

In this section we want to show how the strong form [Postulate 8(b)] of the primitive causality postulate may be derived from the weak form [Postulate 8(a)] and the other postulates. Let us consider some cylindric region in space time, e.g., the collection of all points with

$$|\mathbf{x}| < a; \quad |x_0| < \tau.$$

This region will be denoted by C. Consider next the double light cone which is tangent to C (see Fig. 1.). It divides space-time into various regions. The caps which lie in the time direction of C will be denoted by C_t, those in space direction by C_r. The region which is totally space like to C is called C'. Borchers[29] has proved the interesting theorem

$$R_{C\cup C_r} = R_C. \tag{54}$$

In other words the field operators in C_r are functions of those in C. This result follows essentially from the spectral postulate (Postulate 5).

On the other hand, the primitive causality postulate [Postulate 8(a)] implies that $R_{C'}$ and $R_{C\cup C_r}$ together are complete since the region $C'\cup C_r\cup C'$ contains a full time-slice apart from the neighborhood of the points of contact between C_r and C'. We shall show in a separate paper that the effects from the points of contact may be ignored. Thus

$$\{R_C, R_{C'}\}'' = R(\mathfrak{C}). \tag{55}$$

[29] H. J. Borchers, Nuovo cimento **19**, 797 (1961).

Further we observe that due to Einstein causality $R_{C'}$ is contained in the commutant $(R_C)'$ of R_C:

$$R_{C'} \subset (R_C)'. \tag{56}$$

From (55) and (56) it follows immediately that the center of R_C consists only of multiples of the identity, i.e., that R_C is a "factor" (see Appendix). Beyond this if one makes certain assumptions about the type of the factor in von Neumann's classification, then one gets[30] from Eqs. (55) and (56) the following theorem.

Duality Theorem

$$R_{C'} = (R_C)'. \tag{57}$$

Let us denote the interior of the double cone in the figure by $\hat{C}$:

$$\hat{C} = C \cup C_r \cup C_t. \tag{58}$$

By the same line of argument as above we also get the duality relation

$$R_C = (R_{C'})'. \tag{59}$$

Taking the commutant of (57) and comparing with (59) we have

$$R_{\hat{C}} = R_C. \tag{60}$$

This implies in particular

$$R_{C_t} \subset R_C \tag{61}$$

which is the expression of the hyperbolic propagation character [Postulate 8(b)].

ACKNOWLEDGMENTS

We found the atmosphere of the Summer Institute of Theoretical Physics at the University of Wisconsin in Madison very stimulating and we thank Dr. Robert Sachs for the hospitality extended to us.

APPENDIX

For the convenience of the reader we collect here the few definitions concerning von Neumann rings which are used in this paper.

Definition 1: A von Neumann ring R is a collection of bounded operators in a Hilbert space with the following properties:

(a) With X and Y, R contains also a_1X+a_2Y, XY, $X^\dagger$.

(b) R contains the unit operator.

(c) R is weakly (and strongly) closed. In other words, if X_n is a sequence of operators from R which converges weakly (or strongly) then also the limit belongs to R.

Definition 2: Let T be any self-adjoint set of bounded operators, i.e., a set which contains with every operator also its adjoint. Then the collection of all bounded

[30] J. von Neumann, Ann. Math. **41**, 94 (1940). We intend to discuss the questions relating to the "type" of the factors $R_{\mathcal{O}}$ in a separate paper.

operators which commute with the elements of T is called the commutant of T and denoted by T'.

Theorem: If T is any self-adjoint set of bounded operators then T' is a von Neumann ring. T'' is the smallest von Neumann ring which contains the set T.

Definition 3: A ring R is called a factor if its center consists only of multiples of the identity, i.e., if

$$R \bigcap R' = \{\lambda 1\}.$$

Definition 4: An unbounded operator is said to be "associated with a von Neumann ring" if all its spectral projections (and hence all its bounded functions) belong to the ring. Conversely, if we are given a collection of self-adjoint unbounded operators we speak of the associated von Neumann ring. It is the ring which is generated by the spectral projections of the unbounded operators.

Note added in proof. In connection with the completeness postulate 6 the following interesting fact has been found recently by D. Ruelle and by H. Borchers (private communications):

If the theory satisfies Postulates 1 through 5(b), then Postulate 6 is implied already if we require that the (unique) vacuum is a cyclic vector of $\mathcal{P}_\infty$.[21] In other words, if one starts with the Wightman functions, then Postulate 6 is already implied by the requirement that the vacuum state shall be unique.

For the Sudarshan-Bardakci models Jost and coworkers have shown in a very simple way the existence of several vacua. It may be noted that these degenerate vacua also modify the linked-cluster decomposition of the Wightman functions $W(x_1 \cdots x_n)$ at equal times for large spatial separation and hence the asymptotic condition.[13,14]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

Analytic Properties of Products of Field Operators*

R. F. STREATER†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received July 17, 1961)

It is shown that any function of $n-1$ four-vectors which is analytic in the forward tube, the backward tube, and some neighborhood of the Jost points has a one-valued continuation into the extended tube. The theorem is applied to products of field operators between arbitrary states. It is shown that each matrix element of the product is analytic in the envelope of holomorphy of the union of the permuted extended tubes (the domain of analyticity of the most general causal vacuum expectation value). Thus the product of field operators may be regarded as having analytic properties.

I. INTRODUCTION

IN a fundamental paper[1] in 1956, Wightman showed that a quantized field is determined by the vacuum expectation values of products of field operators, now called W functions. They are defined by

$$W_{12\ldots n}(x_1,x_2,\cdots,x_n) = \langle 0 | \varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n) | 0 \rangle \quad (1.1)$$

where $\varphi_j(x)$ are quantized fields. The W functions are boundary values of analytic functions of the relative coordinates ξ_j:

$$\xi_j = x_j - x_{j+1}, \quad j=1,2,\cdots,n-1. \quad (1.2)$$

The W function given by (1.1) is analytic in the forward tube $T_{n-1}(\xi_j)$ consisting of those points for which $\mathrm{Im}\xi_j = \zeta_j$ is a forward time-like vector:

$$\zeta_j^2 = (\zeta_j^{(0)})^2 - \sum_{i=1}^{3} (\zeta_j^{(i)})^2 > 0; \quad \zeta_j^{(0)} > 0, \quad (1.3)$$

where $\zeta_j^{(\mu)}$ is the μ component of the four-vector ζ_j. If a vector ζ satisfies (1.3) we will sometimes write $\zeta > 0$.

The W functions can be analytically continued beyond the domain T if two further properties are used. First, since the functions are Lorentz invariant they can be continued into the "extended tube" T'_{n-1} obtained from T by all complex Lorentz transformations.[2] It should be stressed that this result is not obvious; the difficult part is to show that the functions are one valued in T'. Now, Jost has shown[3] that $T'_n(x_j)$ contains the real point $(x_1,\cdots,x_n) \epsilon R^{4n}$, if and only if,

$$(\sum_{i=1}^{n} \lambda_i x_i)^2 < 0 \quad (1.4)$$

for all $\lambda_i > 0$. We call a set $(x_1,\cdots,x_n)$ of real four-vectors satisfying (1.4) a Jost point, shortened here to J point.

* Work supported by the U. S. Air Force, Office of Scientific Research, A. R. D. C.
† Now at the Physics Department, Imperial College, London.
[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[2] D. Hall and A. S. Wightman, Mat. Fys. Medd. Kgl. Dan. Vid. Selsk. **31**, No. 5 (1957).
[3] R. Jost, Helv. Phys. Acta. **30**, 409 (1957).

operators which commute with the elements of T is called the commutant of T and denoted by T'.

Theorem: If T is any self-adjoint set of bounded operators then T' is a von Neumann ring. T'' is the smallest von Neumann ring which contains the set T.

Definition 3: A ring R is called a factor if its center consists only of multiples of the identity, i.e., if

$$R \bigcap R' = \{\lambda 1\}.$$

Definition 4: An unbounded operator is said to be "associated with a von Neumann ring" if all its spectral projections (and hence all its bounded functions) belong to the ring. Conversely, if we are given a collection of self-adjoint unbounded operators we speak of the associated von Neumann ring. It is the ring which is generated by the spectral projections of the unbounded operators.

Note added in proof. In connection with the completeness postulate 6 the following interesting fact has been found recently by D. Ruelle and by H. Borchers (private communications):

If the theory satisfies Postulates 1 through 5(b), then Postulate 6 is implied already if we require that the (unique) vacuum is a cyclic vector of $\mathcal{P}_\infty$.[21] In other words, if one starts with the Wightman functions, then Postulate 6 is already implied by the requirement that the vacuum state shall be unique.

For the Sudarshan-Bardakci models Jost and coworkers have shown in a very simple way the existence of several vacua. It may be noted that these degenerate vacua also modify the linked-cluster decomposition of the Wightman functions $W(x_1 \cdots x_n)$ at equal times for large spatial separation and hence the asymptotic condition.[13,14]

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

Analytic Properties of Products of Field Operators*

R. F. STREATER†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received July 17, 1961)

It is shown that any function of $n-1$ four-vectors which is analytic in the forward tube, the backward tube, and some neighborhood of the Jost points has a one-valued continuation into the extended tube. The theorem is applied to products of field operators between arbitrary states. It is shown that each matrix element of the product is analytic in the envelope of holomorphy of the union of the permuted extended tubes (the domain of analyticity of the most general causal vacuum expectation value). Thus the product of field operators may be regarded as having analytic properties.

I. INTRODUCTION

IN a fundamental paper[1] in 1956, Wightman showed that a quantized field is determined by the vacuum expectation values of products of field operators, now called W functions. They are defined by

$$W_{12\ldots n}(x_1,x_2,\cdots,x_n) = \langle 0 | \varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n) | 0 \rangle \quad (1.1)$$

where $\varphi_j(x)$ are quantized fields. The W functions are boundary values of analytic functions of the relative coordinates ξ_j:

$$\xi_j = x_j - x_{j+1}, \quad j=1,2,\cdots,n-1. \quad (1.2)$$

The W function given by (1.1) is analytic in the forward tube $T_{n-1}(\xi_j)$ consisting of those points for which $\mathrm{Im}\xi_j = \zeta_j$ is a forward time-like vector:

$$\zeta_j^2 = (\zeta_j^{(0)})^2 - \sum_{i=1}^{3} (\zeta_j^{(i)})^2 > 0; \quad \zeta_j^{(0)} > 0, \quad (1.3)$$

where $\zeta_j^{(\mu)}$ is the μ component of the four-vector ζ_j. If a vector ζ satisfies (1.3) we will sometimes write $\zeta > 0$.

The W functions can be analytically continued beyond the domain T if two further properties are used. First, since the functions are Lorentz invariant they can be continued into the "extended tube" T'_{n-1} obtained from T by all complex Lorentz transformations.[2] It should be stressed that this result is not obvious; the difficult part is to show that the functions are one valued in T'. Now, Jost has shown[3] that $T'_n(x_j)$ contains the real point $(x_1,\cdots,x_n) \epsilon R^{4n}$, if and only if,

$$(\sum_{i=1}^{n} \lambda_i x_i)^2 < 0 \quad (1.4)$$

for all $\lambda_i > 0$. We call a set $(x_1,\cdots,x_n)$ of real four-vectors satisfying (1.4) a Jost point, shortened here to J point.

* Work supported by the U. S. Air Force, Office of Scientific Research, A. R. D. C.
† Now at the Physics Department, Imperial College, London.
[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[2] D. Hall and A. S. Wightman, Mat. Fys. Medd. Kgl. Dan. Vid. Selsk. **31**, No. 5 (1957).
[3] R. Jost, Helv. Phys. Acta. **30**, 409 (1957).

On the other hand, Dyson[4] defines weak local commutativity by the equation

$$\langle 0|\,\varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n)\,|0\rangle = \langle 0|\,\varphi_n(x_n)\cdots\varphi_2(x_2)\varphi_1(x_1)\,|0\rangle \quad (1.5)$$

and proves that the W functions are analytic at all points where (1.5) holds. In particular, if we assume the usual local property of the fields

$$[\varphi_i(x_1),\ \varphi_j(x_2)]=0 \quad (1.6)$$

if x_1-x_2 is space-like, then the W functions are analytic at J points. Since this same property of analyticity follows also from Lorentz invariance, we might ask whether there is any connection between Lorentz invariance and local commutativity; the axioms of quantum field theory[1,5] might not be logically independent. As an example of this, it is possible to prove that the condition (1.6) is implied by the apparently weaker condition of commutativity at J points, using a result of D. Ruelle.[6] This shows that the system of axioms contains redundant information, and can be replaced by a slightly weaker set leading to the same results.

In the classical theory of special relativity no signal can travel faster than light, and so events separated by a space-like vector cannot interfere with one another. Thus a relativistic classical theory is necessarily local in this sense. In the quantized version we would expect that field operators should commute at space-like separated points, since then the two operators are simultaneously measurable. However it is possible to set up Lorentz invariant field theories[7] not satisfying (1.6). The main result of this paper is the following theorem, which shows that there is a peculiar and surprising connection between commutativity and the Lorentz group.

II. THE MAIN THEOREM

Theorem: The envelope of holomorphy $\mathfrak{C}(\mathfrak{A}')$ of the set

$$\mathfrak{A}'\equiv T_n(x)\bigcup T_n(-x)\bigcup J\text{-points}$$

is $T'_n(x)$. [By $T_n(-x)$ we mean the set of points in C^{4n} such that $-x_j\epsilon T_n$, $j=1,\cdots,n$.] The meaning of the theorem is the following. The set $\mathfrak{A}'$ is not a domain because, although connected, it is not open. The envelope of holomorphy $\mathfrak{C}(\mathfrak{A}')$ is defined as the intersection of the domains of analyticity of all functions analytic on $\mathfrak{A}'$ (and therefore in some neighborhood of $\mathfrak{A}'$). Now a function f analytic on $\mathfrak{A}'$, satisfies the conditions of the edge-of-the-wedge theorem[4,8–10] and so will be analytic in some neighborhood of the J points (of course), but where the neighborhood is characterized purely geometrically and does not depend on the function f. The class of functions analytic on $\mathfrak{A}'$ is invariant under the real Lorentz group $x\rightarrow\lambda x$, hereafter denoted by Λ^r. Hence the intersection of the domains of analyticity of all functions analytic on $\mathfrak{A}'$ is invariant under Λ^r and contains a certain neighborhood of the J points, which may therefore be taken to be invariant under Λ^r. In the following, $\mathfrak{A}$ will denote the set $\mathfrak{A}'$ enlarged in this way to include this neighborhood of the J points, and so $\mathfrak{A}$ is open. Obviously $\mathfrak{C}(\mathfrak{A}')=\mathfrak{C}(\mathfrak{A})$. The proof of the theorem is based on a few lemmas in the theory of many complex variables.

Lemma 1: (Bremermann[11]). If $\{d_\sigma\}$ is a set of analytic discs depending continuously on a parameter σ, $0<\sigma<1$, such that every point inside and on the boundary of $d_\sigma(\sigma<1)$ is in the domain D, and also such that at least one point in the interior of the limit disc d_1 $(\sigma=1)$ is in D, then the envelope of holomorphy of D contains a neighborhood of the interior of d_1.

This is a strong form of the continuity theorem, or protruding horn theorem,[12,13] which essentially says that singularities cannot suddenly start or stop; they must come in from the edge of a disc d_σ, or else all of d_1 is outside $\mathfrak{C}(D)$. For the present paper it is sufficient to consider an "analytic disc" as the interior of a piece-wise analytic nonselfintersecting closed curve in an analytic manifold of complex dimension one. Unlike the usual protruding-horn theorem, the version of Bremermann is local; we cannot directly perform any finite analytic completion, since we are not told how large a neighborhood of the interior of the limit disc is in the envelope of homomorphy. In order to use the theorem to perform finite analytic completion, we combine it with the Cartan-Thullen lemma.[14]

Lemma 2:[14] If D is invariant under an analytic group Λ, then $\mathfrak{C}(D)$ is invariant under Λ. For suppose $f(z)$ is analytic if $z\epsilon D$ [and therefore if $z\epsilon\mathfrak{C}(D)$]. Then $f(\lambda z)$ is analytic in $\lambda D=D$ for all $\lambda\epsilon\Lambda$. Hence $f(\lambda z)$ is analytic in $\mathfrak{C}(D)$, so $f(z)$ is analytic in $\Lambda\mathfrak{C}(D)$. Hence any function analytic in $\mathfrak{C}(D)$ is also analytic in $\Lambda\mathfrak{C}(D)\supset\mathfrak{C}(D)$. But $\mathfrak{C}(D)$ is a natural domain for some function, so $\mathfrak{C}(D)=\Lambda\mathfrak{C}(D)$, proving $\mathfrak{C}(D)$ is invariant under Λ.

Lemma 3: If a set D is invariant under the three abelian subgroups $\Lambda_i{}^c$ of the complex Lorentz group [$\Lambda_i{}^c$ consists of rotations in the $(0,i)$ plane] then it is invariant under the complex Lorentz group Λ^c.

---

[4] F. J. Dyson, Phys. Rev. **110**, 579 (1958).
[5] W. Schmidt and K. Baumann, Nuovo cimento **4**, 860 (1956).
[6] D. Ruelle, Helv. Phys. Acta **32**, 135 (1959).
[7] V. Glaser, H. Lehmann and W. Zimmermann, Nuovo cimento **6**, 1122 (1957).
[8] H. J. Bremermann, R. Oehme, and J. G. Taylor, Phys. Rev. **109**, 2178 (1958).
[9] L. Gårding and A. Beurling (to be published).
[10] H. Epstein, J. Math. Phys. **1**, 524 (1960).
[11] H. J. Bremermann, Math. Ann **127**, 406 (1954); generalized by W. S. Brown, thesis (Ph.D.), Princeton (1961).
[12] G. Källén and A. S. Wightman, Mat. Fys. Skr. Dan. Vid. Selsk **1**, No. 6 (1958).
[13] A. S. Wightman in *Relations de Dispersion et Particules Elémentaire* (Hermann Cie, Paris, France, and John Wiley and Sons, Inc., New York, 1960), p. 229.
[14] H. Carthan and P. Thullen, Math. Ann. **106**, 617 (1932).

Proof: Let M_{0i} be the generator of $\Lambda_1{}^c$. Since

$$[M_{0i},M_{0j}]=iM_{ij},$$

D is invariant under the special rotations generated by $M_{ij}(i.j=1,2,3)$ and so is invariant under Λ^c.

We will now prove that functions analytic in $\mathfrak{A}$ have one-valued continuation into $\Lambda_1{}^c\mathfrak{A}$.

Proof of the Main Theorem

We say $(x_1,\cdots,x_n)\epsilon R^{4n}$ is a two-dimensional J point if

$$\{\sum_{j=1}^{n}\lambda_j x_j{}^{(0)}\}^2-\{\sum_{j=1}^{n}\lambda_j x_j{}^{(1)}\}^2<0 \qquad (2.1)$$

for all $\lambda_j>0$. Clearly, a two-dimensional J point is also a J point [that is, satisfies (1.4)]. Define "light-cone coordinates" in the (0,1) space as follows:

$$\alpha_j=x_j{}^{(0)}-x_j{}^{(1)};\quad \beta_j=x_j{}^{(0)}+x_j{}^{(1)}\quad (j=1,2,\cdots,n).$$

Then if $x_j{}^{(2)}$, $x_j{}^{(3)}$ are real $(j=1,2,\cdots,n)$ the set $\mathfrak{A}$ contains the set

$$(\mathrm{Im}\alpha_j>0,\ \mathrm{Im}\beta_j>0)\cup(\mathrm{Im}\alpha_j<0,\ \mathrm{Im}\beta_j<0)\cup \text{two-dimensional } J \text{ points.} \qquad (2.2)$$

Define a new set of variables

$$\begin{aligned} &z=\alpha_1\beta_1\\ &\lambda_j=\alpha_1\beta_j, \quad j=2,3,\cdots,n\\ &\mu_j=\alpha_j/\alpha_1, \quad j=2,3,\cdots,n\\ &x_j{}^{(2)},\ x_j{}^{(3)}, \quad j=1,2,\cdots,n\\ &\alpha=\alpha_1 \end{aligned}$$

These are independent and so will do as analytic coordinates. The point $\alpha_1=0$ is the only singularity of the transformation.

Lemma 4: The point $(x_1,\cdots,x_n)\epsilon R^{4n}$ is a two-dimensional J point of T'_n if and only if $z<0$, $\lambda_j<0$, $\mu_j>0$, α is real, $\neq 0$, and $x_j{}^{(2)}$, $x_j{}^{(3)}$ are real.

Proof: If $x\epsilon R^{4n}$ is a two-dimensional J point, we have

$$(x_i{}^{(0)}+\lambda x_j{}^{(0)})^2-(x_i{}^{(1)}+\lambda x_j{}^{(1)})^2<0$$

for all $\lambda>0$. This can be written

$$(\alpha_i+\lambda\alpha_j)(\beta_i+\lambda\beta_j)<0$$

for all $\lambda>0$. Therefore $\alpha_i\beta_j<0$ and the solutions λ of

$$(\alpha_i+\lambda\alpha_j)(\beta_i+\lambda\beta_j)=0$$

are negative; this means $\alpha_i\beta_j<0$, which gives the conditions $z<0$, $\lambda_j<0$, $\mu_j>0$. Conversely, if $z<0$, $\lambda_j<0$, $\mu_j>0$ we get $\alpha_i\beta_j<0$, and so any sum of such terms with positive coefficients, as in (2.1), is negative, as was to be shown. This proves Lemma 4.

Suppose now $z<0$, $\lambda_j<0$, $\mu_j>0$, $x_j{}^{(2)}$, $x_j{}^{(3)}$ real. The set (2.2) is then

$$(\mathrm{Im}\alpha>0)\cup(\mathrm{Im}\alpha<0)\cup(\alpha \text{ real}, \neq 0)$$

that is, the entire α plane except for $\alpha=0$. In these coordinates the domain is not simply connected; a function could have a branch point at $\alpha=0$. But we are considering functions which are given as being one valued and analytic in the domain (2.2), and hence are one valued in the α-plane $(\alpha\neq 0)$ if $z<0$, $\lambda_j<0$, $\mu_j>0$; $x_j{}^{(2)}$, $x_j{}^{(3)}$ real.

The set (2.2) is invariant under the real Lorentz transformations in the (0,1) plane (the group $\Lambda_1{}^r$), which take the form $\alpha\to\rho\alpha$, $\rho>0$ in the new coordinates.

Lemma 5: There exists a polycylinder P centered on a point $z'<0$, $\lambda'_j<0$, $\mu'_j>0$, $x'_j{}^{(2)}$, $x'_j{}^{(3)}$ real, say

$$(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)})\epsilon P\langle=\rangle |z-z'|<R,\quad |\lambda_j-\lambda'_j|<R,$$
$$|\mu_j-\mu'_j|<R,\quad |x_j{}^{(2)}-x'_j{}^{(2)}|<R,\quad |x_j{}^{(3)}-x'_j{}^{(3)}|<R,$$

such that $\mathfrak{K}(\mathfrak{A})$ contains $P\times\{\alpha/\alpha\neq 0\}$.

Proof: Let z', $\lambda'_j<0$, $\mu_j>0$, $x'_j{}^{(2)}$, $x'_j{}^{(3)}$ be real. Then $\mathfrak{A}$ is the set $\{\alpha/\alpha\neq 0\}$. Let K be a compact annulus surrounding $\alpha=0$. Then for each $\alpha'\epsilon K$ the point $(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)},\alpha')$ lies in $\mathfrak{A}$ provided

$$(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)})\epsilon P_{\alpha'},$$

where $P_{\alpha'}$ is some polycylinder of radius $R(\alpha')$ (which may depend on α') about the point $(z',\lambda'_j,\mu'_j,x'_j{}^{(2)},x'_j{}^{(3)})$.

This is the statement that $\mathfrak{A}$ is open. Let R be the lower bound of all the radii of $P_{\alpha'}$ as α' varies over K. Since K is compact, a well-known argument assures us that $R>0$. For if $R=0$ there would exist a point $\alpha''\epsilon K$ such that points arbitrarily close to $(z',\lambda'_j,\mu'_j,x'_j{}^{(2)},x'_j{}^{(3)},\alpha'')$ are outside $\mathfrak{A}$, a contradiction. Let P be the polycylinder about $(z',\lambda'_j,\mu'_j,x'_j{}^{(2)},x'_j{}^{(3)})$ with radius R. Then $\mathfrak{A}$ contains the points $P\times\{\alpha/\alpha\epsilon K\}$. Since $\mathfrak{A}$ is invariant under $\Lambda_1{}^r$, that is, $\alpha\to\rho\alpha$, so is its holomorphy envelope (Lemma 2). Clearly $\Lambda_1{}^rK=\{\alpha/\alpha\neq 0\}$. Hence $\mathfrak{K}(\mathfrak{A})$ contains the points $\{z,\lambda_j,\mu_j,x_j{}^{(3)}\epsilon P;\alpha\neq 0\}$. This proves Lemma 5.

A transformation $\lambda\epsilon\Lambda_1{}^c$ leaves z, λ_j, μ_j, $x_j{}^{(2)}$, $x_j{}^{(3)}$ invariant, but changes $\alpha\to\lambda\alpha(\lambda\neq 0)$. Therefore the set $\Lambda_1{}^c\mathfrak{A}$ contains $\{\alpha/\alpha\neq 0\}$ if it contains any point in the α plane. We now show that $\mathfrak{K}(\mathfrak{A})$ contains $\Lambda_1{}^c\mathfrak{A}$.

Let N be the union of all connected open sets containing the point z, λ_j, μ_j, $x_j{}^{(2)}$, $x_j{}^{(3)}$ (where $z<0$, $\lambda_j<0$, $\mu_j>0$ and $x_j{}^{(2)}$, $x_j{}^{(3)}$ are real) such that $\mathfrak{K}(\mathfrak{A})$ contains $N\times\{\alpha/\alpha\neq 0\}$. By Lemma 5 N is nonempty. Suppose if possible that there is a point π on the boundary of N such that $(\pi,\alpha')\epsilon\mathfrak{A}$ for some α'. We will now obtain a contradiction. With the hypothesis, there exists a neighborhood N_1 of π and neighborhood N_2 of α' such that $\mathfrak{A}$ contains $N_1\times N_2$, since $\mathfrak{A}$ is open. Hence $\mathfrak{K}(\mathfrak{A})$ contains $N_1\times N_2$. Since $\Lambda_1{}^r\mathfrak{A}=\mathfrak{A}$, $\mathfrak{A}$ also contains $N_1\times\Lambda_1{}^rN_2$, that is, $\mathfrak{A}$ contains a sector containing N_2 as shown in Fig. 1. Now join π by an analytic arc γ with parameter $\sigma(0<\sigma<1)$ to a two-dimensional J point, such that if $\sigma<1$ the arc lies in N. Then each disc d_σ in the α plane (each d_σ being the same curve as d_1 shown in Fig. 1, but associated with a point σ of γ) lies in $\mathfrak{K}(\mathfrak{A})$ if $\sigma<1$, by definition of N, since if $\sigma<1$,

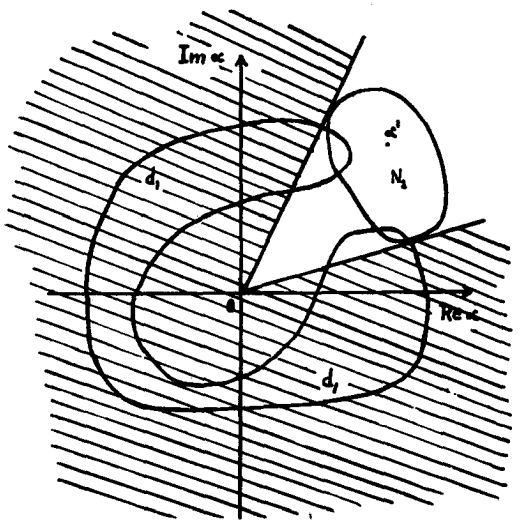


FIG. 1. Picture of T (unshaded) at π.

$\mathcal{H}(\mathcal{A})$ contains $\{\alpha/\alpha\neq 0\}$. Also the limit disc $\sigma=1$ at π (shown in Fig. 1) contains at least one point of $\mathcal{A}$. Hence (Lemma 1) the interior of d_1 in Fig. 1) contains at least one point of $\mathcal{A}$. Hence (Lemma 1) the interior of d_1 in Fig. 1 lies in $\mathcal{H}(\mathcal{A})$. Then just as in the proof of Lemma 5 there is a polycylinder P about π such that $P\times d_1\subset\mathcal{H}(\mathcal{A})$, and using the invariance of $\mathcal{H}(\mathcal{A})$ under $\Lambda_1{}^r$, $\mathcal{H}(\mathcal{A})\supset P\times\{\alpha/\alpha\neq 0\}$, contradicting the assumption that π lies on the boundary of the largest open set with this property. Hence if π is on the boundary of N, no point α' lies in $\mathcal{A}$. So N contains all points $\pi=(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)})$ such that there exists α such that $(\pi,\alpha)\epsilon\mathcal{A}$, and $\mathcal{H}(\mathcal{A})$ contains the set $N\times\{\alpha/\alpha\neq 0\}$. But this is just the definition of $\Lambda_1{}^c\mathcal{A}$. Hence $\mathcal{H}(\mathcal{A})=\mathcal{H}(\Lambda_1{}^c\mathcal{A})\supset\Lambda_1{}^c\mathcal{A}$.

We must show that the continuation is one valued. It is possible that if we continue from $\mathcal{A}$ to a point in $\Lambda_1{}^c\mathcal{A}$ along two different paths in $\Lambda_1{}^c\mathcal{A}$ we will arrive at two different answers. This is possible even if $\Lambda_1{}^c\mathcal{A}$ is simply connected. So far we have shown that it is possible to continue to each point in $\Lambda_1{}^c\mathcal{A}$ along some path in $\Lambda_1{}^c\mathcal{A}$; we have not shown that it is possible to continue along every path in $\Lambda_1{}^c\mathcal{A}$. It could be that $\mathcal{H}(\mathcal{A})$ is a many-sheeted domain over a subset in C^{4n} (containing $\Lambda_1{}^c\mathcal{A}$) such that, while for any point in $\Lambda_1{}^c\mathcal{A}$ there is a sheet on which it is a point of analyticity, there may be another sheet on which it is a ramification point. We now prove that the continuation into $\Lambda_1{}^c\mathcal{A}$ is one valued.

Let Γ_1 be a path of finite length in $\Lambda_1{}^c\mathcal{A}$ joining a given point (π,α) to a two-dimensional J point. Change to the coordinates $(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)},\alpha)$. Then for fixed $z,\lambda_j,\mu_j,x_j{}^2,x_j{}^3$, a point on Γ_1 is a point in the α plane, and $\mathcal{A}$ consists of two disjoint sectors. As we move along Γ_1 the sectors may move about, and open and close, but never *completely* close, since $\Gamma_1\subset\Lambda_1{}^c\mathcal{A}$. Define γ_1 to be the projection of Γ_1 onto the coordinates $(z,\lambda_j,\mu_j,x_j{}^{(2)},x_j{}^{(3)})$. Then γ_1 is a possibly self-intersecting path in the space of these variables. Starting from the two-dimensional J point we can continue to any point on γ_1, obtaining analyticity in $\{\alpha/\alpha\neq 0\}$, by the method of the theorem. Of course we must show that given π on γ_1 we have a one-valued function in the α plane, since two different paths around $\alpha=0$ could give different continuations, or the continuations out of the two different sectors might be different. Let f_1 and f_2 be two different branches of a function f analytic in $\Lambda_1{}^c\mathcal{A}$, defined in the domain $\{\alpha/\alpha\neq 0\}$ for any point on γ_1 by continuation from $\mathcal{A}$ along paths in the α plane encircling the origin in two different directions, or as continuations from the different sectors which make up $\mathcal{A}$. At the beginning of γ_1 we are given a one-valued function in $\{\alpha/\alpha\neq 0\}$, and so $f_1=f_2$. Hence $f_1=f_2$ at all points along γ_1, by continuation. Therefore the value of the function at (π,α) can be defined by fixing π and continuing from $\mathcal{A}$ along a fixed path, say Γ, in $\{\alpha/\alpha\neq 0\}$. Clearly, if we had started with a different path Γ_2 instead of Γ_1 we could still continue along the same Γ, arriving at the same value for the function at (π,α). This shows that there exists a one-valued continuation into $\Lambda_1{}^c\mathcal{A}$.

In a similar manner we can show that there is a one-valued continuation into $\Lambda_2{}^c\Lambda_1{}^c\mathcal{A}$. The proof is the same, except that we work in the coordinates

$$\alpha_j=x_j{}^{(0)}-x_j{}^{(2)},\quad \beta_j=x_j{}^{(0)}+x_j{}^{(2)}$$

and form z, λ_j, μ_j with the new α_j, β_j. Change to the coordinates $(z,\lambda_j,\mu_j,x_j{}^{(1)},x_j{}^{(3)},\alpha=\alpha_1)$. Then as we move along a path $\Gamma_1\subseteq A_2{}^c\Lambda_1{}^c\mathcal{A}$ joining a given point (π,α) with a two-dimensional J point, the domain $\Lambda_1{}^c\mathcal{A}$ is a nonempty region in the α plane; in general it will not be a sector since $\Lambda_1{}^c\mathcal{A}$ is not invariant under $\Lambda_2{}^r$. But $\mathcal{H}(\mathcal{A})$ is invariant under $\Lambda_2{}^r$ so we enlarge $\Lambda_1{}^c\mathcal{A}$ to a sector in the α plane. If $\Lambda_1{}^c\mathcal{A}$ (for π fixed) is a disconnected set in the α plane, we might get several sheets when we do this part of the completion. But the proof that the continuation into $\Lambda_2{}^c\Lambda_1{}^c$ is one valued goes through as above.

Similarly we can show that there is a one-valued continuation into an increasing sequence $(\mathcal{A},\Lambda_1{}^c\mathcal{A},\Lambda_2{}^c\Lambda_1{}^c\mathcal{A},\Lambda_3{}^c\Lambda_2{}^c\Lambda_1{}^c\mathcal{A},\cdots)$. To show that this exhausts $T'=\Lambda^c\mathcal{A}$, we remark that the union R of all sets of the form $(\Pi\Lambda_i{}^c)\mathcal{A}$ (where $\Pi\Lambda_i{}^c$ is any finite product of $\Lambda_1{}^c$, $\Lambda_2{}^c$, $\Lambda_3{}^c$ in any order) is invariant under $\Lambda_1{}^c$, $\Lambda_2{}^c$, $\Lambda_3{}^c$, and therefore (Lemma 3) is invariant under Λ^c, and so contains $\Lambda^c\mathcal{A}=T'$. Since R is contained in $\Lambda^c\mathcal{A}$, it is T'. To show we can continue into R, we note that the set of products $\Pi\Lambda_i{}^c$ is denumerable. Order them, say $\Pi_1=\Lambda_1{}^c$, $\Pi_2=\Lambda_2{}^c$, $\Pi_3=\Lambda_3{}^c$, $\Pi^4=\Lambda_1{}^c\Lambda_2{}^c$, $\cdots$. We have proved that we can continue into the sequence of sets

$$\Pi_1\mathcal{A}\subset\Pi_2\Pi_1\mathcal{A}\subset\Pi_3\Pi_2\Pi_1\mathcal{A}\subset\Pi_4\Pi_3\Pi_2\Pi_1\mathcal{A}\subset\cdots$$

whose limit is R, and the continuation is one-valued. This proves the main theorem.

As a corollary, the following conjecture might be true.

Conjecture: Suppose $f(z)$ is analytic in $\{T_n(z)\bigcup T_n(-z)\bigcup J$ points$\}$, $(z=x+iy,\ z\epsilon C^{4n},\ x\epsilon R^{4n},\ y\epsilon R^{4n})$ and suppose $\lim_{y\to 0} f(z)=f(x)$ in the sense of distribution theory,[15] where $f(x)$ is tempered[15]; then $f(z)$ is a polynomial in the components of the four-vectors; the coefficients in the polynomial are Lorentz invariant functions analytic in $T_n'(z)$.

For $n=1$ the conjecture has been proved by Bogoliubov and Vladiminov.[16] I am indebted to Wightman for this reference.

III. ANALYTIC PROPERTIES OF PRODUCTS OF FIELD OPERATORS

Consider the matrix element

$$f(x_1,\cdots,x_n)=\langle P|\,\varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n)\,|Q\rangle \quad (3.1)$$

where $|P\rangle$, $|Q\rangle$ are any states with momenta P^μ, Q^μ respectively. Change the coordinates to $x=x_1+x_n$ and the $n-1$ relative coordinates ξ_j given by (1.2). Define

$$\tilde{f}(x;\,p_1,\cdots,p_{n-1})=\int \prod_{i=1}^{n-1} e^{-ip_j\xi_j}d^4x_j f(x,\xi_i). \quad (3.2)$$

Then in the usual way we find

$$\begin{aligned} f(x;\,p_1,\cdots,p_{n-1})=&\exp[i\tfrac{1}{2}(Q-P)x](2\pi)^{4(n-1)}\\ &\times\sum_{\beta_j}\langle P|\,\varphi_1(0)\,|\beta_1\rangle\langle\beta_1|\cdots|\beta_{n-1}\rangle\\ &\times\langle\beta_{n-1}|\,\varphi_n(0)\,|Q\rangle\\ &\times\prod_{j=1}^{n-1}\delta^4[p_j-\beta_j+(P+Q/2)] \end{aligned} \quad (3.3)$$

where $|\beta_j\rangle$ is a complete set of states with momentum β_j^μ. This shows that the spectrum of f is forward time-like from the point $-(P+Q)/2$, and we have $f(x,p_1,\cdots,p_{n-1})=0$ unless $p_j+(P+Q)/2>0$, $j=1,\ \cdots,\ n-1$. Then the inverse Laplace transform

$$\begin{aligned} f(x,\xi_j)&=(2\pi)^{4(1-n)}\int f(x;\,p_1,\cdots,p_{n-1})\prod_{j=1}^{n-1}e^{ip_j\xi_j}d^4p_j\\ &=(2\pi)^{4(1-n)}\int f(x;\,p_1,\cdots,p_{n-1})\\ &\qquad\times\prod_{j=1}^{n-1}e^{i\xi_j[P_j+(P+Q)/2]}d^4p_j e^{-i\xi_j(P+Q)/2} \end{aligned}$$

converges if $\xi_j\epsilon T_{n-1}(\xi_j)$, and defines an analytic function there. The variable x enters $f(x,\xi_1,\cdots,\xi_{n-1})$ only in a simple exponential dependence, and so $f(x,\xi_1,\cdots,\xi_{n-1})$ is an entire function of x for fixed ξ_j. If we had smoothed the states $|P\rangle$, $|Q\rangle$ with a test function $h(P,Q)$ with compact support, and instead of (3.1) had considered

$$\int h(P,Q)\langle P|\,\varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n)\,|Q\rangle d^4P d^4Q = f_h(x_1,\cdots,x_n)$$

we should still obtain an entire function of x. In a similar manner

$$g(x,\xi_j)=\langle P|\,\varphi_n(x_n)\cdots\varphi_2(x_2)\varphi_1(x_1)\,|G\rangle \quad (3.4)$$

is analytic in $T_{n-1}(-\xi_j)$. Then as a consequence of (1.6), $f(x,\xi_j)=g(x,\xi_j)$ if $(\xi_1,\cdots,\xi_{n-1})$ is a J point. (Actually, since not every "completely space-like point," such that every difference x_i-x_j is space-like, is a J point a slightly weaker assumption than (1.6) would lead to the same result.) By the edge-of-the-wedge theorem,[4,8-10] f is analytic at J points. Hence $f(x,\xi)$ is analytic if $\xi_j\epsilon T'_{n-1}(\xi_j)$ for all x (by the theorem of Sec. II) if we regard f as a function of the $4n$ variables (x^μ,ξ_j^μ). Similarly a permuted function $\langle P|\,\varphi_2(x_2)\varphi_1(x_1)\,\varphi_3(x_3)\cdots\varphi_n(x_n)\,|Q\rangle$ is analytic if ξ_j lie in the "permuted extended tube" $T'_{n-1}(-\xi_1,\xi_1+\xi_2,\xi_3,\cdots,\xi_{n-1})$, and coincides with f at some common J points where they are both analytic. (Actually instead of $x=(x_1+x_n)$ we first get (x_2+x_n) as the nth independent vector, but a linear transformation back to x is possible, since the function is entire in these variables.) Hence the domain of analyticity of an arbitrary matrix element is $\{x,\xi_j/\xi_j\epsilon\mathfrak{U}_{n-1}\}$, where $\mathfrak{U}_{n-1}$ is the envelope of holomorphy of the union of the permuted extended tubes corresponding to the n-point function. The domain $\mathfrak{U}_1$ is the cut plane $\{x_1,x_2/(x_1-x_2)^2\neq\rho^2\}$, and $\mathfrak{U}_2$ was found by Källén and Wightman.[12] Of course, in any particular theory the actual domain of analyticity of (3.1) might be larger than $\mathfrak{U}_{n-1}$, and might depend on the states $|P\rangle$ and $|Q\rangle$, but must always contain $\{x,\xi_j/\xi_j\epsilon\mathfrak{U}_{n-1}\}$.

The above result has an interesting physical meaning. We may say that the operator product $\varphi_1(x_1)\varphi_2(x_2)\cdots\varphi_n(x_n)$ is analytic in $\{x,\xi_j/\xi_j\epsilon\mathfrak{U}_{n-1}\}$, in the sense that every matrix element is analytic. Now the domain $\mathfrak{U}_{n-1}$ has a natural kernel function $K_{n-1}(\xi_j,\Omega)$, the Bergmann-Weil function, which is Lorentz invariant and analytic inside $\mathfrak{U}_{n-1}$. Here Ω is a set of variables $\omega_1,\ \cdots,\ \omega_k$ which parametrize a subset of the boundary of $\mathfrak{U}_{n-1}$ called the distinguished boundary. These notions are explained by Källén and Toll[17] who compute K_2. Every function analytic in $\mathfrak{U}_{n-1}$ can be expressed as an integral, in terms of its boundary values on the distinguished boundary, using the kernel K_{n-1}. It may be possible to separate the kinematics from the dynamics in the following way. The analytic properties can be proved using only general principles, and may be regarded as a consequence of the kinematics. Thus heuristically the function (3.1) will have the

[15] L. Schwartz, *Théorie des Distributions* (Hermann & Cie, Paris, France, 1950–1951).

[16] N. N. Bogoliubov and V. S. Vladiminov, Phys. Math. Science 1958, No. 3, pp. 26–35 (in Russian).

[17] G. Källén and J. S. Toll, Helv. Phys. Acta **33**, 753 (1960).

representation

$$f(x;\xi_j)=\int d\Omega\Phi(x_1,\cdots,x_n;\omega_1,\cdots,\omega_n)\times K_{n-1}(\xi_1,\cdots,\xi_{n-1},\Omega). \quad (3.5)$$

The singularities arise only from K_{n-1}, which may be regarded as a sort of generalized propagator for the product $\varphi_1(x_1)\cdots\varphi_n(x_n)$. The weight function Φ contains the dynamics, that is, states the details of the particular theory within the general framework; it should be an entire function of the variables x_j (and therefore have a compact spectrum) since the kernel is supposed to take care of the singularities. Very often we can, by a transformation, find a kernel function for which the parameters Ω can be interpreted as mass variables, in which case we should expect that a representation (3.5) is possible in which only physical values of the mass variables enter the integral.

These ideas generalize the well-known Jost-Lehmann-Dyson representation[13,18,19] for the case of one four-vector. In that case we have two fields, and the product in either order can be obtained from the commutator. This has the representation

$$\int\Phi(x_1,x_2,K)\Delta(x_1-x_2;K)dK.$$

The parameters Ω are represented by the mass parameter K. $\Phi(x_1,x_2,K)$ is an entire function of x_1 and x_2. Only physical momenta enter the integral (this is the theorem concerning admissible hyperboloids.[19]

[18] R. Jost and H. Lehmann, Nuovo cimento **5**, 1598 (1957).
[19] F. J. Dyson, Phys. Rev. **110**, 1460 (1958).

The author[20] has proposed that the generalized singular function[21]

$$\Delta(x_1,\cdots,x_n;l_{11},l_{12},\cdots,l_{nn})=\int e^{ip_jx_j}d^4x_j\prod_{i,j}\delta(p_ip_j-l_{ij})\epsilon(p_1^{(0)})$$

should be the kernel for functions vanishing at J points, as a generalization of the Jost-Lehmann-Dyson kernel $\Delta(x,k)$. The connection with the main theorem is obvious. This leads to representations for multiple commutators[22]

$$\langle P|[\varphi_n(x_n),[\cdots[\varphi_2(x_2),\varphi_1(x_1)]\cdots]]|Q\rangle$$

in terms of a weight function and a kernel function. It has not yet been possible to prove the necessity of the condition that only physical momenta enter the integral with nonzero weight (the generalization of Dyson's admissible hyperboloid condition). If such a result could be proved for the representation (3.5) it would be possible to build in the details of the mass-threshold conditions once the kernel K_n has been found.

ACKNOWLEDGMENTS

I would like to thank Dr. V. Glaser of CERN for many interesting discussions concerning the main theorem, and Professor A. Wightman for encouraging me to find a rigorous proof, and for suggesting some important improvements. I would also like to thank Professor M. Goldberger for the hospitality of the Physics Department, Princeton University, during 1960–61.

[20] R. F. Streater, Ph.D. thesis, London, 1959 (unpublished).
[21] G. Källén and H. Wilhelmsson, Mat. Fys. Skv. Dansk. Vid. Selsk. **1**, 9 (1959).
[22] R. F. Streater, Nuovo cimento **15**, 937 (1960).

Some Properties of Impedance as a Causal Operator*

TAI TSUN WU†
Gordon McKay Laboratory of Applied Science, Harvard University, Cambridge, Massachusetts
(Received September 25, 1961)

As a model for the various problems connected with domains of analyticity and dispersion relations, the concept of impedance (self-impedance) has been studied in general under the physical requirements of linearity, passivity, reproducibility, and causality. After identifying impedance functions with suitably defined positive functions, any impedance function may be decomposed to the sum of a minimum reactive part and two reactive parts, one of which is of unfamiliar nature. In the time domain, the convolution representation of the current in terms of the voltage and vice versa are obtained rigorously. As an example, a reactance obtained from the Cantor function is studied in detail. Its peculiar delta-function response is computed explicitly, and it is also used to construct a counter example to a representation by van Kampen of the S matrix of a Maxwellian field. The general results are also applied to study the Kronig-Kramers dispersion relations for dielectric constants. It is proved that one of the two relations is true under very general conditions, but the other is false in general.

1. INTRODUCTION

IN recent years, there has been a great deal of effort made in attempting to understand, through various forms of dispersion relations, the strong interactions of elementary particles. Because of the extreme complexity of this problem, many connections between the various concepts have not been rigorously clarified. For example, in getting a single dispersion relation for an analytic function, the contour of integration is sometimes pushed to the boundary of the domain of analyticity, which is mathematically not permitted without some assumption on continuity. Furthermore, there is always the problem of the number of subtractions, if finite; indeed, in connection with the Mandelstam conjecture of double dispersion relations, no estimate seems to exist about the rate of growth of the scattering amplitude as a function of complex variables, at least in certain sectors.

It is therefore the purpose here to study rigorously the simple model provided by an electric circuit. Here the causal function, whose properties need to be studied, arises from the linear connection between the current and the voltage. No space variable is considered, and both the current and the voltage are functions of time only. The precise assumptions to be used are stated in Sec. 2.

For the clarity of analysis, let the terms, "impedance" and "impedance function" be used somewhat differently. An impedance function is a complex function of one real variable, the frequency, with certain properties to be found later; while an impedance is a linear operator which operates on the current into a network to give the more or less arbitrarily defined voltage across the network. Thus, an impedance function is the frequency representation of an impedance. A most fundamental question is in what sense an impedance may be represented by an impedance function. This will be answered in Secs. 3 and 4.

The fundamental attribute of an impedance is that of a causual operator between two quantities whose product gives the rate of energy transfer. The adjective "causal" is used to describe the fact that the voltage is zero as long as the current remains zero. Since only this fundamental property of an impedance is going to be used, the results are by no means restricted to electric-circuit theory. For example, the results may be used to study the frequency dependence of magnetic permeabilities and dielectric constants. In this connection, the Kronig-Kramers relations will be considered.

Throughout this paper, the term "function" is used to mean a single-valued function of a small number of variables, excluding such pseudofunctions as the Dirac delta function. This is necessary in order to get integrals in the usual sense, not in the sense of distribution theory.

2. FORMULATION OF THE PROBLEM

The following system of symbols is to be used. Let $j(t)$ be the current in the two terminals of a network as a function of the time, $v(t)$ be the voltage across the terminals, and F the impedance. Thus,

$$v=Fj. \tag{2.1}$$

Corresponding lower case and capital letters are used to denote pairs of Fourier transforms in the sense of the Plancherel theorem; thus, for instance,

$$J(x)=\text{l.i.m.}_{A\to\infty}\int_{-A}^{A} j(t)e^{ixt}dt,$$

and

$$j(t)=(1/2\pi)\ \text{l.i.m.}_{A\to\infty}\int_{-A}^{A} J(x)e^{-ixt}dx, \tag{2.2}$$

provided $j(t)$ is Lebesgue square integrable (symbolically ϵL^2) over $(-\infty, \infty)$. Here the symbol "l.i.m." stands for "limit in the mean," and the factor $1/2\pi$ is distributed such that the transforms of convolution integrals take a simple form. Finally, the complex conjugate of a function $f(t)$ is denoted by $f^*(t)$.

Since it is not permissible to have $v(t)=v_0\delta(t)$, where

* Based on a doctoral dissertation, Harvard University (1956).
† Alfred P. Sloan Foundation Fellow.

$\delta(t)$ is the Dirac delta function, there is some implicit restriction of the form of $j(t)$. For example, it is not admissible to use any step function as the current into a pure inductance. However, it is observed that a step function would be admissible if a pure resistance were connected in parallel with the inductance. Thus, a first step is to put this notion of parallel connection in a definite form. Consider two impedances F_1 and F_2. It is natural to define the impedance F of their parallel connection, which is denoted symbolically by $F_1 \otimes F_2$, as follows. The equation $v=Fj$ means that there exist unique functions $j_1(t)$ and $j_2(t)$ such that $j=j_1+j_2$ and $v=F_1 j_1=F_2 j_2$. The domain of F is clearly the set of j for which the unique functions j_1 and j_2 exist.

Basically, the concept of an impedance function is associated only with a passive network. For the present purpose, this is considered to be a fundamental limitation on the concept of impedance. In order to ascertain the properties of a given passive network, it is necessary to have available a sufficiently large collection of test inputs. In other words, it is necessary to know beforehand a sufficiently large set of admissible current-input functions. Since it is unsatisfactory physically to specify this set of functions by their local behavior such as continuity, it is perhaps simplest to describe the permissible input functions for the parallel combination of the given passive network with a pure resistance. To be realistic, all functions of the time under consideration should be zero before some time t_0, because of the impossibility of operating on the past history. Beyond this, it is hereby postulated that, if a current-input function generates only a finite amount of heat when applied to a pure resistance, it can be applied to any parallel combination of a general impedance and a pure resistance, described by $R \otimes F$ with $R>0$.

In any network problem in the usual sense, it is, of course, true that both of the functions $j(t)$ and $v(t)$ are real. However, most of the theory may be carried through for complex functions without any extra effort. To show that this not only of purely mathematical interest, consider the ideal frequency-independent gyrator. Voltages and currents are connected by the formulas

$$v_1(t)=-aj_2(t) \quad \text{and} \quad v_2(t)=aj_1(t), \tag{2.3}$$

where a is a real number. If $v(t)=v_1(t)+iv_2(t)$, and $j(t)=j_1(t)+ij_2(t)$, then $v(t)=iaj(t)$. The constant ia may be considered to be the impedance function of this ideal gyrator. Actually, this simple function has a special place in the general representation of impedance functions, as developed in Sec. 6.

Mathematically, the above idea may be formulated as follows. An impedance is defined to be an operator F which operates on the current function $j(t)$ to give the voltage function $v(t)=(Fj)(t)$, where both j and v are measurable, such that:

(i) $F(j_1+j_2)=Fj_1+Fj_2$ and $F\alpha j=\alpha Fj$, for any complex constant α;

(ii) $FT_a j=T_a Fj$, where T_a is the operator for time translation, i.e., $(T_a j)(t+a)=j(t)$ with a real;

(iii) $$\mathrm{Re}\int_{-\infty}^{T} j(t)v^*(t)dt \geqslant 0$$ for all T;

(iv) for any positive number R, the domain of $R \otimes F$ includes any function $j(t)$ such that $j(t) \epsilon L^2(-\infty, \infty)$ and $j(t)=0$ if $t<t_0$ for some t_0; and

(v) if $j(t)$ satisfies the above conditions, then $(Fj)(t)=0$ for $t<t_0$.

In the next few sections, the general impedance as defined is studied under no further assumption.

3. IMPEDANCE FUNCTION FOR $R \otimes F$

The first question to be answered is this: In what way can an impedance function be associated with $R \otimes F$ where $R>0$? This is actually not entirely trivial to answer.

Let

$$(R \otimes F)j=v, \tag{3.1}$$

where j satisfies the conditions of (iv), then

$$F(j-v/R)=v. \tag{3.2}$$

From (iii) and the equality

$$\int_{-\infty}^{T} |j|^2 dt=\frac{1}{R^2}\int_{-\infty}^{T} |v|^2 dt+\int_{-\infty}^{T} |j-v/R|^2 dt + \frac{2}{R}\mathrm{Re}\int_{-\infty}^{T} (j-v/R)^* F(j-v/R)dt, \tag{3.3}$$

it follows that each of the integrals on the right-hand side approaches a non-negative limit as $T \to \infty$. In particular

$$\int_{-\infty}^{\infty} |v|^2 dt \leqslant R^2 \int_{-\infty}^{\infty} |j|^2 dt. \tag{3.4}$$

It is natural to define the impedance function $Z(x)$ by the relation $Z(x)=V(x)/J(x)$. It remains to be proved that $Z(x)$ is indeed independent of $j(t)$. In order to do this, let

$$Z_1(x)=V_1(x)/J_1(x), \tag{3.5}$$

where

$$j_1(t)=\begin{cases} 1 & \text{for } 0 \leqslant t \leqslant 1, \\ 0 & \text{otherwise,} \end{cases} \tag{3.6}$$

and

$$v_1(t)=[(R \otimes F)j_1](t). \tag{3.7}$$

By a suitable procedure, $Z_1(x)$ indeed satisfies

$$Z_1(x)=V(x)/J(x)$$

provided $j(t)$ satisfies the conditions of (iv) and is also a step function with a finite number of steps at rational points of t. Then it follows from (3.4) that $Z_1(x)$ is essentially bounded by R. With this result, it follows that $Z_1(x)=Z(x)$. More precisely,

$$V(x)=Z(x)J(x) \tag{3.8}$$

almost everywhere for any $j(t)$ that satisfies the conditions of (iv).

Furthermore, since $v(t)=j(t)=0$ for $t<t_0$, the functions V and J are also defined in the upper half of the $\omega=x+iy$ plane. The term "upper half plane" refers to the set of $y>0$. In fact, there is a $Z(\omega)$ such that

$$V(\omega)=Z(\omega)J(\omega) \tag{3.9}$$

for $\mathrm{Im}\omega>0$. In order to see this, it is sufficient to note that from Phragmén-Lindelöf theorem it follows that $Z_1(\omega)=V_1(\omega)/J_1(\omega)$ is bounded by R from a consideration of some $y=$constant lines, and further that a Paley-Wiener theorem[1] leads to the identity $Z(\omega)=Z_1(\omega)$. The relation between $Z(\omega)$ and $Z(x)$ is

$$Z(x)=Z(x+i0) \tag{3.10}$$

for almost all x. By extension, the term "impedance function" is used to denote a complex function of one complex variable such as $Z(\omega)$ defined everywhere in the upper half plane and almost everywhere on the real axis.

It remains to show that the real part of $Z(\omega)$ for the impedance $R\otimes F$ is never negative. This is intuitively obvious but not entirely trivial to prove. In order to show that $\mathrm{Re}Z(x)\geqslant 0$ almost everywhere, assume the contrary. That is, assume that there is a positive number ϵ, such that

$$\mathrm{Re}Z(x)<-\epsilon \tag{3.11}$$

on a set P of positive measure. Without loss of generality, let P be of finite measure $\mu(P)$. Define a function $C(x)$ by

$$C(x)=\begin{cases}1 & \text{on } P,\\ \dfrac{\epsilon\mu(P)}{4R(1+x^2)} & \text{otherwise.}\end{cases} \tag{3.12}$$

It is clear that both of the integrals $\int_{-\infty}^{\infty}C(x)dx$ and $\int_{-\infty}^{\infty}[|\log C(x)|/(1+x^2)]dx$ are finite. Therefore, by another theorem of Paley and Wiener,[2] there exists a function $j(t)$ defined over all t, such that $j(t)=0$ for $t<0$, and such that

$$|J(x)|=[C(x)]^{\frac{1}{2}}. \tag{3.13}$$

Let $v(t)$ be the corresponding voltage function. On the one hand, by the Parseval theorem, (3.8), and (3.11)–(3.13), it follows that

$$\begin{aligned}\mathrm{Re}\int_{-\infty}^{\infty}j(t)v^*(t)dt &= (1/2\pi)\,\mathrm{Re}\int_{-\infty}^{\infty}J(x)V^*(x)dx\\ &= (1/2\pi)\,\mathrm{Re}\int_{-\infty}^{\infty}C(x)Z(x)dx\\ &= (1/2\pi)\left[\int_P \mathrm{Re}Z(x)dx+\int_{P'}\frac{\epsilon\mu(P)}{4R(1+x^2)}\mathrm{Re}Z(x)dx\right]\\ &\leqslant (1/2\pi)\left[-\epsilon\mu(P)+\int_{-\infty}^{\infty}\frac{\epsilon\mu(P)}{4(1+x^2)}dx\right]<0;\end{aligned}$$

and on the other hand, by the definition of $\otimes$ and (iii),

$$\mathrm{Re}\int_{-\infty}^{\infty}j(t)v^*(t)dt=\mathrm{Re}\int_{-\infty}^{\infty}(j-v/R)^*F(j-v/R)dt+(1/R)\int_{-\infty}^{\infty}|v|^2dt\geqslant 0.$$

This is a contradiction and therefore

$$\mathrm{Re}Z(x)\geqslant 0 \tag{3.14}$$

for almost all x. If the upper half plane is mapped onto the unit circle and the Poisson integral representation is applied, it can be seen that

$$\mathrm{Re}Z(\omega)\geqslant 0 \tag{3.15}$$

in the entire upper half plane. Finally, it should be remarked that (3.14) can be proved equally easily without using the theorem of Paley and Wiener.

4. IMPEDANCE FUNCTION FOR F

So far, an impedance function has been associated with $R\otimes F$. In this section, such a function is associated with F itself. This is a new situation since as yet nothing is known about the domain of F.

Consider the combination $R'\otimes[R\otimes F]$. Let $v=\{R'\otimes[R\otimes F]\}j$, where j satisfies the conditions of (iv). Since the operation $\otimes$ is clearly associative, this is the same as $[R'\otimes R]\otimes F$. Then it follows from (v) and (3.4) that v also satisfies the conditions of (iv). Let $j=j_1+j_2$ and $v=R'j_1=[R\otimes F]j_2$, then j_1 and j_2 must also satisfy the condition of (iv). This incidentally shows the consistency between the definition of $\otimes$ and the assumption (iv). Since all Fourier transforms are well defined, it follows that

$$V(\omega)=\frac{R'Z(\omega)}{R'+Z(\omega)}J(\omega). \tag{4.1}$$

[1] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (American Mathematical Society, Providence, Rhode Island, 1934), Theorem V.

[2] Reference 1, Theorem XII.

In particular, if

$$R''=\frac{RR'}{R+R'}, \tag{4.2}$$

then

$$R''=R\otimes R'. \tag{4.3}$$

Note that (4.2) is a numerical equation, while (4.3) is an operator equation. In order to avoid confusion, let the impedance function associated with $R\otimes F$ be designated by $Z_R(\omega)$. Then, in view of (4.1) and (4.3),

$$Z_{R''}(\omega)=\frac{R'Z_R(\omega)}{R'+Z_R(\omega)}.$$

And hence, by symmetry,

$$\frac{RZ_R(\omega)}{R-Z_R(\omega)}=\frac{R'Z_{R'}(\omega)}{R'-Z_{R'}(\omega)}=Z(\omega). \tag{4.4}$$

Since this quantity is independent of R, let it be called $Z(\omega)$ as indicated. Note that pointwise

$$Z(\omega)=\lim_{R\to\infty} Z_R(\omega). \tag{4.5}$$

Physically, this means that $Z(\omega)$ is the impedance function for the parallel combination of F with an infinite resistance, and hence should be that for F itself. Therefore, the function $Z(\omega)$ is defined to be the impedance function for F.

In what part of the ω plane is Z defined? Clearly Z is defined in the upper half plane and on the real axis except the point set where $Z_R=R$ for all R. If this set contains the entire upper half plane, then, with $[R\otimes F]j=v$,

$$j=v/R=(1/R)F(j-v/R)=(1/R)F(0).$$

This is impossible because $j(t)$ is quite arbitrary. Therefore $Z_R\neq R$. Physically, this corresponds to the case of an open circuit which is not considered here. The function Z is thus defined in the upper half plane except possibly at a denumerable number of poles. However, it follows (3.14) that

$$\mathrm{Re}1/Z(\omega)\geqslant -1/R$$

for all $R>0$. Hence, $\mathrm{Re}Z(\omega)\geqslant 0$. Accordingly, $Z(\omega)$ cannot even have a pole. The conclusion is therefore that, everywhere in the upper half plane, and almost everywhere on the real axis, $Z(\omega)$ is well defined and

$$\mathrm{Re}Z(\omega)\geqslant 0. \tag{4.6}$$

Furthermore, it follows from (4.6) that either $Z(\omega)\equiv 0$ or else $Z(\omega)\neq 0$ at all in the upper half plane. The former situation corresponds to the physical case of a short circuit.

In order to justify calling $Z(\omega)$ the impedance function for F, it is necessary to prove the following. If $v=Fj$ and both j and v satisfy the conditions of (iv), then

$$V(\omega)=Z(\omega)J(\omega). \tag{4.7}$$

This can be shown as follows. From

$$[R\otimes F][j+v/R]=v, \tag{4.8}$$

it follows that

$$\frac{RZ(\omega)}{R+Z(\omega)}[J(\omega)+V(\omega)/R]=V(\omega),$$

and hence (4.7).

In the analysis of electric networks that consist of only a finite number of ideal resistances, ideal inductances, and ideal capacitances, it is well known that every admissible impedance function is a rational positive real function, that is, a function $f(\omega)$ which is a quotient of two polynomials such that $\mathrm{Im}f(iy)=0$ and $\mathrm{Re}f(\omega)\geqslant 0$ for $\mathrm{Im}\omega\geqslant 0$. In the present more general formulation, it can only be said that every admissible impedance function is a positive function, where a positive function is defined to be an analytic function of one complex variable in the upper half plane such that its real part is non-negative there.

Since $Z_R(x)$ cannot equal R on a set of positive measure, it follows from Fatou's theorem that

$$Z(x+i0)=Z(x) \tag{4.9}$$

almost everywhere. In this connection, the following remark may be worthwhile. In this paper, results are stated in terms of normal approach to the real axis, such as (4.9). Actually, however, all equations of this form are valid uniformly in any angular sector away from the real axis. For the present purpose, no elaboration of this sort is necessary.

5. IMPEDANCE AS DETERMINED BY $Z(\omega)$

It has been shown that with every impedance a function $Z(\omega)$ may be associated. In this section a few remarks are made about the impedance when $Z(\omega)$ is known. In the next sections, a general representation for $Z(\omega)$ is given, with the purpose of getting the convolution representation of impedance as given in Sec. 7.

The specific questions to be considered here are as follows. Given $j(t)$ and $Z(\omega)$, how can $v(t)$ be found? Furthermore, does every positive function represent an impedance?

Unfortunately, these questions cannot be answered in complete generality. The reason is that some assumption must be made about the function $j(t)$. For the present purpose, $j(t)$ is assumed to satisfy the condition that $\int_{-\infty}^{T}|j(t)|^2dt$ exists for all finite T. Physically, this means that up to any given instant, a finite amount of heat is generated if $j(t)$ is applied to a pure resistance. This assumption is true in most physical situations. Mathematically, this means that only the restriction of an impedance to this class of functions $j(t)$ is being considered.

Given $Z(\omega)$ and $j(t)$ in this class, let $v(t)$ be found as follows. Let

$$j_T(t)=\begin{cases} j(t) & \text{for } t<T \\ 0 & \text{for } t\geqslant T \end{cases} \tag{5.1}$$

and similarly for $v_T(t)$. Then by definition

$$v_T(t)=\lim_{R\to\infty} v_{T,R}(t) \quad \text{for } t<T \tag{5.2}$$

where $v_{T,R}(t)$ is defined through Fourier transform by

$$V_{T,R}(\omega)=\frac{RZ(\omega)}{R+Z(\omega)}J_T(\omega), \tag{5.3}$$

and the limit in the mean should be taken over the interval $(-\infty, T)$.

Before showing that the function $v(t)$ found this way has all the desired properties, a few remarks are appropriate. Note first that, as defined by (5.2),

$$v_T(t)=v_{T'}(t) \quad \text{(almost everywhere)}$$

for $t<\min(T,T')$. Indeed, the procedure would be entirely nonsensical if this were not true. For simplicity, let $v_T(t)$ be chosen such that

$$v_T(t)=v_{T'}(t)$$

for $t<\min(T,T')$. This can then give definite meaning to $v(t)$. The function $v(t)$ is then well defined if all $v_T(t)$ exist as a limit in the mean. Secondly, the limit in the mean is taken over the interval $(-\infty, T)$ instead of the interval $(-\infty, \infty)$ because the limit in the mean may fail to exist due to resonance phenomena. Thirdly, to ascertain the voltage response for $t<T$, it is physically unreasonable to require a knowledge of the current for $t\geqslant T$. Therefore, it is simplest to imagine that the current is zero for $t\geqslant T$.

It remains to show that the operator on $j(t)$ to give $v(t)$ as defined indeed satisfies all the conditions set forth in Sec. 2. The conditions (i) and (ii) are satisfied trivially. To verify (iii), note first that the integral indeed exists because of (5.2). Moreover,

$$\operatorname{Re}\int_{-\infty}^{T} j_T(t)v_{T,R}^*(t)dt=\operatorname{Re}\int_{-\infty}^{\infty} j_T(t)v_{T,R}^*(t)dt$$

$$=\int_{-\infty}^{\infty}|J_T(x)|^2\operatorname{Re}\frac{RZ(x)}{R+Z(x)}dx\geqslant 0.$$

This yields (iii) for all T. Any extension of the impedance is required to satisfy this condition also. Since the condition (v) is also trivially satisfied provided $v(t)$ is appropriately chosen, it only remains to verify (iv).

Suppose $j(t)\epsilon L^2(-\infty, \infty)$. First, let it be verified that the voltage function corresponding to $j-(1/R)F_Rj$ is F_Rj, where the Fourier transform of F_Rj is $\{RZ(\omega)/[R+Z(\omega)]\}J(\omega)$. To do this, it is sufficient to note that the function

$$\frac{R'Z}{R'+Z}\left[J-\frac{1}{R}\frac{RZ}{R+Z}J\right]-\frac{RZ}{R+Z}J=\frac{Z}{R'+Z}\frac{RZ}{R+Z}J$$

indeed approaches 0 in the limit in the mean as $R'\to\infty$. It only remains to show that if the $v(t)$ corresponding to $j(t)$ is $-Rj(t)$, then $j(t)=0$ almost everywhere. This follows directly from (iii). Note that the fact $j(t)=0$ for $t<t_0$ is not used in this proof.

6. PROPERTIES OF $Z(\omega)$

In this section, a few simple properties of $Z(\omega)$ are found.

A theorem of Herglotz gives a representation of any analytic function in the unit circle with a non-negative real part. This theorem can be easily applied to the upper half plane, as given, for example, by Shohat and Tamarkin.[3] $Z(\omega)$ may be written as

$$Z(\omega)=-iA\omega+iB-i\int_{-\infty}^{\infty}\frac{1+s\omega}{s-\omega}d\alpha(s) \tag{6.1}$$

where A is a positive constant, B is real, and $\alpha(s)$ is a nondecreasing bounded real function. The meaning of $\alpha(s)$ is clearly provided by the formula

$$\operatorname{Re}Z(x)=\pi(1+x^2)\alpha'(x) \tag{6.2}$$

almost everywhere. Furthermore, $Z(\omega)$ determines A, B, and $\alpha(s)$ uniquely.

In the synthesis of audio networks, an impedance is often separated into a reactance and a minimum-reactive part. This result may be obtained from (6.1) by the Lebesgue decomposition theorem. Thus, for any nondecreasing function $\alpha(s)$,

$$\alpha(s)=\alpha_1(s)+\alpha_2(s)+\alpha_4(s), \tag{6.3}$$

where $\alpha_1(s)$ is nondecreasing and absolutely continuous, and both $\alpha_2(s)$ and $\alpha_4(s)$ are nondecreasing and have zero derivative almost everywhere; $\alpha_2(s)$ is a step function and $\alpha_4(s)$ is continuous. With (6.3), let

$$Z(\omega)=\sum_{j=1}^{5} Z_i(\omega), \tag{6.4}$$

where

$$Z_3(\omega)=-iA\omega, \quad Z_5(\omega)=iB, \tag{6.5}$$

and

$$Z_j(\omega)=-i\int_{-\infty}^{\infty}\frac{1+s\omega}{s-\omega}d\alpha_j(s) \tag{6.6}$$

for $j=1$, 2, or 4. The identification is as follows. The part $Z_1(\omega)$ is minimum reactive, and all the others represent reactances. The part $Z_3(\omega)$ represents the

[3] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments* (American Mathematical Society, Providence, Rhode Island, 1943), Lemma 2.1.

inductance, and $Z_2(\omega)+Z_3(\omega)$ constitutes every reactance encountered in conventional circuit analysis. The part $Z_4(\omega)$ is heretofore unknown and to be discussed in Sec. 9; $Z_5(\omega)$ is the frequency-independent ideal-gyrator part, referred to before.

It is still necessary to define a reactance. Intuitively, this should be done as follows. If F is a reactance, and $(R\otimes F)j=v$, then

$$\mathrm{Re}\int_{-\infty}^{\infty} j(t)v^*(t)dt=(1/R)\int_{-\infty}^{\infty}|v(t)|^2dt. \qquad (6.7)$$

Physically, this means that the energy put into the parallel combination of a resistance and reactance is all dissipated in the resistance. However, this should be made more precise. Certainly, it is necessary that $v\epsilon L^2(-\infty,\infty)$. Some condition has to be put on $j(t)$. At first it may seem reasonable to require that $j(t)$ is Lebesgue square integrable over every compact set. However, a moment of thought indicates that, in general, this is not enough to insure (6.7). So the simplest thing to do is to assume that $j(t)\epsilon L^2(-\infty,\infty)$. Therefore, a reactance F is defined by the condition that if both $j(t)$ and $v(t)$ are Lebesgue square integrable over $(-\infty,\infty)$ and $(R\otimes F)j=v$, then (6.7) holds.

It remains to show that an impedance is a reactance if and only if the corresponding $Z(\omega)$ satisfies

$$\mathrm{Re}Z(x)=0 \qquad (6.8)$$

almost everywhere. The "only if" part is completely trivial and the "if" part follows from the equation

$$\mathrm{Re}\frac{R+Z(x)}{RZ(x)}=\frac{1}{R}$$

almost everywhere if (6.8) holds.

Finally, it may be stated that if both $j(t)$ and $v(t)$ are required to be real functions of the time t, then

$$Z_j(-\omega^*)=Z_j^*(\omega), \qquad (6.9)$$

$$d\alpha_j(s)=-d\alpha_j(-s), \qquad (6.10)$$

and

$$B=0. \qquad (6.11)$$

7. ANALYSIS OF THE TRANSIENT

In elementary electric-circuit textbooks, it is often stated that

$$v(t)=\int_{0-}^{t+} j(\tau)f(t-\tau)d\tau,$$

where $f(t)$ is called the δ-function response of the impedance. Of course this equation is not strictly true, but a formula of this sort does, indeed, exist. In this section, a description of any impedance in the time domain is to be derived.

First of all, it should be proved that

$$\int_{-\infty}^{\infty}\left|\frac{Z(\omega)}{\omega^2}\right|^2 dx$$

is uniformly bounded for $y=\mathrm{Im}\omega>y_0>0$, where the integration is performed along a line parallel to the real axis. This follows from (6.1) with the observation that

$$\int_{-\infty}^{\infty}\left|\frac{1+s\omega}{\omega^2(s-\omega)}\right|^2 dx$$

is uniformly bounded for $y>y_0>0$ and for all s.

Therefore, there exists a function $h(t)$ such that $h(t)=0$ for $t<0$, $h(t)e^{-\epsilon t}$ is square integrable over $(-\infty,\infty)$ for any $\epsilon>0$, and the Fourier transform of $h(t)$ is $-Z(\omega)/\omega^2$.

The following result is to be proved. If $j(t)=0$ for $t<0$ and $j''(t)$ is square integrable over every bounded interval, then an impedance may be represented through

$$v(t)=\int_{0-}^{t+} j(\tau)d^2h(t-\tau). \qquad (7.1)$$

Of course, the integral here should be interpreted by integration by parts. Physically, this means that a modified Stieljes convolution may be used to study transient behavior provided the current function is sufficiently smooth. Incidentally, this also asserts that this special class of current functions is admissible to every impedance.

To prove this result, it may be assumed without loss of generality that $j''(t)$ is square integrable over $(-\infty,\infty)$. First, if $v(t)$ is defined as in (7.1), $v(t)e^{-\epsilon t}$ is square integrable over $(-\infty,\infty)$ for any $\epsilon>0$, for

$$\int_{-\infty}^{\infty}\left|\int_{0-}^{t+} j(\tau)d^2h(t-\tau)\right|^2 e^{-\epsilon t}dt$$

$$\leqslant\int_{0}^{\infty}\int_{0}^{t}|j''(\sigma)|^2d\sigma\int_{0}^{t}|h(\tau)|^2d\tau e^{-\epsilon t}dt$$

$$\leqslant\int_{-\infty}^{\infty}|j''(\sigma)|^2d\sigma\int_{-\infty}^{\infty}|h(\tau)|^2e^{-\epsilon\tau/2}d\tau$$

$$\times\int_{0}^{\infty}e^{-\epsilon t/2}dt<\infty. \qquad (7.2)$$

In particular

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|j''(\tau)|\,|h(t-\tau)|\,d\tau dt<\infty. \qquad (7.3)$$

Therefore, by Fubini's theorem, if $\mathrm{Im}\omega<0$,

$$\int_{-\infty}^{\infty}\int_{0-}^{t+} j(\tau)d^2h(t-\tau)e^{i\omega t}dt$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} j''(\tau)h(t-\tau)e^{i\omega t}dtd\tau$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} j''(\tau)h(\sigma)e^{i\omega(\tau+\sigma)}d\sigma d\tau=Z(\omega)J(\omega). \quad (7.4)$$

Since

$$ZJ-\frac{RZ}{R+Z}J=ZJ\frac{Z}{R+Z}, \quad (7.5)$$

it follows from (7.2) that

$$e^{-\epsilon t}\int_{0-}^{t+} j(\tau)d^2h(t-\tau)=\underset{R\to\infty}{\text{l.i.m.}}\ e^{-\epsilon t}(R\otimes F)j \quad (7.6)$$

for $\epsilon>0$. Comparison with (5.2) gives the desired result.

Equation (7.1) indicates that $h(t)$ is an important and interesting function. If it has a second derivative, then $h''(t)$ is the δ-function response of the impedance. Otherwise, the δ-function response has to be considered to be a distribution.

It is natural to ask for a direct relation between $h(t)$ and $\alpha(s)$. Formal calculation from (6.1) gives the result

$$h(t)=-g(t)+''g(t)+g(0)+A+iBt \quad (7.7)$$

for $t\geqslant 0$, where

$$g(t)=\int_{-\infty}^{\infty} e^{-ist}d\alpha(s), \quad (7.8)$$

and

$$''g(t)=\int_0^t\int_0^{t_1} g(t_2)dt_2dt_1. \quad (7.9)$$

The simplest way to prove (7.7) is by a direct calculation of Fourier transform. Since $g(t)$ is bounded by the total variation of $\alpha(s)$, all changes of the order of integration are justifiable by Fubini's theorem. The rest of the proof is straightforward.

In particular, (7.7) asserts that $h(t)$ may be easily calculated from the simpler entities A, B and $g(t)$.

8. INVARIANCE AND NORMALIZED IMPEDANCES

The function $g(t)$ is important because of its simplicity in form on the one hand and its simple relation to $h(t)$ on the other hand. Except for a real multiplying factor, $g(t)$ is the Fourier transform of a probability distribution, and thus gives a curious connection between circuit theory and probability theory. From this point of view, impedances with the property that the total variation of $\alpha(s)$ is equal to one are particularly important. Before studying these "normalized" impedances, the following question should be considered. Is the separation of $Z(\omega)$ as given in Sec. 6 really physically meaningful, or, in other words, is the separation anything more than a mathematical fiction?

A reference to (6.2) indicates that the representation (6.1) puts a special emphasis on the point $\omega=i$; indeed

$$Z(i)=A+iB+[\alpha(\infty)-\alpha(-\infty)]. \quad (8.1)$$

By shifting this arbitrary emphasis from $\omega=i$ to another point $\omega=\omega_0$ with $\mathrm{Im}\omega_0>0$, (6.1) may be generalized to yield

$$Z(\omega)=-iA\omega+iB_{\omega_0}$$

$$-i\int_{-\infty}^{\infty}\frac{s(\omega-2\,\mathrm{Re}\omega_0)+|\omega_0|^2}{s-\omega}d\alpha_{\omega_0}(s). \quad (8.2)$$

Note that the constant A is independent of the choice of ω_0. The validity of (6.2) implies that

$$(1+x^2)d\alpha(x)=(|\omega_0|^2+x^2-2x\,\mathrm{Re}\omega_0)d\alpha_{\omega_0}(x). \quad (8.3)$$

Since the decomposition of $Z(\omega)$ is based on the local structure of $\alpha(s)$, it is easy to see that the real parts of $Z_j(\omega)$, $j=1\cdots5$, are invariant under a change of ω_0. This is not true of the imaginary parts; they may change by additive constants.

In the special case that $v(t)$ is real whenever $j(t)$ is real, it is true that $Z_j(\omega)$ are invariant provided $\mathrm{Re}\omega_0=0$, for

$$Z_j(\omega)-Z_{\omega_0,j}(\omega)=i\int_{-\infty}^{\infty} s[d\alpha_j(s)-d\alpha_{\omega_0,j}(s)]$$

$$=|\omega_0{}^2-1|\int_{-\infty}^{\infty}\frac{s}{1+s^2}d\alpha_{\omega_0,j}(s)=0, \quad (8.4)$$

$j=1, 2, 4$. The requirement that $\mathrm{Re}\omega=0$ is reasonable because of symmetry.

An impedance is said to be normalized at ω_0 if the total variation of $\alpha_{\omega_0}(s)$ is 1. It is clear that this is not a very useful physical concept because of its lack of invariance with respect to ω_0. For simplicity, consider only the case $\omega_0=i$. Now $g(t)$ is truly the characteristic function of a probability distribution, and thus existing theorems for probability theory may be directly adopted, for example from Wintner.[4] Out of the infinite variety of possible problems, only the following one is considered. Suppose $h(t)$ is known for small t, is it possible that this information is enough to determine $h(t)$ for all t? This cannot be true in general, for the knowledge about $h(t)$ over any finite interval is usually not enough to determine $h(t)$ for all t. The simplest example is the reflection from the end of a transmission line. For this question, note the theorem that if $g(t)=1+O(t^2)$ as $t\to 0$, then $g(t)=1$. When this is applied to a

[4] A. Wintner, "The Fourier Transforms of Probability Distributions." Lectures at Johns Hopkins University, Baltimore, Maryland, 1947 (unpublished).

normalized impedance with (7.7), it asserts that if $B=0$ and $h(t)=A+\frac{1}{2}t^2+O(t^2)$ as $t\to 0$, then $h(t)=A+\frac{1}{2}t^2$. Physically this means that if $j(t)$ and $v(t)$ are real, if the impedance is normalized, and if, for small t, $h(t)$ behaves sufficiently like that of a series combination of an inductance and a unit capacitance, then the impedance is indeed electrically identical to that series combination. In particular, for any impedance, not necessarily normalized, if $h(t)=A+O(t^2)$ as $t\to 0$, then $h(t)=A$. This is to be expected physically for $h(t)$ is the voltage function corresponding to the current input

$$j(t)=\begin{cases}0 & t\leqslant 0,\\ t & t\geqslant 0.\end{cases}$$

Next, consider the case $B=0$ and $h(t)=A+O(t^2)$. In this case $g(t)=g(0)+O(t^2)$, and hence $\int_{-\infty}^{\infty} s^2 d\alpha(s)$ exists. This is interesting in considering reactances because for $L-C$ impedances, all moments exist for $\alpha(s)$. In general, if the second moment exists for $\alpha(s)$, then $h(t)$ has a second derivative for $t>0$ and for $B=0$

$$\frac{d^2}{dt^2}[h(t)-h(0)]=\int_{-\infty}^{\infty} e^{-ist}d\beta(s) \tag{8.5}$$

for $t\geqslant 0$, where

$$d\beta(s)=(1+s^2)d\alpha(s). \tag{8.6}$$

This is essentially the δ-function response.

9. AN EXAMPLE

Referring to the decomposition of $Z(\omega)$ in Sec. 6, by far the most peculiar part is $Z_4(\omega)$. Since this type of impedance is hitherto unknown, it is instructive to work out an example. Perhaps the simplest one may be constructed from the Cantor function, which is defined over $[0,1]$ and satisfies the relation

$$C\left(\frac{s}{3}\right)=\begin{cases}1/2C(s) & \text{for } 0\leqslant s\leqslant 1,\\ 1/2 & \text{for } 1\leqslant s\leqslant 2,\\ 1-1/2C(3-s) & \text{for } 2\leqslant s\leqslant 3.\end{cases} \tag{9.1}$$

Its approximate shape is given in Fig. 1. In order to satisfy (6.10), choose for simplicity $\alpha(s)$ as follows, in view of (8.5)

$$\beta(s)=\int(1+s^2)d\alpha(s)$$
$$=\begin{cases}-1/2 & \text{for } s\leqslant -1,\\ -1/2C(-s) & \text{for } -1\leqslant s\leqslant 0,\\ 1/2C(s) & \text{for } 0\leqslant s\leqslant 1,\\ 1/2 & \text{for } 1\leqslant s.\end{cases} \tag{9.2}$$

With this choice of $\alpha(s)$,

$$Z(\omega)=-i\int_0^1 \frac{dC(s)}{s^2-\omega^2}, \tag{9.3}$$

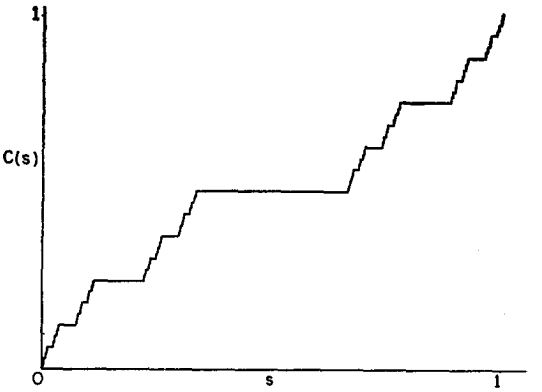


FIG. 1. The Cantor function.

and

$$h''(t)=\int_0^1 \cos st\,dC(s) \quad \text{for } t\geqslant 0. \tag{9.4}$$

Since $Z(\omega)$ is not even continuous for ω real, only $h''(t)$ is to be evaluated numerically. For this purpose, let

$$k(t)=\int_0^1 e^{-is(t-1/2)}dC(s). \tag{9.5}$$

Since $k(t)$ is real, it follows that

$$h''(t)=k(t)\cos\tfrac{1}{2}t. \tag{9.6}$$

A recurrence relation may be obtained by substituting (9.1) into (9.5). The result is simply

$$k(3t)=k(t)\cos t. \tag{9.7}$$

Since a power series expansion for $k(t)$ may be obtained from (9.7), Eqs. (9.6) and (9.7) are enough for the numerical determination of $h''(t)$.

A few simple properties of $h''(t)$ may be obtained by inspection. It is clear from (9.4) that

$$|h''(t)|\leqslant 1, \tag{9.8}$$

with equality sign only for $t=0$. From (9.6) and (9.7), the principal peaks of $h''(t)$ must be in the neighborhood of

$$t_n=2\cdot 3^n\cdot\pi, \tag{9.9}$$

where n is an integer. Equation (9.7) also indicates that the peaks become narrower as n increases. From

$$h''(t_n)=h''(2\pi)\neq 0 \tag{9.10}$$

and (9.8), it is seen that the maximum values of $|h''(t)|$ for these principal peaks are bounded above and below by positive numbers. Physically, this means that the transient for a δ-function current input never dies out, and thus there is no steady state in the conventional sense. This is to be expected because the function defined in (9.2) is not the integrated spectrum of an almost periodic function. The function $h''(t)$ is given graphically in Fig. 2.

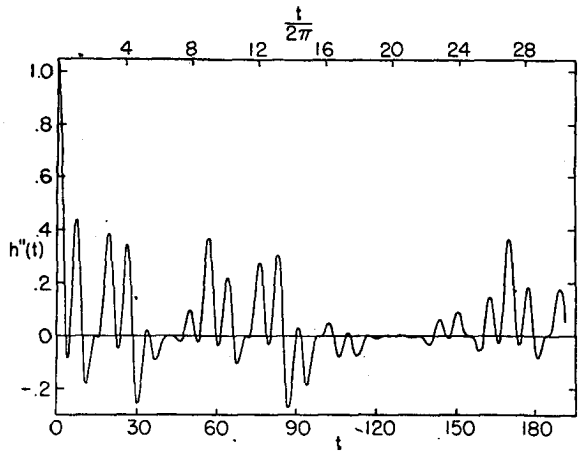


FIG. 2. A delta-function response.

10. THE KRONIG-KRAMERS RELATIONS

The Kronig-Kramers relations are the integral representations of the real and imaginary parts of $Z(x)$ in terms of each other. One of the relations may be obtained easily from the results of Sec. 6, without the many defects of the conventional proof by contour integration[5]; the other one turns out to be false in general. It is of interest to note that Kronig gave only the correct one.

To begin with, let Eqs. (6.9)–(6.11) be satisfied. Without the assumption that $\alpha(s)$ is absolutely continuous, the integral representations are impossible because of (6.2). To get the Kronig-Kramers relations, it is thus further assumed that $\alpha(s)$ is absolutely continuous but the constant A in (6.1) is not necessarily zero. With

$$Z(\omega)=R(\omega)+iX(\omega), \tag{10.1}$$

it follows from (6.10), (6.11), (6.1), and (8.6) that

$$X(\omega)=-2\omega\int_0^\infty d\beta(s)\,\mathrm{Re}\frac{1}{s^2-\omega^2}-A\,\mathrm{Re}\omega \tag{10.2}$$

for $\mathrm{Im}\omega>0$. It remains to take the limit $\mathrm{Im}\omega\to 0$.

Since $\alpha(s)$ is bounded, $\int_0^\infty d\alpha(s)/(1+s^2)$ exists. Therefore, by Fatou's theorem, it follows[6] from (10.2) that

$$X(x)=-Ax-2x\int_0^\infty\frac{d\beta(s)}{s^2-x^2}, \tag{10.3}$$

where Cauchy principal value is taken at $s=|x|$. By (6.2)

$$X(x)=-Ax-\frac{2x}{\pi}\int_0^\infty\frac{R(x)dx}{s^2-x^2}. \tag{10.4}$$

[5] J. H. Van Vleck, Massachusetts Institute of Technology Radiation Laboratory Report 735, 1945 (unpublished).

[6] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford University Press, New York, 1948), 2nd ed., p. 124, Theorem 92.

Now, let this result be applied to the complex dielectric constant. Let it be separated into the real and imaginary parts by

$$\epsilon=\epsilon'+i\epsilon''. \tag{10.5}$$

Since the electric energy dissipation per unit time is proportional to $\omega\epsilon''\|E\|^2$, where E is the electric field vector, the combination $-i\omega\epsilon(\omega)$ is an impedance function. Application of (10.4) to this function yields

$$\epsilon'(x)=A+\frac{2}{\pi}\int_0^\infty\frac{s\epsilon''(s)ds}{s^2-x^2}, \tag{10.6}$$

still with the principal value taken at $s=|x|$. This is one of the Kronig-Kramers relations except that the constant A need not equal $\lim_{x\to\infty}\epsilon'(x)$, which may not exist.

It may be noted that (10.6) is sufficient to insure a causal relation between **D** and **E**.

The second Kronig-Kramers relation is usually given as

$$\epsilon''(x)=-\frac{2x}{\pi}\int_0^\infty\frac{\epsilon'(s)-A}{s^2-x^2}ds, \tag{10.7}$$

with principal value at $s=|x|$. However, in general the integral on the right does not converge. To see this, it is sufficient to consider

$$\beta'(s)=\sum_{n=1}^{\infty}\frac{a_n}{(s-n)^2+a_n^2},$$

with $a_n>0$ for all n. It is clear that the integral diverges if a_n approaches zero sufficiently rapidly as $n\to\infty$. This represents the physical situation of a sequence of absorption lines with rapidly increasing Q. Although there are a number of possible sufficient conditions for (10.7) to hold, none of them seems to make sense physically.

Even so, it seems to be reasonable to think that the invalidity of (10.7) has more to do with the interpretation of the integral than with the actual physical situation. From the work of Titchmarsh,[7] it seems to be true that (10.7) is valid provided the integral is interpreted as follows:

$$\int f(x)dx=\lim_{n\to\infty}\int f_n(x)dx,$$

where

$$f_n(x)=\begin{cases} n & \text{for } f(x)\geqslant n,\\ f(x) & \text{for } |f(x)|\leqslant n,\\ -n & \text{for } f(x)\leqslant -n.\end{cases}$$

It should be particularly noted that the present proof of one of the Kronig-Kramers relations depends on the observation that $\mathbf{D}\cdot\mathbf{E}$ represents energy storage, and

[7] E. C. Titchmarsh, Proc. London Math. Soc. **29**, 49 (1929).

that with this observation none of the difficulties encountered by Toll[8] arises.

11. DISCUSSION

Except in a few examples, this paper is concerned with a general impedance under the conditions of linearity, passivity, reproducibility, and causality. Although linearity in terms of the superposition of two functions only is postulated, this fortunately leads to the stronger linearity in terms of Fourier transforms. The assumption of reproducibility, or the commutation with time translation, is necessary for the existence of a frequency response, in particular of the impedance function. Of course this does not necessarily imply that the physical entity under consideration is mechanically stationary. Suppose that a two-terminal black box is given as a physical realization of an impedance. It should be particularly emphasized that the present work is exclusively concerned with the external behavior of the black box; what happens inside is of no interest at all. In particular, this should be noticed in connection with the definition of a reactance. If the given black box does not represent a reactance, it is still not necessarily true that electric energy is converted into heat inside the black box. The energy that cannot be extracted again may be stored in a number of other forms, such as electric, mechanical, or chemical energy. A simple example is the semi-infinite lossless transmission line. Its external behavior is identical with that of a resistance, and hence it is certainly not reactive. But no heat is generated.

The Lebesgue decomposition theorem has been used to separate a general impedance function into several simpler parts. Of these the example given in Sec. 9 is of special interest. Let $Z(\omega)$ denote that particular impedance function defined there, then $e^{-Z(\omega)}$ is indeed admissible as a 1×1 scattering matrix. This is a counterexample to a result of van Kampen,[9] in that $e^{-Z(\omega)}$ cannot be extended to be a meromorphic function. This is due to an erroneous application of the Schwartz reflection principle. In general, this example is useful in testing a number of conjectures. Furthermore, it is an open question whether the behavior of $h''(t)$ in this case, in particular the boundedness of the peaks, is generally valid in any nontrivial subset of reactance functions constructed from continuous singular functions.

The following two anomalies may be of particular interest:

From the point of view of physics, it is inconceivable and against all intuition for a continuous singular function to appear. However, it seems to take a new physical principle to exclude these functions. For example, it cannot be argued that these functions should be excluded because no experiment can be performed in a finite time to ascertain their existence, since this same argument may be used to exclude a number of lossless systems.

In the last section, one of the Kronig-Kramers relations was obtained, including the number of subtractions. However, this was done at the expense of the other relation. It may be worth reemphasizing that the subtraction constant is not necessarily the limit of the imaginary part, more generally the absorptive part. If one more subtraction is used, both relations may be expected to hold. Loosely, it may be said that the failure of one of the Kronig-Kramers relations is due to to the possibility of extremely violent oscillations of the real part, whose occurrence is also against physical intuition.

It is of course completely unknown whether the difficulties encountered here also occur in problems of field theory or not. The assumption of passivity is of fundamental importance here; without this assumption the whole derivation fails. In field theory of elementary particles, it may be hoped that unitarity can play a similar role. Whether it does and whether it can be used to exclude the two anomalies mentioned in the last two paragraphs are completely open questions.

ACKNOWLEDGMENTS

For guidance and helpful discussions, I am greatly indebted to Professor P. C. Rosenbloom, Professor R. W. P. King, Professor A. M. Gleason, Professor J. H. Van Vleck, Professor R. J. Eden, Professor F. W. Gehring, Professor E. M. Purcell, and Professor R. H. Dicke.

[8] J. S. Toll, Ph.D. Thesis, Princeton University (1952).
[9] N. G. van Kampen, Phys. Rev. 89, 1072 (1953).

Convergence of Perturbation Expansions in Cutoff Meson Theories

WILLIAM M. FRANK
U. S. Naval Ordnance Laboratory, Silver Spring, Maryland
(Received June 12, 1961)

It is proved that the perturbation expansion of matrix elements for a class of meson theories has a nonzero radius of convergence. Included are theories which describe bosons of nonvanishing mass, coupled linearly or bilinearly to fermions (but not antifermions) via an interaction which cuts off or attenuates sufficiently rapidly for high boson momenta. By bounding the propagators by their values for zero boson momentum one obtains a majorizing series to the perturbation expansion which converges geometrically.

INTRODUCTION

THE applicability of perturbation methods to field theoretic problems has not been theoretically justified. A fundamental outstanding question is the convergence of perturbation expansions. In this paper a demonstration is presented of the convergence of perturbation expansions for cutoff meson theories if the coupling is sufficiently weak. This includes theories where a field of bosons (mesons) of nonzero mass interacts linearly and/or bilinearly with a fermion field (without antiparticles) and where the interaction attenuates sufficiently rapidly for high momenta of the mesons. The proof will apply to scalar, pseudoscalar, vector and axial vector, neutral, charged, or symmetric mesons interacting with one or more nucleons which may recoil or be recoilless. A detailed proof will be presented for the system of one fixed source nucleon interacting via linear coupling with any kind of meson. The modifications necessary in other cases will then be indicated. The quantities to be expanded are a class of matrix elements in terms of which nearly all quantities of physical interest can be evaluated. The perturbation expansion is a sum over all possible diagrams for the process described by this matrix element. The amplitude for each diagram will be majorized in such a manner that the majorizing series, a power series in the coupling constant, converges geometrically. No effort will be made in this paper to obtain sharp or realistic upper bounds for the radius of convergence beyond mere demonstration of a nonzero lower bound. The majorizing scheme consists in merely replacing all propagators by their value for zero momentum of the mesons. The generally large number of mesons in propagation in a large order diagram provides sufficient suppression of the amplitude to result in a total amplitude for all nth order diagrams of the order of $(\text{constant})^n$.

This naturally will imply analyticity of the matrix element as a function of coupling constant f within some circle in the complex plane about $f=0$. The sense of this statement, however, should be made clear. The derivation of the power series expansion in f is based upon certain assumptions concerning the spectrum. These may fail to hold for unreal f, undermining the physical basis for such an expansion. Nevertheless, the expansion analytically continued from real values will converge.

HAMILTONIAN AND MATRIX ELEMENTS

The system to be discussed, a fixed source nucleon with linear coupling to mesons, is described by the Hamiltonian ($\hbar=c=$ normalization volume $=1$)

$$H=H_0+H_1. \tag{1}$$

$$H_0=\sum_k \omega_k a_k^* a_k \tag{2}$$

is the free meson Hamiltonian, $\omega_k=(k^2+\mu^2)^{\frac{1}{2}}$ where μ is the mass of the meson and $\mu\neq 0$. The interaction Hamiltonian is

$$H_1=f\sum_k v(k)S_k(a_k+a_{-k}^*), \tag{3}$$

where f is the coupling constant. $v(k)$ is a real "autonomous" cutoff function, i.e., a real c-number function of k that guarantees convergence at the high-momentum end of all integrals in which it appears, independently of other factors present. This is to be distinguished from an "auxiliary" cutoff function which guarantees convergence only in the presence of certain other factors. S_k in a recoilless theory is a nucleon interaction operator describing the transition in nucleon state accompanying the absorption of a meson of momentum and internal state (charge, spin) described by k. S_k is a finite-dimensional matrix $S_k^*=S_{-k}$ and we denote an upper bound to the diagonal matrix elements of $S_k^*S_k$ by $|\mathcal{S}(\mathbf{k})|^2$, $[\mathcal{S}(\mathbf{k})=\mathcal{S}(|\mathbf{k}|)]$. We assume there are one physical particle states $|\alpha\rangle$, $|\beta\rangle$ in the discrete spectrum which go over into bare states $|\alpha)$, $|\beta)$ as $f\to 0$.[1] The presence of the discrete states which approach a bare state when $f\to 0$ is a sufficient, possibly necessary condition for derivation of perturbation expansions in f. If f were complex, the Hamiltonian would be non-Hermitian, and stationary states in the sense of time-separable normalization-conserving states will in general not exist. Whether the states, stationary in the sense that they undergo a factorable exponential decay in time (due to the complex frequencies), become bare states in the limit of $f\to 0$, is not clear. The physical assumptions which seem to be required for the justification of the perturbation expansion may not justify expansion for complex values of the coupling constant, although convergence for complex f can be proven. For f real, more than one discrete point in the energy spectrum may exist, as in the case of more than one

[1] W. Frank, Nuovo cimento (to be published).

fixed nucleon, so long as these states all approach bare states as $f \to 0$. For simplicity they are assumed degenerate.

Such a discrete state, denoted by $|\alpha\rangle$, would satisfy

$$(H-E_0)|\alpha\rangle=0, \tag{4}$$

and $|\alpha\rangle$ is a normalizable state. It is assumed for the present, but proven in the Appendix, that $E_0 \to 0$ as $f \to 0$. The algorithm of the Chew-Low formalism applicable to this system, permits the matrix elements of all operators between eigenstates of the total Hamiltonian to be written as the matrix elements of an operator function of the total Hamiltonian and nucleon variables alone between bound states of the system. Matrix elements of physical interest (scattering and production amplitudes, bound-state energy, form factors) all seem to involve operators of a special form: a product of an alternating sequence of nucleon operators and propagators. (By a propagator is understood the reciprocal of a polynominal in H, with possible projection or rejection operators present for certain states.) The convergence will therefore be studied for matrix elements of the canonical form

$$M(\mathbf{\Omega};\mathbf{K})\equiv\langle\beta|S_{k_l}(\Omega_l+H)^{-1}S_{k_{l-1}}\cdots \times S_{k_1}(\Omega_1+H)^{-1}S_{k_0}|\alpha\rangle. \tag{5}$$

$|\alpha\rangle$, $|\beta\rangle$ are bound meson eigenstates of H, $\mathbf{\Omega}$ is a set of Ω_i $(i=1, 2, \cdots, l)$ which are given (or "external") energy parameters, and $\mathbf{K}$ is a set of k_i $(i=0, 1, 2, \cdots, l)$, a given set of meson indices. This also constitutes a general canonical form for physically interesting matrix elements for theories other than Eq. (1).[2] In more general theories the S_k may denote the interaction currents with an incident or created meson, and the Ω_i $(i=1, \cdots, l)$ the energies of these external interacting quanta. If the above stated requirements hold for such S_k, convergence can be demonstrated. To prevent singular behavior at $f=0$, it is assumed none of the $\Omega_i(\Omega_i \in \mathbf{\Omega})$, is equal to zero. For situations of physical interest we first limit ourselves to Ω_i real, but having an infinitesimal imaginary part if the denominator is singular. Without loss of generality we may take all these imaginary parts to have the same sign, which we choose to be (say) positive. The convergence arguments will be applied with small modifications for Ω_i complex.

THE EXPANSIONS IN DIAGRAMS

A perturbative expansion of $M(\mathbf{\Omega};\mathbf{K})$ as is well known, is easily represented as a sum of the amplitudes corresponding to diagrams for the emission and absorption of virtual mesons relative to the external vertices S_k. Such a result can be obtained either by expansion of the term in the S matrix corresponding to $M(\mathbf{\Omega};\mathbf{K})$,[3] by expanding the propagators and initial and final state vectors of Eq. (5) in an explicit power series in f,[4] or by evaluation of the derivatives of $M(\mathbf{\Omega};\mathbf{K})$ with respect to f at $f=0$.[5] In each derivation one assumes a discrete state which becomes a bare state when $f=0$. The rules for writing down the amplitude for any such diagram are here summarized: (see Appendix)

(1) All diagrams corresponding to any total number of mesons emitted and absorbed are admissible, if they have the following structure: (a). Only one meson may be emitted or absorbed at a vertex. (b). The vertices may not coincide with any of the vertices for an external interaction S_k. (c). All emitted mesons must be absorbed. (d). The mesons emitted preceding (in time) the first vertex S_{k_0} may not all be absorbed preceding that vertex. (e). Similarly, the mesons absorbed following the final vertex S_{k_l} may not all have been emitted following it.

(2) Starting with the earliest vertex and proceeding consecutively in time through all the vertices, one forms an operator by multiplying in sequence (to the left) operators $fv(p)S_p^*$ for a vertex at which a meson p is emitted, $fv(p)S_p$ at a vertex at which a meson p is absorbed, and S_{k_i} for the vertex where an external interaction $S_{k_i}(k_i \in \mathbf{K})$ of Eq. (5) takes place. One evaluates the matrix element of the operator so formed between the initial bare nucleon state $|\alpha\rangle$ and the final nucleon state $|\beta\rangle$.

(3) For each interval between two consecutive vertices (whether for a virtual or an external interaction), one writes the energy denominator $(\sum\omega+\Omega)^{-1}$, where $\sum\omega$ is the sum of the energies of all the virtual mesons in propagation in that interval and Ω is generally a particular one of the $\Omega_i(\Omega_i \in \mathbf{\Omega})$ in Eq. (5). If the interval in question lies between the vertices for the external interactions $S_{k_{i-1}}$ and $S_{k_i}(k_i, k_{i-1} \in \mathbf{K})$, then Ω_i is chosen for its energy denominator. If the particular interval precedes S_{k_0} or follows S_k, then $-E_0$ is written for its value of Ω.

(4) One multiplies the quantity described in rule 2 by the quantity described in rule 3, and sums over the momenta and other indices of all the virtual mesons. One also multiplies by Z_2, the wave function renormalization.

(5) One sums these amplitudes over all admissible diagrams described in (1) to get the terms of the perturbation expansion.

(6) If $l=-1$, i.e., no external vertices are present, an additional factor of -1 is present.

For the purpose of clarification we cite an illustration. For

$$M(\mathbf{\Omega};\mathbf{K})=\langle\beta|S_{k_2}(\Omega_2+H)^{-1}S_{k_1}(\Omega_1+H)^{-1}S_{k_0}|\alpha\rangle, \tag{6}$$

[2] For pion nucleon scattering see e.g., F. E. Low, Phys. Rev. **97**, 1392 (1955). In the general case it is a consequence of the reduction procedure. H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo cimento **1**, 205 (1955).

[3] M. Gell-Mann and F. Low, Phys. Rev. **84**, 350 (1951).

[4] See e.g., G. C. Wick, Revs. Modern Phys. **27**, 339 (1955).

[5] W. Frank (unpublished).

the diagram in Fig. 1 is admissible. Its amplitude would be

$$A=Z_2f^4\sum_{p_1,p_2}\frac{v^2(p_1)v^2(p_2)(\beta|S_{k_2}S_{p_2}S_{p_1}S_{k_1}S_{p_2}{}^*S_{k_0}S_{p_1}{}^*|\alpha)}{(\omega_{p_1}-E_0)(\omega_{p_1}+\Omega_1)(\omega_{p_1}+\omega_{p_2}+\Omega_1)(\omega_{p_1}+\omega_{p_2}+\Omega_2)(\omega_{p_2}+\Omega_2)\Omega_2}. \tag{7}$$

CONVERGENCE FOR Re $\Omega>-\mu$

The pessimism with regard to convergence of perturbation expansions in field theories is partly based on the consideration that the number of possible diagrams in which n virtual mesons are involved is of order $n!$ for large n. One may verify for the theory under discussion, that the number of admissible diagrams in which n virtual mesons appear in different time orderings relative to the $l+1$ external vertices, behave for large n like $2^{-\frac{1}{2}}e^{-n}(2n)^{n+l+1}$. The presumption that all amplitudes are of the same order of magnitude and roughly constant would imply divergence. Such a presumption is unrealistic as it fails to consider depression of the amplitudes resulting from propagation of many virtual mesons. If it is assumed that the average number of mesons present in the field at any time is of the order $\alpha n^{\frac{1}{2}}(\alpha>0)$, then the suppression factor due to the $2n$ propagators is sufficient to give geometrical convergence. That the average number of mesons in the field goes like $\alpha n^{\frac{1}{2}}$ seems plausible, if one notes that emission and absorption of mesons is basically a one-dimensional random walk process whose statistics show such a result.

We first restrict ourselves to the case where all Ω are real, greater than $-\mu$, and unequal to zero, so that all the propagators in $M(\mathbf{\Omega};\mathbf{K})$ are nonsingular. The amplitude for a given diagram of order $2n$, labeled by an index l, is of the form

$$A_l=Z_2f^{2n}\sum_{p_1,p_2\cdots p_n}v^2(p_1)\cdots v^2(p_n)\times(\beta|S_{r_{2n+l}}S_{r_{2n+l-1}}\cdots S_{r_1}S_{r_0}|\alpha)D_l, \tag{8}$$

where $r_0, r_1, r_2, \cdots, r_{2n+l}$ is some permutation of the meson indices $k_0, k_1, \cdots, k_l, p_1, p_2, \cdots, p_n, -p_1, -p_2, \cdots, -p_n$, and D_l is the product of the energy denominators appropriate to the diagram l.

We proceed to find bounds for this amplitude. From the upper bound $|\mathcal{S}^2(k)|$ on the diagonal matrix elements of $S_k{}^*S_k$, and the Schwarz inequality one can easily derive the following upper bound for the modulus of the matrix element appearing on the right side of Eq. (8).

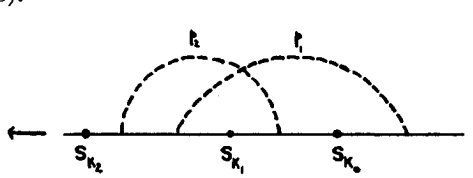


FIG. 1. One of the diagrams in the expansion of the $M(\mathbf{\Omega};\mathbf{K})$ given by Eq. (6). The arrow shows the direction of time.

$$|(\beta|S_{r_{2n+l}}\cdots S_{r_1}S_{r_0}|\alpha)|\leqslant\prod_{j=0}^{2n+l}|\mathcal{S}(r_j)|=|\mathcal{S}(k_0)\cdots\mathcal{S}(k_l)\mathcal{S}^2(p_1)\cdots\mathcal{S}^2(p_n)|, \tag{9}$$

so that we can write

$$|A_l|\leqslant Z_2sf^{2n}\sum_{p_1,p_2,\cdots p_n}v^2(p_1)\cdots v^2(p_n)\times|\mathcal{S}^2(p_1)\cdots\mathcal{S}^2(p_n)||D_l|, \tag{10}$$

where

$$s\equiv|\mathcal{S}(k_0)\mathcal{S}(k_1)\cdots\mathcal{S}(k_l)|. \tag{11}$$

We shall majorize $|D_l|$ by a quantity $\bar{D}_l$ independent of the momenta $p_1, p_2, \cdots, p_n$ so that we can write

$$|A_l|\leqslant Z_2sf^{2n}K^n\bar{D}_l \tag{12}$$

where

$$K\equiv\textstyle\sum_p v^2(p)|\mathcal{S}^2(p)| \tag{13}$$

and as we have assumed, all integrals containing $v(p)$ converge. Z_2 is shown to be convergent in the Appendix.

The quantity D_l is of the form

$$D_l=\prod_{\rho=1}^{2n+l}\frac{1}{\omega_\rho{}^l+\Omega_\rho{}^l}, \tag{14}$$

where ρ is an index enumerating the energy denominators which are $2n+l$ in number, $\omega_\rho{}^l$ is the sum of the energies of all the virtual mesons propagating in the interval ρ, and $\Omega_\rho{}^l$ is $-E_0$ or one of the $\Omega_i(\Omega_i\in\mathbf{\Omega})$ appropriate to the interval ρ. We underestimate each denominator by replacing each $\omega_\rho{}^l$ by $n_\rho{}^l\mu$, where $n_\rho{}^l$ is the number of mesons in propagation over the interval ρ. This is equivalent to setting the momenta of all mesons equal to zero in the propagators. The $\Omega_\rho{}^l$ are also replaced by a lower bound $\tilde{\Omega}^l$. If all the Ω_i are positive, we choose

$$\tilde{\Omega}^l=\min_{\Omega_i\in\mathbf{\Omega}}\{\Omega_i\},$$

and then use the estimate if $\Omega_\rho{}^l\neq-E_0$,

$$1/(\omega_\rho{}^l+\Omega_\rho{}^l)\leqslant1/[\mu(n_\rho{}^l+z)], \tag{15}$$

where $\tilde{\Omega}^l\equiv z\mu$, and $z>0$. If $\Omega_\rho{}^l=-E_0$, since E_0 is negative we underestimate the denominator by setting $\tilde{\Omega}^l=0$. If some of the Ω_i are negative but greater than $-\mu$, then in the denominators where $n_\rho{}^l\geqslant1$ we choose $\tilde{\Omega}^l$ as that value Ω' of the $\Omega_i(\Omega_i\in\mathbf{\Omega})$, for which the quantity $\mu+\Omega_i$ is minimum, while if $n_\rho{}^l=0$ we choose as a lower bound to its $\Omega_\rho{}^l$ the quantity Ω_0 which is the smallest of the quantities $|\Omega_i|$ $(\Omega_i\in\mathbf{\Omega})$. Thus, in this

case we use the estimate

$$1/|\omega_\rho{}^t+\Omega_\rho{}^t| \leqslant 1/[\mu(n_\rho{}^t+z)] \quad \text{for} \quad n_\rho{}^t \geqslant 1$$
$$1/\mu z_0 \quad \text{for} \quad n_\rho{}^t=0, \tag{16}$$

where $z=\Omega'/\mu(-1<z<0)$, and $z_0=\Omega_0/\mu(z_0>0)$. Without loss of generality we assume $z<1$. The ensuing analysis applies with slightly fewer complications if $z\geqslant 1$. However, if $\min\{\Omega_i\}\geqslant\mu$ one can always choose $\tilde{\Omega}^t<\mu$.

The quantity $|D_t|$ is therefore majorized by

$$|D_t|<\bar{D}_t\equiv\mu^{-2n-l}\prod_{\rho=1}^{2n+l}\frac{\zeta_\rho{}^t}{|n_\rho{}^t+z|}, \tag{17}$$

where $\zeta_\rho{}^t=1$ if $\Omega_\rho{}^t\not\equiv -E_0$ when all the Ω_i are positive, and if $n_\rho{}^t\neq 0$ when some of the Ω_i are negative. If all the Ω_i are positive and $\Omega_\rho{}^t\equiv -E_0$, then $\zeta_\rho{}^t=(1+z/n_\rho{}^t)$. If some Ω_i are negative and $n_\rho{}^t=0$, $\zeta_\rho{}^t=|z|/z_0$. Thus,

$$|A_t|<\frac{Z_2{}^s f^{2n}K^n\zeta^{2n+l}}{\mu^{2n+l}}\prod_{\rho=1}^{2n+l}\frac{1}{|n_\rho{}^t+z|}, \tag{18}$$

where $\zeta\equiv\max[1+|z|,\ |z|/z_0]$. Summing now over all diagrams t of order $2n$, the total amplitude for processes entailing $2n$ virtual mesons, $A^{(2n)}$, can be majorized by

$$|A^{(2n)}|\leqslant\sum_{\substack{t \text{ of}\\ \text{order } 2n}}|A_t|<\frac{Z_2{}^s f^{2n}K^n\zeta^{2n+l}}{\mu^{2n+l}}\times\sum_{\substack{t \text{ of}\\ \text{order } 2n}}\prod_{\rho=1}^{2n+l}\frac{1}{|n_\rho{}^t+z|}. \tag{19}$$

Further estimation is facilitated by classification of the structure of diagrams by means of an index which will be termed a "word." The "word" corresponding to a diagram is a sequence of $2n+l+1$ letters consisting of the three letters "e," "a," and "o," where the kth letter in the word is an "e," "a," or "o" depending on whether the kth vertex (ordered timewise) of the diagram is an emission vertex, an absorption vertex or an outside (external) vertex. Thus, the word corresponding to the diagram in Fig. 1 would be "eoeoaao." More than one diagram may correspond to the same word, as the word does not specify the order in which mesons are absorbed. However, for each ρ, the $n_\rho{}^t$ are the same for all diagrams t belonging to the same word. This is clear, since the $n_\rho{}^t$ depend on how many emissions and absorptions have taken place preceding and up to the ρ interval, and this quantity is common to all diagrams having the same word. One can thus rewrite the sum over diagrams in Eq. (19) as a sum over all possible words.

$$\mathfrak{D}_n\equiv\sum_{\substack{t \text{ of}\\ \text{order } 2n}}\prod_{\rho=1}^{2n+l}\frac{1}{|n_\rho{}^t+z|}=\sum_{\substack{\text{words } w \text{ of}\\ \text{length } 2n+l+1}}N_w\prod_{\rho=1}^{2n+l}\frac{1}{|n_\rho{}^w+z|}, \tag{20}$$

where N_w is the number of diagrams belonging to a given word w, and $n_\rho{}^w$ is the common value of $n_\rho{}^t$ for all diagrams t belonging to the word w. We find it convenient to further classify all words having the same sequence of "e's" and "a's" but differing in the position of the $l+1$ "o's" into a class indexed by a "wrd," viz. a word without the "o's". A "wrd" is thus a sequence consisting only of "e's" and "a's." The number of words corresponding to a given wrd is smaller than or equal to $\binom{2n+l+1}{l+1}$. The latter is the number of ways of selecting the $l+1$ "o's" from the $2n+l+1$ vertices, and not all locations of "o's" are permissible since absorption of all mesons before the first "o" is not allowable, and a similar restriction holds for the final "o." We may factor from the denominator product in Eq. (20), the $(l+1)$ denominators following the external vertices (the "o's") in a diagram, and we may bound these $(l+1)$ terms by $1/z_1{}^{l+1}$ where $z_1=\min(|z|,z_0,1)$. We can thus write

$$\mathfrak{D}_n\leqslant\frac{1}{z_1{}^{l+1}}\binom{2n+l+1}{l+1}\sum_{\substack{\text{wrds } u \text{ of}\\ \text{length } 2n}}N_u\prod_{\rho=1}^{2n}\frac{1}{|n_\rho{}^u+z|}, \tag{21}$$

where $n_\rho{}^u$ is the number of mesons present in the ρ denominator after the $(l+1)$ denominators following an external vertex have been eliminated. N_u is the number of diagrams containing no external vertices, which correspond to a wrd u. The quantity

$$d_n\equiv\sum_{\substack{\text{wrds } u \text{ of}\\ \text{length } 2n}}N_u\prod_{\rho=1}^{2n}\frac{1}{|n_\rho{}^u+z|}\equiv\sum_{\substack{\text{wrds } u \text{ of}\\ \text{length } 2n}}\frac{N_u}{\Delta_u} \tag{22}$$

is easily bounded. Each wrd in Eq. (20) is a sequence of $2n$ letters, consisting only of "e's" and "a's". The first letter must clearly be an "e" and the last one an "a." Let a wrd u consist of a block of r_1 "e's" followed by a block of r_2 "a's," followed by r_3 "e's," followed by r_4 "a's," etc. Let there be q blocks altogether. Clearly q is even and $\leqslant 2n$. Moreover,

$$r_1+r_3+\cdots+r_{q-1}=r_2+r_4+\cdots+r_q=n.$$

Let ν_m be the excess of "e's" over "a's" (viz. the number of mesons present) after the mth block. Clearly, $\nu_{2k-1}\geqslant 1$, $\nu_{2k}\geqslant 0$, $\nu_{2k-1}\geqslant r_{2k}$, $\nu_{2k+1}>\nu_{2k}$ (k an integer).

$$\nu_m=\sum_{j=1}^{m}(-)^{j+1}r_j. \tag{23}$$

We use the notation

$$(c)_m\equiv c(c-1)\cdots(c-m+1)=\Gamma(c+1)/[\Gamma(c-m+1)], \tag{24}$$

where m is an integer. Then by counting the number of different ways in which each meson to be absorbed

can be chosen from the number of surviving mesons at any time, one finds for a given wrd u

$$N_u=(\nu_1)_{r_2}(\nu_3)_{r_4}\cdots(\nu_{q-1})_{r_q}. \tag{25}$$

Similarly the product of the denominators in Eq. (20) is easily found to be

$$\Delta_u=|(z+\nu_1)_{r_1}(z+\nu_1-1)_{r_2}(z+\nu_3)_{r_3}(z+\nu_3-1)_{r_4}\cdots \times(z+\nu_{q-1})_{r_{q-1}}(z+\nu_{q-1}-1)_{r_q}|. \tag{26}$$

Thus, for each wrd u

$$\frac{N_u}{\Delta_u}=\frac{(\nu_1)_{r_2}(\nu_3)_{r_4}\cdots(\nu_{q-1})_{r_q}}{(\nu_1+z-1)_{r_2}(\nu_3+z-1)_{r_4}\cdots(\nu_{q-1}+z-1)_{r_q}} \cdot\frac{1}{(\nu_1+z)_{r_1}(\nu_3+z)_{r_3}\cdots(\nu_{q-1}+z)_{r_{q-1}}}. \tag{27}$$

The terms in this expression are easily bounded by quantities independent of u. All the factors in N_u can be "canceled" by corresponding terms in Δ_u.

$$\left|\frac{(\nu_{2k-1})_{r_{2k}}}{(\nu_{2k-1}+z-1)_{r_{2k}}}\right|\leqslant[c(z)]^{r_{2k}} \tag{28}$$

where

$$c(z)\equiv\max[1/|z|,\ 2/(1+z)] \tag{29}$$

and is a fixed finite number. We note an excess of terms in the denominator contributing to further diminution of N_u/Δ_u, which we shall not take maximal advantage of.

$$\left|\frac{1}{(\nu_{2k-1}+z)_{r_{2k-1}}}\right|\leqslant\left(\frac{1}{1+z}\right)^{r_{2k-1}}, \tag{30}$$

$$\therefore\left|\frac{N_u}{\Delta_u}\right|\leqslant\left[\frac{c(z)}{1+z}\right]^n. \tag{31}$$

Therefore

$$d_n<\left[\frac{4c(z)}{1+z}\right]^n, \tag{32}$$

since the number of wrds is smaller than 4^n. From Eqs. (19)–(22) and (32), one obtains the estimate

$$|A^{(2n)}|<\frac{Z_2 s\zeta^l}{\mu^l z_1^{l+1}}\binom{2n+l+1}{l+1}\left[\frac{4f^2Kc(z)\zeta^2}{\mu^2(1+z)}\right]^n. \tag{33}$$

This clearly demonstrates the geometrical convergence of the perturbation expansion for $M(\boldsymbol{\Omega};\mathbf{K})$. A lower bound on the radius of convergence found from Eq. (33) is

$$f_{r.c.}\geqslant\left[\frac{\mu^2(1+z)}{4K\zeta^2c(z)}\right]^{\frac{1}{2}}. \tag{34}$$

This quantity approaches zero as z approaches 0 or -1, suggesting singularities in $M(\boldsymbol{\Omega};\mathbf{K})$ if one of the Ω_i should be 0 or $-\mu$. This is indeed so for $z=0$, as a propagator H^{-1} in $M(\boldsymbol{\Omega};\mathbf{K})$ is equal to E_0^{-1} and therefore has a pole in f for propagation of the intermediate physical state $|\alpha\rangle$. If a rejection operator for the bound "zero meson" states accompanies the H^{-1}, convergence will follow from the above reasoning. The suggested singularity is, however, spurious for $z=-1$. This will be discussed in a succeeding paper dealing with the singularities of the perturbation expansions.

It is to be noted that this convergence result will continue to hold under "variation of parameters," viz., if in the terms of the expansion of a particular $M(\boldsymbol{\Omega};\mathbf{K})$, the positive Ω_i appearing in the propagators are varied independently from term to term over a range of positive values bounded away from zero. Similarly the negative Ω_i can be varied independently over a range of values $\geqslant-\mu$ but excluding some finite internal about 0. The radius of convergence may vary under these conditions but will be bounded away from zero.

Convergence also follows by slight modification of the proof for Ω_i complex-valued in the half-plane $\mathrm{Re}\Omega_i>-\mu$ excluding the point $\Omega_i=0$. Denominators in this case are estimated by

$$1/|\Omega+\sum_{i=1}^{m}\omega_p|\leqslant\sqrt{2}/(\mathrm{Re}\,\Omega+|\mathrm{Im}\,\Omega|+m\mu). \tag{35}$$

Case Re $\Omega_i\leqslant-\mu$

This estimation procedure must be modified when for some of the Ω_i, $\mathrm{Re}\Omega_i\leqslant-\mu$ so that the minimum value of the energy denominators is not bounded away from zero. Let the set of $\Omega_i\in\boldsymbol{\Omega}$ for which $\mathrm{Re}\Omega_i\leqslant-\mu$ be denoted by $\boldsymbol{\Omega}^{R-}$. We first consider the situation where all $\Omega_i\in\boldsymbol{\Omega}^{R-}$ are real. The difficulty of vanishing denominators can be circumvented by modification of the path of momentum integration in the complex plane. We assume that all $\Omega_i\in\boldsymbol{\Omega}^{R-}$ have an infinitesimal imaginary part of the same sign which is say positive. Such is the structure of all matrix elements arising from an S matrix describing scattering processes. Let the most negative of the Ω_i say Ω' obey the inequality $-(I+1)\mu<\Omega'\leqslant-I\mu$ (I an integer). Then, the propagation of more than the number I of mesons at a time results in a nonvanishing energy denominator, which suggests that the convergence characteristics of the expansion would not be much affected by the modifications due to these vanishing denominators. We also assume that no two of the $\Omega_i\in\boldsymbol{\Omega}^{R-}$ are equal. This is not a restriction in view of the quantity $M(\boldsymbol{\Omega};\mathbf{K})$ being a distribution in the variables Ω_i, where the omission of a finite number of values for Ω_i is of no consequence. Physically, as a result of the finite energy width of the incident beam necessary to properly define scattering states, one never speaks of the exact equality in energy of two asvmptotic particles in a scattering process. We moreover assume that the function $v(k)$ and S_k are analytic in $k(k=|\mathbf{k}|)$ in the strip $0\leqslant\mathrm{Im}k\leqslant\mu$, $\mathrm{Re}k>0$, and that all integrals are bounded which contain $v(k)$

and run along the path $\mathrm{Im}k=\mu$ from some positive $\mathrm{Re}k$ to $\mathrm{Re}k=\infty$. If $v(k)$ should be abruptly cut off for $k>k_{\max}$, it is assumed that $v(k)$ is analytic for $0\leqslant \mathrm{Im}k\leqslant\mu$, $0<\mathrm{Re}k\leqslant k_{\max}$. Then for a selected number of momentum variables, for that matter for all of them, one deforms the path of integration from the positive real axis to (say) the path C shown in Fig. 2. Along such a path $\mathrm{Re}\omega_p\geqslant\mu$, *a fortiori* $|\omega_p|\geqslant\mu$, and all our previous estimates based on such a relation still hold. If in particular, none of the $\Omega_i\in\mathbf{\Omega}^{R-}$ are an integral multiple of $-\mu$, clearly all formerly vanishing energy denominators are bounded away from zero and the modulus of each may be estimated by this lowest bound which we shall call $b\mu$. We now estimate the matrix element (with its complex valued meson indices) as in Eq. (9), and obtain the expression Eq. (12) with the two modifications, that the path of integrations will be different and the estimates for the vanishing energy denominators in D_t shall have been replaced by the quantity $b\mu$. Estimation of D_t takes place as before except for the just mentioned replacement for the vanishing denominators. The integration in Eq. (13) can be performed as before to yield by analyticity the same constant K along the deformed path of integration. One finally arrives at the expression Eq. (17) with the only change that $n_p{}^t+z$ is to be replaced by b if its corresponding exact denominator vanished along the undeformed path. As previously suggested, although the radius of convergence may change, the convergence of the perturbation expansion for sufficiently small f will not. Replacing the vanishing energy denominators by $b\mu$ modifies the previous estimate of $\mathfrak{D}_n$ in essentially two ways. Firstly, the factor $z_1{}^{-l-1}$, removed in Eq. (21) as an overestimate of the propagators following the external vertices, should trivially be replaced by $z_2{}^{-l-1}$ where

$$\frac{1}{z_2}=\max\left[\frac{1}{b},\frac{1}{|z|},\frac{1}{z_0},\frac{1}{z'},\frac{1}{z'_0},1\right], \tag{36}$$

where we interpret z and z_0 as before in terms of $\Omega_i\in\mathbf{\Omega}-\mathbf{\Omega}^{R-}$, and z', z_0', respectively, represent the absolute value of the minimum fractional part of

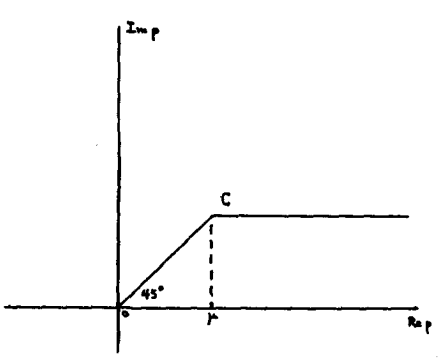


FIG. 2. Deformed contour C for virtual momentum integration when some Ω_i are real and negative.

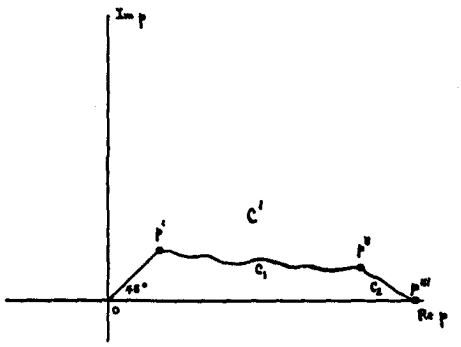


FIG. 3. Deformed contour C' for virtual momentum integration when some $\Omega_i\in\mathbf{\Omega}^{c-}$.

$1-\Omega_i/\mu$ and $\Omega_i/\mu(\Omega_i\in\mathbf{\Omega}^{R-})$. The z and z_0 of (15) and (16) are also to be redefined to apply to this more general case. Secondly, the quantity d_n in Eq. (22) should also be multiplied by a factor

$$\xi\equiv\prod_{\substack{p' \text{ of vanishing}\\ \text{denominators}}}\left(\frac{|n_{p'}{}^t+z|}{b}\right). \tag{37}$$

In view of $n_{p'}{}^t<I+1$ for vanishing denominators, one easily finds the overestimate

$$\xi<[(I+z+1)/b]^{n-1}. \tag{38}$$

These changes though they may affect the radius of convergence in f do not affect convergence for sufficiently small f. If all $\Omega_i\in\mathbf{\Omega}^{R-}$ have nonvanishing imaginary parts, contour deformation is not necessary, and for real virtual momenta, the denominators are estimated by

$$\frac{1}{|\Omega_i+\sum_{i=1}^{m}\omega_{p_i}|}\leqslant\begin{cases}\dfrac{1}{|\Omega''|} & \text{if } m<I+1\\[2ex] \dfrac{1}{\Omega'+m\mu} & \text{if } m\geqslant I+1,\end{cases} \tag{39}$$

where

$$\begin{aligned}\Omega'&\equiv\min_{\Omega_i\in\mathbf{\Omega}}\{\mathrm{Re}\Omega_i\}\\ \Omega''&\equiv\min_{\Omega_i\in\mathbf{\Omega}^{R-}}\{\mathrm{Im}\Omega_i\}.\end{aligned} \tag{40}$$

Convergence then follows as it did for $\Omega_i\in\mathbf{\Omega}^{R-}$ real, with $|\Omega''|$ playing the role that $b\mu$ did there.

If some of the $\Omega_i\in\mathbf{\Omega}^{R-}$ are real (but for the infinitesimal imaginary increment) and some are complex, so long as none of the real $\Omega_i\in\mathbf{\Omega}^{R-}$ are non-negative integer multiples of $-\mu$, a bounding procedure can be established by appropriately deforming the contour of the momentum integrations. If the imaginary parts of all the complex $\Omega_i\in\mathbf{\Omega}^{R-}$ are positive, the contour of Fig. 2 will do. If some of the complex $\Omega_i\in\mathbf{\Omega}^{R-}$ have negative imaginary parts, another contour C' can be chosen. Let $\mathbf{\Omega}^{c-}$ be the set of $\Omega_i\in\mathbf{\Omega}^{R-}$ whose imaginary

part is negative. The contour C', illustrated in Fig. 3, starts at $p=0$ and rises at an angle of 45° till $p=p'$ continues along the curve C_1 to $p=p''$ then follows the curve C_2 to $p=p'''$ (real) and continues along the real axis. The point p' is chosen according to the specification $\mathrm{Re}p'=\mathrm{Im}p'$, $0<\mathrm{Im}\omega_{p'}<b_1\mu$ where

$$b_1\mu \equiv \min_{\Omega_i\in\mathbf{\Omega}^{c-}} \{\Omega_i/I_1\} \tag{41}$$

where

$$-(I_1+1)\mu < \min_{\Omega_i\in\mathbf{\Omega}^{c-}} \mathrm{Re}\Omega_i \leqslant -I_1\mu. \tag{42}$$

The curve C_1 is chosen so that everywhere along it $\mathrm{Im}\omega_p<b_1\mu$ and $|\omega_p|\geqslant\mu$. This qualification guarantees that there is no cancellation of the negative imaginary part of an $\Omega_i\in\mathbf{\Omega}^{c-}$ against the positive imaginary parts of an ω_p along the deformed contour, so long as there are fewer than I_1+1 mesons in propagation. The contour is continued along C_1 to a point p'' such that $\mathrm{Re}\omega_{p''}>(I_1+1)\mu$ and then continues along a more or less arbitrary curve for which $|\omega_p|\geqslant\mu$ returning (say) to the real axis at $p'''>p''$. Along this deformed contour, all denominators having at least I_1+1 mesons in propagation are bounded in modulus by an inequality like Eq. (15). Propagators with I_1 or fewer mesons in propagation are in consequence of the contour chosen, bounded in modulus. Convergence for this case then follows as before.

Convergence can also be demonstrated when some of the $\mathrm{Re}\Omega_i(\Omega_i\in\mathbf{\Omega}^{R-})$ are positive integer multiples of $-\mu$. The demonstration entails a greater complexity of analysis and will be deferred to a subsequent paper discussing the singularities of these expansions.

EXTENSIONS—BILINEAR COUPLINGS

This analysis constructed for a theory with linear coupling of recoilless nucleons to mesons is easily extended to include theories with couplings bilinear in the meson field and nucleons that can recoil. We first discuss the extension to bilinear couplings. Such bilinear couplings give rise to a term in the Hamiltonian of the form

$$H_2=\sum_{k,k'}\Big\{u(k,k')S_{kk'}a_k a_{k'}+u^*(k,k')S_{kk'}{}^*a_k{}^*a_{k'}{}^* +w(k,k')T_{kk'}a_k{}^*a_{k'}\Big\}. \tag{43}$$

The $S_{kk'}$, $T_{kk'}$ are again finite dimensional matrices and it is assumed that their bounds and the $u(k,k')$, $w(k,k')$ are such as to allow the same kind of bounding procedure on the nuclear matrix element and the momentum integrations as before. Energy denominators are estimated as before, by their minimum value if $\mathrm{Re}\Omega_i>-\mu$ for all $\Omega_i\in\mathbf{\Omega}$, and the procedure outlined when this fails to hold. The principle of the estimation procedure carried out on D_t, the product of the energy denominators, continues to hold with the following modifications. One now indexes each diagram by a "biword" constructed of six possible letters: "e_1" for the emission of one meson; "e_2" for the emission of two; "a_1" for absorption of one meson; "a_2" for the absorption of two; "n" for the bilinear vertex which both emits and absorbs one meson making no change in their number; and "o" for the outside vertex. Again, the biwords are classified into classes of "biwrds"; the $l+1$ denominators following an "o" vertex are estimated by z_1^{-l-1} and a factor $\binom{2n+l+1}{l+1}$ appears as an upper bound to the number of biwords in a biwrd. In the estimation of the quantity N_b/Δ_b for the biwrds of $2n$th-order diagrams, we note that the number of factors in the denominator Δ_b is still $2n-1$, the same as for wrds. They are modified in that the denominator for an interval will change by $+2$, 0, -2 if that interval respectively follows an "e_2," "n," or "a_2" vertex. The number of factors in N_b, which counts the number of diagrams corresponding to a biwrd is, however, changed in a significant way. For an interval ρ following an "n" vertex, a factor $n_\rho{}^b$ appears counting the number of ways a meson can be absorbed, and for an interval following an "a_2" vertex a factor $\frac{1}{2}n_\rho{}^b(n_\rho{}^b-1)$ appears enumerating the number of ways of selecting two mesons to be absorbed. A bound for N_b/Δ_b similar to that constructed for N_u/Δ_u for wrds can be found. As in the estimation of N_u/Δ_u in Eq. (27), to every factor in the numerator a factor can be found in the denominator for which the ratio can be bounded independently of the value of the numerator. As before, for each $n_\rho{}^b$ in the numerator counting the number of ways of absorbing a meson after an "a_1" or "n" vertex or the number of ways of absorbing the first meson after an "a_2" vertex, we correspond the denominator for the interval following that absorption vertex. This is $n_\rho{}^b+z-1$ if the denominator is bounded away from zero, and b otherwise. To the factors $\frac{1}{2}(n_\rho{}^b-1)$ in the numerator which enumerate number of ways of choosing the second meson after an "a_2" vertex (in this case $n_\rho{}^b\geqslant 2$) we correspond another denominator containing $n_\rho{}^b$, that which follows the last interval σ which precedes ρ when at least that number of mesons first appeared. These are the propagators

$$\prod_{j=1}^{\frac{1}{2}q}[(\nu_{2j-1}+z)_{r_{2j-1}}]^{-1} \tag{44}$$

in Eq. (26) which were not "canceled" by any numerator in Eq. (27), and were individually bounded by $(1+z)^{-1}$ in Eq. (30). In place of this for biwrds, if the denominator for creation of the $n_\rho{}^{b*}$meson is nonvanishing, we find a ratio

$$\frac{\frac{1}{2}(n_\rho{}^b-1)}{(n_\sigma{}^b+z+\gamma_\sigma{}^b)} \tag{45}$$

where $\gamma_\sigma{}^b=0$, 1 depending on whether vertex preceding the interval were "e_1" or "e_2," respectively. The quantity in Eq. (45) can be bounded by a quantity $c_1(z)$ which is independent of b and ρ. An upper bound for N_b/Δ_b under these conditions is easily seen to be

$$|N_b/\Delta_b| < [c(z)c_1(z)]^n \tag{46}$$

and convergence easily follows. If the denominator estimated by $n_\sigma{}^b+z+\gamma_\sigma{}^b$ was a vanishing one and was to be replaced by b, the same procedures as before show convergence.

This method of pairing off of factors in the numerator against those in the denominator in N_b/Δ_b will not follow through for trilinear and higher multilinear interactions. This, of course, does not prove the divergence of such theories.

RECOIL

The inclusion of recoil necessitates two changes in the structure of perturbation expansions which has been described. A matrix element for the recoiling transition of the nucleon appears at each vertex and it depends on the momenta of all mesons propagating in the two adjoining intervals. The propagators for an interval are also modified by the energy of the recoiling nucleon. If an assumption is made of "autonomous cutoff" like behavior of the recoil transition matrix elements, guaranteeing convergence of all integrals after disregard of the momentum dependence of the propagators, then the methods and results established will continue to apply with slight modification. The situation of auxiliary cutoffs and theories with convergent integrals will be dealt with in a subsequent paper.

DISCUSSION

The convergence of static meson theory is a conclusion of interest, as static meson theory has provided an adequate model for the elucidation of certain physical phenomena. This includes the two and three nucleon systems which can be discussed fruitfully within the framework of static meson theory.[6] Another approach to the convergence of cutoff theories has been investigated by Edwards.[7] This convergence justified the prescription of Castillejo, Dalitz, and Dyson[8] that the solution of the Low integral equation becomes unique and physical when one demands analyticity in the coupling constant near zero.

It should be noted that it was not necessary to expand, as we have done, about $f=0$. One might write a perturbation expansion using $H_0+\epsilon H_1$, ($\epsilon<1$) as the unperturbed Hamiltonian and $(1-\epsilon)H_1$ as the perturbation term. The amplitude for a diagram contains a matrix element of the form Eq. (5) (with H in the denominator representing $H_0+\epsilon H_1$) between eigenstates of $H_0+\epsilon H_1$. By means of the Schwarz inequality, such a matrix element can be estimated by a bound of the form Eq. (9), and convergence of the resulting power series in $(1-\epsilon)f$ (with coefficients depending on ϵf) can be concluded. Thus no real significance has been given to the point $f=0$. It should further be noted that the estimated radius of convergence is proportional to μ, clearly a result of the estimation procedure. The presumption that quantum mesodynamics may have a finite radius of convergence proportional to μ for small μ suggests divergence for a zero mass boson theory. Such a result has recently been indicated for quantum electrodynamics with a further indication that the nonanalyticity for $e=0$ may be entirely contained in the functional dependence of the renormalization constants on e.[9] Application of the methods of this paper to relativistic field theories is currently under investigation. The theories investigated in this paper differ from relativistic field theories in three respects apart from the problem of the zero mass boson. They are, firstly, the restriction to theories with autonomous cutoff functions in this paper, secondly the presence of antifermions in the relativistic theory, and thirdly, the necessity for renormalization procedures in field theories, leading to a φ^4 term in meson theory. The restriction to autonomous cutoff functions is not essential, so long as the momentum integrals converge. This will be discussed in a subsequent paper. The inclusion of anti-fermions introduces greater complexity in the structure of the diagrams. This situation is under investigation. The presence of a φ^4 term in meson theory might not lead to convergent results. This is basically the same problem as the convergence in the presence of antifermions. In a cutoff or regularized theory the φ^4 term is obviated, and the linear interaction can be studied without such a renormalization.

ACKNOWLEDGMENTS

The author would like to thank Dr. John Toll for the opportunity of discussing this material in seminar at the University of Maryland, and Dr. Z. Fried and Dr. H. Reiss for helpful discussions. The author would like to acknowledge the help of Mrs. D. Cox in preparing the manuscript.

APPENDIX

The Perturbation Expansion and Convergence of E_0 and Z_2

The proof for the convergence of $M(\Omega;\mathbf{K})$ has relied on the assumed convergence of E_0 and Z_2, and the consequent properties of E_0, that E_0 is negative and goes to zero continuously as $f\to 0$. The convergence of E_0 and Z_2 will now be shown by essentially the same reasoning applied to $M(\Omega;\mathbf{K})$, but with no prior assumptions. The complete proof of the convergence

[6] W. Frank, Ann. Phys. (to be published).
[7] S. F. Edwards, Phil. Mag. 46, 569 (1955).
[8] L. Castillejo, R. H. Dalitz and F. J. Dyson, Phys. Rev. 101, 453 (1956).
[9] I. Bialynicki-Birula, Phys. Rev. 122, 1942 (1961).

of $M(\mathbf{\Omega};\mathbf{K})$ thus consists of a chain of two similar sets of steps applied first to demonstrate the convergence of E_0 and Z_2 then reapplied to $M(\mathbf{\Omega};\mathbf{K})$. Unnecessary repetition has been avoided by first assuming convergence of E_0 and Z_2 in the demonstration for $M(\mathbf{\Omega};\mathbf{K})$.

The rules for the expansion in diagrams presented in this paper correspond to a mixture of the explicit, standard form of perturbation theory and the implicit or Brillouin-Wigner form of perturbation theory. The dependence of the matrix element $M(\mathbf{\Omega};\mathbf{K})$ on the coupling constant f is contained in three places: in the H in the propagators, and in the initial and final state vectors. The propagators are expanded in an explicit perturbation expansion, i.e., by writing

$$\frac{1}{\Omega+H}=\frac{1}{\Omega+H_0}-\frac{1}{\Omega+H_0}H_1\frac{1}{\Omega+H_0}+\cdots. \quad \text{(A1)}$$

The initial and final state vectors are expanded implicitly from the formula[10]

$$|\alpha\rangle=Z_2^{\frac{1}{2}}[1+(E_0-H_0-\Lambda H_1)^{-1}\Lambda H_1]|\alpha) \quad \text{(A2)}$$

where Λ is a rejection operator for the bare no-meson state. In the perturbation expansion of $|\alpha\rangle$, $Z_2^{\frac{1}{2}}$ and E_0 are treated as parameters independent of f and not expanded. One thus expands $|\alpha\rangle$ as

$$|\alpha\rangle=Z_2^{\frac{1}{2}}[1+(1/E_0-H_0)\Lambda H_1 +(1/E_0-H_0)\Lambda H_1(1/E_0-H_0)\Lambda H_1+\cdots]|\alpha). \quad \text{(A3)}$$

E_0 as a function of f is determined[11] as that solution of

$$E=\sum(E,f)\equiv(\alpha|H_1(E-H_0-\Lambda H_1)^{-1}H_1|\alpha) \quad \text{(A4)}$$

which is $O(f^2)$ as $f\to 0$. Z_2 is determined[12] by

$$Z_2^{-1}=1-\partial\sum/\partial E|_{E=E_0}. \quad \text{(A5)}$$

[10] G. C. Wick, Revs. Modern Phys. **27**, 4 (1955), Eqs. (2.5′) and (2.6′).
[11] Reference 10, Eq. (2.7).
[12] Reference 10, Eq. (2.11).

The perturbation expansion of $\sum(E,f)$ as a function of f is identical with the perturbation expansion described for $M(\mathbf{\Omega};\mathbf{K})$ in the case $l=-1$, i.e., no external vertices, where the self energy E_0 is replaced by the parameter E and the Z_2 factor is omitted. The proof presented in this paper shows that this perturbation expansion is convergent for sufficiently small f if $E<\mu$ and is in fact uniformly convergent in E for f interior to its circle of convergence. For this range of values $\sum(E,f)$ is analytic function of both f and E. The function

$$T(E,f)\equiv E-\sum(E,f) \quad \text{(A6)}$$

for this range of values is therefore analytic in f and E and clearly has a root at $E=0$, $f=0$. Since $\partial T/\partial E\neq 0$ at this point, the implicit function theorem[13] implies that the equation $T(E,f)=0$ defines E as an analytic function of f in some neighborhood of $f=0$. The consequence is that E_0 is an analytic function of f for sufficiently small f, demonstrating the convergence of E_0. From this and the analyticity of $\partial\sum/\partial E$ as a function of both f and E, follows the analyticity of Z_2^{-1} in the neighborhood of $f=0$. Since $Z_2^{-1}\neq 0$ at $f=0$, the analyticity in f and hence the convergence of Z_2 follows.

The matrix element

$$\langle\beta|S_k|\alpha\rangle \quad \text{(A7)}$$

represents the renormalized vertex, and its convergence implies the analyticity of the renormalized coupling constant f_r as a function of the unrenormalized coupling constant f. From $df_r/df\neq 0$ at $f=0$, follows the analyticity of f as a function of f_r, and consequently the convergence of perturbation expansions in the renormalized coupling constant.

Analyticity of $M(\mathbf{\Omega};\mathbf{K})$ as a function of f implies analyticity in the case where the Ω_i are analytic functions of f in the neighborhood of $f=0$.

[13] L. Bieberbach, *Lehrbuch der Funktionentheorie* (Chelsea Publishing Company, New York, 1945), pp. 190–191.

Some Examples of Canonical Transformations in Quantum Mechanics*

FRANCIS R. HALPERN†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received March 29, 1961)

The set of all pairs of polynomials $X(x,p)$ and $P(x,p)$ in the position x and momentum p are found such that X and P have the same commutator as x and p. Several other types of representations of the commutation relations are studied. The reducibility, time, and space inversion properties of these are discussed. The unitary transformations connecting different representations are exhibited. The use of different representations to solve scattering and bound state problems is indicated. The application of one of the transformations to a partial diagonalization of a perturbed harmonic oscillator Hamiltonian is made.

I. INTRODUCTION

A possible approach to the problem of diagonalizing the Hamiltonian is to construct a pair of operators x and p that have the proper commutation relations and such that the Hamiltonian is diagonal in this representation. If there is available a large catalog of representations, it may be possible to select one that diagonalizes or nearly diagonalizes the energy. The same transformations may be applied to the calculation of the S matrix and the method is illustrated below.

Several methods of constructing representations are considered. It should be noted that all representations are connected by unitary transformations but that it may frequently be simpler to construct different representations than the unitary transformations connecting them. In general, it is useful to have sequences of transformations of increasing complexity. This will permit the presentation of solutions as the limits of simple approximations.

This observation suggests that representations that are polynomials in existing representations might be useful and this case is treated first. There exists an $n+3$ parameter family of such representations if n is the degree of the polynomials. Unfortunately because of their time and space inversion properties, the usefulness of these representations for physical problems is limited. This may be illustrated by observing that $x'=x+p$ and $p'=p$ are satisfactory representations of the commutation rules if x and p are. However x' does not behave in a reasonable manner under time inversion. If an attempt is made to remedy this by using even powers of p in x (for example, $x'=x+p^2$, $p'=p$) the space inversion properties of x are spoiled. This situation is completely general for polynomial representations except for those of the type $x'=\alpha x$, $p'=p/\alpha$. In the Appendix it is shown that x and p form a basis for all operators in the Hilbert space and thus all representations, aside from questions of convergence, are series in x and p.

Alternatively, one may attempt to find matrix representations. The conventional representations (harmonic oscillator) have only the two lines adjacent to the principal diagonal occupied. Suppose that more, but still a finite number, are occupied. The algebraic difficulties have prevented a definitive answer to this question. If there are only two nonvanishing elements in each row of the matrix, the representation is a standard one up to phase changes in the basis states or a direct sum of standard representations.

Finally, any complete set of functions $u_n(x)$ orthonormal in $(-\infty, \infty)$ gives rise to a representation,

$$x_{nm}=\int u_n(x)xu_m(x)dx$$

$$p_{nm}=\frac{1}{i}\int u_n(x)\frac{d}{dx}u_m(x)dx.$$

Changes in the variable of integration give rise to a wide variety of possible x and p.

The expansion of the S matrix after a canonical transformation has been made can be given in a manner similar to the conventional scattering theory. If the scattering states $\psi_a^{(\pm)}$ are given by

$$\psi_a^{(\pm)}=\frac{i\epsilon}{E_a-H+i\epsilon}\phi_a$$

A unitary transformation W may be introduced as follows

$$\psi_a^{(\pm)}=WW^*\frac{i\epsilon}{E_a-H\pm i\epsilon}WW^*\phi_a$$

$$=W\frac{i\epsilon}{E_a-\bar{H}+i\epsilon}W^*\phi_a$$

where $\bar{H}=W^*\bar{H}W$. The S matrix given conventionally by

$$S_{ab}=\langle\psi_a^{(-)}|\psi_b^{(+)}\rangle,$$

becomes after the transformation,

$$S_{ab}=\left\langle\phi_a\left|W\frac{-i\epsilon}{E_a-\bar{H}+i\epsilon}\frac{i\epsilon}{E_b-\bar{H}+i\epsilon}W^*\right|\phi_b\right\rangle.$$

* Supported in part by the U. S. Air Force Office of Scientific Research.
† Present address: Department of Physics University of California, San Diego, La Jolla, California.

Finally, if $\bar{H}=\bar{K}+\bar{V}$, and $\bar{\phi}_c$ are a complete set of eigenstates of $\bar{K}$, S is given by

$$S_{ab}=\langle\phi_a|W|\bar{\phi}_c j\rangle \times\left\langle\bar{\phi}_c\left|\frac{-i\epsilon}{E_a-\bar{H}+i\epsilon}\frac{i\epsilon}{E_b-\bar{H}+i\epsilon}\right|\bar{\phi}_d\right\rangle\langle\bar{\phi}_d|W^*|\phi_b\rangle.$$

The outside terms are transformations from the eigenstates of K to $\bar{K}$. In general they are not delta functions because K is not the transform of $\bar{K}$. This middle term is almost the expression for the S matrix in the barred representation, except that the energy denominators are appropriate to the unbarred representation. The utility of this result is that with a proper choice of W, the ratio $\bar{V}/\bar{K}$ may be less than V/K.

The same result can also be obtained starting from the time-dependent theory. Let $\psi(t)$ be the time-dependent Schrödinger state vector. Then $\phi(t)$, defined by

$$\phi(t)=e^{-iKt}\psi(t),$$

is the interaction representation state vector. The time translation operator $U(t,t_0)$ is defined by

$$\phi(t)=U(t,t_0)\phi(t_0)$$

and satisfies the differential equation

$$i\partial U/\partial t=V(t)U(t,t_0).$$

Similar quantities are defined in a new representation related to the original one by

$$\psi=W\chi$$

The vector χ satisfies the differential equation

$$i\dot{\chi}=\bar{H}\chi.$$

Again it is possible to introduce an interaction representation and a time translation operator $\bar{U}(t,t_0)$. The two time translation operators are related by

$$U(t,t_0)=e^{iKt}We^{-iKt}\bar{U}(t,t_0)e^{iKt_0}W^*e^{-iKt_0}.$$

The Matrix element of $U(t,t_0)$ between the states ϕ_a and ϕ_b is given by

$$\langle\phi_a|U(t,t_0)|\phi_b\rangle=\langle\phi_a|W|\bar{\phi}_c\rangle e^{i(E_a-E_c)t} \times\langle\bar{\phi}_c\bar{U}(t,t_0)|\bar{\phi}_d\rangle e^{i(E_d-E_b)t}\langle\bar{\phi}_d|W^*|\phi_b\rangle.$$

S is calculated by allowing $t\to\infty$ and $t_0\to-\infty$. This gives the same result as above when proper account is taken of the oscillatory factors.

II. POLYNOMIAL REPRESENTATIONS

In this section we assume that one representation of the commutation rule $[x,p]=i$ is known and we look for others that are polynomials in the original pair. Instead of x and p it is more convenient to use the non-Hermitean pair

$$a=(mx+ip/m)/\sqrt{2},\quad a^*=(mx-ip/m)/\sqrt{2}$$

where m is a parameter having the dimensions of mass. The operators a and a^* have the commutation rule $[a,a^*]=1$. New operators A and A^* are defined by

$$A=\sum_{t=0}^{T}\sum_{p=0}^{t}A_{tp}(a^*)^p a^{t-p},$$

where the A_{tp} are coefficients to be determined so that $[A,A^*]=1$.

After a lengthy but straightforward calculation that involves numerous changes in the orders of iterated sums and redefinitions of the indices of summation, the commutator of A and A^* is found to be

$$[A,A^*]=\sum_{W=0}^{2T-2}\sum_{V=0}^{W}f(W,V)(a^*)^V a^{W-V},$$

where $f(w,v)$ is given by

$$f(W,V)=\sum_t\sum_r\sum_q\left\{\frac{(q+r)!(V-t+q+r)!}{q!r!(V-t+q)!}A_{t+r\ t-q}\times A^*_{W-t+r\ W-V-q}\frac{(t-q+r)!(W-V-q+r)!}{(t-q)!r!(W-V-q)!}\times A_{t+r\ t-q+r}A^*_{W-t+r\ W-V-q+r}\right\}.$$

In order to satisfy $[A,A^*]=1$, it is necessary that $f(W,V)=0$ when $W\neq 0$, and that $f(0,0)=1$.

The solution of this lengthy system of equations begins by considering the case $W=2T-2$. The case $T=2$ is illustrated as the general case proceeds in the same manner and is considerably more complicated. The f's for $T=2$ are

$$f(0,0)=A_{10}A_{10}^*-A_{11}A_{11}^*+2A_{20}A_{20}^*-2A_{22}A_{22}^*$$
$$f(1,0)=A_{10}A_{21}^*+2A_{20}A_{10}^*-2A_{11}A_{22}^*-A_{21}A_{11}^*$$
$$f(2,0)=2A_{20}A_{21}^*-2A_{21}A_{22}^*$$
$$f(2,1)=4A_{20}A_{20}^*-4A_{22}A_{22}^*.$$

Since $[A,A^*]$ is an Hermitean operator, $f^*(W,V)=f(W,W-V)$ and no new information is gained from the remaining f's. An examination of $f(2.1)$ shows that A_{20} and A_{22} have the forms,

$$A_{20}=Re^{i\alpha}\quad A_{22}=Re^{i\beta}.$$

Similarly from $f(2,0)$ it follows that A_{21} is given by

$$A_{21}=2R'e^{i(\alpha+\beta)/2}.$$

The equation $f(1,0)$ presents the first serious problem. It should be noted at this point that because of $f(2,1)$ we have that $f(0,0)=A_{10}A_{10}^*-A_{11}A_{11}^*$, and consequently the magnitudes of A_{10} and A_{11} must be different.

The equation $f(1,0)$ is to be considered as two real

linear equations in two unknowns $\mathrm{Re}\{A_{10}\}$ and $\mathrm{Im}\{A_{10}\}$, to be solved for in terms of the other A's. To solve a complex linear equation the following lemma is of use. The equation

$$ax+bx^*=c$$

has the solution

$$x=(ca^*-c^*b)/(|a|^2-|b|^2),$$

provided

$$|a|^2\neq|b|^2.$$

If this condition is not satisfied, then there may still exist a one-parameter family of solutions provided that

$$a^*c=bc^*.$$

If this condition is satisfied then x is given by

$$x=c/2a+ir(b/a)^{\frac{1}{2}}.$$

The application of this lemma to $f(1,0)$ gives the result

$$A_{10}=e^{i(\alpha+\beta)}A_{11}^*,$$

provided that

$$|A_{21}|^2\neq 4|A_{20}|^2.$$

This result is not consistent with

$$f(0,0)=1,$$

hence

$$R'=R.$$

From part two of the lemma, there still may exist a solution of the equation $f(1,0)=0$ and it is easily shown that the necessary condition is satisfied. Thus A_{10} is given by

$$A_{10}=\tfrac{1}{2}A_{11}e^{i(\alpha-\beta)/2}+\tfrac{1}{2}A_{11}^*e^{i(\alpha+\beta)}+ire^{i(3\alpha+\beta)/4}.$$

Finally, using this result to satisfy $f(0,0)=1$ gives the following set of values for the A's, where further use had been made of the lemma.

$$A_{21}=(\rho+ir)e^{i(4\alpha+\beta)/4}$$
$$A_{11}=[\rho\pm i(r^2-1)^{\frac{1}{2}}]e^{i(3\beta+\alpha)/4}$$
$$A_{20}=Re^{i\alpha}$$
$$A_{21}=2Re^{i(\alpha+\beta)/2}$$
$$A_{22}=Re^{i\beta}.$$

The solution for the general case is

$$A_{10}=(r_1+ir_2)e^{i(\chi+\phi)}$$
$$A_{11}=[r_1\pm i(r_2^2-1)^{\frac{1}{2}}]e^{i(\chi-\phi)}$$
$$A_{tp}=R_t\binom{t}{p}e^{i\chi+i(t-2p)\phi}.$$

The detailed argument has been carried through for $T=3$ and it is simple to show that these A's really solve the equations for any T.

Under space inversion both x and p change sign. Since a and a^* are linear combinations of x and p they also change sign. If then A and A^* are to have the same behavior, their expansions can contain only odd t. Under time inversion, x is invariant and p changes sign, but so does i. Hence, a and a^* are time inversion invariants. For A and A^* to preserve this property it is necessary that the coefficients A_{tp} all be real. A brief inspection shows that the only satisfactory solution having these properties is

$$A=\cosh\alpha a+\sinh\alpha a^*$$
$$A^*=\sinh\alpha a+\cosh\alpha a^*,$$

or, equivalently,

$$X'=(\cosh\alpha+\sinh\alpha)x$$
$$P'=(\cosh\alpha-\sinh\alpha)p.$$

Thus in any system in which time and space reversal are allowed operations, these representations will not suffice to diagonalize the Hamiltonian. Since the limit of any sequence of transformation of this character is in the set, it is not possible to approach arbitrarily closely to the solution of such a problem with a representation of this character. The time development of some systems may be represented by an operator of the polynomial type with time dependent parameters. That is a solution of the Heisenberg equation of motion may be found in this form, since there is no reality requirement on the Heisenberg operator.

Aside from the phase factors, the polynomial representation may be considered as the product of two simpler transformations

$$a'=a+\sum R_t\sum\binom{t}{p}(a^*)^p a^{t-p}$$
$$a''=\cosh\alpha a'+\sinh\alpha a'^*.$$

The Unitary operators that generate these two transformations are

$$\exp\Big[\sum R_t\sum\binom{t}{p}(a^*)^p a^{t-p}\Big]$$

and

$$\exp[\alpha(a'^2-a'^{*2})].$$

III. CONNECTION WITH CLASSICAL CANONICAL TRANSFORMATIONS

A classical canonical transformation is specified by a generating function which may be several different types. For one-dimensional mechanics it is a function of two variables. If it is separable, that is the generating function is the product of two functions each of one variable, it is possible to give a reasonably explicit solution for the new variables in terms of the old. One may then verify that if the appropriate Hermitean combinations are formed, at least formally a quantum mechanical canonical transformation is also generated. Since one is dealing with functions of operators it is necessary to give some definition to them and at this point many of the classical transformations fail to go

over to genuine quantum transformations. We restrict ourselves to simple generating functions of the type

$$F(p',x)=p'f(x).$$

Then we have

$$x'=f(x)$$
$$p'=[f'(x)]^{-\frac{1}{2}}p[f'(x)]^{-\frac{1}{2}}.$$

Because of the parity and time reversal requirements we find that $f(x)$ ashould be an odd increasing function of x and f'(x) should be a positive even function of x. The simplest choice is that $f'(x)$ is a constant. This gives the linear transformation $x'=\alpha x$. Next in order of increasing complexity is $f'(x)=(a+bx^2)^2$. The explicit construction for $(a+bx^2)^{-1}$ is discussed at length in the appendix. Because of the difficulties encountered in this construction, the present method does not appear suitable for producing quantum mechanical canonical transformations.

IV. MATRIX REPRESENTATIONS

It is possible to choose any real Hermitean matrix with zeros in the places dictated by parity as a possible x. If the equation $px-xp=-i$ can be solved, a canonical pair has been exhibited. We begin with x that have only the position adjacent to the main diagonal different from 0:

$$x_{ij}=\delta_{i+1j}a_j+\delta_{ij+1}a_{i+1}.$$

The matrix p can be solved row by row. The first three are

$$-ip_{ij}=\delta_{ij+1}i(2a_j)^{-1}+\delta_{ij+3}(2a_ja_{j+1}a_{j+2})^{-1}\Sigma_{j+1}$$
$$+\delta_{ij+5}(2a_ja_{j+1}a_{j+2}a_{j+3}a_{j+4})^{-1}$$
$$\times\sum_{s=0}^{j}(a_{s+2}{}^2\Sigma_i-a_{s-1}{}^2\Sigma_{s+1})+\cdots$$

where

$$+\Sigma_{s+1}=(s+1)a_{s+1}{}^2-2\sum_{t=0}^{s}a_t{}^2.$$

If we require that p stop after a certain number of rows, it is only necessary that one row vanish to make all succeeding rows vanish. (Row is used here to denote the set of elements on a diagonal not a horizontal row). For example, for p to have only one row, $\Sigma_i=0$ and this generates the harmonic oscillator representation. If p is to have two rows,

$$s_{s+2}{}^2\Sigma_s=a_{s-1}{}^2\Sigma_{s+1}.$$

$a_0{}^2$, $a_1{}^2$, and $a_2{}^2$ are optional and all others are determined, but $a_i{}^2$ must be greater than 0. If the restrictions $a_i{}^2>0$ are imposed, each successive equation puts a more stringent demand on a_0, a_1, a_2 and it appears not unlikely that there is no set $a_0{}^2$, $a_1{}^2$, $a_2{}^2$ different from 1, 2, 3 that satisfy these equations.

Matrices in which the row adjacent to the diagonal is not occupied are unsatisfactory as they are reducible. For example the matrices with the sequence of values $\sqrt{1}$, $\sqrt{1}$, $\sqrt{2}$, $\sqrt{2}$, etc. two rows above and below the main diagonal, is the direct sum of standard x's.

V. APPLICATION TO A PERTURBED HARMONIC OSCILLATOR

The application of the linear transformation to the perturbed harmonic oscillator with Hamiltonian

$$H=p^2/2m+\tfrac{1}{2}m^2x^2+\lambda x^4$$

is considered. The result of making the transformation $x'=\alpha x$, $p'=p/\alpha$ is

$$H'=p^2/2m\alpha^2+\tfrac{1}{2}m^2\alpha^2x^2+\lambda a^4x^4$$
$$=\frac{1}{\alpha^2}\left(\frac{p^2}{2m}+\frac{m^2x^2}{2}\right)+\lambda a^4x^4+\left(\alpha^2-\frac{1}{\alpha^2}\right)\frac{m^2x^2}{2}.$$

A value of the parameter α less than one can be chosen so that there is cancellation between the x^4 and x^2 parts. This makes the unperturbed part larger and the perturbed part smaller.

APPENDIX

In this Appendix a variety of associated algebraic results are derived.

I. Any operators may be expressed in terms of a and a^*. If a and a^* are taken in the harmonic-oscillator representation, then the matrix e_{ij} which has a one in the ith row and the jth column is given by

$$e_{ij}=\frac{1}{(i!j!)^{\frac{1}{2}}}\sum_{n=0}^{\infty}(-1)^n\frac{(a^*)^{n+i}a^{n+j}}{n!}. \tag{A1}$$

The r, s matrix element of e_{ij} is

$$(e_{ij})_{rs}=\frac{1}{(i!j!)^{\frac{1}{2}}}\sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\left(\frac{r!s!}{(r-n-i)!(s-n-j)!}\right)^{\frac{1}{2}}\times\delta_{r-n-js-n-j}$$
$$=\left(\frac{r!s!}{i!j!}\right)^{\frac{1}{2}}\delta_{r-is-j}\sum_{n=0}^{r-i}\frac{(-1)^n}{n!(r-i-n)!}$$
$$=\delta_{ri}\delta_{sj}.$$

Since any operator can be expressed in terms of the e_{ij}, it can also be expressed in terms of a and a^*. As an example the parity operator P may be written as

$$P=\sum(-1)^re_{rr}$$
$$=\sum(-2)^r\frac{(a^*)^ra^r}{r!}. \tag{A2}$$

II. The normal products of order t are defined as

$$N_t(\alpha,\beta)=\sum_{p=0}^{t}\binom{t}{p}(\alpha^*a^*)^p(\beta a)^{t-p}. \tag{A3}$$

The Hermitean conjugate of $N_t(\alpha,\beta)$ is given by

$$N_t^*(\alpha,\beta)=N_t(\beta,\alpha). \tag{A4}$$

If $\alpha=\beta$, N is Hermitean. The commutator of a and N_t is

$$[a,N_t(\alpha,\beta)]=\alpha^* tN_{t-1}(\alpha,\beta). \tag{A5}$$

Similarly the commutator of N and a^* is

$$[N_t(\alpha,\beta),a^*]=\beta tN_{t-1}(\alpha,\beta). \tag{A5'}$$

By induction it is possible to prove that

$$(\alpha^*a^*+\beta a)^t=\sum_{s=0}^{[t/2]}\frac{t!N_{t-2s}(\alpha^*\beta)^s}{s!(t-2s)!2^s}. \tag{A6}$$

Assume the formula is true for t. Then we have

$$\begin{aligned}
&(\alpha^*a^*+\beta a)^{t+1}\\
&=\sum_{s=0}^{[t/2]}(\alpha^*a^*+\beta a)N_{t-2s}\frac{t!(\alpha^*\beta)^s}{s!(t-2s)!2^s}\\
&=\sum_{s=0}^{[t/2]}\frac{t!(\alpha^*\beta)^s}{s!(t-2s)!2^s}\Big\{\sum_{p=0}^{t-2s}\binom{t-2s}{p}(\alpha^*a^*)^{p+1}(\beta a)^{t-2s-p}\\
&\qquad+(\alpha^*a^*)^p(\beta a)^{t-2s-p+1}+\beta\alpha^*(t-2s)N_{t-2s}\Big\}\\
&=\sum_{s=0}^{[t/2]}\frac{t!(\alpha^*\beta)^s}{s!(t-2s)!2^s}(N_{t-2s+1}+\beta\alpha^*(t-2s)N_{t-2s-1})\\
&=\sum_{s=0}^{[t/2]}\frac{t!(\alpha^*\beta)^s}{s!(t-2s)!2^s}N_{t+1-2s}\\
&\qquad+\sum_{s=1}^{[(t+1)/2]}\frac{t!(\alpha^*\beta)^s}{(s-1)!(t+1-2s)!2^{s-1}}N_{t-2s+1}\\
&=\sum_{s=0}^{[(t+1)/2]}\frac{(t+1)!(\alpha^*\beta)^s}{s!(t+1-2s)!2^s}N_{t+1-2s}.
\end{aligned}$$

It is possible to invert (A6) and express the N's in terms of the powers of $(\alpha^*a^*+\beta a)$,

$$N_t=\sum_{s=0}^{t/2}\frac{(-1)^s t!(\alpha^*a^*+\beta a)^{t-2s}(\alpha^*\beta)^s}{s!(t-2s)!2^s}. \tag{A7}$$

If (A6) is used for $(\alpha^*a^*+\beta a)^{t-2s}$, this result may easily be checked. The multiplication rule for the N's N's may be deduced from that for the $(\alpha^*a^*+\beta a)$

$$N_tN_u=\sum_r\frac{t!u!}{(t-r)!r!(t-u)!}(\alpha^*\beta)^rN_{t+u-2r}, \tag{A8}$$

where the sum is over values of r compatible with the factorials in the denominator. This is proved by expressing both N's in terms of $(\alpha^*a^*+\beta a)$ by use of (A7) and then re-expressing the product in terms of a single N with the help of (A6). The only difficult point is the formula

$$\begin{aligned}
F(u,t,r)&=\sum_a\sum_b\frac{(-1)^{a+b}(u+t-2a-2b)!}{a!b!(u-2a)!(t-2b)!(r-a-b)!}\\
&=\frac{2^s(u+t-2r)!}{(u-n)!(t-r)!r!},
\end{aligned}$$

which is proved by showing that both expressions for $F(u,t,r)$ satisfy the difference equation

$$F(u,t,r)=(2/r)F(u-1,t-1,r-1).$$

In particular we note that

$$N_1N_t=N_{t+1}+\alpha^*\beta tN_{t-1}.$$

The most general function of $\alpha^*a^*+\beta a$ is a series in the N's. Let us determine some functions. The reciprocal of $\alpha^*a^*+\beta a=N_1$ is given by

$$\frac{1}{\alpha^*a^*+\beta a}=\sum\rho_tN_t.$$

The coefficients can be determined by multiplication with N_1

$$\begin{aligned}
1=N_0&=\sum\rho_tN_1N_t\\
&=\sum\rho_t(N_{t+1}+\alpha^*\rho_tN_{t-1})\\
&=\sum\rho_{t-1}+\rho_{t+1}(t+1)(\alpha^*\beta)N_t.
\end{aligned}$$

The coefficients of like N on both sides are equated and the recursion relations

$$\begin{aligned}
\rho_1&=1/\alpha^*\beta\\
\rho_{t+1}&=-\rho_{t-1}/(t+1)\alpha^*\beta
\end{aligned}$$

result. These may easily be solved

$$\begin{aligned}
\frac{1}{\alpha^*a^*+\beta a}&=\sum\frac{(-1)^tN_{2t+1}}{1\,3\,5\cdots 2t+1(\alpha^*\beta)^t}+A\sum\frac{(-1)^tN_{2t}}{2^t t!(\alpha^*\beta)^t}\\
&=R_1+AR'.
\end{aligned}$$

The first operator R is a particular solution of the inhomogeneous equation

$$(\alpha^*a^*+\beta a)x=1,$$

while AR' is the general solution of the homogenous equation

$$(\alpha^*a^*+\beta a)x=0.$$

These formulas may be slightly generalized to find the inverse of $\alpha^*a^*+\beta a+\gamma$ by introducing the quantity $N_t(\alpha\beta\gamma)$ defined by

$$N_t(\alpha,\beta,\gamma)=\sum_{s=0}^{t/2}\frac{t!(-1)^s(\alpha^*a^*+\beta a+\gamma)^{t-2s}(\alpha^*\beta)^s}{s!(t-2s)!2^s}. \tag{A3'}$$

These new N's are related to the old ones by

$$N_t(\alpha,\beta,\gamma)=\sum_{r=0}^{t}\binom{t}{r}N_r(\alpha,\beta)^{t-r}, \tag{A10}$$

and they have the same multiplication rule (A8). Thus the reciprocal of $\alpha^*a^*+\beta a+\gamma$ is given by

$$1/(\alpha^*a^*+\beta a+\gamma)=R+AR', \tag{A9'}$$

where R and R' are now given by $N(\alpha\beta\gamma)$ instead of $N(\alpha\beta)$. The difficult point with both of these equations is the determination of the constant A so that the reciprocal is more than formally correct.

Equation (A9') may be expressed entirely in terms of $N(\alpha\beta)$ by means of Eq. (A10).

$$\frac{1}{\alpha^*a^*+\beta a+\gamma}=\sum_{u=0}^{\infty}\frac{N_u}{u!}\sum_{t=[u/2]}^{\infty}\frac{t!2^t(-1)^t\gamma^{2t+1-u}}{(2t+1-u)!(\alpha^*\beta)^t}$$
$$+A(\gamma)\sum_{u=0}^{\infty}\frac{N_u}{u!}\sum_{t=[u/2]}^{\infty}\frac{(-1)^t(2t)!\gamma^{2t-u}}{t!2^t(\alpha^*\beta)^t(2t-u)!}.$$

It has been shown that

$$\exp[N_t(\alpha,\beta)]a\exp[-N_t(\alpha,\beta)]=a-\alpha^*tN_{t-1}(\alpha,\beta).$$

Thus the commutator of $\exp[N_t(\alpha,\beta)]$ with a is

$$\{\exp[N_t(\alpha,\beta)],a\}=-\alpha^*tN_{t-1}(\alpha,\beta)\exp N_t(\alpha,\beta). \tag{A11}$$

On the other hand $\exp[N_t(\alpha,\beta)]$ has a series expansion of the form

$$\exp[N_t(\alpha,\beta)]=\sum\rho_sN_s(\alpha,\beta).$$

Both sides of Eq. (A11) may be calculated using the series expansion and then the coefficients of identical N's may be equated. This leads to

$$\rho_{s+1}(s+1)=t!\sum_r\frac{(s+2s+1-t)!(\alpha^*\beta)^r}{(t-1-r)!r!(s+r-t+1)!}\rho_{s+2s+1-t}. \tag{A12}$$

It is only practicable to solve this relation when $t=1$ or 2. In these cases it becomes

$$(s+1)\rho_{s+1}=\rho_s \tag{A12a}$$

$$(s+1)\rho_{s+1}=2(\rho_{s-1}+(s+1)\rho_{s+1}\alpha^*\beta), \tag{A12b}$$

The solution of these is

$$\rho_s=\rho_0/s! \tag{A13a}$$

$$\rho_{2s}=\rho_0/s!(1-2\alpha^*\beta)^s \tag{A13b}$$

so that

$$\exp[\alpha^*a^*+\beta a]=\rho_0\sum\frac{N_s(\alpha,\beta)}{s!} \tag{A14a}$$

and

$$\exp[\alpha^{*2}a^{*2}+2\alpha^*\beta a^*a+\beta^2a^2]=\rho_0\sum\frac{N_{2s}}{(1-\alpha^*\beta)^ss!}. \tag{A14b}$$

The ρ's can be found by using the relation

$$\exp[\lambda N_t]\exp[\mu N_t]=\exp(\lambda+\mu)N_t.$$

Thus for N_1 we find $\rho_0=e^{\alpha^*\beta/2}$ and for N_2,

$$\rho_0=e^{-\alpha^*\beta}\sum_{r=0}^{\infty}\binom{2r}{r}(\alpha^*\beta/2)^r.$$

Bounds on Elements of the S Matrix for Elastic Scattering: One-Dimensional Scattering

RALPH BARTRAM*†
General Telephone and Electronics Research Laboratories, Inc., Bayside, New York, and Physics Department, New York University, New York, New York

AND

LARRY SPRUCH‡
Physics Department, Washington Square College, and Institute of Mathematical Sciences, New York University, New York, New York

(Received July 5, 1961)

For the quantum-mechanical scattering of a particle by a static potential, a method due to Kato determines upper and lower bounds on $\cot\eta$, where η is the phase shift for a given angular momentum. The method is readily generalizable to the case of elastic multi-channel scattering *if* one can *a priori* diagonalize the scattering matrix **S** to determine the standing-wave eigenmodes (by symmetry considerations, for example). We here consider the case when this can *not* be done. Both bounds are obtained on the independent elements B_{ij} of **B**, where $\mathbf{B}=i(\mathbf{1}+\mathbf{S})(\mathbf{1}-\mathbf{S})^{-1}$, which then determine bounds on the elements of **S**, and on the eigenphase shifts and mixing parameters. One-dimensional scattering by $V(x)\neq V(-x)$ provides a concrete two-channel example. The numerical results obtained are very encouraging.

The Kato method is useful only if one can obtain solutions of related scattering problems. The most significant feature of the present approach is that the related scattering problems can be single-channel scattering problems.

1. INTRODUCTION

THE possibility of determining upper and lower bounds on $\cot\eta$, where η is the phase shift for arbitrary angular momentum for the quantum-mechanical scattering of a particle by a static central potential, has been demonstrated by Kato.[1] The basic interest of the method derives from the fact that the expression for either bound includes a variational trial function, and gives an error which is of second order in the difference between the trial function and the exact function. The bounds therefore are also variational estimates and might be termed variational bounds. In subsequent studies,[2–4] the case of angular momenta greater than zero was considered more intensively, some slight extensions and improvements in the formalism were effected, and some preliminary results were obtained for the applicability of the method to scattering by compound systems with emphasis on the low energy scattering of positrons by hydrogen atoms. An application to waveguides was also given.[5]

Static central-potential scattering problems involve an infinite number of channels which are however uncoupled, a consequence of conservation of angular momentum. In other words, an expansion into partial waves leads to a diagonalization of the scattering matrix **S** *before* the detailed analysis is begun. Furthermore, each angular-momentum channel wave function is completely characterized by one real number, the real phase shift for that angular momentum. More generally, the Kato method can be applied whenever the S matrix can be diagonalized on some *a priori* basis (generally this will be through symmetry considerations); the resultant eigenmodes can then each be characterized by one real number. Examples[5] include the one-dimensional scattering problem for an even potential $V(x)=V(-x)$ and the problem of electromagnetic waves in a dominant-mode rectangular waveguide which has an obstacle symmetric with respect to some plane perpendicular to the axis of the waveguide. The latter case involves vector wave functions of three independent variables rather than scalar wave functions of one independent variable, but this causes no essential changes. In both cases, the eigenmodes are the even and odd standing-wave solutions.

In this paper, we will consider multi-channel scattering in which the scattering matrix cannot be diagonalized *a priori*. More formally, we are concerned with elastic multi-channel scattering processes. We shall cast the Schrödinger equation into a form in which the potential is a real symmetric $N\times N$ matrix and the wave function is an $N\times 1$ matrix, though it is not necessary to do so. The equation for neutron-proton scattering with tensor forces is conveniently written in matrix form[6] with $N=2$, and we shall see in Sec. 2 that the same is true for the problem of one-dimensional

* Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at New York University.

† Present address: Physics Department, University of Connecticut, Storrs, Connecticut.

‡ Supported in part by the Geophysics Research Directorate of the Air Force Cambridge Research Laboratories, Air Research and Development Command, by the Army Research Office, Durham, North Carolina, and by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

[1] T. Kato, Progr. Theoret. Phys. (Kyoto) **6**, 394 (1951).

[2] L. Spruch and M. Kelly, Phys. Rev. **109**, 2144 (1958).

[3] L. Spruch, Phys. Rev. **109**, 2149 (1958).

[4] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959).

[5] L. Spruch and R. Bartram, J. Appl. Phys. **31**, 905 (1960); R. Bartram and L. Spruch, *ibid.* **31**, 913 (1960).

[6] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952) p. 102.

scattering. It is our purpose to determine upper and lower bounds on the elements of the **S** matrix, the $N\times N$ matrix which relates the N outgoing waves to the N incoming waves. The elements of **S** are in general complex, so that there are $2N^2$ real quantities in all, but it is well known that only $\frac{1}{2}N(N+1)$ of these real quantities are independent.[7]

The procedure, in analogy with that of Kato, is to derive an integral variational principle which is in fact an identity and then to bound the explicitly exhibited second-order term. Upper and lower bounds are first obtained on the elements B_{ij} of **B**, where the susceptance matrix **B** is the inverse of the Cayley transform of **S**. One can then also obtain bounds on linear combinations of the elements of **S**, or on still another set of $\frac{1}{2}N(N+1)$ quantities which characterize the scattering, the N eigenphase shifts and the $\frac{1}{2}N(N-1)$ mixing parameters.

Numerical examples are calculated in Sec. 6 for a one-dimensional attractive square well, asymmetric with respect to the origin.

Because of the enormous advantages of variational bounds over variational estimates, it is of great interest to study the limits to which the Kato method can be extended. There are a number of obvious directions in which to proceed. We will list some systems of interest, starting with those already considered.

(1) Systems which involve the scattering of a particle by a static central potential. We are then of course concerned with one-channel processes.

(2) Systems which for sufficiently low energy of the incident particle involve, as does (1), only one (elastic) channel, or *a priori*-separable (elastic) channels, but which are compound. We will refer to a system as compound if it has spatial (as opposed to finite-dimensional spin) internal degrees of freedom.

(3) Systems which involve the scattering of a particle by a potential which is static (the scattering system has no spatial degrees of freedom) and which can be represented by a Hermitian $N\times N$ matrix. It is assumed that the $N>1$ elastic channels cannot be uncoupled in any *a priori* way.

(4) Systems which are compound *and* which involve $N>1$ elastic channels.

(5) Systems which involve inelastic scattering.

From a purely formal point of view, case (2) is included in the original paper of Kato.[1] Unfortunately, the application of the method requires the evaluation of lower bounds on two eigenvalues α and β, which arise in an associated eigenvalue problem. While these eigenvalue bounds are relatively simple to obtain for scattering by a static central potential, no methods for obtaining bounds on α for scattering by a compound system have been developed.

The determination of a lower bound on β gives an upper bound on $(-k\cot\eta)^{-1}$. The determination of this upper bound through the Kato approach has however now been rendered largely academic by the development of a technique[8,9] which does not require the evaluation of matrix elements of the square of the Hamiltonian, as does the Kato approach. This technique is applicable to all five systems.

The Kato method as originally formulated was explicitly restricted to one-channel processes, and could therefore not be applied to (3), (4), or (5). The present paper concerns itself with the development of extensions of the Kato method which will allow it to be applied to (3); furthermore, explicit methods will be given which will often allow one to obtain lower bounds on α and on β.[10] [These extensions can also be used, in principle, for (4), and can probably be extended to (5), but these cases have not as yet been seriously examined.] We noted in the preceding paragraph that there is now available a very much simpler method for obtaining an upper bound on $(-k\cot\eta)^{-1}$ for case (3) that can be realized by the present paper. The present work is still justified because one would of course like very much to be able to determine the *other* bound. Since it entails very little additional labor, we will actually obtain both bounds for (3).

2. MATRIX FORMULATION OF THE VARIATIONAL PRINCIPLE

We begin by showing how the one-dimensional scattering problem with potential $V(x)$ can be written as a differential equation in matrix form. We have

$$-\frac{\hbar^2}{2m}\frac{d^2u(x)}{dx^2}+[V(x)-E]u(x)=0, \qquad (1)$$

where we take $V(x)$ to be real and to vanish more rapidly than $1/x$ for $|x|\to\infty$. It is evident that there are two channels, corresponding to $x>0$ and to $x<0$, respectively, which are coupled by the potential, and consequently three independent parameters are required to specify the asymptotic effects of the scattering process.

It is convenient to recast Eq. (1) in matrix form so that the three unknown parameters appear as the three independent elements of a single matrix quantity. This can be accomplished by expressing $u(x)$ and $V(x)$ as

[7] Reference 6, p. 530.

[8] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959); **117**, 1095 (1960); L. Rosenberg, L. Spruch, and T. F. O'Malley, *ibid* **118**, 184 (1960); L. Rosenberg and L. Spruch, *ibid.* **120**, 474 (1960); **121**, 1720 (1961); **125**, 1407 (1962). The last paper contains the generalization of the basic identity of the present paper, Eq. (16), to inelastic multi-channel scattering processes. A review of this material is contained in L. Spruch, *Lectures in Theoretical Physics, Boulder, 1961* [Interscience Publishers, Inc., New York (to be published)].

[9] Y. Hahn, T. F. O'Malley, and L. Spruch (to be published). This paper removes the need ever to have to truncate any of the potentials.

[10] Lower bounds on β can be determined by the methods of references 8 and 9.

sums of even and odd functions,

$$u(x)=u_e(x)+u_0(x), \quad V(x)=V_e(x)+V_0(x), \quad (2)$$

where

$$u_{e,0}(x)\equiv\tfrac{1}{2}[u(x)\pm u(-x)] \quad (3)$$

and

$$V_{e,0}(x)\equiv\tfrac{1}{2}[V(x)\pm V(-x)]. \quad (4)$$

V_e and V_0 will clearly also vanish more rapidly than $1/x$ as $|x|\to\infty$. If we replace x by $-x$ in Eq. (1), add and subtract the derived equation and Eq. (1), and use Eqs. (3) and (4), we obtain the single matrix equation[11]

$$(\mathbf{H}-E)\mathbf{u}\equiv-\frac{\hbar^2}{2m}\frac{d^2\mathbf{u}}{dx^2}+[\mathbf{V}-E]\mathbf{u}=0 \quad (5)$$

where

$$u=\begin{pmatrix}u_e\\u_0\end{pmatrix}, \quad V=\begin{pmatrix}V_e & V_0\\V_0 & V_e\end{pmatrix}. \quad (6)$$

Note that the channels, which are still coupled in general, now correspond to the even and odd portions of the wave function rather than to $x>0$ and $x<0$. We can restrict the range of the independent variable to $x\geqq 0$ since the components of u are even and odd functions of the coordinate. The boundary conditions at the origin are, obviously,

$$u_0=0, \quad du_e/dx=0. \quad (7)$$

We see then that one can conveniently formulate one-dimensional scattering in terms of a Schrödinger equation with $\mathbf{V}(x)$ a real Hermitian 2×2 matrix. For the formal development which follows, we will not restrict ourselves to matrix potentials of the above form but rather we will take $\mathbf{V}(x)$ to be an arbitrary real Hermitian 2×2 matrix whose elements fall off faster than $1/x$. The indices will therefore be 1 and 2 rather than e and 0. The results to be obtained are actually valid for $\mathbf{V}$ an $N\times N$ matrix, that is, for an N-channel process; we set $N=2$ only as a matter of convenience since it slightly simplifies the discussion and the notation. Many elastic scattering processes involving compound systems can be written in a slightly generalized form of Eq. (5).

We will now derive a variational principle for scattering which can be described by Eq. (5). The asymptotic from of $\mathbf{u}(x)$ can be written as

$$\mathbf{u}(x)\to\mathbf{a}_\theta e^{-i(kx+\theta)}+\mathbf{b}_\theta e^{+i(kx+\theta)}, \quad (8a)$$

$$k^2\equiv 2mE/\hbar^2. \quad (8b)$$

The utility of the parameter θ, where $0\leqq\theta<\pi$ but is otherwise arbitrary, will become apparent later. It is convenient to rewrite Eq. (8a) in a form that involves standing waves rather than traveling waves, since we will then be able to deal with Hermitian rather than with unitary matrices,

$$\mathbf{u}(x)\to\mathbf{e}_\theta\cos(kx+\theta)-i\mathbf{j}_\theta\sin(kx+\theta), \quad (9a)$$

$$\mathbf{e}_\theta\equiv\mathbf{a}_\theta+\mathbf{b}_\theta, \quad \mathbf{j}_\theta\equiv\mathbf{a}_\theta-\mathbf{b}_\theta. \quad (9b)$$

The amplitude vector for the outgoing waves $\mathbf{b}_\theta$ is related to that for the incoming waves $\mathbf{a}_\theta$ by the scattering $\mathbf{S}_\theta$ matrix through

$$\mathbf{b}_\theta=-\mathbf{S}_\theta\mathbf{a}_\theta. \quad (10)$$

The vectors $\mathbf{e}_\theta$ and $\mathbf{j}_\theta$ are related by the susceptance matrix $\mathbf{B}_\theta$ (so called because it is analogous to the susceptance matrix in network theory),

$$\mathbf{j}_\theta=-i\mathbf{B}_\theta\mathbf{e}_\theta. \quad (11)$$

From Eqs. (9)–(11), it can be seen that $\mathbf{B}_\theta$ is the inverse of the Cayley transform of $\mathbf{S}_\theta$,

$$\mathbf{B}_\theta=i(1+\mathbf{S}_\theta)(1-\mathbf{S}_\theta)^{-1}. \quad (12)$$

It can be shown that $\mathbf{S}_\theta$ is unitary and symmetric[12] and as a consequence that $\mathbf{B}_\theta$ is Hermitian and symmetric and therefore real. Equation (11) can now be used to eliminate $\mathbf{j}_\theta$ in Eq. (9a), giving rise to

$$\mathbf{u}(x)\to\mathbf{e}_\theta\cos(kx+\theta)-\mathbf{B}_\theta\mathbf{e}_\theta\sin(kx+\theta). \quad (13)$$

We now introduce a trial wave function $\mathbf{u}_t(x)$ which is required to have the asymptotic form

$$\mathbf{u}_t(x)\to\mathbf{e}_\theta\cos(kx+\theta)-\mathbf{B}_{\theta t}\mathbf{e}_\theta\sin(kx+\theta) \quad (14)$$

as $x\to+\infty$, that is, the amplitude vector $\mathbf{e}_\theta$ is the same as that of the exact function, but the unknown $\mathbf{B}_\theta$ is replaced by $\mathbf{B}_{\theta t}$. In addition, we impose the boundary condition

$$\mathbf{u}^\dagger d\mathbf{u}_t/dx=(d\mathbf{u}^\dagger/dx)\mathbf{u}_t$$

at $x=0$, and we require that $\mathbf{u}_t$ and $d\mathbf{u}_t/dx$ be continuous in the range $0<x<\infty$. It is to be noted that all of the boundary conditions imposed on $\mathbf{u}_t$ are also satisfied by the exact function. In order to arrive at a variational principle, consider the relation (all integrals extend from 0 to ∞, and the limits will henceforth be omitted)

$$-(2m/\hbar^2)\int[\mathbf{u}^\dagger\mathbf{H}\mathbf{u}_t-(\mathbf{H}\mathbf{u})^\dagger\mathbf{u}_t]dx=\left[\mathbf{u}^\dagger\frac{d}{dx}\mathbf{u}_t-\left(\frac{d}{dx}\mathbf{u}^\dagger\right)\mathbf{u}_t\right]\Bigg|_0^\infty. \quad (15)$$

Using Eqs. (13) and (5), the boundary condition at the origin, and the Hermitian character of $\mathbf{B}_\theta$, we obtain from Eq. (15)

$$\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta=\mathbf{e}_\theta^\dagger\mathbf{B}_{\theta t}\mathbf{e}_\theta-(2m/\hbar^2k)\int\mathbf{u}_t^\dagger(\mathbf{H}-E)\mathbf{u}_t dx+(2m/\hbar^2k)\int\mathbf{w}^\dagger(\mathbf{H}-E)\mathbf{w}dx, \quad (16a)$$

$$\mathbf{w}\equiv\mathbf{u}_t-\mathbf{u}. \quad (16b)$$

[11] Matrix quantities will be denoted by boldface type. Other symbols employed are † for Hermitian adjoint, ~ for transpose, and * for complex conjugate.

[12] Reference 6, pp. 525, 529.

The first two terms on the right-hand side of Eq. (16a) can be evaluated directly from the trial function, and the third term is of second order in the difference between the trial function and the exact function. The first two terms therefore constitute a variational approximation for $\mathbf{e}_\theta^\dagger \mathbf{B}_\theta \mathbf{e}_\theta$. However, the approximation is in general neither an upper nor a lower bound on the exact value. In the next section we will derive rigorous second-order upper and lower bounds on the error term, the third term on the right-hand side of Eq. (16a).

3. RIGOROUS BOUNDS ON THE ERROR TERM

In this section we will derive rigorous upper and lower bounds on the error term in Eq. (16a), following a procedure analogous to that derived by Kato.[1] We begin by considering the associated eigenvalue problem,

$$(\mathbf{H}-E)\boldsymbol{\phi}(x)-\mu\boldsymbol{\varrho}(x)\boldsymbol{\phi}(x)=0. \tag{17}$$

In Eq. (17), $\boldsymbol{\varrho}(x)$ is required to be a real, positive definite Hermitian matrix each of whose elements must fall off faster than $1/x$ as $|x|\to\infty$. At the origin, $x=0$, $\boldsymbol{\phi}(x)$ is required to satisfy the condition $\boldsymbol{\phi}^\dagger d\boldsymbol{\phi}/dx = (d\boldsymbol{\phi}^\dagger/dx)\boldsymbol{\phi}$. In particular, we are interested in discrete eigenvalues μ_n, of which there will be an infinite number, defined by the condition that H is a Hermitian operator when operating on the corresponding eigenfunctions, $\boldsymbol{\phi}_n$, or on $\mathbf{w}$ [defined by Eq. (16b)], that is,

$$\int[\boldsymbol{\phi}_m^\dagger \mathbf{H}\boldsymbol{\phi}_n-(\mathbf{H}\boldsymbol{\phi}_m)^\dagger\boldsymbol{\phi}_n]dx=0 \tag{18a}$$

$$\int[\mathbf{w}^\dagger \mathbf{H}\boldsymbol{\phi}_n-(\mathbf{H}\mathbf{w})^\dagger\boldsymbol{\phi}_n]dx=0. \tag{18b}$$

It follows from the Hermitian character of $\mathbf{H}$ that the associated eigenfunctions are orthogonal with respect to the weight factor $\boldsymbol{\varrho}$ and that the associated eigenvalues are real,

$$\int \boldsymbol{\phi}_m^\dagger \boldsymbol{\varrho}\boldsymbol{\phi}_n dx=(\hbar^2k/2m)\delta_{mn}, \tag{19a}$$

$$\mu_n^*=\mu_n. \tag{19b}$$

The normalization in Eq. (19a) has been chosen for later convenience. In addition we require that the $\boldsymbol{\phi}_n$ satisfy the boundary condition[13] at $x=0$,

$$\boldsymbol{\phi}_m^\dagger d\boldsymbol{\phi}_n/dx=(d\boldsymbol{\phi}_m^\dagger/dx)\boldsymbol{\phi}_n, \tag{20a}$$

$$\mathbf{w}^\dagger d\boldsymbol{\phi}_n/dx=(d\mathbf{w}^\dagger/dx)\boldsymbol{\phi}_n. \tag{20b}$$

[13] It is presumed that Eqs. (20) can be satisfied. In the case of one-dimensional scattering, $\mathbf{u}$ is required to satisfy Eq. (7) at the origin; if the same conditions are imposed on $\mathbf{u}_t$ and $\boldsymbol{\phi}$, Eqs. (20) will be satisfied. If the problem involved tensor forces or a non-central potential, Eq. (7) would be replaced by the requirement that all components of $\mathbf{u}$ vanish at the origin. Again, Eqs. (20) could be satisfied by imposing the same conditions on $\mathbf{u}_t$ and $\boldsymbol{\phi}$.

In analogy with Eq. (13), $\boldsymbol{\phi}_n$ has the asymptotic form as $x\to\infty$,

$$\boldsymbol{\phi}_n \to \mathbf{e}_n\cos kx-\mathbf{B}(\mu_n)\mathbf{e}_n\sin kx. \tag{21}$$

From Eqs. (13), (14), and (16b) it follows that the asymptotic form of $\mathbf{w}$ is

$$\mathbf{w}\to(\mathbf{B}_\theta-\mathbf{B}_{\theta t})\mathbf{e}_\theta\sin(kx+\theta). \tag{22}$$

Using Green's theorem and Eqs. (5) and (20)–(22) to evaluate the integrals in Eqs. (18), we find that (18) can be satisfied only if $\mathbf{B}(\mu_n)$ satisfies

$$\mathbf{B}(\mu_n)\mathbf{e}_n=-\cot\theta\,\mathbf{e}_n. \tag{23}$$

Thus the associated eigenfunctions are also eigenmodes, i.e., their amplitude vectors $\mathbf{e}_n$ are eigenvectors of the susceptance matrix $\mathbf{B}(\mu_n)$. The μ_n are just the values of μ for which one or another of the eigenvalues of the susceptance matrix is equal to $-\cot\theta$.

We will now proceed to derive a monotonicity theorem required in subsequent developments. In principle, we can find solutions $\boldsymbol{\phi}(x)$ of Eq. (17) for arbitrary μ if we no longer impose the boundary conditions (18). In particular, we can find 2 eigenmodes $\boldsymbol{\phi}_i(x)$ $(i=1, 2)$ for an arbitrary value of μ which are distinguished by their asymptotic forms,

$$\boldsymbol{\phi}_i(\mu)\to\mathbf{e}_i(\mu)[\cos kx-\zeta_i(\mu)\sin kx]. \tag{24}$$

In Eq. (24) $\zeta_i(\mu)$ is the ith eigenvalue of $\mathbf{B}(\mu)$. [It should be noted that the eigenmode solutions of Eq. (17) are employed here only in the formal development. Presumably, it is more difficult to find such solutions than to solve the original problem, Eq. (5).] At this point it is convenient to relabel the associated eigenfunctions with two indices. The associated eigenvalues μ_{ni} are then defined by the relation

$$\zeta_i(\mu_{ni})=-\cot(\theta+n\pi) \tag{25}$$

together with the boundary conditions specified by Eqs. (20), where the index i labels the eigenmodes and the corresponding associated eigenfunctions have the asymptotic form

$$\boldsymbol{\phi}_{ni}\to\mathbf{e}_i(\mu_{ni})[\cos kx+\cot(\theta+n\pi)\sin kx]. \tag{26}$$

Consider the eigenmode solutions of Eq. (17) for two different values of μ, to be denoted by μ' and μ'', without applying the boundary conditions (18), and form the expression

$$\begin{aligned}-(2m/\hbar^2)\int\{\boldsymbol{\phi}_i^\dagger(\mu')\mathbf{H}\boldsymbol{\phi}_j(\mu'')-[\mathbf{H}\boldsymbol{\phi}_i(\mu')]^\dagger\boldsymbol{\phi}_j(\mu'')\}dx\\ =(2m/\hbar^2)(\mu'^*-\mu'')\int\boldsymbol{\phi}_i^\dagger(\mu')\boldsymbol{\varrho}\boldsymbol{\phi}_j(\mu'')dx\\ =\left[\boldsymbol{\phi}_i^\dagger(\mu')\frac{d}{dx}\boldsymbol{\phi}_j(\mu'')-\left\{\frac{d}{dx}\boldsymbol{\phi}_i^\dagger(\mu')\right\}\boldsymbol{\phi}_j(\mu'')\right]\Bigg|_0^\infty,\end{aligned} \tag{27}$$

where we have made use of Eqs. (5) and (17). Using boundary conditions at the origin analogous to Eqs. (20) and the asymptotic form of $\phi_i(\mu)$ given by Eq. (24), we may rewrite Eq. (27) as

$$(2m/\hbar^2)(\mu'^*-\mu'')\int \phi_i^\dagger(\mu')\varrho\phi_j(\mu'')dx = k\mathbf{e}_i^\dagger(\mu')\mathbf{e}_j(\mu'')[\zeta_i(\mu')-\zeta_j(\mu'')]. \quad (28)$$

Consider Eq. (28) with μ' and μ'' real but otherwise arbitrary and with $j=i$. Let the eigenphase shift $\delta_i(\mu)$ be defined by

$$\zeta_i(\mu) = -\cot\delta_i(\mu). \quad (29)$$

Substituting Eq. (29) in Eq. (28) and taking the limit $\mu'' \to \mu'$, we obtain

$$d\delta_i/d\mu = (2m/\hbar^2)\sin^2\delta_i \int \phi_i^\dagger\varrho\phi_i dx/(k\mathbf{e}_i^\dagger\mathbf{e}_i) \geqq 0. \quad (30)$$

The inequality in Eq. (30) follows from the positive definite character of ϱ. Thus the eigenphase shifts are monotonic increasing functions of μ (for $k\neq 0$).

We now proceed to derive rigorous bounds on the error term in Eq. (16a). Suppose that $\mathbf{f}$ and $\mathbf{g}$ are functions which either have the same phase shift as the ϕ_{ni} or fall off asymptotically faster than $1/x$. We define expansion coefficients a_{ni} and b_{ni} by the relations

$$a_{ni} \equiv (2m/\hbar^2k)\int \phi_{ni}^\dagger\varrho\mathbf{f}dx, \quad (31a)$$

$$b_{ni} \equiv (2m/\hbar^2k)\int \phi_{ni}^\dagger\varrho\mathbf{g}dx. \quad (31b)$$

It can then be shown that the associated eigenfunctions form a complete set[14] in the sense that

$$\int \mathbf{f}^\dagger\varrho\mathbf{f}dx = (\hbar^2k/2m)\sum_{n,i} a_{ni}^* a_{ni} \quad (32a)$$

$$\int \mathbf{f}^\dagger\varrho\mathbf{g}dx = (\hbar^2k/2m)\sum_{n,i} a_{ni}^* b_{ni} \quad (32b)$$

$$\int \mathbf{g}^\dagger\varrho\mathbf{g}dx = (\hbar^2k/2m)\sum_{n,i} b_{ni}^* b_{ni}. \quad (32c)$$

If we now set $\mathbf{f}=\mathbf{w}$ and $\mathbf{g}=-\varrho^{-1}(\mathbf{H}-E)\mathbf{w}$, it follows from Eqs. (31) that a_{ni} and b_{ni} are related by

$$b_{ni} = -(2m/\hbar^2k)\int \phi_{ni}^\dagger(\mathbf{H}-E)\mathbf{w}dx = -(2m/\hbar^2k)\int [(\mathbf{H}-E)\phi_{ni}]^\dagger\mathbf{w}dx = -\mu_{ni}(2m/\hbar^2k)\int \phi_{ni}^\dagger\varrho\mathbf{w}dx = -\mu_{ni}a_{ni}, \quad (33)$$

provided ϱ is chosen such that $\mathbf{g}$ falls off asymptotically faster than $1/x$. Equations (32) can now be written in the form

$$\int \mathbf{w}^\dagger\varrho\mathbf{w}dx = (\hbar^2k/2m)\sum_{n,i} \mu_{ni}^{-2} b_{ni}^* b_{ni}, \quad (34a)$$

$$\int \mathbf{w}^\dagger(\mathbf{H}-E)\mathbf{w}dx = (\hbar^2k/2m)\sum_{n,i} \mu_{ni}^{-1} b_{ni}^* b_{ni}, \quad (34b)$$

$$\int [(\mathbf{H}-E)\mathbf{w}]^\dagger\varrho^{-1}(\mathbf{H}-E)\mathbf{w}dx = (\hbar^2k/2m)\sum_{n,i} b_{ni}^* b_{ni}. \quad (34c)$$

Let α_θ be the smallest positive eigenvalue and $-\beta_\theta$ be the smallest (in absolute value) negative eigenvalue. It follows from Eqs. (34) that

$$-\beta_\theta^{-1}\epsilon_\theta^2 \leqq \int \mathbf{w}^\dagger(\mathbf{H}-E)\mathbf{w}dx \leqq \alpha_\theta^{-1}\epsilon_\theta^2, \quad (35a)$$

where

$$\epsilon_\theta^2 \equiv \int [(\mathbf{H}-E)\mathbf{w}]^\dagger\varrho^{-1}(\mathbf{H}-E)\mathbf{w}dx. \quad (35b)$$

We have thus obtained rigorous upper and lower bounds on the error term in Eq. (16a). The quantity ϵ_θ^2 is of second order in the difference between the trial function and the exact function, and can be calculated from the trial function since it follows from Eqs. (5) and (16b) that

$$(\mathbf{H}-E)\mathbf{w} = (\mathbf{H}-E)\mathbf{u}_t. \quad (36)$$

The problem still remains of course of obtaining lower bounds on α_θ and β_θ. It is to be recalled that it is possible to obtain close bounds on $\int \mathbf{w}^\dagger(\mathbf{H}-E)\mathbf{w}dx$ even with quite crude lower bounds on α_θ and β_θ, since ϵ_θ^2 is a second-order quantity.

4. BOUNDS ON α_θ AND ON β_θ

In this section we will indicate a rather general method for obtaining lower bounds on α_θ and on β_θ for cases covered by Eq. (5). The monotonicity theorem, Eq. (30), states that the eigenphase shifts $\delta_i(\mu)$ are monotonic increasing functions of μ. The labeling of the eigenphase shifts is essentially arbitrary, but the application of this theorem requires that we keep track of the labeling as μ is varied. Since μ is a continuous variable, this presents no difficulty except at crossing points where $\delta_i(\mu)=\delta_j(\mu)$, $i\neq j$. However, we are interested in the weaker statement

$$\mu_{n-1,i} \leqq \mu_{ni} \leqq \mu_{n+1,i}, \quad (37)$$

where μ_{ni} is the value of μ for which $\delta_i(\mu_{ni})=\theta+n\pi$. Equation (37) holds in spite of the ambiguity in labeling at crossing points, since no matter which curve is continued into which, both continue to rise monotonically. The relationship between α_θ and β_θ and the

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

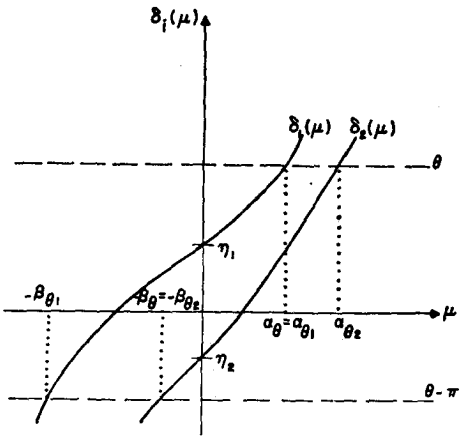


FIG. 1. Relationship between α_θ and β_θ and the eigenphase shifts of the associated eigenvalue problem, $\delta_i(\mu)$. The eigenphase shifts of the original problem are η_1 and η_2.

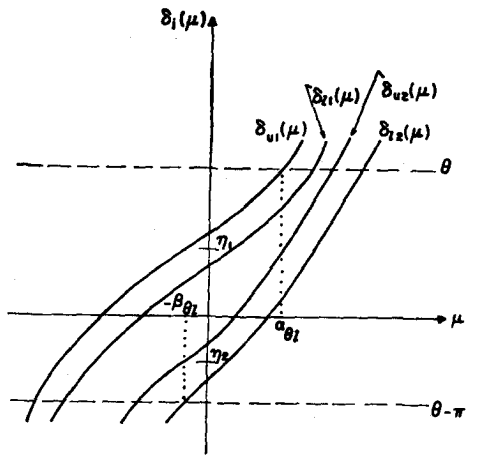


FIG. 2. Use of comparison potentials to obtain lower bounds, $\alpha_{\theta l}$ and $\beta_{\theta l}$, on α_θ and β_θ, respectively. $\delta_{ui}(\mu)$ and $\delta_{li}(\mu)$ are derived from the comparison potentials and are upper and lower bounds, respectively, on $\delta_i(\mu)$.

eigenphase shifts is illustrated for a typical case in Fig. 1. If we define $\alpha_{\theta i}$ and $-\beta_{\theta i}$ to be the smallest positive eigenvalue and smallest (in absolute value) negative eigenvalue, respectively, associated with the ith eigenmode, then α_θ is the smaller of the $\alpha_{\theta i}$ and β_θ the smaller of the $\beta_{\theta i}$. Thus there is only one α_θ and one β_θ.

Figure 1 demonstrates the relationship between α_θ and β_θ and the eigenphase shifts, but these eigenphase shifts are presumably not known. We must therefore devise a method for obtaining lower bounds on α_θ and β_θ. To this end, we introduce a Hermitian comparison potential $\mathbf{V}_c$ such that $\mathbf{\Delta V}=\mathbf{V}_c-\mathbf{V}$ is nonpositive and for which the associated eigenvalue problem can be solved exactly, or for which bounds can be obtained. We will now show that the eigenphase shifts corresponding to the comparison potential are upper bounds on the eigenphase shifts corresponding to the true potential for any given value of μ. Under the substitutions

$$\mathbf{H}\rightarrow\mathbf{H}-\mu\boldsymbol{\varrho} \tag{38a}$$

$$\mathbf{H}-\mu\boldsymbol{\varrho}\rightarrow\mathbf{H}-\mu\boldsymbol{\varrho}+\lambda\mathbf{\Delta V}, \tag{38b}$$

the derivation which led to Eq. (30) gives, with $\bar{\delta}_i=\bar{\delta}_i(\mu,\lambda)$ the ith eigenphase shift for the Hamiltonian on the right in expression (38b) and $\bar{\boldsymbol{\phi}}_i$ the corresponding function.

$$d\bar{\delta}_i/d\lambda=-(2m/\hbar^2)\sin^2\bar{\delta}_i\int\bar{\boldsymbol{\phi}}_i^\dagger\mathbf{\Delta V}\bar{\boldsymbol{\phi}}_i dx/(k\mathbf{e}_i^\dagger\mathbf{e}_i)\geqq 0. \tag{39}$$

The inequality follows from the nonpositive character of $\mathbf{\Delta V}$. The eigenphase shift $\delta_{ui}(\mu)$ corresponding to the comparison potential is obtained when $\lambda=1$, and $\delta_i(\mu)$ when $\lambda=0$. It follows from Eq. (39) that $\delta_{ui}(\mu)$ is an upper bound on $\delta_i(\mu)$,

$$\delta_{ui}(\mu)\geqq\delta_i(\mu). \tag{40}$$

As before, we are required to keep track of the eigenphases as λ is varied continuously from zero to one. The only difficulty occurs at a possible crossing point, but $\bar{\delta}_i(\mu,\lambda)=\bar{\delta}_j(\mu,\lambda)$ already implies that these eigenphase shifts are greater than the true values and they can only increase as λ approaches one. Equation (40) is therefore valid in any case.

By introducing another comparison potential such that $\mathbf{\Delta V}$ is nonnegative, it can be shown in a similar fashion that the corresponding eigenphase shifts $\delta_{li}(\mu)$ are lower bounds on the $\delta_i(\mu)$,

$$\delta_{li}(\mu)\leqq\delta_i(\mu). \tag{41}$$

The way in which the bounding curves obtained from the comparison potentials can be used to calculate lower bounds $\alpha_{\theta l}$ and $\beta_{\theta l}$ on α_θ and on β_θ, respectively, is illustrated for a typical case in Fig. 2. It is clear from Fig. 2 that if the comparison potentials do not approximate the actual potential fairly closely, it may not be possible to obtain lower bounds on α_θ and on β_θ. For example, if the curve labeled $\delta_{u1}(\mu)$ in Fig. 2 were to cross the dashed line $\delta_i(\mu)=\theta$ at a point where μ is negative, then it would not be possible to obtain a lower bound on α_θ.

In general it is possible, as we shall show, to calculate lower bounds on α_θ and on β_θ by using comparison potentials and a weight function $\boldsymbol{\varrho}$ which are diagonal. In that case, the 4 bounding curves $\delta_{ui}(\mu)$ and $\delta_{li}(\mu)$ are obtained by solving 4 single-channel problems rather than 2 two-channel problems. The consequences of this result are far reaching in that very few truly multi-channel problems can be solved at all, even with simple comparison potentials. On the other hand, the possibility of using diagonal comparison potentials implies that it is no more difficult to obtain bounds in the multi-channel problem than in the single-channel problem, though it may be much more tedious, and thus the power of the method is greatly enhanced. Furthermore, the Kato method for obtaining bounds

in the single-channel case can be used to obtain bounds on the eigenphase shifts of the comparison potentials, or the single-channel comparison potential problems could even be solved numerically.

We will now consider the requirements that must be imposed on the elements of $V_c(x)$ if it is taken to be diagonal. Writing

$$V_c=\begin{pmatrix} V_{c1} & 0 \\ 0 & V_{c2} \end{pmatrix}, \tag{42}$$

we have

$$\mathbf{\Delta V}=\begin{pmatrix} V_{c1}-V_{11} & -V_{12} \\ -V_{21} & V_{c2}-V_{22} \end{pmatrix}. \tag{43}$$

It is then trivial to show that if $\mathbf{\Delta V}$ is to be non-negative, we must have

$$V_{c1}\geqq V_{11}+|V_{12}|, \quad V_{c2}\geqq V_{22}+|V_{12}|, \tag{44}$$

while if $\mathbf{\Delta V}$ is to be nonpositive, we must have

$$V_{c1}\leqq V_{11}-|V_{12}|, \quad V_{c2}\leqq V_{22}-|V_{12}|. \tag{45}$$

The above conditions are not sufficient for obtaining lower bounds on α_θ and on β_θ, since we have already seen that it is possible that no diagonal comparison potential may be sufficiently close to $\mathbf{V}$ to be able to determine bounds.

5. METHOD OF OBTAINING BOUNDS

We are now in a position to apply the foregoing results to the problem of obtaining rigorous upper and lower bounds on the independent elements of the scattering matrix. From Eqs. (16) and (35), we can obtain expressions which are upper and lower bounds on a quadratic form of the susceptance matrix,

$$-(2m\epsilon_\theta^2/\hbar^2k\beta_{\theta l})\leqq \mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta-\mathbf{e}_\theta^\dagger\mathbf{B}_{\theta t}\mathbf{e}_\theta + (2m/\hbar^2k)\int \mathbf{u}_t^\dagger(\mathbf{H}-E)\mathbf{u}_t dx \leqq (2m\epsilon_\theta^2/\hbar^2k\alpha_{\theta l}), \tag{46}$$

$$\epsilon_\theta^2\equiv\int[(\mathbf{H}-E)\mathbf{u}_t]^\dagger\varrho^{-1}(\mathbf{H}-E)\mathbf{u}_t dx, \tag{47}$$

where $\alpha_{\theta l}$ and $\beta_{\theta l}$ are lower bounds on α_θ and on β_θ, respectively. We thus have variational bounds on $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$. The bound property enables us to improve the approximations by incorporating variational parameters into $\mathbf{u}_t$. As opposed to the situation in the usual variational calculation, we do *not* have to find the stationary values of the parameters (it is preferable but not essential) but can rather evaluate the bound for a few values of the parameter and unambiguously choose the best result. This makes feasible the use of variational parameters which are not linear. θ can be thought of as such a parameter, though in the present paper we will not take full advantage of this property.

As an alternative procedure, we could adjust the variational parameters to minimize the quantity ϵ_θ^2 given by Eq. (47). While this procedure is simpler, in general it yields inferior bounds and consequently it will not be followed in this paper. With reference to Fig. 2, θ should be chosen to provide the greatest lower bounds on α_θ and on β_θ. In the single-channel problem θ can be chosen differently for α_θ and for β_θ, since the upper and lower bounds on the phase shift are obtained by solving two distinct problems. However, in the multi-channel case the susceptance matrices $\mathbf{B}_\theta$ corresponding to two different values of θ are related in such a complex manner that it does not appear feasible to obtain upper and lower bounds on the 3 independent parameters of the scattering matrix by using different values of θ for the upper and lower bounds on the quadratic forms $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$.

In order to obtain bounds on the 3 distinct elements of $\mathbf{B}_\theta$, we must evaluate the right-hand sides of Eqs. (46) with 3 different choices of the trial function which differ in the amplitude vector $\mathbf{e}_\theta$. Note that the different values of $\mathbf{e}_\theta$ are *not* required to be linearly independent (they of course could not be) since $\mathbf{e}_\theta$ appears quadratically in $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$. Rather, the resulting inequalities in the 3 independent elements of $\mathbf{B}_\theta$ must be linearly independent. Since there are two bounds for each choice of $\mathbf{e}_\theta$, there are 6 problems in all.

Since the scattering matrix $\mathbf{S}$ (the scattering matrix $\mathbf{S}$ as normally defined is $\mathbf{S}_\theta$ with $\theta=0$) can be calculated from $\mathbf{B}_\theta$, the 3 independent elements of $\mathbf{B}_\theta$ are sufficient to determine $\mathbf{S}$. $\mathbf{S}$ can also be calculated from an alternative set of 3 parameters, the 2 eigenphase shifts and the mixing parameter.[15] This alternative mode of description is useful in many applications.

We now examine the procedure for obtaining bounds on the eigenphase shifts and mixing parameters. Since $\mathbf{B}_\theta$ is Hermitian and real, it can be diagonalized by an orthogonal matrix, $\mathbf{T}$. The same matrix $\mathbf{T}$ also diagonalizes $\mathbf{S}$, since $\mathbf{B}_\theta$ is related to $\mathbf{S}$ by a rational fraction. Because $\mathbf{S}$ is unitary, its diagonal form can be written as $\exp(2i\boldsymbol{\eta})$ where $\boldsymbol{\eta}$ is a diagonal matrix whose elements are the eigenphase shifts. Note that when the eigenphase shifts are defined in this way, they do not necessarily vanish in the absence of a potential but rather assume values consistent with the boundary conditions at the origin. The mixing parameter is just the angle of rotation on which $\mathbf{T}$ depends. It follows from Eqs. (8a), (10), and (12) that $\mathbf{B}_\theta$ is related to $\mathbf{S}$ by

$$\mathbf{B}_\theta=i(1+\mathbf{S}e^{-2i\theta})(1-\mathbf{S}e^{-2i\theta})^{-1}. \tag{48}$$

From the foregoing discussion, $\mathbf{S}$ can be represented by

$$\mathbf{S}=\mathbf{T}\exp(2i\boldsymbol{\eta})\tilde{\mathbf{T}}. \tag{49}$$

[15] For a discussion of eigenphase shifts and mixing parameters, see for example W. Kohn, Phys. Rev. **74**, 1763 (1948); J. M. Blatt and L. C. Biedenharn, *ibid.* **86**, 399 (1952); L. C. Biedenharn and J. M. Blatt, *ibid.* **93**, 1387 (1954).

Therefore, from Eqs. (48) and (49) $\mathbf{B}_\theta$ can be written as

$$\mathbf{B}_\theta = -\mathbf{T}\cot(\eta-\theta)\tilde{\mathbf{T}}. \tag{50}$$

The orthogonal matrix $\mathbf{T}$ can be written as

$$\mathbf{T}=\begin{pmatrix}\cos\psi & \sin\psi\\ -\sin\psi & \cos\psi\end{pmatrix}, \tag{51}$$

where ψ is the mixing parameter. Solving Eqs. (50) and (51) for ψ, η_1, and η_2, we obtain

$$\psi=\tfrac{1}{2}\cot^{-1}[(B_{\theta 22}-B_{\theta 11})/(2B_{\theta 12})]. \tag{52a}$$

$$\eta_1=\theta+\cot^{-1}[\tfrac{1}{2}(B_{\theta 22}-B_{\theta 11})\cos 2\psi+B_{\theta 12}\sin 2\psi -\tfrac{1}{2}(B_{\theta 22}+B_{\theta 11})] \tag{52b}$$

$$\eta_2=\theta+\cot^{-1}[-\tfrac{1}{2}(B_{\theta 22}-B_{\theta 11})\cos 2\psi-B_{\theta 12}\sin 2\psi -\tfrac{1}{2}(B_{\theta 22}+B_{\theta 11})]. \tag{52c}$$

It is clear that by choosing appropriate combinations of upper and lower bounds on the elements of $\mathbf{B}_\theta$ in Eqs. (52), we can derive upper and lower bounds on ψ, η_1, and η_2.

6. NUMERICAL EXAMPLES

In order to test the practicality of the method, we have applied it to the problem of one-dimensional scattering by an attractive square well which is asymmetrically placed with respect to the origin, that is, $V(x)=0$ for $x<0$ and for $x>a$, and $V(x)=(-\hbar^2/2m)U$ for $0\leqq x\leqq a$, where U is a positive constant. (It should be emphasized that our method is *not* restricted to potentials which vanish identically as $|x|\to\infty$.) Equations (4) and (6) then give

$$\mathbf{V}(x)=-(\hbar^2/4m)U\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}, \quad 0\leqq x\leqq a \tag{53a}$$

$$=\mathbf{0}, \quad a<x<\infty \tag{53b}$$

where the range of x is now 0 to ∞. The trial function within the region of the square well was chosen to be

$$\mathbf{u}_t(x)=\begin{bmatrix}\sum_{p=0,2,\cdots}^{2P} b_p(x/a)^p\\ \sum_{p=1,3,\cdots}^{2P+1} b_p(x/a)^p\end{bmatrix}, \quad 0\leqq x\leqq a. \tag{54}$$

The illogicality of the unnecessarily restricted form of power series was not recognized until after the numerical calculations had been completed, and it was not then considered worthwhile to repeat the calculations with a less restricted form.

For all $x>a$, $\mathbf{u}_t$ was chosen to have the form given by Eq. (14). $\mathbf{u}_t$ must of course be continuous in slope and value at $x=a$.

The comparison potentials which generate non-positive and non-negative matrices $\Delta\mathbf{V}$ were taken to be

$$\mathbf{V}_c(x)=-(\hbar^2/2m)U\mathbf{1}, \quad 0\leqq x\leqq a$$
$$=\mathbf{0}, \quad a<x<\infty \tag{55}$$

and

$$\mathbf{V}_c(x)=\mathbf{0}, \quad 0\leqq x\leqq\infty, \tag{56}$$

respectively. ϱ was chosen as

$$\varrho=\rho_0\mathbf{1}, \quad 0\leqq x\leqq a, \quad \varrho=\mathbf{0}, \quad a<x<\infty,$$

where ρ_0 is a constant. The eigenphase shifts $\delta_{el}(\mu)$ and $\delta_{0l}(\mu)$ (in discussing the example, we will use subscripts e and 0 rather than 1 and 2) corresponding to the lower bound are just the phase shifts associated with the even and odd solutions, respectively, of the scalar

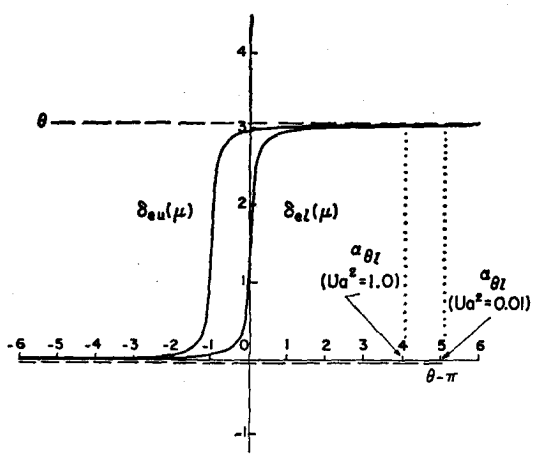


FIG. 3. Upper and lower bounds on $\delta_e(\mu)$ for $ka=0.1$, $Ua^2=1.0$. The lower bound, $\delta_{el}(\mu)$, also applies to $ka=0.1$, $Ua^2=0.01$. The corresponding upper bound, $\delta_{eu}(\mu)$, cannot be distinguished from $\delta_{el}(\mu)$. The lower bounds on α_θ were actually chosen from the numerical data rather than from the curves.

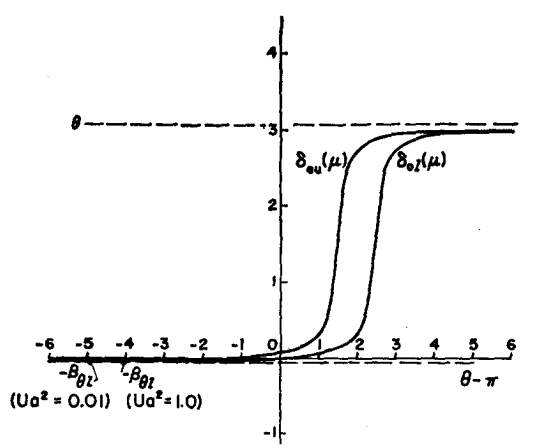


FIG. 4. Upper and lower bounds on $\delta_0(\mu)$ for $ka=0.1$, $Ua^2=1.0$. The lower bound, $\delta_{0l}(\mu)$, also applies to $ka=0.1$, $Ua^2=0.01$. The lower bounds on β_θ were actually chosen from the numerical data rather than from the curves. Both bounds were derived from the same curve and could have been chosen equal.

differential equations

$$\frac{d^2\phi(x)}{dx^2}+k^2\phi(x)+(2m/\hbar^2)\mu\rho_0\phi(x)=0, \quad 0\leqq x\leqq a \quad (57a)$$

$$\frac{d^2\phi(x)}{dx^2}+k^2\phi(x)=0, \quad a<x<\infty, \quad (57b)$$

while the eigenphase shifts corresponding to the upper bound are obtained from the same equation under the substitution

$$\mu\rho_0 \rightarrow \mu\rho_0+(\hbar^2/2m)U. \quad (58)$$

The phase shifts corresponding to Eqs. (57) are given by

$$\delta_{el}(\mu)=\pi/2-ka+\cot^{-1}[(ka/\kappa a)\cot\kappa a], \quad (59a)$$

$$\delta_{0l}(\mu)=-ka+\tan^{-1}[(ka/\kappa a)\tan\kappa a], \quad (59b)$$

$$\kappa\equiv(k^2+2m\mu\rho_0/\hbar^2)^{\frac{1}{2}}. \quad (60)$$

If we plot $\delta_{el}(\mu)$ and $\delta_{0l}(\mu)$ against $2m\mu\rho_0a^2/\hbar^2$, we obtain curves each of which lies entirely below the corresponding curve for the exact eigenphase shift. If instead of Eq. (60) we define κ by

$$\kappa=(k^2+2m\mu\rho_0/\hbar^2+U)^{\frac{1}{2}}, \quad (61)$$

we obtain curves which lie entirely above the corresponding exact curves. These pairs of bounding curves are then used to obtain lower bounds on α and on β in accordance with the procedure described in Sec. 4. The particular value chosen for ρ_0 plays no role whatever. (μ and ϱ appear only in the combination $\mu\varrho$. Choosing ρ_0 larger would then give smaller values of α_θ and of β_θ, but of course the bounds on $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$ would not be affected.)

We will now present some numerical results obtained with the use of the Bendix G-15 digital computer at the General Telephone and Electronics Laboratories, Bayside, Long Island. Bounding curves on $\delta_e(\mu)$ and $\delta_0(\mu)$ calculated from Eqs. (59), (60), and (61) are plotted in Figs. 3 and 4 for $ka=0.1$, $Ua^2=1.0$ and in Figs. 5 and 6 for $ka=\pi/4$, $Ua^2=\pi^2/4$. The lower bounding curves in Figs. 3 and 4 apply also to $ka=0.1$, $Ua^2=0.01$, but in this case the upper bounding curves are so close to the lower bounding curves that the two curves cannot be distinguished from one another on the scale of the drawings. Values of θ and lower bounds on α_θ and β_θ consistent with those curves are listed in Table I. In each case, θ was somewhat arbitrarily chosen to provide approximately equal lower bounds on α_θ and β_θ. For each choice of ka and of Ua^2, three choices were made for $\mathbf{e}_\theta$, and upper and lower bounds were calculated on the corresponding values of $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$. The three choices of $\mathbf{e}_\theta$ were

$$\mathbf{e}_\theta=\begin{pmatrix}1\\0\end{pmatrix}, \quad \begin{pmatrix}0\\1\end{pmatrix}, \quad \text{and} \quad \begin{pmatrix}1\\1\end{pmatrix}.$$

The corresponding values of $\mathbf{e}_\theta^\dagger\mathbf{B}_\theta\mathbf{e}_\theta$ are B_{11}, B_{22}, and $B_{11}+B_{22}+2B_{12}$. While it has been noted that three different values of $\mathbf{e}_\theta$ are required in order to obtain bounds on all three independent elements of $\mathbf{B}_\theta$, no criterion has been given for selecting these values. It is probably desirable to bound B_{11} and B_{22} separately but the remaining value of $\mathbf{e}_\theta$ could, in principle, be selected to provide the closest bounds on B_{12}. However, in the calculations presented here no attempt has been made to optimize the choice of $\mathbf{e}_\theta$.

The calculations were based on the trial function given by Eq. (54) with terms up to x^3, so that there were four linear variational parameters. Due to the two

TABLE I. Values of θ and lower bounds $\alpha_{\theta l}$ and $\beta_{\theta l}$ for three choices of ka and Ua^2, obtained from Figs. 3–6.

ka	Ua^2	θ	$\alpha_{\theta l}$	$\beta_{\theta l}$
0.1	0.01	3.08359	5.0	5.0
0.1	1.0	3.08359	4.0	4.0
$\pi/4$	$\pi^2/4$	2.775	3.3	3.3

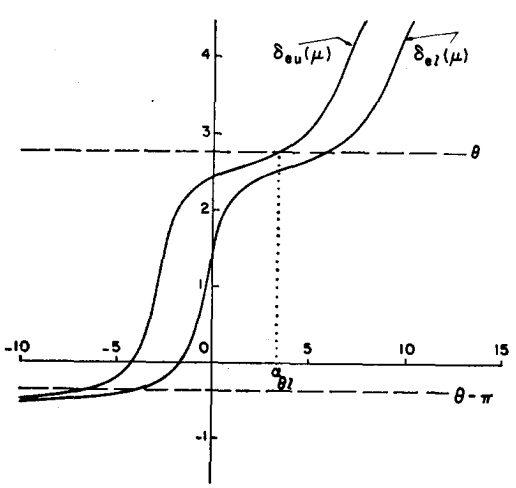


FIG. 5. Upper and lower bounds on $\delta_e(\mu)$ for $ka=\pi/4$, $Ua^2=\pi^2/4$.

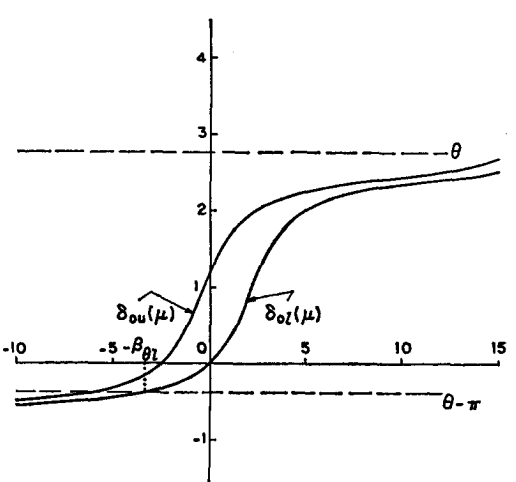


FIG. 6. Upper and lower bounds on $\delta_0(\mu)$ for $ka=\pi/4$, $Ua^2=\pi^2/4$.

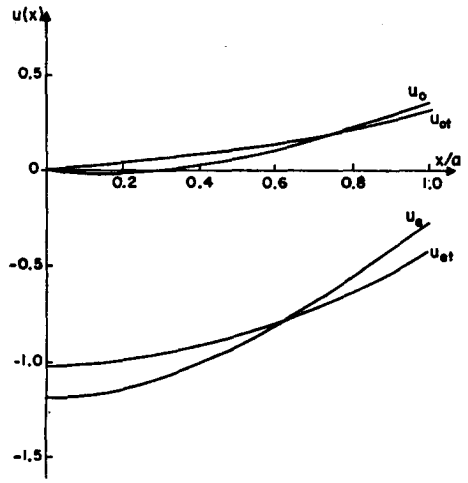


FIG. 7. Comparison of trial function with exact function. The trial function, which contains two free parameters, was used to calculate an upper bound on B_{11} for the case in which $ka=\pi/4$, $Ua^2=\pi^2/4$.

boundary conditions at $x=a$, this corresponds to two *free* parameters. The free parameters were varied to give the best bounds. The bounds obtained on the various combinations of the elements of $\mathbf{B}_\theta$ are listed in Table II. The bounds are seen to be very much closer in the case in which $ka=0.1$, $Ua^2=0.01$ than in the other two cases. This result is to be expected, for considering k to be fixed, this case involves a less effective potential than either of the other two. Note that for the second and third sets of values, the potential is by no means negligible as can be seen (Table IV) from the exact values of the phase shifts and the mixing parameters; for zero potentials, one would have $\eta_e=90°$, $\eta_0=0°$, and $\psi=0°$.

The effect of increasing the number of parameters is shown in Table III, which shows bounds on B_{11} for the case $ka=\pi/4$, $Ua^2=\pi^2/4$, calculated with 2, 4, and 6 free parameters. The bounds must improve as the number of free parameters is increased, and they do, but not as rapidly as might be expected. This result is attributed to the poor choice of the form of the trial function. The trial function used in calculating an upper bound on B_{11} with 2 parameters is compared with the exact function in Fig. 7.

Finally, bounds on the eigenphase shifts and mixing parameters are listed in Table IV. These quantities were calculated from the bounds on combinations of the elements of $\mathbf{B}_\theta$ listed in Table II, using Eqs. (52).

TABLE II. Upper and lower bounds on combinations of the elements of the susceptance matrix for three choices of ka and Ua^2 compared with the exact values. Two parameters were used in the calculations, and lower bounds on α_θ and β_θ were determined from Figs. 3–6.

Quantity	Upper bound	Lower bound	Exact value
$ka=0.1$, $Ua^2=0.01$			
B_{11}	0.1080314	0.1080083	0.1080210
B_{22}	−17.1726221	−17.1726309	−17.1726230
$B_{11}+B_{22}+2B_{12}$	−16.9785186	−16.9786503	−16.9786300
$ka=0.1$, $Ua^2=1.0$			
B_{11}	4.87112	4.48705	4.71216
B_{22}	−12.95737	−13.08907	−12.98583
$B_{11}+B_{22}+2B_{12}$	−0.64432	−1.14168	−0.85319
$ka=\pi/4$, $Ua^2=\pi^2/4$			
B_{11}	1.76488	1.23977	1.58228
B_{22}	−1.22869	−1.31261	−1.25519
$B_{11}+B_{22}+2B_{12}$	2.32112	1.75493	2.09091

TABLE III. Bounds on B_{11} for $ka=\pi/4$, $Ua^2=\pi^2/4$ for different numbers of parameters. The values of θ, $\alpha_{\theta l}$, and $\beta_{\theta l}$ are as shown in Table I.

Number of free parameters	Upper bound	Lower bound	Exact value
2	1.76488	1.23977	1.58228
4	1.69009	1.45289	1.58228
6	1.65803	1.49891	1.58228

It would be a trivial matter to also obtain bounds on the elements of $\mathbf{S}$, but the title of the paper notwithstanding, we have not bothered to do so.

From the foregoing results, it is clear that useful bounds can be calculated on the independent quantities which determine the scattering matrix in a multichannel problem, even with relatively extreme values for the depth and range of the potential. Furthermore, since diagonal comparison potentials can be used, the calculation requires no more than the exact solution of related single-channel problems. Thus it is anticipated that the method presented here can provide useful results in a number of multi-channel problems for which there would otherwise be no practical approach.

TABLE IV. Upper and lower bounds on the eigenphase shifts and mixing parameters for three choices of ka and Ua^2 compared with the exact values. These quantities were calculated from the bounds on the combinations of the elements of $\mathbf{B}_\theta$ listed in Table II, using Eqs. (52).

Quantity	Upper bound	Lower bound	Exact value
$ka=0.1$, $Ua^2=0.01$			
ψ	−0° 8′ 32.90″	−0° 8′ 33.29″	−0° 8′ 33.08″
η_e	92° 50′ 56″	92° 50′ 50″	92° 50′ 53″
η_0	0° 0′ 34.28″	0° 0′ 34.24″	0° 0′ 34.26″
$ka=0.1$, $Ua^2=1.0$			
ψ	−10° 34′	−12° 16′	−11° 23′
η_e	167° 12′	165° 7′	166° 18′
η_0	1° 0′	0° 41′	0° 51′
$ka=\pi/4$, $Ua^2=\pi^2/4$			
ψ	−10° 48′	−22° 4′	−15° 56′
η_e	137° 27′	116° 3′	130° 24′
η_0	27° 33′	1° 30′	11° 34′

ACKNOWLEDGMENTS

Special thanks are due to Dr. Martin Kelly and Dr. Leonard Rosenberg for a number of helpful discussions.

Application of the Singular Wave Function Operator in Scattering Theory. II

ROBERT A. MAPLETON
Geophysics Research Directorate, Air Force Cambridge Research Laboratories, Office of Aerospace Research, L. G. Hanscom Field, Bedford, Massachusetts

(Received July 18, 1961)

In a recent paper by this author, the singular wave function matrix of Møller was used to obtain the scattering solution for the nonrelativistic Coulomb operator. In the present paper this singular wave function operator is defined differently. This definition is deduced from the relation of the characteristic form of the Green's function to the branch cut in the continuous part of the spectrum of the Green's function. As a consequence of this interpretation, the calculation difficulties of the previous paper do not occur. This method is also used to obtain the scattering solution for the attractive square-well potential.

IN a recent paper the scattering solution of the three-dimensional nonrelativistic Coulomb operator was calculated from the following formal solution.[1]

$$\psi_E^{(\pm)} = \lim_{\epsilon \to 0} \frac{\pm i\epsilon}{E \pm i\epsilon - H} \phi_E. \tag{A}$$

In Eq. (A), E is the system energy, H is the Hamiltonian operator, ϕ_E is the incident plane wave, and the superscripts + and − refer to the diverging and converging wave solutions, respectively. In the position space representation, the relevant part of Eq. (A) is given by Eq. (B).

$$\psi^{(\pm)}(\mathbf{r}) = \pm i\epsilon \sum_{l=0}^{\infty} \int \frac{d\mathbf{k}\phi_{kl}(\mathbf{r})\phi_{kl}^*(\mathbf{r}')d\mathbf{r}' e^{i\mathbf{K}\cdot\mathbf{r}'}}{K^2 \pm i\epsilon - k^2}. \tag{B}$$

Since the discrete part of the spectrum is not needed, it has been omitted, as was discussed in I. In I, the convergence of the integral was forced by the introduction of the factor,

$$-\lim_{\lambda \to 0} \frac{\partial}{\partial \lambda} \frac{e^{-\lambda r}}{r},$$

under the integral sign. After the integration was performed, and the limit,

$$-\lim_{\lambda \to 0} (\partial/\partial\lambda),$$

evaluated, it was necessary to compare the relative largeness of two terms in order to determine which term should be retained. [See the discussion following Eq. (4aI).] In this paper, Eq. (B) is evaluated more straightforwardly. This more consistent calculation is suggested by the structure of the characteristic form of the Green's function[2] G and its relation to the eigenfunction expansion of G. The structure of this first form of G can be preserved in the second form of G by utilizing the relation of the wave boundary condition to the branch cut in the continuous part of the spectrum of G. This relation is used to rewrite Eq. (B) (exclusive of the $\mathbf{r}'$ integration) in a different form; then the quantity,

$$R^{(\pm)} = \pm i\epsilon G^{(\pm)},$$

is calculated by integration over the $\mathbf{k}$ variables of this new form. $R^{(\pm)}$ now explicitly contains the same sign of the imaginary component of the energy as the corresponding characteristic form of $G^{(\pm)}$. As a consequence of this fact, the r' integrals are convergent, and only that solution which is appropriate to a given boundary condition emerges; i.e., no argument such as the one following Eq. (4aI) is necessary. This method is best described by the calculation.

A. ATTRACTIVE COULOMB POTENTIAL

The quantities of Eq. (B) are written for convenient reference. There are some slight changes from the notation that was used in I.

$$\phi_{kl} = N(k)L_l(kr)P_l(\cos\theta_{r,k}),$$

$$\phi_{kl}^* = N^*(k)L_l^*(kr')P_l(\cos\theta_{k,r'}),$$

$$\alpha = (ka_0)^{-1}, \quad N(k) = \frac{1}{2\pi\Gamma(1-i\alpha)}\left[\frac{\alpha}{1-e^{-2\pi\alpha}}\right]^{\frac{1}{2}}.$$

The notation, $\theta_{a,b}$, signifies the angle between the vectors $\mathbf{a}$ and $\mathbf{b}$.

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

$$L_l(kr) = \frac{\Gamma(l+1-i\alpha)}{(2l)!}(i2kr)^l e^{ikr} \times {}_1F_1(l+1-i\alpha,\ 2l+2;\ -i2kr).$$

The confluent hypergeometric function of the variable r' is expressed in terms of the two linearly independent solutions of the confluent hypergeometric equation that

[1] Robert A. Mapleton, J. Math. Phys. 2, 482 (1961). Future reference to this paper is denoted by I. Equations of this reference are written as Eq. (xyI).

[2] Robert A. Mapleton, J. Math. Phys. 2, 478 (1961). Future reference to this paper is denoted by II. Equations of this reference are written as Eq. (xyII).

are used in II.

$$ {}_1F_1(a,c;z)=W_1(a,c;z)+W_2(a,c;z),$$

$$a=l+1+i\alpha,\quad c=2l+2;\quad z=i2kr,$$

$$W_1=\frac{(2l+1)!\,e^z}{2\pi i}\int_{-\infty}^{(0+)} dt e^t\, t^{-a}_{(-\pi,\pi)}\,(t+z)^{(a-c)}_{(\pi,\pi)},$$

$$W_2=\frac{(2l+1)!}{2\pi i}\int_{-\infty}^{(0+)} dt e^t\, t^{(a-c)}_{(-\pi,\pi)}\,(t-z)^{-a}_{(-\pi,-\pi)}.$$

The initial and terminal values of the phases are placed below the corresponding factors of the integrand. The functions $W_{1,2}$ are defined so that the following behavior is exhibited:

$$r\to\infty$$

$$r^l e^{-ikr}W_1 \to \text{diverging};$$

$$r^l e^{-ikr}W_2 \to \text{converging}.$$

The preceding relations are used to define the functions, $L_l^{(\pm)}$.

$$L_l^*=L_l^{(+)}+L_l^{(-)}.$$

In this expression the superscripts have the previously assigned meaning. The procedure of II is used to establish the two identities.

$$W_1(ke^{\pi i})=W_2(k)e^{(2\pi\alpha-z)},$$
$$W_2(ke^{-\pi i})=W_1(k)e^{(2\pi\alpha-z)}.$$

Now, with the aid of another identity,[3]

$$ {}_1F_1(a,c;z)e^{-z/2}={}_1F_1(c-a,c;-z)e^{z/2},$$
$$z=-i2kr,\quad a=l+1-i\alpha,\quad c=2l+2,$$

it is readily verified that the two additional relations are valid.

$$k'=ke^{\pi i},$$
$$|N(k')|^2L_l(k'r)L_l^{(+)}(k'r')=|N(k)|^2L_l(kr)L_l^{(-)}(kr');$$
$$k'=ke^{-\pi i},$$
$$|N(k')|^2L_l(k'r)L_l^{(-)}(k'r')=|N(k)|^2L_l(kr)L_l^{(+)}(kr').$$

With this information and Eq. (B), the quantity, $R^{(\pm)}$, is defined by Eq. (1a) below.

$$R^{(\pm)}=\pm 4\pi i\epsilon\sum_{l=0}^{\infty}\frac{P_l(\cos\theta)}{2l+1}\times\int_{-\infty}^{\infty}\frac{dkk^2|N(k)|^2}{K^2\pm i\epsilon-k^2}L_l(kr)L_l^{(\pm)}(kr'),$$

$$\theta=\theta_{r,r'}. \tag{1a}$$

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

In the derivation of Eq. (1a) the integration over the solid angle of $\mathbf{k}$ space was effected. It is easily demonstrated that $R^{(\pm)}$ is still symmetric in r and r'. The k integral of Eq. (1a) must be interpreted correctly since the negative real k axis is a branch cut with the phases defined by the following scheme.

$$k<0,\quad \arg k=\begin{cases}\pi, & \text{above cut}\\ -\pi, & \text{below cut}\end{cases}$$

There is no contribution to the k integral from the branch point at

$$k=0,$$

and this fact can be verified by using Taylor's results[4] for the Whittaker functions which are related to $W_{1,2}$. It should be observed that this phase scheme is consistent with the restrictions of II which are quoted for reference.

$$\text{diverging wave},\qquad 0<\arg k<\pi,$$
$$\text{converging wave},\qquad -\pi<\arg k<0.$$

The k integration is performed by utilizing the smallness of ϵ. [See the discussion following Eq. (4aI) which relates to this.] With this integration effected, Eq. (1a) is given by Eq. (1b).

$$R^{(\pm)}=\pm 4\pi^2\epsilon\sum_{l=0}^{\infty}\frac{P_l(\cos\theta)}{2l+1}\gamma|N(\gamma)|^2L_l(\gamma r)L_l^{(\pm)}(\gamma r'),$$

$$\gamma^{(\pm)}=(K^2\pm i\epsilon)^{\frac{1}{2}}. \tag{1b}$$

In Eq. (1b), it is understood that γ has the same superscript as $R^{(\pm)}$. This function, $R^{(\pm)}$, which projects the incident state into the scattering solution, contains the same value of γ as the corresponding characteristic form of the Green's function $G^{(\pm)}$.

Attention is directed to the fact that Eq. (B) is obtained by calculating the position space representative of Eq. (A) using Dirac transformation methods.[5,6] However, it is the relation between the two forms of the Green's function that provides the clue to transforming the expansion of Eq. (B) into the form given by Eq. (1b). Thus, without this knowledge, the procedure for the calculation of Eq. (A) is not well defined, as is amply illustrated in I. However, the procedure of calculation is unambiguous once Eq. (A) is interpreted by means of Eq. (1b). The calculation of $\psi^{(\pm)}$ is now continued.

The plane wave $\exp(i\mathbf{q}\cdot\mathbf{r})$ is represented by the Rayleigh expansion as in I, and the angular integrations

[4] W. C. Taylor, J. Math. and Phys. **18**, 34 (1939).

[5] P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, New York, 1947), 3rd edition.

[6] B. A. Lippman and Julian Schwinger, Phys. Rev. **79**, 469 (1950). [See Eq. (2.32).]

are effected. The result is given by Eq. (1c).

$$\psi^{(\pm)}=16\pi^3\epsilon\sum_{l=0}^{\infty}\frac{i^lP_l(\cos\theta)}{2l+1}\gamma|N(\gamma)|^2 \times L_l(\gamma r)\int_0^\infty dr r^2 L_l^{(\pm)}(\gamma r)j_l(qr),$$

$$j_l=\left(\frac{\pi}{2qr}\right)^{\frac{1}{2}}J_{l+\frac{1}{2}}(qr),\quad \theta=\theta_{r,K}. \tag{1c}$$

The next step in the calculation is the transformation of the integrand of Eq. (1c) by the use of the two differential equations labeled by (L) and (J).

$$\frac{d}{dr}\left[r^2\frac{d}{dr}L_l^{(\pm)}(\gamma r)\right]+[\gamma^2r^2+2\alpha\gamma r-l(l+1)]L_l^{(\pm)}(\gamma r)=0. \tag{L}$$

$$\frac{d}{dr}\left[r^2\frac{d}{dr}j_l(qr)\right]+[q^2r^2-l(l+1)]j_l(qr)=0. \tag{J}$$

Multiply (L) by j_l and (J) by $L_l^{(\pm)}$; subtract these two equations, and divide the result by

$$(\gamma^2-q^2);$$

integrate this relation over the interval $(0,\infty)$, and rearrange the terms to obtain the following two relations.

$$\int_0^\infty dr r^2 L_l^{(\pm)}(\gamma r)j_l(qr) = -\frac{2\alpha\gamma}{\gamma^2-q^2}\int_0^\infty dr r L_l^{(\pm)}(\gamma r)j_l(qr) + \lim_{r\to 0}\frac{r^2}{\gamma^2-q^2}\left[j_l\frac{d}{dr}L_l^{(\pm)}-L_l^{(\pm)}\frac{d}{dr}j_l\right].$$

The integrated terms vanish at the upper limit in accordance with the discussion following Eq. (1b). This last relation is used to get Eq. (1d).

$$\psi^{(\pm)}=\pm\frac{16\pi^3\epsilon\gamma|N(\gamma)|^2}{\gamma^2-q^2}\sum_{l=0}^{\infty}\frac{i^lP_l(\cos\theta)}{2l+1} \times L_l(\gamma r)\left(-2\alpha\gamma\int_0^\infty dr r L_l^{(\pm)}j_l + \lim_{r\to 0}r^2\left[j_l\frac{d}{dr}L_l^{(\pm)}-L_l^{(\pm)}\frac{d}{dr}j_l\right]\right). \tag{1d}$$

The term which is evaluated at the origin is omitted until later in the calculation. The relations between $W_{1,2}$ and the Whittaker functions are written for reference.[7,8]

$$W_1(a,c;z)=\frac{(2l+1)!}{\Gamma(l+1+i\alpha)}e^{i[\gamma r-\pi(l+1-i\alpha)]}\times(-i2\gamma r)^{-(l+1)}W_{-\mu,m}(-z)$$
$$=\frac{(2l+1)!}{\Gamma(l+1+i\alpha)}\Psi(c-a,c;-z)e^{i[2\gamma r-\pi(l+1-i\alpha)]},$$

$$W_2(a,c;z)=\frac{(2l+1)!}{\Gamma(l+1-i\alpha)}e^{i[\gamma r+\pi(l+1+i\alpha)]}\times(i2\gamma r)^{-(l+1)}W_{\mu,m}(z)$$
$$=\frac{(2l+1)!}{\Gamma(l+1-i\alpha)}\Psi(a,c;z)e^{i\pi(l+1+i\alpha)},$$

$$W_{\mu,-m}(z)=W_{\mu,m}(z),\quad a=l+1+i\alpha,$$
$$c=2l+2,\quad z=i2\gamma r,\quad \mu=-i\alpha,\quad m=l+\tfrac{1}{2}.$$

These relations are consistent with the expression which is given below.[9]

$${}_1F_1(a,c;z)=W_1+W_2$$
$$=(2l+1)!\left\{\frac{e^{i[2\gamma r-\pi(l+1-i\alpha)]}}{\Gamma(l+1+i\alpha)}\Psi(c-a,c;-z) + \frac{e^{i\pi(l+1+i\alpha)}}{\Gamma(l+1-i\alpha)}\Psi(a,c;z)\right\}.$$

There is the following integral representation[10] for the Whittaker function.

$$W_{-\mu,m}(-z)=(-i2\gamma r)^{(l+1)}e^{i\gamma r}\times\int_0^\infty dt t^{(l-i\alpha)}(1+t)^{(l+i\alpha)}e^{i2\gamma rt},$$
$$\mathrm{Re}(l+\tfrac{1}{2}-i\alpha)>-\tfrac{1}{2},\quad \mathrm{Im}\gamma>0,$$
$$\mathrm{Re}(i2\gamma r)<0.$$

(Re signifies real part and Im imaginary part.) This leads to the integral representation for W_1.

$$W_1(a,c;z)=\frac{(2l+1)!}{|\Gamma(l+1+i\alpha)|^2}e^{i[2\gamma r-\pi(l+1-i\alpha)]}\times\int_0^\infty dt t^{(l-i\alpha)}(1+t)^{(l+i\alpha)}e^{i2\gamma rt}.$$

[7] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (The MacMillan Company, New York, 1946), Chapt. 16, p. 339.

[8] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 264, Eq. (4).

[9] Reference 8, p. 259, Eq. (7).

[10] Reference 8, p. 274, Eq. (18).

A similar expression is derived for W_2.

$$W_2(a,c;z)=\frac{(2l+1)!}{|\Gamma(l+1+i\alpha)|^2}e^{i\pi(l+1+i\alpha)}\times\int_0^\infty dt t^{(l+i\alpha)}(1+t)^{(l-i\alpha)}e^{-i2\gamma rt},$$

$$\mathrm{Re}(l+\tfrac{1}{2}+i\alpha)>-\tfrac{1}{2},\quad \mathrm{Im}\gamma<0,$$

$$\mathrm{Re}(-i2\gamma r)<0.$$

The calculation of $\psi^{(+)}$ is effected first.

The factors of $L_l^{(+)}(\gamma r')$ that are independent of r' are absorbed into a factor F. An integral that was used in I is written for reference.

$$\int_0^\infty dr r^l j_l(qr)e^{-ar}=\frac{(2q)^l l!}{[(\lambda-iK)^2(1+2t)^2+q^2]^{(l+1)}},$$

$$\gamma\approx K+i\epsilon/2K=K+i\lambda,\quad a=(\lambda-iK)(1+2t).$$

Let $\psi_1^{(+)}$ be that part of Eq. (1d) exclusive of the contribution from the origin. This function is given by Eq. (2a).

$$\psi_1^{(+)}=FT,\quad a=q/\gamma,$$

$$F=-\frac{32\pi^3\epsilon\alpha|N(\gamma)|^2e^{-\pi\alpha}}{\gamma^2-q^2}\sum_{l=0}^\infty\frac{l!(4a)^l}{\Gamma(l+1-i\alpha)}\times P_l(\cos\theta)L_l(\gamma r),$$

$$T=\int_0^\infty\frac{dt t^{(l-i\alpha)}(1+t)^{(l+i\alpha)}}{[(1+2t)^2-a^2]^{(l+1)}}. \tag{2a}$$

In the derivation of Eq. (2a) the relation,

$$\lambda-iK=-i(K+i\lambda)\approx-i\gamma,$$

was used. The change of variable $t=v/(1-v)$ is made, and T of Eq. (2a) is given by Eq. (2b).

$$T=(1-a^2)^{-(l+1)}\times\int_0^1\frac{dt t^{(l-i\alpha)}}{\{[t+(1-a)/(1+a)][t+(1+a)/(1-a)]\}^{(l+1)}}. \tag{2b}$$

The factor of the integrand,

$$[t+(1+a)/(1-a)]^{-(l+1)},$$

is now expanded in the convergent power series.

$$\left(t+\frac{1+a}{1-a}\right)^{-(l+1)}=\left(\frac{1-a}{1+a}\right)^{(l+1)}\sum_{n=0}^\infty\frac{(-1)^n(n+l)!}{n!l!}\times\left(\frac{1-a}{1+a}\right)^n t^n.$$

This expansion is used to rewrite Eq. (2b) as Eq. (2c).

$$T=(1+a)^{-2(l+1)}\sum_{n=0}^\infty\frac{(-1)^n(l+n)!}{l!n!}\left(\frac{1-a}{1+a}\right)^n+\int_0^1\frac{dt t^{(l+n-i\alpha)}}{[t+(1-a)/(1+a)]^{(l+1)}}. \tag{2c}$$

The integral in Eq. (2c) is first evaluated subject to the restriction $|a|<1$. A contour integral is constructed by joining the end points of the loop integral,

$$\int_1^{(0-)}dt$$

to the integral over the unit circle $|t|=1$. The phase of t above and below the branch cut $0<t<1$ is selected according to the following scheme.

$$\arg t=\begin{cases}-2\pi, & \text{above cut}\\ 0, & \text{below cut}\end{cases}$$

With this specification of the phase of t, it is evident that

$$\lim_{\epsilon\to0}\arg[-(1-a)/(1+a)]\to-\pi.$$

The Cauchy' residue theorem is applied to this closed contour integral to give the following result.

$$\int_0^1\frac{dt t^{(l+n-i\alpha)}}{[t+(1-a)/(1+a)]^{(l+1)}}=\frac{2\pi i\Gamma(l+n+1-i\alpha)}{l!\Gamma(n+1-i\alpha)(1-e^{-2\pi\alpha})}\left(\frac{1-a}{1+a}e^{-\pi i}\right)^{(n-i\alpha)}-i(e^{-2\pi\alpha}-1)^{-1}\int_{-2\pi}^0\frac{d\phi e^{i(n-i\alpha)\phi}}{[1+(1-a)/(1+a)e^{-i\phi}]^{(l+1)}}.$$

The integrand of the remaining integral is expanded in a convergent power series, and the order of summation and integration is interchanged as before. After some simplification, the result can be written as Eq. (2d).

$$\int_0^1\frac{dt t^{(l+n-i\alpha)}}{[t+(1-a)/(1+a)]^{(l+1)}}=\frac{2\pi i\Gamma(l+n+1-i\alpha)(-1)^n e^{-\pi\alpha}}{l!\Gamma(n+1-i\alpha)(e^{-2\pi\alpha}-1)}\left(\frac{1-a}{1+a}\right)^{(n-i\alpha)}+\sum_{m=0}^\infty\frac{(-1)^m(l+m)!}{l!m!(n-m-i\alpha)}\left(\frac{1-a}{1+a}\right)^m. \tag{2d}$$

Equation (2d) and the two relations[11] listed below are

[11] Reference 3, pp. 1 and 8.

used to rewrite Eq. (2c) as Eq. (2e).

$$\frac{2\pi i}{e^{-2\pi\alpha}-1}=\frac{\pi e^{\pi\alpha}}{\sin(\pi i\alpha)}=\Gamma(i\alpha)\Gamma(1-i\alpha)e^{\pi\alpha},$$

$${}_2F_1(a,b,c;z)=(1-z)^{-b}{}_2F_1(b,c-a,c;z/[z-1]).$$

$$T=T_1^{(+)}+T',$$

$$T_1^{(+)}=\frac{\Gamma(i\alpha)\Gamma(l+1-i\alpha)}{l!(4a)^{(l+1)}}\left(\frac{1+a}{1-a}\right)^{i\alpha}\times{}_2F_1(-l,l+1,1-i\alpha;-(1-a)^2/4a),$$

$$T'=2^{-2(l+1)}\sum_{m=0}^{\infty}\frac{(-1)^m(l+m)!}{l!m!(-m-i\alpha)}\left(\frac{1-a}{1+a}\right)^m\times{}_2F_1(l,l+1,1-m-i\alpha;\tfrac{1}{2}(1-a)). \quad (2e)$$

The contribution from the term of Eq. (1d) that is evaluated at the origin is now calculated. For this purpose, $L_l^{(+)}$ is expressed in terms of

$$\Psi(c-a,c;-z),$$

$$L_l^{(+)}=(2l+1)e^{i[\gamma r-\pi(l+1-i\alpha)]}(-i2\gamma r)^l\Psi(c-a,c;-z).$$

The leading term of Ψ is[12]

$$\Psi\to\frac{(2l)!}{\Gamma(l+1-i\alpha)(-i2\gamma r)^{(2l+1)}},$$

and the leading term of $j_l(qr)$ is[13]

$$j_l\to\frac{(2qr)^l l!}{(2l+1)!}.$$

This information is used to obtain the quantity C for the contribution from the origin.

$$C=\lim_{r\to 0}r^2[j_l(d/dr)L_l^{(+)}-L_l^{(+)}(d/dr)j_l]$$

$$=-\frac{(2l+1)l!i^{-(l+1)}q^l e^{-\pi\alpha}}{2\Gamma(l+1-i\alpha)\gamma^{(l+1)}}.$$

After multiplication of C by the remaining factors of Eq. (1d), the result can be written as $F/2^{2(l+1)}(i\alpha)$. This term, exclusive of the factor F, can be combined with T' of Eq. (2e) to define $T_2^{(+)}$.

$$T_2^{(+)}=2^{-2(l+1)}\sum_{m=0}^{\infty}\frac{(-1)^m(l+m)!}{l!m!(-m-i\alpha)}\left(\frac{1-a}{1+a}\right)^m\times[{}_2F_1(1,l+1,1-m-i\alpha;\tfrac{1}{2}[1-a])-\delta_{m0}].$$

[12] Reference 8, p. 261, Eq. (13).

[13] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc. New York, 1953), Vol. II, pp. 1573–1574.

The solution $\psi^{(+)}$, is written as Eq. (2f).

$$\psi^{(+)}=\lim_{\epsilon\to 0}\lim_{q\to K}F(T_1^{(+)}+T_2^{(+)}),\quad q<K. \quad (2f)$$

It is evident that $T_2^{(+)}$ does not contribute to $\psi^{(+)}$ since each term of $T_2^{(+)}$ contains an integral power of $1-a$ as a factor, which factor vanishes in the preceding limit. It is now a simple matter to obtain Eq. (2g) as the final result.

$$\psi^{(+)}=\Gamma(1+i\alpha)\left(\frac{4}{\beta}\right)^{i\alpha}\sum_{l=0}^{\infty}P_l(\cos\theta)L_l(Kr)$$

$$=\frac{\pi\alpha}{\sinh(\pi\alpha)}\left(\frac{4}{\beta}\right)^{i\alpha}{}_2F_1(i\alpha,1;i[Kr-\mathbf{K}\cdot\mathbf{r}])e^{i\mathbf{K}\cdot\mathbf{r}}, \quad (2g)$$

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

This agrees with Eq. (4bI).

The derivation of the solution for the case that

$$|a|>1.$$

is sketched. In the power series expansion of Eq. (2c) the relevant factor is rewritten as

$$(-1)^n\left(\frac{1-a}{1+a}\right)^n=\left(\frac{a-1}{a+1}\right)^n.$$

At the pole,

$$t=(a-1)/(a+1),$$

the factor $e^{-i\pi}$ is not needed, since for real γ this pole approaches the t axis from below in the interval

$$0<t<1.$$

These changes alter $T_1^{(+)}$ of Eq. (2e) by the additional factor,

$$e^{\pi\alpha}.$$

Since $T_2^{(+)}$ vanishes in the limit, $a\to 1$, the alterations of $T_2^{(+)}$ are unimportant. It is now evident that the result is

$$e^{\pi\alpha}\psi^{(+)},\quad (\beta=1-K^2/q^2)$$

with $\psi^{(+)}$ defined by Eq. (2g).

The converging wave solution can be calculated in the same manner. The essential difference is that W_2 is used in place of W_1. One readily derives the following expression for $T^{(-)}$.

$$T^{(-)}=T_1^{(-)}+T_2^{(-)},\quad |a|<1,$$

$$T_1^{(-)}=\frac{\Gamma(-i\alpha)\Gamma(l+1+i\alpha)}{l!(4a)^{(l+1)}}\left(\frac{1+a}{1-a}\right)^{-i\alpha}\times{}_2F_1(-l,l+1,1+i\alpha;-(1-a)^2/4a),$$

$$T_2^{(-)}=-2^{-2(l+1)}\sum_{m=0}^{\infty}\frac{(-1)^m(l+m)!}{l!m!(-m-i\alpha)}\left(\frac{1-a}{1+a}\right)^m\times[{}_2F_1(1,l+1,1-m-i\alpha;\tfrac{1}{2}[1-a])-\delta_{m0}].$$

If the limiting process, $a \to 1$, is performed, the quantity $FT^{(-)}$ can be written as Eq. (4cI). Another point of interest is the following relation.

$$T_2^{(+)}+T_2^{(-)}=0, \quad \mathrm{Im}\gamma=0.$$

Thus, if both signs of ϵ are admitted in the factor F, the function

$$F(T_1^{(+)}+T_1^{(-)})$$

agrees in form with Eq. (4aI). Another example of the application of Eq. (A) to obtain the scattering solution is presented in Sec. B.

B. ATTRACTIVE SQUARE-WELL POTENTIAL

The differential equation for the Green's function of this scattering problem is

$$\{\nabla^2+U_0[\eta(r)-\eta(r-a)]+K^2\}G=\delta(\mathbf{r}-\mathbf{r}'),$$

and the radial part of this equation is

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d}{dr}G_l\right)+\left\{U_0[\eta(r)-\eta(r-a)]-\frac{l(l+1)}{r^2}+K^2\right\}G_l=\frac{\delta(r-r')}{r^2},$$

$$U_0=2V_0/a_0, \quad K^2=(2m/\hbar^2)E.$$

η is the Heaviside unit step function, V_0 is the depth of the well, and E is the system energy. For each region, $r \gtrless a$, the Green's function is given by[14]

$$G=\sum_{l=0}^{\infty}\frac{(2l+1)}{4\pi}P_l(\cos\theta_{r,r'})G_l(r,r').$$

The Green's function, G_l, is now constructed for the diverging wave case.[15,16]

$$r<a$$

$$\begin{aligned}
&G_l=-iKU_i(r>)V_i(r<),\\
&U_i=C_l^{(+)}j_l(\alpha r)+D_l^{(+)}h_l^{(1)}(\alpha r),\\
&V_i=A_l^{(+)}j_l(\alpha r), \quad \alpha^2=K^2+U_0,\\
&A_l^{(+)}=[h_l^{(1)}(Ka)J_l(K,\alpha)-j_l(Ka)H_l^{(+)}(K,\alpha)]R_1^{-1},\\
&C_l^{(+)}=[h_l^{(1)}(\alpha a)H_l^{(+)}(K,\alpha)-h_l^{(1)}(Ka)H_l^{(+)}(\alpha,\alpha)]R_2^{-1},\\
&D_l^{(+)}=[h_l^{(1)}(Ka)-j_l(\alpha a)H_l^{(+)}(K,\alpha)]R_2^{-1},\\
&R_1=h_l^{(1)}(Ka)-j_l(\alpha a)H_l^{(+)}(K,\alpha),\\
&R_2=h_l^{(1)}(\alpha a)-j_l(\alpha a)H_l^{(+)}(\alpha,\alpha),\\
&J_l(K,\alpha)=Kj_l'(Ka)/\alpha j_l'(\alpha a),\\
&H_l^{(+)}(K,\alpha)=Kh_l^{(1)\prime}(Ka)/\alpha j_l'(\alpha a).
\end{aligned} \tag{3a}$$

$$r>a$$

$$\begin{aligned}
&G_l=-iKU_e(r>)V_e(r<),\\
&U_e=h_l^{(1)}(Kr),\\
&V_e=j_l(Kr)+B_l^{(+)}h_l^{(1)}(Kr),\\
&B_l^{(+)}=[j_l(\alpha a)J_l(K,\alpha)-j_l(Ka)]R_1^{-1}.
\end{aligned} \tag{3b}$$

In Eqs. (3a) and (3b), $h_l^{(1)}$ and j_l are the spherical Hankel functions of the first kind and the spherical Bessel functions, respectively. The superscript (+) signifies that $h_l^{(1)}$ is used, and the primes indicate differentiation with respect to the argument. The eigenfunction expansion form of G is now derived for the region, $r>a$. (Only this region is treated since the results will suffice to support the conjectures of this paper.) It is emphasized that the Green's function in question requires that $\mathrm{Im}K>0$.

One proceeds as in II. The substitution, $K=(\lambda)^{\frac{1}{2}}$, is made, and the following contour integral is formed on a circle at infinity in the complex λ plane.

$$\frac{1}{2\pi i}\oint\frac{d\lambda G_l^{(+)}(\lambda)}{\lambda-\gamma^2}, \quad \gamma^2=K^2+i\epsilon,$$

$$0<\arg\lambda<2\pi, \quad r>r'.$$

It is not difficult to show that this integral vanishes. Then, corresponding to Eq. (2bII), the following result is derived.

$$G_l(\gamma)=\frac{1}{2\pi i}\int_{C_1}\frac{d\lambda G_l^{(+)}(\lambda)}{\gamma^2-\lambda}+R$$

R is the bound state contribution from the poles on the negative λ axis. This term is omitted as in Sec. A. C_1 is a contour that encircles the branch cut, $0<\lambda<\infty$. The leading terms of the Hankel and Bessel function at the origin are[13]

$$\lambda=0, \quad j_l \propto (\lambda)^{l/2}, \quad h_l^{(1)} \propto (\lambda)^{-(l+1)/2}.$$

With this information, one demonstrates straightforwardly that there is no contribution from the branch point at $\lambda=0$. The contour integral on the path C_1 can now be written as the sum of two integrals.

$$\begin{aligned}
&\frac{1}{2\pi i}\int_{C_1}\frac{d\lambda G_l^{(+)}(\lambda)}{\gamma^2-\lambda}\\
&=\frac{-1}{2\pi}\left\{\int_{\infty}^{0}\frac{d\lambda\lambda^{\frac{1}{2}}}{\gamma^2-\lambda}h_l^{(1)}(\lambda^{\frac{1}{2}}r)[j_l(\lambda^{\frac{1}{2}}r')\right.\\
&\quad+B_l^{(+)}(\lambda)h_l^{(1)}(\lambda^{\frac{1}{2}}r')]-\int_0^{\infty}\frac{d\lambda\lambda^{\frac{1}{2}}}{\gamma^2-\lambda}h_l^{(1)}(\lambda^{\frac{1}{2}}e^{\pi i}r)\\
&\quad\left.\times[j_l(\lambda^{\frac{1}{2}}e^{\pi i}r')+B_l^{(+)}(\lambda e^{2\pi i})h_l^{(1)}(\lambda^{\frac{1}{2}}e^{\pi i}r')]\right\}.
\end{aligned}$$

[14] E. C. Titchmarsh, *Eigenfunction Expansion Associated With Second Order Differential Equations* (Oxford University Press, New York, 1958), Vol. II, Chapter XV.
[15] Reference 13, Vol. I, p. 826.
[16] Nathan Marcuvitz, Communs. Pure and Appl. Math. 4, 263 (1951).

For real values of z, the following relations are used to simplify the last integral.[13]

$$j_l(ze^{\pm i\pi})=e^{\pm i\pi l}j_l(z),$$
$$n_l(ze^{\pm i\pi})=-e^{\mp i\pi l}n_l(z),$$
$$h_l^{(1)}(ze^{i\pi})=e^{i\pi l}h_l^{(2)}(z),$$
$$h_l^{(2)}(ze^{-i\pi})=e^{-i\pi l}h_l^{(1)}(z),$$
$$h_l^{(\frac{1}{2})}(z)=j_l(z)\pm in_l(z).$$

With the aid of this list of formulas, $G_l^{(+)}$ can be written as Eq. (3c).

$$G_l^{(+)}(\gamma)=\frac{1}{2\pi}\int_0^\infty \frac{d\lambda\lambda^{\frac{1}{2}}}{\gamma^2-\lambda}\{2j_l(\lambda^{\frac{1}{2}}r)j_l(\lambda^{\frac{1}{2}}r')$$
$$+B_l^{(+)}h_l^{(1)}(\lambda^{\frac{1}{2}}r)h_l^{(1)}(\lambda^{\frac{1}{2}}r')$$
$$+B_l^{(-)}h_l^{(2)}(\lambda^{\frac{1}{2}}r)h_l^{(2)}(\lambda^{\frac{1}{2}}r')\},$$
$$B_l^{(+)}(\lambda e^{i2\pi})=B_l^{(-1)}(\lambda). \quad (3c)$$

The quantity $B_l^{(-)}$ is derived from $B_l^{(+)}$ by replacing $h_l^{(1)}$ by $h_l^{(2)}$. If this calculation is repeated for the converging wave case,

$$-2\pi<\arg\lambda<0,$$

the same result is obtained, apart from the sign that precedes $i\epsilon$. Furthermore, since the expression in Eq. (3c) is symmetric in r and r', Eq. (3c) is valid for $r\gtrless r'$. The relation of B_l to the phase shifts δ_l, is now demonstrated.

$$D^{(+)}(\lambda)=D_2+iD_1,$$
$$D_2=\alpha j_l'(\alpha a)j_l(\lambda^{\frac{1}{2}}a)-\lambda^{\frac{1}{2}}j_l(\alpha a)j_l'(\lambda^{\frac{1}{2}}a),$$
$$D_1=\alpha j_l'(\alpha a)n_l(\lambda^{\frac{1}{2}}a)-\lambda^{\frac{1}{2}}j_l(\alpha a)n_l'(\lambda^{\frac{1}{2}}a),$$
$$B_l^{(+)}(\lambda)=\frac{-D_2}{D_2+iD_1}=\frac{iD_2(D_1+iD_2)}{|D^{(+)}|^2},$$
$$\sin\delta_l=D_2|D^{(+)}|^{-1},\quad \cos\delta_l=D_1|D^{(+)}|^{-1},$$
$$B_l^{(+)}=i\sin\delta_l e^{i\delta_l},$$
$$B_l^{(-)}=-i\sin\delta_l e^{-i\delta_l}.$$

These are the conventional phase shifts.[17] The change of variable $\lambda=k^2$ is made; the angular part of G is recovered, and some algebraic manipulation is performed to obtain Eq. (3d) for $R^{(\pm)}$.

$$R^{(\pm)}(\gamma)=\pm i\epsilon G^{(\pm)}(\gamma)$$
$$=\pm i\epsilon\sum_{l=0}^\infty\frac{2l+1}{4\pi}P_l(\cos\theta_{r,r'})\frac{2}{\pi}\int_0^\infty\frac{dkk^2}{\gamma^2-k^2}$$
$$\times[j_l(kr)\cos\delta_l-n_l(kr)\sin\delta_l]$$
$$\times[j_l(kr')\cos\delta_l-n_l(kr')\sin\delta_l]. \quad (3d)$$

[17] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc. New York, 1949), Chap. 5.

The sign of the imaginary part of γ is defined as in Sec. A. Again, as in Sec. A, the following two relations are easily verified.

$$k'=ke^{i\pi},$$
$$[j_l(kr)\cos\delta_l(k)-n_l(kr)\sin\delta_l(k)]h_l^{(2)}(kr')e^{-i\delta_l(k)}$$
$$=[j_l(k'r)\cos\delta_l(k')-n_l(k'r)\sin\delta_l(k')]h_l^{(1)}(k'r')e^{i\delta_l(k')};$$
$$k'=ke^{-i\pi},$$
$$[j_l(kr)\cos\delta_l(k)-n_l(kr)\sin\delta_l(k)]h_l^{(1)}(kr')e^{i\delta_l(k)}$$
$$=[j_l(k'r)\cos\delta_l(k')-n_l(k'r)\sin\delta_l(k')]h_l^{(2)}(k'r')e^{-i\delta_l(k')}.$$

The k integration now covers the entire real k axis with the phases above and below the branch cut, $-\infty<k<0$, defined as in Sec. A. The smallness of ϵ is exploited and the integration is performed. Equation (3e) displays the final result for $R^{(\pm)}$.

$$R^{(\pm)}=\pm\epsilon\gamma\sum_{l=0}^\infty\frac{2l+1}{4\pi}P_l(\cos\theta_{r,r'})$$
$$\times[j_l(\gamma r)\cos\delta_l(\gamma)-n_l(\gamma r)\sin\delta_l(\gamma)]$$
$$\times h_l^{(\pm)}(\gamma r')e^{\pm i\delta_l(\gamma)},\quad h_l^{(\pm)}=h_l^{(\frac{1}{2})}. \quad (3e)$$

The calculation of $\psi^{(+)}$ is easily accomplished, and the result for $\psi^{(+)}$ is presented first. The plane wave

$$\exp(i\mathbf{K}\cdot\mathbf{r}')$$

is represented by the Rayleigh expansion, and the integration over the solid angle of r' space is effected. The radial integral $T^{(\pm)}$ is given by[13]

$$T^{(\pm)}=\int_a^\infty drr^2h_l^{(1)}(\gamma r)j_l(Kr)$$
$$=\frac{a^2}{\gamma^2-K^2}[\gamma h_{(l+1)}^{(1)}(\gamma a)j_l(Ka)$$
$$-Kh_l^{(1)}(\gamma a)j_{(l+1)}(Ka)].$$

The upper limit vanishes by reason of the condition,

$$\mathrm{Im}\gamma>0.$$

The quantities, γ, $h_{(l+1)}^{(1)}$ and $h_l^{(1)}$ are expanded in a Taylor's series about the point, K, and the terms that contain ϵ are discarded in view of the limiting process $\epsilon\to 0$ which occurs subsequently. The relation[13]

$$n_{(l+1)}(z)j_l(z)-n_l(z)j_{(l+1)}(z)=-z^{-2}$$

is then used so that the expression for $T^{(+)}$ becomes

$$T^{(+)}=\frac{i}{\gamma(\gamma^2-K^2)}=\frac{1}{\gamma\epsilon}.$$

All of these results are used, and the limit process $\epsilon\to 0$

is invoked; the result is Eq. (4a).

$$\psi^{(+)}=\sum_{l=0}^{\infty}(2l+1)i^lP_l(\cos\theta)[j_l(Kr)\cos\delta_l-n_l(Kr)\times\sin\delta_l]e^{i\delta_l},\quad\theta=\theta_{r,K}. \tag{4a}$$

Similarly, the converging wave solution is given by Eq. (4b).

$$\psi^{(-)}=\sum_{l=0}^{\infty}(2l+1)i^lP_l(\cos\theta)[j_l(Kr)\times\cos\delta_l-n_l(Kr)\sin\delta_l]e^{-i\delta_l}. \tag{4b}$$

These last two results and the results of Sec. A support the interpretation that has been given to the position space representative of the operator in Eq. (A). Several other facts concerning the calculations of this paper deserve clarification.

The scattering solutions of Sec. B, $\psi^{(\pm)}$, are unique functions of the wave parameter, K; their normalizations are correct, and their asymptotic forms contain the same incident state from which they are calculated. In contrast to these correct properties of the solutions of Sec. B are the corresponding anomalies in the solutions of Sec. A; viz., these solutions are not unique functions of the wave parameter K; their normalizations are incorrect, and their asymptotic forms do not contain the same incident state from which they are calculated. These anomalies are discussed in I and elsewhere.[18]

This article is concluded by stating a working rule for the calculation of the representative of Eq. (A) for the three-dimensional case. "The positive (negative) sign of $i\epsilon$ gives the diverging (converging) wave solution provided that the incident state is projected into the diverging (converging) wave components of the eigenfunction expansion." This rule differs somewhat from the conventional rule in which only the sign of $i\epsilon$ is specified.[19]

[18] S. Okubo and D. Feldman, Phys. Rev. **117**, 292 (1960).
[19] M. Gell-Mann and M. L. Goldberger, Phys. Rev. **91**, 398 (1953).

Derivation of Admissibility Conditions for Wave Functions from General Quantum-Mechanical Principles

DAVE PANDRES, JR.
The Martin Company, Denver Colorado
(Received September 19, 1961)

Admissibility conditions for wave functions are derived from general quantum-mechanical principles. The conditions apply in general curvilinear coordinates to functions that describe general N-particle systems. They restrict the behavior of wave functions so severely as to suggest that the Hilbert space of ordinary quantum mechanics cannot be enlarged unless the fundamental principles of quantum mechanics are modified. One of the conditions implies that all particles must be either fermions or bosons, i.e., that no particles obeying intermediate statistics can exist.

I. INTRODUCTION

ALTHOUGH the predictions of wave mechanics depend on the admissibility conditions that are imposed on wave functions, there has been no general agreement among different writers[1–4] as to what conditions should be used. The conditions that have been proposed apply only to single-particle functions, only to functions of Cartesian coordinates, or only to functions describing systems with special Hamiltonians. We shall show that such limitations are unnecessary. There are, however, more serious difficulties. The various sets of conditions differ in physical content because they include requirements adopted for reasons of mathematical convenience rather than physical necessity.

There has recently been a renewed interest[5–10] in this problem because the behavior of field operators in quantum field theory depends on the behavior of the wave functions from which they are formed. Dresden[11] has suggested that it may be necessary to enlarge the Hilbert space of ordinary quantum mechanics (i.e., relax admissibility conditions) in order to give consistent definitions of the field operators.

We shall show that certain conditions must be imposed on wave functions because of the following general quantum mechanical principles: (1) Momenta must be represented by Hermitian operators; (2) The standard commutation relations for coordinates and momenta must apply; (3) An observable must lead from one admissible wave function to another within a domain which is everywhere dense; (4) Matrix mechanics and wave mechanics must agree in every particular; and (5) In the case of a system of identical particles, the same observable that leads from ψ to φ must also lead from $P\psi$ to $P\varphi$, where P is the general permutation operator.

Our conditions are applicable in general curvilinear coordinates to functions that describe general N-particle systems. They restrict the behavior of wave functions so severely as to suggest that Dresden's proposed enlargement of Hilbert space cannot be obtained unless the fundamental principles of quantum mechanics are modified.

One of our conditions implies that wave functions must be single valued. This is, of course, a standard condition on wave functions, but a number of incorrect statements have appeared in the literature concerning the reasons such a condition must be used. Schrödinger[12] originally used this requirement because he believed that wave functions were directly associated with some experimentally observable property of a physical system. Pauli[13] realized that since wave functions are only indirectly associated with observables, there is no *a priori* reason why they must be single valued. However, he argued that multi-valued functions give rise to sources and sinks of probability current and so are physically inadmissible. It is this argument of Pauli's to which Blatt and Weisskopf[14] refer, although they attribute it to Nordsieck. The argument appears in other current textbooks,[15] although Pauli[16] later recognized that it was incorrect. He noted that there exists a large class of multi-valued angular momentum eigenfunctions with well-behaved probability currents.

Pauli then gave another argument in which he concluded that only single-valued functions of position in ordinary space are physically admissible. The essential point of Pauli's second argument is that double-valued functions generate angular momentum matrices that do not faithfully represent infinitesimal rotations. Pauli considered only single- and double-

[1] E. C. Kemble, *The Fundamental Principles of Quantum Mechanics* (Dover Publications, New York, 1958), pp. 79, 196.
[2] W. Pauli, *Handbuch der Physik*, edited by Geiger Sheele, (Verlag Julius Springer, Berlin, Germany, 1933), Vol. 24, p. 121.
[3] G. Jaffe, Z. Phys. **66**, 770 (1930) and references cited therein.
[4] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).
[5] J. M. Cook, J. Math. and Phys. **36**, 82 (1957).
[6] J. M. Jauch, Helv. Phys. Acta **31**, 127, 661 (1958).
[7] J. M. Jauch and I. I. Zines, Nuovo cimento **11**, 553 (1959).
[8] M. N. Hack, Nuovo cimento **9**, 731 (1958).
[9] S. J. Kuroda, Nuovo cimento **12**, 431 (1959).
[10] T. A. Green and O. E. Langford III, J. Math. Phys. **1**, 139 (1960).
[11] M. Dresden, *Lectures in Theoretical Physics*, edited by W. E. Brittin and B. W. Downs (Interscience Publishers, Inc., New York, 1960), p. 389.
[12] E. Schrödinger, Ann. Physik **79**, 361, 489 (1926).
[13] Reference 2, p. 26.
[14] J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (John Wiley & Sons, Inc., New York, 1952), p. 783.
[15] R. M. Sillitto, *Quantum Mechanics* (Edinburgh University Press, Edinburgh, Scotland, 1960), p. 51.
[16] W. Pauli, Helv. Phys. Acta **12**, 147 (1939).

valued functions as possibilities because Schrödinger[17] had shown that only these behave properly under time reversal. One can also exclude all other functions by noting that it follows from purely matrix methods[18] that angular momentum eigenvalues must be integer or half-integer multiples of $\hbar$ and that only single- and double-valued functions are associated with such eigenvalues.

Pauli's second argument has been generally accepted, although Bohm[19] has argued that logically one can only demand that wave functions satisfy conditions that insure that the average value of every observable is single valued. We show that if multi-valued wave functions are used in quantum mechanics, a violation of the fundamental commutation relations results.

In the case of systems of identical particles, one of our conditions implies that all particles must be either fermions or bosons, i.e., that no particles which obey intermediate statistics can exist. Our argument leading to this result does not depend on field theoretical assumptions (as does an argument of Pauli's[20]) or on the questionable assumption that $\psi^*\psi$ must play the role of a probability density function (as does an earlier argument due to Pauli[21]). Tolman[22] and others have noted that it is possible to define probability densities in other ways, and have speculated that particles obeying intermediate statistics might exist. We shall show that the existence of such particles would violate certain general quantum-mechanical principles.

II. SINGLE-PARTICLE WAVE FUNCTIONS

Consider a particle whose state is described classically by the Cartesian coordinates x_1, x_2, x_3, and their conjugate momenta p_1, p_2, p_3. In wave mechanics, x_j is diagonal while p_j is represented by the operator $P_j = -i\hbar\partial/\partial x_j$. Let ∇ denote the gradient operator, and let $\mathbf{f}$ denote a vector whose Cartesian components f_1, f_2, f_3, are arbitrary real (and therefore Hermitian) functions of the coordinates only. The operator P_j must be Hermitian, so it follows that $f_jP_j+P_jf_j$ is Hermitian and, hence, that $i(\mathbf{f}\cdot\nabla+\nabla\cdot\mathbf{f})$ must be Hermitian. It is easily seen from the definition of the Hermitian property and the formula for the divergence of a scalar multiplied by a vector that this operator is Hermitian only if the relation

$$\int_R \nabla\cdot(\psi^*\varphi\mathbf{f})d\tau=0 \tag{1}$$

is satisfied. Here ψ and φ denote any two wave functions, while R denotes the region of space which is accessible to the particle. Let $\boldsymbol{\sigma}$ be an outwardly-directed unit vector, normal to the surface S which bounds R. If we choose $\mathbf{f}$ so that $\psi^*\varphi\mathbf{f}$ is continuous everywhere in R and on S with continuous first partial derivatives everywhere in R, then we may use the divergence theorem in the form given by Kellogg[23] to obtain

$$\int_S \psi^*\varphi\mathbf{f}\cdot\boldsymbol{\sigma}dS=0 \tag{2}$$

from Eq. (1). Since $\mathbf{f}$ may vary in an arbitrary way on S, it is clear that $\psi^*\varphi$ must vanish there. By taking $\psi=\varphi$, we obtain the usual requirement[24] that wave functions vanish on any surface which bounds the accessible region of space.

We may obtain other admissibility conditions by lightening our restrictions on $\mathbf{f}$. Suppose we choose $\mathbf{f}$ so that $\psi^*\varphi\mathbf{f}$ is continuous everywhere in R and on S with continuous first partial derivatives everywhere in R except possibly on a surface $g(x_1,x_2,x_3)=0$ which separates two regions of space. We may use the divergence theorem together with Eq. (1) and the fact that $\psi^*\varphi$ vanishes on S, to obtain the relation

$$\lim_{\epsilon\to 0}\left[\int_{g=|\epsilon|}\psi^*\varphi\mathbf{f}\cdot\boldsymbol{\gamma}dS-\int_{g=-|\epsilon|}\psi^*\varphi\mathbf{f}\cdot\boldsymbol{\gamma}dS\right]=0, \tag{3}$$

where $\boldsymbol{\gamma}$ is a unit vector in the direction of ∇g. The integrals in Eq. (3) appear with opposite signs because $\boldsymbol{\gamma}$ is outwardly directed on one of the surfaces but is inwardly directed on the other. Since $\mathbf{f}$ may vary in an arbitrary way on $g=0$, it is clear that $\psi^*\varphi$ must be continuous across this surface. We have only to take $\psi=\varphi$ in order to see that the amplitude of a wave function must be continuous across $g=0$. Now, if a function possesses a discontinuous phase, its gradient will possess a discontinuous amplitude. But the momentum operator (since it represents an observable) must lead from one admissible function to another, so we obtain the usual requirement[25] that the phase as well as the amplitude of a wave function must be continuous across every surface which separates two regions of space.

Now, suppose that we express the position of the particle in terms of some general curvilinear coordinates q_1, q_2, q_3 such that all of space corresponds to the interval $a_k\leqslant q_k\leqslant b_k$ where a_k and b_k are constants. The Jacobian determinant $|\partial x/\partial q|$ may vanish on either or both of the surfaces $q_k=a_k$ and $q_k=b_k$. This means that these surfaces may have zero area, so they do not separate two regions of space. We shall now consider the behavior of wave functions near such degenerate

[17] E. Schrödinger, Ann. Physik, 32, 49 (1938).
[18] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955), 2nd ed., p. 141.
[19] D. Bohm, *Quantum Theory* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1951), p. 389.
[20] W. Pauli, Phys. Rev. 58, 716 (1940).
[21] See e.g., G. Temple, *Quantum Theory* (Methuen and Company, Ltd., London, England, 1934), p. 104.
[22] R. C. Tolman, *Principles of Statistical Mechanics* (Oxford University Press, New York, 1938), p. 318.
[23] O. D. Kellogg, *Foundations of Potential Theory* (Frederick Ungar Publishing Company, New York, 1929), p. 119.
[24] Reference 18, p. 30.
[25] Reference 18, p. 29.

surfaces. We select $\mathbf{f}$ so that $\psi^*\varphi\mathbf{f}$ is continuous everywhere in R and on S with continuous first partial derivatives everywhere in R except possibly on $q_k=a_k$ and $q_k=b_k$. We may use the divergence theorem together with Eq. (1) and the fact that ψ vanishes on S to obtain the relation

$$\lim_{\epsilon\to 0}\left[\int_{q_k=b_k-|\epsilon|}\psi^*\varphi\mathbf{f}\cdot\boldsymbol{\eta}dS-\int_{q_k=a_k+|\epsilon|}\psi^*\varphi\mathbf{f}\cdot\boldsymbol{\eta}dS\right]=0, \quad (4)$$

where $\boldsymbol{\eta}$ is a unit vector in the direction of ∇q_k. Now, if the surface element dS with unit normal $\boldsymbol{\eta}$ is moved through the distance $d\mathbf{r}=\sum_n(\partial\mathbf{r}/\partial q_n)dq_n$, the volume element $d\tau$ which it sweeps out is $d\mathbf{r}\cdot\boldsymbol{\eta}dS$. But, $d\tau$ is also given by $|\partial x/\partial q|dq_1dq_2dq_3$. Upon equating these expressions for $d\tau$ and using the fact that $\nabla q_k\cdot\partial\mathbf{r}/\partial q_n$ reduces to the Kronecker symbol δ_{kn}, we find that $\boldsymbol{\eta}dS=|\partial x/\partial q|dQ_k\nabla q_k$, where $dQ_kdq_k=dq_1dq_2dq_3$. We may substitute this expression for $\boldsymbol{\eta}dS$ into Eq. (4) to obtain the relation

$$\lim_{\epsilon\to 0}\left[\int_{q_k=b_k-|\epsilon|}\psi^*\varphi\left|\frac{\partial x}{\partial q}\right|\nabla q_k\cdot\mathbf{f}dQ_k - \int_{q_k=a_k+|\epsilon|}\psi^*\varphi\left|\frac{\partial x}{\partial q}\right|\nabla q_k\cdot\mathbf{f}dQ_k\right]=0. \quad (5)$$

If $q_k=a_k$ and $q_k=b_k$ refer to different surfaces in space, then $\mathbf{f}$ may vary arbitrarily and independently over these surfaces. Hence, it is clear that the condition

$$\lim_{q_k\to a_k,\,b_k}\psi^*\psi\left|\frac{\partial x}{\partial q}\right|\nabla q_k=0 \quad (6)$$

must be satisfied. For the usual spherical polar coordinates, the condition at $r=0$ which follows from Eq. (6) is just Dirac's condition[26] that ψ cannot possess a singularity as strong as r^{-1}. The conditions at $\theta=0$ and $\theta=\pi$ are that ψ cannot possess a singularity as strong as $(\sin\theta)^{-\frac{1}{2}}$. It should be noted that these conditions, restricting the behavior of ψ near any point or any line, are stronger than the conditions required to make $\psi^*\psi$ locally integrable. The condition for large r is that ψ must approach zero more rapidly than r^{-1}.

We now inquire whether ψ must be a single-valued function of q_k if this coordinate is periodic in the sense that $q_k=a_k$ and $q_k=b_k$ refer to the same surface. Since $\mathbf{f}$ may vary in an arbitrary way over this surface, it follows from Eq. (4) that the product of any wave function with the complex conjugate of another must be single valued. We shall now go on to show that the functions themselves must be single valued.

A wave function is single valued if it returns to its original value when a particle is moved from an arbitrary point around an arbitrary closed curve back to the original point. Thus, it is clear that ψ is single-valued if the line integral

$$Q=\int_C\nabla\psi\cdot d\mathbf{r} \quad (7)$$

vanishes for an arbitrary closed curve C. We may use Stokes theorem to write Eq. (7) in the form

$$Q=\int_s(\nabla\times\nabla\psi)\cdot\boldsymbol{\eta}ds \quad (8)$$

where $\boldsymbol{\eta}$ is a unit vector normal to the surface bounded by C. Now, $\nabla\times\nabla\psi$ vanishes provided that ψ satisfies conditions which insure that the commutator $[(\partial/\partial x_i),(\partial/\partial x_j)]$ vanishes for all i and j. The vanishing of this commutator is required if P_i and P_j commute, i.e., if p_i and p_j are simultaneously observable. This implies that Q must vanish, and hence that ψ must be single valued.

Finally, we show that each component of a spinor wave function must satisfy the conditions which have been derived for the total wave function. For this, it is only necessary to prove that the operator which leads from the spinor ψ with components $\psi_1,\cdots,\psi_4$ to the spinor φ with components $0,\cdots,0,\psi_k,0,\cdots,0$ represents an observable. If we regard the spinors as column matrices, then $\varphi=\Omega\psi$, where Ω is a square matrix whose general element is $\delta_{ik}\delta_{kj}$. Since Ω is Hermitian and satisfies the algebraic equation $\Omega^2=\Omega$, it represents an observable.[27] Hence φ (whose only nonvanishing component is ψ_k) must be admissible if ψ is admissible.

III. N-PARTICLE WAVE FUNCTIONS

In the interest of clarity we have discussed only single-particle wave functions in Sec. II. It is clear, however, that the entire discussion applies also to N-particle functions when one works in a $3N$-dimensional configuration space.

Let ∇ denote the gradient operator in the $3N$-dimensional configuration space for the system, and let $\mathbf{f}$ denote a vector whose Cartesian components $f_1,\cdots,f_{3N}$ are arbitrary real functions of the coordinates only. Further, let R be the region of configuration space which is accessible to the system, and let S be a $(3N-1)$-dimensional hypersurface which bounds R. Let

$$g(x_1,\cdots,x_{3N})=0$$

be a $(3N-1)$-dimensional hypersurface which separates two regions of configuration space. Finally, let $q_1,\cdots,q_{3N}$ be generalized coordinates which describe the configuration of the system, and let $a_k\leqslant q_k\leqslant b_k$ correspond to all of configuration space. With these generalized definitions, we have only to use the divergence theorem and Stokes theorem in $3N$-dimensional space[28]

[26] P. A. M. Dirac, *Quantum Mechanics* (Oxford University Press. New York, 1958), 4th ed., p. 156.

[27] Reference 26, p. 38.

[28] J. L. Synge and A. Schild, *Tensor Calculus* (The University of Toronto Press, Toronto, Canada, 1956), p. 274.

in order to see that all of the conditions derived in Sec. II for single-particle functions apply also to many-particle functions.

Wave functions describing systems of identical particles must also satisfy conditions which insure that the general permutation operator P commutes with each of a complete set of commuting observables $\Omega_1, \cdots, \Omega_Q$. The following argument, due to Bargmann,[29] shows that this is necessary in order to insure that the same operator which leads from ψ to ϕ also leads from $P\psi$ to $P\phi$. If Ω is any operator, corresponding to an observable, such that $\phi=\Omega\psi$, then $P\phi=P\Omega\psi$. If the particles are equivalent then we must have $P\phi=\Omega P\psi$. Upon equating the two expressions for $P\phi$, we see that Ω and P must commute. Since $\Omega_1, \cdots, \Omega_Q$ all commute, we may choose a representation in which they are all diagonal. Now, only a diagonal matrix can commute with each matrix of a complete diagonal set, so we see that P must be diagonal in the representation in which $\Omega_1, \cdots, \Omega_Q$ are diagonal. Since this may be *any* complete set of commuting observables it follows that P must be diagonal in *every* representation. Only a constant multiple of the identity matrix satisfies this requirement. This implies that wave functions must satisfy conditions which insure that

$$P\psi=C_P\psi, \tag{9}$$

where C_P is a number which may depend upon P, but not upon ψ. An arbitrary permutation can be expressed as a product of permutations involving only pairs of particles. Let P_{ij} be a permutation that interchanges the coordinates, including spin, of the ith and jth particles of the system. The defining relations for P_{ij} are

$$\begin{aligned} P_{ij}\psi(i,j)&=\psi(j,i)\\ P_{ij}\psi(j,i)&=\psi(i,j). \end{aligned} \tag{10}$$

Equations in (10) imply that

$$P_{ij}^2\psi(i,j)=\psi(i,j), \tag{11}$$

but Eq. (9) implies that

$$P_{ij}\psi(i,j)=C_{ij}\psi(i,j), \tag{12}$$

from which it follows that

$$P_{ij}^2\psi(i,j)=C_{ij}^2\psi(i,j). \tag{13}$$

Equations (11) and (13) imply that C_{ij}^2 must be unity, and hence that $C_{ij}=\pm 1$. Finally, we note that C_{ij} cannot depend upon the indices i or j. This is clear since the identity $P_{ik}P_{ij}=P_{kj}P_{ik}$ leads to the requirement that $C_{ij}=C_{kj}$. Since C_{ij} cannot depend on i, j, or ψ, we conclude that it can only depend on the nature of the system of particles, and that there can exist only two classes of particles, corresponding to $C_{ij}=+1$ and $C_{ij}=-1$, respectively.

[29] V. Bargmann, lecture notes (unpublished).

IV. CONTINUUM WAVE FUNCTIONS

Although the conditions we have obtained in this paper are consistent with the usual restrictions imposed on wave functions, generally they are not fulfilled by eigenfunctions corresponding to eigenvalues which belong to a continuum. For example, neither plane waves (momentum eigenfunctions) nor Dirac delta functions (position eigenfunctions) satisfy our admissibility conditions. However, these difficulties are easily resolved. As Dirac[30] has remarked, we can measure such quantities with an arbitrary degree of precision, but not with an infinite degree of precision. This implies that we can prepare a system in a state which approximates a plane wave or a delta function arbitrarily closely, but can never prepare a state which is described exactly by either of these functions. We may gain some physical insight into the meaning of this by considering the measurement of the position of a particle. If we observe the particle with electromagnetic radiation, we must use photons with an infinitesimally short wavelength to measure the position of the particle exactly. But the energy of such a photon would be infinite so the measurement is physically impossible.

Dirac has described the situation by saying that states corresponding to eigenvalues which belong to a continuum are mathematical idealizations of what can be achieved in practice. It is not surprising that such idealizations will sometimes violate the boundary conditions which apply to states that actually occur in nature. On the other hand, one can only agree with Dirac that such idealizations play a very useful role in the theory and one could not very well do without them. It therefore seems clear that we must admit into wave mechanics any functions which can be represented as the limit of a sequence of functions which satisfy our boundary conditions. This point of view is consistent with Schwartz's[31] theory of generalized functions and with von Neumann's[32] argument that the domain of an observable need not include all of Hilbert space although it must be everywhere dense.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor Wesley E. Brittin and to Dr. John Walter Clark for a number of helpful discussions of this work. He also wishes to thank Professor Arnold Nordseick for calling attention to Pauli's second paper on single valuedness.[16]

[30] Reference 26, p. 48.

[31] L. Schwartz, *Theorie des distributions* (Herman et Cie, Paris, France, 1950).

[32] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955), p. 90.

An Approach to the Eulerian-Lagrangian Problem*

JOHN L. LUMLEY†
The Pennsylvania State University, University Park, Pennsylvania
(Received July 31, 1961)

The generalization is developed, for n functions of n variables, of the Kac-Rice theorem on the occurrence of zeros of a random function, and is applied to a particular kind of vector integral equation in a single independent variable (time) by discretizing the time axis. In this way the functional probability for the solution of the equation is obtained. This technique is applied to the stochastic integral equation representing the motion of a tagged point in a fluid in turbulent motion.

INTRODUCTION

THE problem of describing the motion of a tagged point in a continuous fluid in turbulent motion can be reduced to the problem of obtaining the statistics of a function $\mathbf{r}(\mathbf{a},t)$ defined by

$$\mathbf{r}(\mathbf{a},t)=\mathbf{a}+\int_0^t \mathbf{u}[\mathbf{r}(\mathbf{a},\tau),\tau]d\tau$$

when the statistics of $\mathbf{u}(\mathbf{x},t)$ are known, where $\mathbf{u}(\mathbf{x},t)$ is the Eulerian velocity field specified in laboratory coordinates, and $\mathbf{r}(\mathbf{a},t)$ is the position at time t of a material point located at $\mathbf{a}$ at $t=0$. The difficulty of the problem is that each realization of $\mathbf{u}(\mathbf{x},t)$ generates a unique $\mathbf{r}(\mathbf{a},t)$, so that the $\mathbf{r}$ are statistically dependent on the $\mathbf{u}$, and it is not clear how one should go about obtaining the statistics of $\mathbf{u}(\mathbf{r},t)$ from those of $\mathbf{u}(\mathbf{x},t)$. The author has been unable to find a treatment of such problems in the literature.

In this paper this difficulty is resolved in the sense that a method is described for obtaining the functional probability of $\mathbf{r}$ in terms of that of $\mathbf{u}$. This is done by showing that the distribution of the solution of the equation is the distribution of the zeros of the set of simultaneous equations formed by discretizing the time axis.

For applications, however, a serious difficulty remains. The functional probability is far more information than is needed or wanted in physical problems. To obtain the distribution for position at a single time, it is necessary to integrate in function space over all paths that lead to the point in question. The author has been able to carry out such an integration only in problems which can be done more easily by other methods, but functional integration has arisen in other problems[1–3] and is under consideration by other authors.

1. EXTENSION OF THE RICE-KAC THEOREM

Under fairly general restrictions, a well-known theorem by Rice[4] and Kac[5] states, that for a random function $f(x)$ for which

$$P(\xi,\eta;x)d\xi d\eta \tag{1}$$

is the probability that the amplitude and slope satisfy the inequalities

$$\begin{aligned}\xi\leqslant f(x)<\xi+d\xi\\ \eta\leqslant f'(x)<\eta+d\eta,\end{aligned} \tag{2}$$

then the probability that $f(x)$ has a zero in the interval x, $x+dx$ is given by

$$dx\int_{-\infty}^{+\infty}|\eta|\,P(0,\eta;x)d\eta. \tag{3}$$

We need a more general theorem applying to a set of n functions of n variables

$$f_i(x_1,\cdots,x_n)\quad i=1,\cdots,n \tag{4}$$

giving us the probability that the f_i simultaneously have zeros in the intervals x_j, x_j+dx_j.

Now, if x_j^0 is a set of zeros, then we can write, close to the zeros,

$$\begin{aligned}f_i(x_1^0,\cdots,x_n^0)=0=f_i(x_1,\cdots,x_n)+f_{i,k}(x_1,\cdots,x_n)\\ \times(x_k^0-x_k)+O(|x_k^0-x_k|^2),\\ |x_k^0-x_k|^2=(x_k^0-x_k)(x_k^0-x_k),\end{aligned} \tag{5}$$

where we have expanded f_i at a zero in a Taylor series about an adjacent point. Ignoring higher-order terms, the solution for $x_k-x_k^0$ is given by

$$f_{ki}^{-1}(x_1,\cdots,x_n)f_i(x_1,\cdots,x_n)=d_k(x_1,\cdots,x_n)\quad\text{(say)} \tag{6}$$

and we are led to consider d_k not only close to a zero,

---

* The work on which the present paper is based in taken from the author's PhD. Thesis, "Some problems connected with the motion of small particles in a turbulent fluid," June, 1957, available from the Department of Mechanics, The Johns Hopkins University. The investigation reported there was supported by the Mechanics Branch, Office of Naval Research. Subsequent work contained in the present paper was performed at the Garfield Thomas Water Tunnel, Ordnance Research Laboratory, The Pennsylvania State University, with the support of the Bureau of Naval Weapons.

† Associate Professor of Aeronautical Engineering, The Pennsylvania State University.

[1] R. P. Feynman, Revs. Modern Phys. **20**, 367 (1948).
[2] E. Hopf, J. Rational Mech. Anal. **1**, 87 (1952).
[3] N. Wiener, Acta. Math. **55**, 117 (1930), note especially pp. 214–234.
[4] S. O. Rice, Am. J. Math. **61**, 409 (1939).
[5] M. Kac, Bull. Am. Math. Soc. **49**, 314 (1943).

but elsewhere. f_{ki}^{-1} is defined by

$$f_{ki}^{-1}f_{i,j}=\delta_{kj} \quad \text{(Kronecker's delta)}, \quad f_{i,j}=\partial f_i/\partial x_j \tag{7}$$

that is, it is the inverse of $f_{i,j}$. Here we must make a basic assumption, which is that neither of the determinants

$$\|f_{i,j}\|\neq 0 \quad \|f_{ij}^{-1}\|\neq 0 \tag{8}$$

is ever zero. This permits us to solve (5) to obtain (6), for instance, and will be used again later. Equation (8) is essentially a restriction that the transformation defined by the f_i is one to one.

Examining (6), we find that, if x_j is a zero of the set f_i, then the set d_i is zero, a conclusion we can reach because of (8). Conversely, if the set d_i is zero, then due to (8), the set f_i must be zero, and hence x_j is a zero of f_i. We can thus state that the set $d_i=0$ if and only if the set x_j is a zero of f_i. Now we must examine the limiting behavior of (6). We have in mind equating the probability of d_i being near zero to the probability of a zero of f_i being near x_j; the question is, how near? If x_i^0 is a zero of f_i, and in the limit $d_i\to x_i-x_i^0$, then the "nearness" in the two probabilities is the same. We can expand (6) about a zero, obtaining

$$d_i(x_1,\cdots,x_n)=x_i-x_i^0+O(|x_i-x_i^0|^2). \tag{9}$$

So that as $|x_i-x_i^0|\to 0$, $d_i\to x_i-x_i^0$, which is not surprising in view of our definition of d_i. Hence we can write

$$\begin{aligned}&W(x_1,\cdots,x_n)dx_1\cdots dx_n\\&\quad=\text{prob}\{\text{the set } f_i(x_1,\cdots,x_n) \text{ has a zero in the interval } x_j,\ x_j+dx_j\}\\&\quad=\text{prob}\{0\leqslant d_i(x_1,\cdots,x_n)<dx_i\}.\end{aligned} \tag{10}$$

Now, the latter quantity is fairly easy to compute. Let us write

$$\begin{aligned}&\text{prob}\{\xi_i\leqslant d_i(x_1,\cdots,x_n)<\xi_i+d\xi_i,\\&\qquad\eta_{ij}\leqslant f_{i,j}(x_1,\cdots,x_n)<\eta_{ij}+d\eta_{ij}\}\\&\quad=Q(\xi_1,\cdots,\xi_n,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)d\xi d\eta\end{aligned} \tag{11}$$

(we will use the symbol without subscript to indicate the volume element, $d\xi=d\xi_1\cdots d\xi_n$). If we write

$$\begin{aligned}&\text{prob}\{\zeta_i\leqslant f_i(x_1,\cdots,x_n)<\zeta_i+d\zeta_i,\\&\qquad\eta_{ij}\leqslant f_{i,j}(x_1,\cdots,x_n)<\eta_{ij}+d\eta_{ij}\}\\&\quad=P(\zeta_1,\cdots,\zeta_n,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)d\zeta d\eta,\end{aligned} \tag{12}$$

then we must define $\zeta_i=\eta_{ij}\xi_j$ (from the definition of d_i), and

$$\begin{aligned}&Q(\xi_1,\cdots,\xi_n,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)d\xi d\eta\\&\quad=P(\eta_{1j}\xi_j,\cdots,\eta_{nj}\xi_j,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)\\&\qquad\times|\|\eta_{ij}\||\,d\xi d\eta,\end{aligned} \tag{13}$$

where $|\|\eta_{ij}\||$ is the absolute value of the determinant, i.e., the absolute value of the Jacobian of the transformation. The quantity we wish, given in (10), is

$$\begin{aligned}&W(x_1,\cdots,x_n)d\xi\\&\quad=d\xi\int_{-\infty}^{+\infty}Q(0,\cdots,0,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)d\eta\\&\quad=d\xi\int_{-\infty}^{\infty}P(0,\cdots,0,\eta_{11},\cdots,\eta_{nn};x_1,\cdots,x_n)\\&\qquad\times|\|\eta_{ij}\||\,d\eta.\end{aligned} \tag{14}$$

We will indicate the n^2 multiple integral by a single integral where no confusion is possible. Expression (14) is the n-dimensional analog of the one-dimensional expression (3), and reduces to it for $n=1$.

2. GENERALIZED RICE-KAC THEOREM FOR VECTOR FUNCTIONS

In what follows we will apply the generalized Rice-Kac theorem obtained above to a set of vector equations. This application is straightforward, since it requires only a convention for ordering. Suppose, for example, that we had a set of n vector functions of m components, the arguments of which were n vectors of m components each:

$$\mathbf{f}_i(\mathbf{x}_j)=0. \tag{15}$$

All we need do is define a new set of functions and arguments:

$$\begin{aligned}&(\mathbf{f}_k)_p=f_{(k-1)m+p}^{*} \quad p=1,\cdots,m;\ k=1,\cdots,n\\&(\mathbf{x}_k)_p=x_{(k-1)m+p}^{*} \quad p=1,\cdots,m;\ k=1,\cdots,n.\end{aligned} \tag{16}$$

The adoption of such a convection, however, tends somewhat to obscure the physics. We will instead use a notation which maintains the distinction. The reader can easily convince himself that the proper form for (14) is given by (with an appropriate interpretation of the notation)

$$\begin{aligned}&W(\mathbf{x}_1,\cdots,\mathbf{x}_n)d\xi\\&\quad=d\xi\int_{-\infty}^{+\infty}P(\mathbf{0},\cdots,\mathbf{0},\mathbf{H}_{11},\cdots,\mathbf{H}_{nn};\mathbf{x}_1,\cdots,\mathbf{x}_n)\\&\qquad\times|\|\mathbf{H}_{ij}\||\,d\mathbf{H}\end{aligned} \tag{17}$$

in which boldface lower-case letters indicate m vectors, boldface upper-case letters indicate $m\times m$ matrices, the determinant $\|\mathbf{H}_{ij}\|$ is a block determinant, each block $m\times m$ consisting of n^2 blocks, and $d\mathbf{H}$ is the volume element in m^2n^2 dimensions,

$$d(\mathbf{H}_{11})_{11}d(\mathbf{H}_{11})_{12}\cdots d(\mathbf{H}_{nn})_{mm}.$$

If we define a gradient operator as

$$(\nabla_i)_p=\partial/\partial(\mathbf{x}_i)_p, \tag{18}$$

we can write the gradient matrix as

$$\nabla_i\mathbf{f}_j=\mathbf{F}_{ij}\ (\mathbf{F}_{ij})_{pq}=\partial(\mathbf{f}_i)_p/\partial(\mathbf{x}_j)_q. \tag{19}$$

3. APPLICATION TO VECTOR INTEGRAL EQUATIONS

Let us consider

$$\mathbf{r}(t)=\int_0^t \mathbf{g}[\mathbf{r}(\tau),t,\tau]d\tau \tag{20}$$

although for our fluid mechanical problem we will have no need for so general an equation. $\mathbf{r}$ and $\mathbf{g}$ are m vectors. Let us divide the time axis by a mesh, $t_{i-1}<t_i<t_{i+1}$, $i=1,\cdots,n$ and write the discrete analog of (20):

$$\mathbf{r}(t_i)=\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\mathbf{g}[\mathbf{r}(\tau),t_i,\tau]d\tau\cong\sum_{j=1}^{n}\mathbf{g}(\mathbf{r}(t_j),t_i,t_j)\Delta_j, \tag{21}$$

where we have written $\Delta_j=t_j-t_{j-1}$. Writing $\mathbf{r}(t_j)=\mathbf{x}_j$, and

$$\mathbf{f}_i(\mathbf{x}_1,\cdots,\mathbf{x}_n)=\mathbf{x}_i-\sum_{k=1}^{n}\mathbf{g}(\mathbf{x}_k,t_i,t_k)\Delta_k, \tag{22}$$

we have a set of n vector functions (each having m components) of n vector arguments (each having m components). Asking for the probability that there is a zero of this set in the vicinity of the points $\mathbf{x}_i$ is the same as asking for the probability that the solution to (21) lies near the set $\mathbf{x}_i$. Except for the discrete approximation, this is the functional probability for the solution of (20). Thus we wish to compute the expression (17) for such a set and the first quantity to be computed is the determinant of the gradient matrix $\|\mathbf{F}_{ij}\|$. If we presume (as we must) that we have all statistical information about $\mathbf{g}(\mathbf{x},t,t')$ for deterministic arguments, then $P(\mathbf{0},\cdots,\mathbf{0},\mathbf{H}_{11},\cdots,\mathbf{H}_{nn};\mathbf{x}_1,\cdots,\mathbf{x}_n)$ can, in principle, be written down.

Now, the determinant $\|\mathbf{F}_{ij}\|$ will be made up of n^2 blocks, each for a fixed pair of values of t_i; the i, j block can be written as

$$\begin{aligned}\mathbf{F}_{ij}&=\mathbf{1}\delta_{ij}-\sum_{k=1}^{n}\nabla_j\mathbf{g}(\mathbf{x}_k,t_i,t_k)\Delta_k\\&=\mathbf{1}\delta_{ij}-\nabla\mathbf{g}(\mathbf{x}_j,t_i,t_j)\Delta_j,\end{aligned} \tag{23}$$

where the subscript has been dropped from ∇, since only one of the vectors appears as argument, and no confusion can arise; δ_{ij} is Kronecker's delta, and $\mathbf{1}$ is the unit matrix. The determinant is

$$\|\mathbf{F}_{ij}\|=\begin{vmatrix}\mathbf{F}_{11} & \cdots & \mathbf{F}_{1n}\\ \vdots & & \\ \mathbf{F}_{n1} & \cdots & \mathbf{F}_{nn}\end{vmatrix}. \tag{24}$$

Now, we are particularly interested in physical systems, which are in general characterized by having their present state depend on past time only, equivalent to writing (20) with the upper limit t the running variable. In such a system, $\mathbf{g}(\mathbf{x}_j,t_i,t_j)=0$ if $j>i$, so that $\mathbf{F}_{ij}=0$ if $j>i$. This means that (24) may be evaluated in a particularly simple way by Laplace's expansion,[6] since all the blocks above the diagonal are zero, and only the diagonal blocks contribute. Thus, making this restriction,

$$\|\mathbf{F}_{ij}\|=\prod_{\alpha=1}^{n}\|\mathbf{F}_{\alpha\alpha}\|=\prod_{\alpha=1}^{n}\det[\mathbf{1}-\nabla\mathbf{g}(\mathbf{x}_\alpha,t_\alpha,t_\alpha)\Delta_\alpha] \tag{25}$$

(no sum on α).

Now, each determinant in this product is a characteristic equation for the matrix $\nabla\mathbf{g}\Delta_\alpha$, with unity eigenvalue.[7] Hence it can be written as

$$\det(\mathbf{1}-\nabla\mathbf{g}\Delta_\alpha)=1-I_1+I_2-I_3\cdots, \tag{26}$$

where each term I_n is the nth principle invariant of $\nabla\mathbf{g}\Delta_j$ (i.e, the sum of the principle minors of order n), and hence is homogeneous of degree n in its terms, and in particular, in Δ_α. Hence, we can write

$$\det(\mathbf{1}-\nabla\mathbf{g}\Delta_\alpha)=1-\nabla\cdot\mathbf{g}\Delta_\alpha+O(\Delta_\alpha^2), \tag{27}$$

$\nabla\cdot\mathbf{g}$ being the first invariant, the trace. Hence, we can write

$$\begin{aligned}\|\mathbf{F}_{ij}\|&=\prod_{i=1}^{n}[1-\nabla\cdot\mathbf{g}(\mathbf{x}_i,t_i,t_i)\Delta_i+O(\Delta_i^2)]\\&=\exp\Big(\sum_{i=1}^{n}\ln[1-\nabla\cdot\mathbf{g}(\mathbf{x}_i,t_i,t_i)\Delta_i+O(\Delta_i^2)]\Big)\\&=\exp\Big(-\sum_{i=1}^{n}[\nabla\cdot\mathbf{g}(\mathbf{x}_i,t_i,t_i)\Delta_i+O(\Delta_i^2)]\Big)\end{aligned} \tag{28}$$

expanding the logarithm; as the mesh is drawn finer and finer, this becomes

$$\|\mathbf{F}_{ij}\|=\exp\left(-\int_0^T\nabla\cdot\mathbf{g}[\mathbf{r}(t),t,t]dt\right). \tag{29}$$

In the particular case of $\nabla\cdot\mathbf{g}=0$ (which will prove to be of importance physically as that of an incompressible fluid) we have

$$\|\mathbf{F}_{ij}\|\equiv 1 \tag{30}$$

so that the integration in (17) can be carried out directly, giving the result that, for this particular case, the probability of the solution to (20) lying in a narrow band about a particular "test function" is the same as the probability that the two sides of (20), evaluated at the "test function," differ by an amount that lies in the same narrow band.

4. MOTION OF A TAGGED POINT IN A FLUID—THE EULERIAN-LAGRANGIAN PROBLEM

For the motion of a tagged material point in a fluid, (for simplicity $\mathbf{a}=0$)

$$\begin{aligned}\mathbf{g}(\mathbf{x},t,\tau)&=\mathbf{u}(\mathbf{x},\tau)\quad \tau\leqslant t\\&=0\qquad\quad \tau>t\end{aligned} \tag{31}$$

[6] See e.g., E. A. Guillemin, *The Mathematics of Circuit Analysis* (John Wiley & Sons, Inc., New York, 1949), p. 7.

[7] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, New York, 1956), Chap. 4.

the Eulerian velocity field. If, furthermore, the fluid is incompressible $\nabla \cdot \mathbf{u}=0$, so that

$$\|\mathbf{F}_{ij}\|=1 \tag{32}$$

and the conclusions following (30) hold. That is, since

$$P(\mathbf{0},\cdots,\mathbf{0},\mathbf{H}_{11},\cdots,\mathbf{H}_{nn};\mathbf{x}_1,\cdots,\mathbf{x}_n)$$

will be nonzero only on those points where

$$\|\mathbf{H}_{ij}\|=1,$$

$$W(\mathbf{x}_1,\cdots,\mathbf{x}_n)=\int_{-\infty}^{+\infty} P(\mathbf{0},\cdots,\mathbf{0},\mathbf{H}_{11},\cdots,\mathbf{H}_{nn};\mathbf{x}_1,\cdots,\mathbf{x}_n) \times|\|\mathbf{H}_{ij}\||\,d\mathbf{H}$$

$$=R(\mathbf{0},\cdots,\mathbf{0};\mathbf{x}_1,\cdots,\mathbf{x}_n) \tag{33}$$

where

$$R(\zeta_1,\cdots,\zeta_n;\mathbf{x}_1,\cdots,\mathbf{x}_n)d\zeta = \text{prob}\{\zeta_i \leqslant \mathbf{f}_i(\mathbf{x}_1,\cdots,\mathbf{x}_n)<\zeta_i+d\zeta_i\} \tag{34}$$

and $m=3$.

Now, $R(\mathbf{0},\cdots,\mathbf{0};\mathbf{x}_1,\cdots,\mathbf{x}_n)$ is the integral of the characteristic functional

$$R(\mathbf{0},\cdots,\mathbf{0};\mathbf{x}_1,\cdots,\mathbf{x}_n) =\frac{1}{(2\pi)^{3n}}\int_{-\infty}^{+\infty} E\{\exp[i\boldsymbol{\phi}_k\mathbf{f}_k(\mathbf{x}_1,\cdots,\mathbf{x}_n)]\}d\boldsymbol{\phi}, \tag{35}$$

where $E\{\ \}$ is expectation. If we substitute from (31) and (22) we can write

$$R(\mathbf{0},\cdots,\mathbf{0};\mathbf{x}_1,\cdots,\mathbf{x}_n) =\frac{1}{(2\pi)^{3n}}\int_{-\infty}^{+\infty}\exp[i\boldsymbol{\phi}_k\mathbf{x}_k] \times E\{\exp[-i\boldsymbol{\phi}_k\sum_{p=1}^{k}\mathbf{u}(\mathbf{x}_p,t_p)\Delta_p]\}d\boldsymbol{\phi}, \tag{36}$$

where a sum over k is understood. The expression (36) can be evaluated if the Eulerian characteristic functional is known; in the limit it is simply the probability density that for a *chosen* $\mathbf{r}(t)$,

$$\mathbf{r}(t)=\int_0^t \mathbf{u}[\mathbf{r}(\tau),\tau]d\tau \quad \text{on an interval } [0,T]. \tag{37}$$

Since $\mathbf{r}(t)$ is fixed, this is an Eulerian quantity.

5. CONCLUSIONS

We have shown how in an incompressible fluid, the functional probability density that a test function is the displacement of a tagged point over a finite time may be obtained from the characteristic functional for the Eulerian velocity field. For such a fluid, the functional density for displacement is identical with the probability density that a test function satisfies the kinematical equation in the same time interval. In order to obtain physically useful information, it is necessary to integrate over all paths which end at a point of interest in order to obtain the ordinary density at a given time. It is possible to make an assumption for the characteristic functional of the Eulerian velocity field which is kinematically consistent, if not dynamically so (the Gaussian); it is still not possible to carry out the functional integrations without other assumptions. Those which suggest themselves to the author (spacially uniform velocity field, or temporally uncorrelated) are not physically meaningful, for turbulence problems, and give results that can be obtained directly in both cases.

ACKNOWLEDGMENT

The author wishes to thank Dr. S. Corrsin for suggesting, and contributing his insight through endless discussion of, the problem.

Symmetries of the Einstein-Maxwell Equations

GERALD ROSEN
The Martin Company, Baltimore, Maryland
(Received April 6, 1961)

Space-time symmetries admissible according to the Einstein-Maxwell equations are analyzed from the standpoint of the groups of motions in Rainich geometry. Necessary and sufficient conditions for a motion are expressed in terms of the Ricci vierbein of principal directions. Normal Rainich geometries, for which the Ricci congruences are orthogonal to four sets of hypersurfaces, are studied in some detail. First integrals to the Rainich conditions are presented for the latter class of geometries. A new solution to the Einstein-Maxwell equations is derived in the form of a spatially homogeneous Rainich geometry.

I. INTRODUCTION

THE gravitational behavior of classical electromagnetic radiation is governed by the Einstein-Maxwell equations. If the square of the Maxwell stress tensor is not zero, the electromagnetic field is said to be *non-null* and the Einstein-Maxwell equations imply that

$$\rho^2 \equiv \tfrac{1}{4} R_\nu{}^\mu R_\mu{}^\nu > 0. \tag{1.1}$$

When this non-null condition is satisfied, the whole content of source-free electrodynamics and Einstein gravitational theory can be recast in a purely geometrical form, as originally shown by Rainich[1] and investigated more recently by Misner and Wheeler[2] and other authors.[3–5]

Such Riemannian manifolds, the so-called *Rainich geometries*, take a simple description in terms of the invariants associated with the Ricci vierbein of principal directions.[3] Let us denote the four unit vectors of the Ricci vierbein by $\lambda_{a|\mu}$, where the Latin subscript labels the vectors $[a=0, 1, 2, 3]$ and the Greek subscript labels the coordinate components of each vector $[\mu=0, 1, 2, 3]$. Then it can be shown that the metric tensor and the Ricci tensor of a Rainich geometry are given by the expansions

$$g_{\mu\nu} = \sum_{a=0}^{3} e_a \lambda_{a|\mu} \lambda_{a|\nu} \tag{1.2}$$

and

$$R_{\mu\nu} = \rho \sum_{a=0}^{3} e_a f_a \lambda_{a|\mu} \lambda_{a|\nu} \tag{1.3}$$

in which

$$g^{\mu\nu} \lambda_{a|\mu} \lambda_{b|\nu} \equiv \lambda_{a|\mu} \lambda_{b|}{}^{\mu} = e_a \delta_{ab} \tag{1.4}$$

$$e_0 = -1, \qquad e_1 = e_2 = e_3 = +1 \tag{1.5}$$

$$f_0 = f_1 = -1, \quad f_2 = f_3 = +1, \tag{1.6}$$

and the ρ in (1.3) is a positive scalar invariant.

One set of additional requirements must be imposed on a Riemannian manifold if it is to be a Rainich geometry. Let the linear operators

$$\partial_a \equiv \lambda_{a|}{}^{\mu} \partial / \partial x^{\mu} \tag{1.7}$$

represent differentiation in the Ricci principal directions and define the associated *structure coefficients* by writing the commutation relations

$$[\partial_a, \partial_b] \equiv \partial_a \partial_b - \partial_b \partial_a = \sum_{c=0}^{3} C_{cab} \partial_c. \tag{1.8}$$

From (1.7) it follows that the structure coefficients are computed according to the formulas

$$C_{cab} = -C_{cba} = e_c (\lambda_{c|\mu,\nu} - \lambda_{c|\nu,\mu}) \lambda_{a|}{}^{\mu} \lambda_{b|}{}^{\nu}. \tag{1.9}$$

In terms of the intrinsic notation established by (1.7) and (1.9), four of the structure coefficients must be derived from a scalar invariant α, the so-called *complexion* of the electromagnetic field. More precisely, we have the conditions

$$\begin{aligned} C_{123} &= \partial_0 \alpha \quad C_{032} = \partial_1 \alpha \\ C_{310} &= \partial_2 \alpha \quad C_{201} = \partial_3 \alpha. \end{aligned} \tag{1.10}$$

The Ricci vierbein of principal directions $\lambda_{a|\mu}$ satisfies (1.2), (1.3), and (1.10) if the Riemannian manifold fulfills the conditions of a Rainich geometry. However, the Ricci vierbein is not uniquely determined by a given Rainich geometry, because (1.2), (1.3), and (1.10) preserve their form if we make the replacements

$$\begin{aligned} \lambda_{0|\mu} &\to \lambda_{0|\mu}{}' = (\cosh\phi)\lambda_{0|\mu} + (\sinh\phi)\lambda_{1|\mu} \\ \lambda_{1|\mu} &\to \lambda_{1|\mu}{}' = (\cosh\phi)\lambda_{1|\mu} + (\sinh\phi)\lambda_{0|\mu} \\ \lambda_{2|\mu} &\to \lambda_{2|\mu}{}' = (\cos\theta)\lambda_{2|\mu} + (\sin\theta)\lambda_{3|\mu} \\ \lambda_{3|\mu} &\to \lambda_{3|\mu}{}' = (\cos\theta)\lambda_{3|\mu} - (\sin\theta)\lambda_{2|\mu}, \end{aligned} \tag{1.11}$$

with ϕ and θ arbitrary differentiable functions. The transformation $\lambda \to \lambda'$, defined by (1.11), is called an *internal gauge transformation*, and a specification of the functions ϕ and θ is called an *internal gauge*. Later we shall describe special internal gauges which facilitate the analysis of Rainich geometries.

A set of important relations is readily derived from (1.3). Substituting this equation into the contracted Bianchi identities

$$(R_\mu{}^\nu - \tfrac{1}{2} R \delta_\mu{}^\nu)_{;\nu} = 0 \tag{1.12}$$

and making use of the definitions (1.7) and (1.9), we

[1] G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).
[2] C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).
[3] G. Rosen, Phys. Rev. **114**, 1179 (1959).
[4] L. Witten, Phys. Rev. **115**, 206 (1959).
[5] L. Witten, Phys. Rev. **120**, 635 (1960).

eventually obtain equations for ρ,

$$\begin{aligned}\partial_0(\ln\rho)&=2(C_{202}+C_{303})\\ \partial_1(\ln\rho)&=2(C_{212}+C_{313})\\ \partial_2(\ln\rho)&=2(C_{020}+C_{121})\\ \partial_3(\ln\rho)&=2(C_{030}+C_{131}).\end{aligned}\tag{1.13}$$

Once we have solved the above conditions for a particular Rainich geometry, the associated electromagnetic field can be expressed in terms of the Ricci vierbein $\lambda_{a|\mu}$ and the scalar invariants ρ and α. The electromagnetic field tensor is

$$F_{\mu\nu}=c^2\left(\frac{\rho}{G}\right)^{\frac{1}{2}}[(\cos\alpha)(\lambda_{0|\mu}\lambda_{1|\nu}-\lambda_{0|\nu}\lambda_{1|\mu})+(\sin\alpha)(\lambda_{3|\mu}\lambda_{2|\nu}-\lambda_{3|\nu}\lambda_{2|\mu})],\tag{1.14}$$

where c and G denote well-known physical constants.

II. GROUPS OF MOTIONS IN RAINICH GEOMETRY

In the preceding section we have reviewed the general properties of Rainich geometry, the geometrical formulation of the Einstein-Maxwell equations. Now we shall study Rainich geometries which can be characterized by a group of motions.[6] Symmetrical solutions of the Einstein-Maxwell equations are associated with Rainich geometries of this type.

The infinitesimal generators of a group of motions are the linear operators

$$X^{(r)}=k^{(r)\mu}\partial/\partial x^\mu\tag{2.1}$$

where the superscript (r) labels the linearly independent solutions to Killing's equation

$$k_{\mu;\nu}{}^{(r)}+k_{\nu;\mu}{}^{(r)}=0.\tag{2.2}$$

If the superscript (r) enumerates a complete set of linearly independent solutions to (2.2), then the commutator of two distinct X's is expressible as a linear combination of a certain number of the X's and the group is said to be *complete*.

The Rainich conditions preclude Riemannian manifolds which support more than four (linearly independent) X's, as shown by applying well-known existence theorems to some of the results below, in particular, the condition (2.16). Thus, the order of a complete group of motions in Rainich geometry is less than or equal to four. Furthermore, a group with four X's must be *intransitive*, that is, the rank of the matrix $k^{(r)\mu}$ is always less than four. The Reissner-Nordstrom solution (for an electrically charged masspoint at rest) is an example of a Rainich geometry with maximal symmetry, characterized by four X's that generate displacements in time and rotations about three mutually perpendicular spatial axes. Other static Rainich geometries with maximal symmetry, and several with symmetry less than maximal, appear in the literature.[7,8] An example of a nonstatic Rainich geometry with maximal symmetry is obtained from the spatially homogeneous Rainich geometry (reported below) by setting $b_2=b_3=1$ in the line element (4.10).

Rainich geometry assumes its most tractable form in terms of the invariants associated with the Ricci vierbein. In order to determine how a group of motions influences the analytic properties of a Rainich geometry, we shall translate Killing's Eq. (2.2) into equivalent conditions on the Ricci vierbein. Our main results concern the invariants $X_{ab}{}^{(r)}$, where

$$[\partial_a,X^{(r)}]\equiv\sum_{b=0}^{3}X_{ab}{}^{(r)}\partial_b.\tag{2.3}$$

Let us also introduce the expansion

$$k_\mu{}^{(r)}=\sum_{a=0}^{3}K_a{}^{(r)}\lambda_{a|\mu},\tag{2.4}$$

and recall (2.1), (1.7), and (1.8). Then according to (2.3),

$$X_{ab}{}^{(r)}\equiv\partial_aK_b{}^{(r)}+\sum_{c=0}^{3}C_{bac}K_c{}^{(r)}.\tag{2.5}$$

Theorem 1:

$$e_aX_{ab}{}^{(r)}+e_bX_{ba}{}^{(r)}=0.\tag{2.6}$$

Proof: As a consequence of (1.9) and (1.4), we find

$$e_aC_{abc}+e_bC_{bac}=(\lambda_{c|\mu;\nu}+\lambda_{c|\nu;\mu})\lambda_a{}^\mu\lambda_b{}^\nu.\tag{2.7}$$

Thus, (2.6) follows by putting (2.4) into (2.2) and invoking the definition (2.5).

Corollary:

$$X_{aa}{}^{(r)}=0.\tag{2.8}$$

Theorem 2:

$$X_{ab}{}^{(r)}=0\quad\text{for}\quad f_a\neq f_b.\tag{2.9}$$

Proof[9]: The identities

$$k_{\mu;\nu\sigma}{}^{(r)}-k_{\mu;\sigma\nu}{}^{(r)}=k_\tau{}^{(r)}R_{\mu\nu\sigma}{}^\tau\tag{2.10}$$

and

$$R_{\mu\nu\sigma}{}^\tau+R_{\nu\sigma\mu}{}^\tau+R_{\sigma\mu\nu}{}^\tau=0\tag{2.11}$$

imply that a solution to (2.2) also satisfies

$$k_{\mu;\nu\sigma}{}^{(r)}=k_\tau{}^{(r)}R_{\sigma\nu\mu}{}^\tau.\tag{2.12}$$

Integrability conditions for the last equation take the form

$$k_{\tau;\mu}{}^{(r)}R_{\nu\omega\sigma}{}^\tau+k_{\tau;\nu}{}^{(r)}R_{\mu\sigma\omega}{}^\tau+k_{\tau;\omega}{}^{(r)}R_{\sigma\mu\nu}{}^\tau+k_{\tau;\sigma}{}^{(r)}R_{\omega\nu\mu}{}^\tau+k^{(r)\tau}R_{\mu\sigma\omega\nu;\tau}=0.\tag{2.13}$$

[6] Lie's theory of continuous groups was first applied to Riemannian geometry by Killing, Fubini, and Ricci. The basic theorems have been surveyed by L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1949), Chap. VI.

[7] W. B. Bonnor, Proc. Phys. Soc. (London) **A66**, 145 (1953).

[8] A. K. Raychaudhuri, Ann. Phys. **11**, 501 (1960).

[9] Equivalent expressions for some of the steps in this proof have been given by G. Ricci, Rend. Lincei **14**, 489 (1905).

Contracting ν and σ, we obtain

$$k_{\tau;\mu}{}^{(r)}R_{\omega}{}^{\tau}+k_{\tau;\omega}{}^{(r)}R_{\mu}{}^{\tau}+k^{(r)\tau}R_{\mu\omega;\tau}=0, \quad (2.14)$$

a relation which indicates that the manifold of solutions to Killing's Eq. (2.2) is constrained by the Ricci tensor.

Next, multiply (2.14) by $\lambda_{a|}{}^{\mu}\lambda_{b|}{}^{\omega}$ and sum over μ and ω. Recalling (1.3), (1.9), (2.1), (2.4), and (2.5), we find

$$\rho(f_a-f_b)\mathrm{X}_{ab}{}^{(r)}-f_a\delta_{ab}(X^{(r)}\rho)=0. \quad (2.15)$$

Hence (2.9) follows from (2.15) by taking $a\neq b$. For $a=b$ we have:

Corollary:

$$X^{(r)}\rho=0. \quad (2.16)$$

In view of (2.8) and (2.9) most of the $\mathrm{X}_{ab}{}^{(r)}$ vanish. The surviving components of $\mathrm{X}_{ab}{}^{(r)}$, which are not necessarily zero, are

$$\mathrm{X}_{-}{}^{(r)}\equiv\mathrm{X}_{01}{}^{(r)}=\mathrm{X}_{10}{}^{(r)}$$

and (2.17)

$$\mathrm{X}_{+}{}^{(r)}\equiv\mathrm{X}_{23}{}^{(r)}=-\mathrm{X}_{32}{}^{(r)}.$$

Putting (2.8), (2.9), and (2.17) into (2.3), we have

Theorem 3: Necessary and sufficient conditions for the X's to be the infinitesimal generators of a group of motions are

$$\begin{aligned}
&[\partial_0,X^{(r)}]-\mathrm{X}_{-}{}^{(r)}\partial_1=0\\
&[\partial_1,X^{(r)}]-\mathrm{X}_{-}{}^{(r)}\partial_0=0\\
&[\partial_2,X^{(r)}]-\mathrm{X}_{+}{}^{(r)}\partial_3=0\\
&[\partial_3,X^{(r)}]+\mathrm{X}_{+}{}^{(r)}\partial_2=0
\end{aligned} \quad (2.18)$$

where $\mathrm{X}_{-}{}^{(r)}$ and $\mathrm{X}_{+}{}^{(r)}$ represent arbitrary scalar invariants.

Proof: The necessary quality of (2.18) is obvious. That (2.18) provides sufficient conditions for the X's to be the infinitesimal generators of a group of motions follows from the converse of Theorem 1, proved by reversing the proof of Theorem 1.

Now we can prove

Theorem 4: In a special internal gauge, a *symmetry gauge* related to $X^{(r)}$, the necessary and sufficient conditions for $X^{(r)}$ become

$$[\partial_a,X^{(r)}]=0. \quad (2.19)$$

Proof: No reference to the internal gauge was required in order to derive (2.18), and therefore (2.18) must be valid in every internal gauge. Consequently, under an internal gauge transformation (1.11), $\mathrm{X}_{-}{}^{(r)}$ and $\mathrm{X}_{+}{}^{(r)}$ transform according to[10]

$$\begin{aligned}
&\mathrm{X}_{-}{}^{(r)}\longrightarrow\mathrm{X}_{-}{}^{(r)\prime}=\mathrm{X}_{-}{}^{(r)}-X^{(r)}\phi\\
&\mathrm{X}_{+}{}^{(r)}\longrightarrow\mathrm{X}_{+}{}^{(r)\prime}=\mathrm{X}_{+}{}^{(r)}-X^{(r)}\theta.
\end{aligned} \quad (2.20)$$

[10] We have taken a shortcut in order to get (2.20). Since the transformation character of $\mathrm{X}_{-}{}^{(r)}$ and $\mathrm{X}_{+}{}^{(r)}$ does not depend on the theorems which precede (2.18), a strict derivation of (2.20) should be based on the relations (1.11), (2.17), (2.5), (1.7), (1.9), and the internal gauge invariance of (2.4). Of course the result is the same.

Hence if the ϕ and θ are solutions of the equations

$$\begin{aligned}
&X^{(r)}\phi=\mathrm{X}_{-}{}^{(r)}\\
&X^{(r)}\theta=\mathrm{X}_{+}{}^{(r)}\quad [r\text{ fixed}],
\end{aligned} \quad (2.21)$$

the internal gauge transformation (2.20) produces

$$\mathrm{X}_{-}{}^{(r)\prime}=\mathrm{X}_{+}{}^{(r)\prime}=0\quad [r\text{ fixed}], \quad (2.22)$$

and (2.19) follows from (2.18) and (2.22) by dropping the primes.

An internal gauge which gives (2.19) is called a *symmetry gauge* related to $X^{(r)}$. Generally speaking, it is not possible to gauge out $\mathrm{X}_{-}{}^{(r)}$ and $\mathrm{X}_{+}{}^{(r)}$ for several values of (r) simultaneously.

Corollary: In a symmetry gauge related to $X^{(r)}$ we have

$$X^{(r)}C_{abc}=0. \quad (2.23)$$

Proof: Using (2.19) and (1.8), compute

$$\sum_{a=0}^{3}(X^{(r)}C_{abc})\partial_a=\sum_{a=0}^{3}[X^{(r)},C_{abc}\partial_a]=$$
$$[X^{(r)},[\partial_b,\partial_c]]=0, \quad (2.24)$$

which implies (2.23).

As an application of (2.19) and (2.23) we shall consider a motion in a "symmetrized" Ricci principal direction, that is, a motion which is everywhere parallel to the vierbein vector $\lambda_{m|\mu}$ in a related symmetry gauge. The infinitesimal generator of the motion is expressed as

$$X=K_m\partial_m. \quad (2.25)$$

So from (2.23), (2.19), and (1.8) we obtain conditions on the structure coefficients,

$$\begin{aligned}
&\partial_mC_{abc}=0\\
&C_{amb}=0\quad [a\neq m]\\
&C_{mmb}=\partial_b\ln|K_m|,
\end{aligned} \quad (2.26)$$

where

$$\partial_mK_m=0. \quad (2.27)$$

If the index m in (2.26) and (2.27) is set equal to zero, we obtain the conditions for static Rainich geometries in a related symmetry gauge

$$\begin{aligned}
&\partial_0C_{abc}=0\\
&C_{a0b}=0\quad [a\neq0]\\
&C_{00b}=\partial_b\ln|K_0|,
\end{aligned} \quad (2.28)$$

where

$$\partial_0K_0=0. \quad (2.29)$$

A comparison of (2.28) and (1.10) shows that α need not be identically constant for static Rainich geometries. However, the conditions in (2.28) do require $\partial_2\alpha=\partial_3\alpha=0$, that is, the normal to the $\alpha=$constant hypersurfaces is necessarily contained in the *negative blade* at every point.[3]

III. NORMAL RAINICH GEOMETRIES

In this section we investigate Rainich geometries for which the Ricci congruences are normal. Although mathematical difficulties cohibit the detailed analysis of most Rainich geometries, normalcy of the Ricci congruence permits immediate integration of the conditions in (1.10) and (1.13). Later we shall consider groups of motions for this special class of mathematically tractable Rainich geometries.

The *congruences* associated with the Ricci vierbein are the four sets of curves which are, respectively, tangent to the vectors $\lambda_{a|\mu}$. A congruence is said to be *normal* if its curves are orthogonal to a set of hypersurfaces. If the four congruences of the Ricci vierbein are normal, the Rainich geometry is said to be *normal*, and the Ricci vierbein takes the form

$$\lambda_{a|\mu}=e_a\omega_a\partial\sigma_a/\partial x^\mu \tag{3.1}$$

where the four ω's and the four σ's are scalar invariants. Observe that the ω's and σ's are not uniquely determined by the normalcy of the Ricci congruences, because (3.1) retains its functional form under the replacements

$$\begin{aligned}\omega_a &\to \omega_a{}^*\equiv(d\sigma_a/d\sigma_a{}^*)\omega_a\\ \sigma_a &\to \sigma_a{}^*\equiv\sigma_a{}^*(\sigma_a)\end{aligned} \tag{3.2}$$

where each $\sigma_a{}^*$ is an arbitrary monotonic function of the corresponding σ_a. However, the normalcy conditions (3.1) are *not* preserved under an internal gauge transformation (1.11), and thus (3.1) presupposes a particular internal gauge, the *normalcy gauge*. It should be emphasized that all Rainich geometries are not normal, in the sense that all Rainich geometries do not admit an internal gauge transformation which brings the vierbein into the form (3.1).

With the help of (1.4) and (1.7), the normalcy conditions (3.1) can be rewritten in terms of invariants,

$$\partial_a\sigma_b=\delta_{ab}/\omega_a. \tag{3.3}$$

Next, the structure coefficients can be computed from (1.9), (3.1), (1.7), and (3.3). In this way we find

$$C_{abc}=(\delta_{ab}\partial_c-\delta_{ac}\partial_b)\ln|\omega_a| \tag{3.4}$$

or, equivalently,

$$C_{abc}=0, \quad [b\neq a\neq c] \tag{3.5}$$

$$C_{aab}=-C_{aba}=\partial_b\ln|\omega_a|, \quad [a\neq b]. \tag{3.6}$$

Notice that Eqs. (3.5), the integrability conditions for the existence of σ's satisfying (3.3), are sufficient conditions for a normal Rainich geometry.

Now by virtue of (3.5), the four structure coefficients in (1.10) vanish. Therefore the Rainich conditions (1.10) are trivially satisfied, and for normal Rainich geometries we have

$$\alpha\equiv\text{const.} \tag{3.7}$$

The geometrical significance of $\alpha\equiv$constant Rainich geometries has been discussed previously.[3]

Geometrical coordinate conditions prepare the way for the integration of Eqs. (1.13). We choose the σ invariants as our coordinates, $x^a\equiv\sigma_a$, and (3.1) becomes

$$\lambda_{a|\mu}=e_a\omega_a\delta_{a\mu}. \tag{3.8}$$

Then the metric tensor (1.2) reduces to

$$g_{ab}=e_a\omega_a{}^2\delta_{ab}. \tag{3.9}$$

This special coordinate system also makes the ∂'s of (1.7) take the form

$$\partial_a=(1/\omega_a)\partial/\partial\sigma_a, \tag{3.10}$$

so that (3.6) becomes

$$C_{aab}=-C_{aba}=\frac{1}{\omega_a\omega_b}\frac{\partial\omega_a}{\partial\sigma_b}, \quad [a\neq b]. \tag{3.11}$$

Integrating (1.13), we obtain the results

$$\rho\omega_0{}^2\omega_1{}^2=\pi_-(\sigma_0,\sigma_1) \tag{3.12}$$

$$\rho\omega_2{}^2\omega_3{}^2=\pi_+(\sigma_2,\sigma_3) \tag{3.13}$$

where π_- and π_+ are arbitrary positive functions of the indicated coordinates.

Equations (3.7), (3.12), and (3.13) constitute first integrals for a normal Rainich geometry; Eq. (1.3) expresses conditions which remain to be fulfilled. Using the coordinate system of (3.8), let us calculate the Ricci tensor associated with the metric (3.9). This somewhat lengthy computation gives the left side of (1.3). Then Eq. (1.3) splits up into the conditions

$$\sum_{c(\neq a,b)=0}^{3}\left[\frac{\partial_a(\partial_b\omega_c)}{\omega_c}-\frac{(\partial_a\omega_c)(\partial_b\omega_a)}{\omega_a\omega_c}\right]=0, \quad [a\neq b] \tag{3.14}$$

$$\sum_{c(\neq a)=0}^{3}\left[\frac{e_a\partial_a{}^2\omega_c}{\omega_c}+\frac{e_c\partial_c{}^2\omega_a}{\omega_a}+\sum_{b(\neq a,c)=0}^{3}\frac{e_b(\partial_b\omega_c)(\partial_b\omega_a)}{\omega_a\omega_c}\right]+f_a\rho=0 \tag{3.15}$$

where the ∂'s are given by (3.10). All solutions of (3.14) and (3.15) satisfy Eqs. (3.12) and (3.13), a fact which facilitates the integration of (3.14) and (3.15).

The best-known example of a normal Rainich geometry is the Reissner-Nordstrom solution.[11,12] With the Ricci vierbein in the form (3.8) (normalcy gauge and σ's as coordinates), the Reissner-Nordstrom solution is

$$\begin{aligned}\omega_0&=\left(1-\frac{2m}{\sigma_1}+\frac{\epsilon^2}{\sigma_1{}^2}\right)^{\frac{1}{2}}\\ \omega_1&=\left(1-\frac{2m}{\sigma_1}+\frac{\epsilon^2}{\sigma_1{}^2}\right)^{-\frac{1}{2}}\\ \omega_2&=\sigma_1\\ \omega_3&=\sigma_1\sin\sigma_2\\ \rho&=\epsilon^2/\sigma_1{}^4,\end{aligned} \tag{3.16}$$

[11] K. Reissner, Ann. phys. **50**, 106 (1916).
[12] L. Nordstrom, Proc. Amsterdam Acad. **20**, 1238 (1918).

where m and ϵ are constants. It is easy to verify that (3.16) satisfies Eqs. (3.12)–(3.15).

Let us consider normal Rainich geometries which support a motion in a Ricci principal direction. In this connection, we assert the results:

Theorem 5: Normalcy gauge is a symmetry gauge for a motion everywhere parallel to a member of the Ricci vierbein. $\lambda_{m|\mu}$. Furthermore, an appropriate transformation (3.2) achieves

$$X^{(m)}=\partial/\partial\sigma_m \tag{3.17}$$

$$X^{(m)}\omega_a=0. \tag{3.18}$$

Proof: Equation (2.25) gives the infinitesimal generator for a motion everywhere parallel to $\lambda_{m|\mu}$. Necessary and sufficient conditions for X to be the infinitesimal generator of a motion are stated in (2.18). The structure coefficients of a normal Rainich geometry are given by (3.5) and (3.6). Combining these equations, we have

$$\begin{aligned}
&\left(\omega_m\partial_0-\frac{K_m}{\omega_m}\right)\partial_m+\left(\frac{K_m}{\omega_0}\partial_m\omega_0\right)\partial_0-\mathrm{X}_-\partial_1=0,\\
&\left(\omega_m\partial_1-\frac{K_m}{\omega_m}\right)\partial_m+\left(\frac{K_m}{\omega_1}\partial_m\omega_1\right)\partial_1-\mathrm{X}_-\partial_0=0,\\
&\left(\omega_m\partial_2-\frac{K_m}{\omega_m}\right)\partial_m+\left(\frac{K_m}{\omega_2}\partial_m\omega_2\right)\partial_2-\mathrm{X}_+\partial_3=0,\\
&\left(\omega_m\partial_3-\frac{K_m}{\omega_m}\right)\partial_m+\left(\frac{K_m}{\omega_3}\partial_m\omega_3\right)\partial_3+\mathrm{X}_+\partial_2=0.
\end{aligned} \tag{3.19}$$

Since the ∂'s are linearly independent operators, Eqs. (3.19) imply

$$\mathrm{X}_-=\mathrm{X}_+=0, \tag{3.20}$$

which shows that the motion is expressed in a symmetry gauge [see (2.22)]. From (3.19) we also deduce the relations

$$\begin{aligned}
&\partial_m\omega_a=0, \quad [a\neq m]\\
&\partial_a(K_m/\omega_m)=0, \quad [a\neq m]\\
&\partial_mK_m=0.
\end{aligned} \tag{3.21}$$

In terms of the σ coordinates introduced above, Eqs. (3.21) become

$$\begin{aligned}
&\partial\omega_a/\partial\sigma_m=0, \quad [a\neq m]\\
&(\partial/\partial\sigma_m)(\gamma\omega_m)=0
\end{aligned} \tag{3.22}$$

where $\gamma\equiv K_m/\omega_m$ is a function of σ_m. Therefore an appropriate transformation (3.2), defined by $d\sigma_m/d\sigma_m{}^*=\gamma$, achieves

$$\begin{aligned}
&\partial/\partial\sigma_m{}^*=\gamma\partial/\partial\sigma_m=K_m\partial_m=X^{(m)}\\
&\omega_m{}^*=\gamma\omega_m.
\end{aligned} \tag{3.23}$$

Dropping the stars, (3.17) and (3.18) follow from (3.23) and (3.22).

Of the four infinitesimal generators associated with the Reissner-Nordstrom solution (3.16),

$$\begin{aligned}
X^{(0)}&=\partial/\partial\sigma_0\\
X^{(1)}&=-(\sin\sigma_3)\frac{\partial}{\partial\sigma_2}-\left(\frac{\cos\sigma_3}{\tan\sigma_2}\right)\frac{\partial}{\partial\sigma_3}\\
X^{(2)}&=(\cos\sigma_3)\frac{\partial}{\partial\sigma_2}-\left(\frac{\sin\sigma_3}{\tan\sigma_2}\right)\frac{\partial}{\partial\sigma_3}\\
X^{(3)}&=\partial/\partial\sigma_3,
\end{aligned} \tag{3.24}$$

we observe that $X^{(0)}$ and $X^{(3)}$ generate motions parallel to the vierbein members $\lambda_{0|\mu}$ and $\lambda_{3|\mu}$, respectively. In accord with (3.17) and (3.18), the ω's in (3.16) do not depend on σ_0 and σ_3. The quantity ρ in (3.16) is a function of σ_1 as a consequence of (2.16) and (3.24).

IV. SPATIALLY HOMOGENEOUS RAINICH GEOMETRY

The preceding theory suggests new solutions to the Einstein-Maxwell equations. As an illuminating example, we shall study a normal Rainich geometry which supports motions in the three space-like Ricci principal directions, $\lambda_{i|\mu}$ with $i=1, 2, 3$. If we choose appropriate σ's as coordinates, then (3.17) and (3.18) states that

$$\partial\omega_a/\partial\sigma_i=0, \quad [i=1, 2, 3]. \tag{4.1}$$

As a consequence, Eqs. (3.14) are automatically satisfied. Moreover, Eqs. (3.15) reduce to

$$\begin{aligned}
&\sum_{j=1}^{3}\frac{\partial_0^2\omega_j}{\omega_j}+\rho=0\\
&\partial_0[\omega_1\omega_2\omega_3(\partial_0\omega_i/\omega_i)]-f_i\rho\omega_1\omega_2\omega_3=0, \quad [i=1, 2, 3]
\end{aligned} \tag{4.2}$$

where, by (3.10), $\partial_0=(1/\omega_0)d/d\sigma_0$.

To integrate the total differential equations (4.2), first recall the general result (3.13). In view of (4.1) and (2.16), the quantities on the left side of (3.13) are functions of σ_0, and therefore (3.13) implies that

$$\rho\omega_2{}^2\omega_3{}^2=a^2\equiv\text{a positive constant}. \tag{4.3}$$

Equation (4.3) is a first integral for the system (4.2).

Next, let us define a new time coordinate,

$$t=a\int^{\sigma_0}\frac{\omega_0d\sigma_0}{\omega_2\omega_3}+(\text{a constant}). \tag{4.4}$$

Introducing (4.4) into (4.2) and making use of (4.3),

we have the system of equations

$$\sum_{j=1}^{3} \frac{\omega_2\omega_3}{\omega_j} \frac{d}{dt}\left(\frac{1}{\omega_2\omega_3}\frac{d\omega_j}{dt}\right)+1=0 \tag{4.5}$$

$$\frac{d}{dt}\left(\frac{\omega_1}{\omega_i}\frac{d\omega_i}{dt}\right)-f_i\omega_1=0, \quad [i=1, 2, 3]. \tag{4.6}$$

Equations (4.6) are easily integrated. With a suitable adjustment of the constant in (4.4), we find

$$\begin{aligned} \omega_1&=a_1 \sin t \\ \omega_2&=\frac{a_2}{\sin t}(\tan\tfrac{1}{2}t)^{b_2} \\ \omega_3&=\frac{a_3}{\sin t}(\tan\tfrac{1}{2}t)^{b_3} \end{aligned} \tag{4.7}$$

in which the a's and the b's are constants. After a computation we learn that (4.5) imposes a simple condition on the integration constants in (4.7), namely,

$$b_2b_3=1. \tag{4.8}$$

Finally, by virtue of (4.3), we write

$$\rho=\frac{(\sin t)^4}{b_1^2(\tan\tfrac{1}{2}t)^{2(b_2+b_3)}} \tag{4.9}$$

with the constant $b_1\equiv a_2a_3/a$.

Equations (4.7), (4.8), and (4.9) describe a new solution to the Einstein-Maxwell equations, a *spatially homogeneous Rainich geometry*. The line element for the geometry is found by combining (3.9), (4.4), and (4.7),

$$\begin{aligned} (ds)^2&=\sum_{a=0}^{3} e_a\omega_a^2(d\sigma_a)^2 \\ &=-\frac{b_1^2(\tan\tfrac{1}{2}t)^{2(b_2+b_3)}}{(\sin t)^4}(dt)^2+(\sin t)^2(dx_1)^2 \\ &\quad+\frac{(\tan\tfrac{1}{2}t)^{2b_2}}{(\sin t)^2}(dx_2)^2+\frac{(\tan\tfrac{1}{2}t)^{2b_3}}{(\sin t)^2}(dx_3)^2 \end{aligned} \tag{4.10}$$

where we have made a scale transformation for each spatial coordinate,

$$x_i\equiv a_i\sigma_i, \quad [i=1, 2, 3]. \tag{4.11}$$

Two independent constants remain in (4.10), b_1 and either b_2 or b_3 $(=1/b_2)$, and they provide an intrinsic parametrization of the solution. Notice that singularities appear in the line element (4.10) at $t=0$ and $t=\pi$.

The associated electromagnetic field is found by putting (3.8), (4.7), and (4.9) into (1.14). Remembering that we have made the coordinate transformations (4.4) and (4.11), the nonvanishing components of the electromagnetic field are found to be

$$\begin{aligned} F_{01}&=-F_{10}=-\frac{c^2\cos\alpha}{(G)^{\frac{1}{2}}}\sin t \\ F_{23}&=-F_{32}=-\frac{c^2\sin\alpha}{(G)^{\frac{1}{2}}b_1} \end{aligned} \tag{4.12}$$

where the constant α is at our disposal [see (3.7)]. Thus, the electric field in the x_1 direction oscillates against a constant magnetic field in the x_1 direction.

Continuum Coulomb Wave Function*

W. R. JOHNSON AND R. T. DECK†
University of Notre Dame, Notre Dame, Indiana

(Received July 5, 1961)

An alternate form for the continuum-state relativistic Coulomb wave function is developed which expresses the spin dependence in terms of the usual Dirac plane wave spinor. The computational convenience of this form is indicated by a direct derivation of the Sommerfeld-Maue approximation.

I. ALTERNATE FORM FOR CONTINUUM-STATE SOLUTION OF DIRAC EQUATION FOR A COULOMB POTENTIAL

THE continuum-state solution of the Dirac equation for a Coulomb potential has been given by Darwin as an eigenfunction of asymptotic momentum and polarization. A solution of this character is not obtainable in closed form. In this paper it is shown, however, that the Darwin solution can be written in a more convenient form in terms of the usual Dirac plane wave spinors.

The Dirac equation for a Coulomb potential is written[1]

$$(ip+m)\psi=(\alpha Z/r)\gamma_4\psi. \tag{1}$$

Darwin's solution which asymptotically exhibits the character of a plane wave plus an outgoing spherical wave has the form[2]

$$\psi=4\pi\left(\frac{W+m}{2W}\right)^{\frac{1}{2}}\sum_{jlm_j}(\Omega_{jlm_j}{}^{\dagger}(\hat{p})v)\begin{pmatrix}iG_\kappa\Omega_{jlm_j}(\hat{r})\\ F_\kappa\Omega_{jl'm_j}(\hat{r})\end{pmatrix},$$

$$l'=2j-l \tag{2}$$

where the summation extends over $j\geq\frac{1}{2}$, $-j\leq m_j\leq j$, and $l=j\pm\frac{1}{2}$. In Eq. (2) we have introduced the following notation:

$$\left.\begin{aligned} iG_\kappa&=ie^{i\delta_\kappa+il\pi+\nu\pi/2}(2pr)^{\gamma-1}\frac{|\Gamma(\gamma+i\nu)|}{\Gamma(2\gamma+1)}\{\ \}_+\\ F_\kappa&=i\frac{p}{W+m}e^{i\delta_\kappa+il\pi+\nu\pi/2}(2pr)^{\gamma-1}\frac{|\Gamma(\gamma+i\nu)|}{\Gamma(2\gamma+1)}\{\ \}_-\end{aligned}\right\} \tag{3}$$

$$\{\ \}_\pm=\{e^{-ipr+i\eta}(\gamma+i\nu)_1F_1(\gamma+1+i\nu;\,2\gamma+1;\,+2ipr)$$
$$\pm e^{ipr-i\eta}(\gamma-i\nu)_1F_1(\gamma+1-i\nu;\,2\gamma+1;\,-2ipr)\}$$

$$\kappa=\begin{cases}-(j+\frac{1}{2}) & \text{for } l=j-\frac{1}{2}\\ +(j+\frac{1}{2}) & \text{for } l=j+\frac{1}{2}\end{cases}$$

$$\Omega_{jlm_j}=\left.\begin{aligned}&\begin{pmatrix}[(j+m_j)2j]^{\frac{1}{2}}Y_{l,m_j-\frac{1}{2}}\\ [(j-m_j)2j]^{\frac{1}{2}}Y_{l,m_j+\frac{1}{2}}\end{pmatrix} && l=j-\tfrac{1}{2}\\ &\begin{pmatrix}[(j+1-m_j)/(2j+2)]^{\frac{1}{2}}Y_{l,m_j-\frac{1}{2}}\\ -[(j+1+m_j)/(2j+2)]^{\frac{1}{2}}Y_{l,m_j+\frac{1}{2}}\end{pmatrix} && l=j+\tfrac{1}{2}\end{aligned}\right\} \tag{4}$$

$$Y_{lm}=Y_{lm}(\theta,\varphi),\quad Y_{l,-|m|}\equiv(-1)^mY_{l,|m|}$$

$$\left.\begin{aligned}&\delta_\kappa=\eta-\tfrac{1}{2}\pi\gamma-\arg\Gamma(\gamma+i\nu)\\ &e^{2i\eta}=-\frac{\gamma-i\nu}{\kappa+i\nu'}=-\frac{\kappa-i\nu'}{\gamma+i\nu}\\ &\gamma=(k^2-\alpha^2Z^2)^{\frac{1}{2}}\\ &k=|\kappa|=(j+\tfrac{1}{2})\\ &\nu=\alpha ZW/p,\quad \nu'=\alpha Zm/p.\end{aligned}\right\} \tag{5}$$

The two component matrix v represents the "large component" of the Dirac plane wave spinor[3]

$$u=\left(\frac{W+m}{2W}\right)^{\frac{1}{2}}\begin{pmatrix}v\\ w\end{pmatrix} \tag{6}$$

$$v=\begin{pmatrix}A\\ B\end{pmatrix},\quad v^{\dagger}v=1 \tag{7}$$

$$w=\frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{W+m}v. \tag{8}$$

We seek to express the spin dependence of ψ in terms of the spinor u. To this end it is sufficient to consider a system of coordinates in which the polar z axis is oriented along the direction of the asymptotic momentum vector $\hat{p}$ (Fig. 1). Since the results to

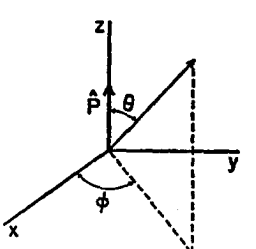


FIG. 1. A convenient choice of coordinate axes.

* Supported in part by the U. S. Atomic Energy Commission.
† Now at Bartol Research Foundation, Swarthmore, Pennsylvania.
[1] Natural units are used in the following with $\hbar=c=1$.
[2] See, for instance, A. I. Akhiezer and V. B. Berestetsky, *Quantum Electrodynamics*, AEC-tr-2876 (Pt. I). The normalization in Eq. (2) is chosen such that the coefficient of the plane wave asymptotically approaches one.
[3] The usual representation for the Pauli σ matrices is used in the following with σ_z chosen to be diagonal.

follow will be expressed in vector form, they will be independent of the orientation of the coordinate system used here.

Equation (2) can be written

$$\psi=\left(\frac{W+m}{2W}\right)^{\frac{1}{2}}\begin{pmatrix}\varphi\\ \chi\end{pmatrix}, \tag{9}$$

with

$$\left.\begin{aligned}\varphi&=4\pi\sum_{jlm_j}[\Omega_{jlm_j}{}^{\dagger}(\hat{p})v]iG_\kappa\Omega_{jlm_j}(\hat{r})\\ \chi&=4\pi\sum_{jlm_j}[\Omega_{jlm_j}{}^{\dagger}(\hat{p})v]F_\kappa\Omega_{jl'm_j}(\hat{r}).\end{aligned}\right\} \tag{10}$$

For the geometry of Fig. 1, $\Omega_{jlm_j}(\hat{p})$ assumes the form

$$\Omega_{jlm_j}(\hat{p})=\Omega_{jlm_j}(0,\varphi_{\hat{p}})=\left(\frac{|\kappa|}{4\pi}\right)^{\frac{1}{2}}\begin{pmatrix}\delta_{m_j,\frac{1}{2}}\\ -(\kappa/|\kappa|)\delta_{m_j,-\frac{1}{2}}\end{pmatrix}. \tag{11}$$

The eigenvalue m_j in Eq. (10) thus has the two possible values $\frac{1}{2}$ and $-\frac{1}{2}$. The corresponding spherical spinors are

$$\left.\begin{aligned}\Omega_{jl\frac{1}{2}}&=\begin{pmatrix}(|\kappa|/4\pi)^{\frac{1}{2}}P_l\\ (1/4\pi|\kappa|)^{\frac{1}{2}}(\kappa/|\kappa|)P_l{}^1e^{i\varphi}\end{pmatrix}\\ \Omega_{jl,-\frac{1}{2}}&=\begin{pmatrix}(1/4\pi|\kappa|)^{\frac{1}{2}}P_l{}^1e^{-i\varphi}\\ -(|\kappa|/4\pi)^{\frac{1}{2}}(\kappa/|\kappa|)P_l\end{pmatrix}.\end{aligned}\right\} \tag{12}$$

It follows that

$$\left.\begin{aligned}\varphi&=\sum_{jl}iG_\kappa\begin{pmatrix}A|\kappa|P_l-B(\kappa/|\kappa|)P_l{}^1e^{-i\varphi}\\ B|\kappa|P_l+A(\kappa/|\kappa|)P_l{}^1e^{i\varphi}\end{pmatrix}\\ \chi&=\sum_{jl}F_\kappa\begin{pmatrix}A|\kappa|P_{l'}-B(\kappa/|\kappa|)P_{l'}{}^1e^{-i\varphi}\\ -B|\kappa|P_{l'}-A(\kappa/|\kappa|)P_{l'}{}^1e^{i\varphi}\end{pmatrix}.\end{aligned}\right\} \tag{13}$$

We note that φ can be written

$$\varphi=a\begin{pmatrix}A\\ B\end{pmatrix}+Ob\begin{pmatrix}A\\ B\end{pmatrix}=(a+Ob)v, \tag{14}$$

where

$$a=\sum_{jl}iG_\kappa|\kappa|P_l,\quad b=\sum_{jl}iG_\kappa\frac{\kappa}{|\kappa|}P_l{}^1 \tag{15}$$

and

$$O=\begin{pmatrix}0 & -e^{-i\varphi}\\ e^{i\varphi} & 0\end{pmatrix}. \tag{16}$$

Further, with

$$w=\frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{W+m}v=\left(\frac{W-m}{W+m}\right)^{\frac{1}{2}}\begin{pmatrix}A\\ -B\end{pmatrix}, \tag{17}$$

χ can be written

$$\chi=c\left(\frac{W-m}{W+m}\right)^{\frac{1}{2}}\begin{pmatrix}A\\ -B\end{pmatrix}-Od\left(\frac{W-m}{W+m}\right)^{\frac{1}{2}}\begin{pmatrix}A\\ -B\end{pmatrix}$$
$$=(c-Od)w, \tag{18}$$

where

$$\left.\begin{aligned}c&=\left(\frac{W+m}{W-m}\right)^{\frac{1}{2}}\sum_{jl}F_\kappa|\kappa|P_{l'},\\ d&=\left(\frac{W+m}{W-m}\right)^{\frac{1}{2}}\sum_{jl}F_\kappa\frac{\kappa}{|\kappa|}P_{l'}{}^1.\end{aligned}\right\} \tag{19}$$

Setting $k=|\kappa|$, the term a in Eq. (15) is expressible as a sum over k alone.

$$a=\sum_{jl}iG_\kappa|\kappa|P_l=\sum_{j=\frac{1}{2}}^{\infty}i(G_\kappa|\kappa|P_{j+\frac{1}{2}}+G_\kappa|\kappa|P_{j-\frac{1}{2}}) \tag{20.1}$$

$$a=\sum_k a_k=\sum_{k=1}^{\infty}(iG_kkP_k+iG_{-k}kP_{k-1}).$$

In a similar manner,

$$b=\sum_k b_k=\sum_{k=1}^{\infty}(iG_kP_k{}^1-iG_{-k}P_{k-1}{}^1) \tag{20.2}$$

$$c=\sum_k c_k=\sum_{k=1}^{\infty}\left(\frac{W+m}{W-m}\right)^{\frac{1}{2}}(F_kkP_{k-1}+F_{-k}kP_k) \tag{20.3}$$

$$d=\sum_k d_k=\sum_{k=1}^{\infty}\left(\frac{W+m}{W-m}\right)^{\frac{1}{2}}(F_kP_{k-1}{}^1-F_{-k}P_k{}^1). \tag{20.4}$$

The functions F_κ and G_κ, as defined in Eq. (3), can be simplified by use of the Kummer transformation,[4]

$$e^{-ipr}{}_1F_1(\gamma+1+i\nu;2\gamma+1;2ipr)=e^{ipr}{}_1F_1(\gamma-i\nu;2\gamma+1;-2ipr),$$

whence

$$\{\ \}_\pm=e^{ipr-i\eta}\{-(\kappa-i\nu'){}_1F_1(\gamma-i\nu;2\gamma+1;-2ipr)\pm(\gamma-i\nu){}_1F_1(\gamma+1-i\nu;2\gamma+1;-2ipr)\}.$$

With $\exp[-i\arg\Gamma(\gamma+i\nu)]$ written as

$$\Gamma(\gamma-i\nu)/|\Gamma(\gamma+i\nu)|,$$

$$iG_{\pm k}=\frac{\Gamma(\alpha)}{\Gamma(\beta)}\exp(\nu\pi/2-i\pi\gamma/2+ik\pi\pm i\pi/2)(ix)^{\gamma-1}e^{-x/2}\times\{-(\pm k-i\nu'){}_1F_1(\alpha;\beta;x)+\alpha_1F_1(\alpha+1;\beta;x)\} \tag{21}$$

$$\left(\frac{W+m}{W-m}\right)^{\frac{1}{2}}F_{\pm k}=\frac{\Gamma(\alpha)}{\Gamma(\beta)}\exp[\nu\pi/2-i\pi\gamma/2+ik\pi\pm i\pi/2](ix)^{\gamma-1}e^{-x/2}\times\{-(\pm k-i\nu'){}_1F_1(\alpha;\beta;x)-\alpha_1F_1(\alpha+1;\beta;x)\}. \tag{22}$$

[4] *Higher Transcendental Functions*, Bateman Manuscript Project, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1, p. 253.

Here

$$\left.\begin{aligned} x&=-2ipr \\ \alpha&=\gamma-i\nu, \quad \beta=2\gamma+1. \end{aligned}\right\} \tag{23}$$

We introduce for convenience the further notation:

$$\left.\begin{aligned} \xi_k&=-k(-1)^k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}{}_1F_1(\alpha;\beta;x) \\ \zeta_k&=i\nu'(-1)^k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}{}_1F_1(\alpha;\beta;x) \\ \lambda_k&=\alpha(-1)^k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}{}_1F_1(\alpha+1;\beta;x), \end{aligned}\right\} \tag{24}$$

in terms of which

$$\left.\begin{aligned} a_k&=\xi_k k(P_k+P_{k-1})+\zeta_k k(P_k-P_{k-1})+\lambda_k k(P_k-P_{k-1}) \\ b_k&=\xi_k(P_k{}^1-P_{k-1}{}^1)+\zeta_k(P_k{}^1+P_{k-1}{}^1)+\lambda_k(P_k{}^1+P_{k-1}{}^1) \\ c_k&=\xi_k k(P_k+P_{k-1})-\zeta_k k(P_k-P_{k-1})+\lambda_k k(P_k-P_{k-1}) \\ d_k&=\xi_k(P_{k-1}{}^1-P_k{}^1)+\zeta_k(P_{k-1}{}^1+P_k{}^1)-\lambda_k(P_{k-1}{}^1+P_k{}^1). \end{aligned}\right\} \tag{25}$$

By means of the relations

$$P_k{}^1\pm P_{k-1}{}^1=\frac{(1\pm\cos\theta)}{\sin\theta}k(P_{k-1}\mp P_k) \tag{26.1}$$

and

$$k(P_k\pm P_{k-1})=-(1\pm\cos\theta)(P_{k-1}{}'\mp P_k'), \tag{26.2}$$

φ and χ, as expressed in Eqs. (14) and (18), can be written

$$\left.\begin{aligned} \varphi&=\{-[1+\cos\theta+O\sin\theta]\sum_{k=1}^{\infty}\xi_k(P_{k-1}{}'-P_k') \\ &\quad-[1-\cos\theta-O\sin\theta]\sum_{k=1}^{\infty}(\lambda_k+\zeta_k)(P_{k-1}{}'+P_k')\}v \\ \chi&=\{-[1+\cos\theta+O\sin\theta]\sum_{k=1}^{\infty}\xi_k(P_{k-1}{}'-P_k') \\ &\quad-[1-\cos\theta-O\sin\theta]\sum_{k=1}^{\infty}(\lambda_k-\zeta_k)(P_{k-1}{}'+P_k')\}w. \end{aligned}\right\} \tag{27}$$

The primes in the above denote derivations with respect to the argument of the Legendre polynomials.

It is seen from Eq. (27) that the coefficients which relate φ and χ to the respective spinors v and w differ solely in the sign of ζ_k. After some rearrangement and use of the relation[5]

$$\alpha_1F_1(\alpha+1;\beta;x)=(\beta-1)_1F_1(\alpha;\beta-1;x)+(\alpha-\beta+1)_1F_1(\alpha;\beta;x), \tag{28}$$

φ and χ are obtained in the form

$$\left.\begin{aligned} \varphi&=\{N+(1-\cos\theta-O\sin\theta)[L+i(\nu-\nu')M]\}v \\ \chi&=\{N+(1-\cos\theta-O\sin\theta)[L+i(\nu+\nu')M]\}w, \end{aligned}\right\} \tag{29}$$

[5] Reference 4, p. 254.

where

$$\left.\begin{aligned} N&=-2\textstyle\sum_k\xi_k(P_{k-1}{}'-P_k') \\ L&+i(\nu\pm\nu')M \\ &=\textstyle\sum_k[(\xi_k-\lambda_k\pm\zeta_k)P_{k-1}{}'-(\xi_k+\lambda_k\mp\zeta_k)P_k']. \end{aligned}\right\} \tag{30}$$

It follows that

$$N=2\sum_{k=1}^{\infty}(-1)^k k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}{}_1F_1(\alpha;\beta;x)\times(P_{k-1}{}'-P_k') \tag{31}$$

$$M=\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}{}_1F_1(\alpha;\beta;x)\times(P_{k-1}{}'+P_k') \tag{32}$$

$$\begin{aligned} L=\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2} \\ \times\{[(\gamma-k)_1F_1(\alpha;\beta;x)-2\gamma_1F_1(\alpha;\beta-1;x)]P_{k-1}{}' \\ +[(\gamma+k)_1F_1(\alpha;\beta;x)-2\gamma_1F_1(\alpha;\beta-1;x)]P_k'\}. \end{aligned} \tag{33}$$

By means of Eq. (16) and the geometry of Fig. 1, the factor $(1-\cos\theta-O\sin\theta)$ in (29) can be expressed in vector form;

$$\begin{aligned} (1-\cos\theta-O\sin\theta)&=1-\hat{r}\cdot\hat{p}-i\boldsymbol{\sigma}\cdot\hat{r}\times\hat{p} \\ &=\boldsymbol{\sigma}\cdot(\hat{p}-\hat{r})\boldsymbol{\sigma}\cdot\hat{p}. \end{aligned} \tag{34}$$

From (9) and the above

$$\begin{aligned} \psi=\left(\frac{W+m}{2W}\right)^{\frac{1}{2}}\begin{pmatrix} Nv+L\boldsymbol{\sigma}\cdot(\hat{p}-\hat{r})\boldsymbol{\sigma}\cdot\hat{p}v \\ Nw+L\boldsymbol{\sigma}\cdot(\hat{p}-\hat{r})\boldsymbol{\sigma}\cdot\hat{p}w \end{pmatrix} \\ +\left(\frac{W+m}{2W}\right)^{\frac{1}{2}}iM\begin{pmatrix} (\nu-\nu')\boldsymbol{\sigma}\cdot(\hat{p}-\hat{r})\boldsymbol{\sigma}\cdot\hat{p}v \\ (\nu+\nu')\boldsymbol{\sigma}\cdot(\hat{p}-\hat{r})\boldsymbol{\sigma}\cdot\hat{p}w \end{pmatrix}. \end{aligned} \tag{35}$$

The first term here has the desired form in that it can be written directly in terms of the plane wave spinor u defined in (6). The second term can be expressed in a similar manner via relations (5) and (8) and the following:

$$\nu\pm\nu'=\alpha Z(W\pm m)/p=\alpha Zp/(W\mp m) \tag{36}$$

$$\frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{W+m}v=w, \quad \frac{\boldsymbol{\sigma}\cdot\mathbf{p}}{W-m}w=v \tag{37}$$

$$(\nu-\nu')\boldsymbol{\sigma}\cdot\hat{p}v=\alpha Zw, \quad (\nu+\nu')\boldsymbol{\sigma}\cdot\hat{p}w=\alpha Zv. \tag{38}$$

We obtain

$$\psi=[N+i\alpha ZM\boldsymbol{\alpha}\cdot(\hat{p}-\hat{r})+L\boldsymbol{\alpha}\cdot(\hat{p}-\hat{r})\boldsymbol{\alpha}\cdot\hat{p}]u \tag{39}$$

where

$$\boldsymbol{\alpha}=\begin{pmatrix} 0 & \boldsymbol{\sigma} \\ \boldsymbol{\sigma} & 0 \end{pmatrix}, \tag{40}$$

and the coefficients L, M, N are given by Eqs. (31), (32), and (33), with

$$P_k = P_k(\hat{r}\cdot\hat{p}). \tag{41}$$

Considered in terms of an expansion in the Coulomb parameter αZ, the coefficients N and M are of order unity while the coefficient L is of order $(\alpha Z)^2$. The latter is demonstrated in Sec. II below. It follows that the three terms on the right in (39) are, respectively, of order unity, (αZ), and $(\alpha Z)^2$.

The function ψ expressed by Eq. (39) has the asymptotic character associated with description of an incident electron state. In particular, it has for large r the form of a plane wave plus outgoing spherical waves. A function ψ' with the asymptotic character associated with a final electron state can be determined from (39) via time reversal.

The adjoint of the function ψ' can be written

$$\bar{\psi}' = \bar{u}[N' + i\alpha Z M'\boldsymbol{\alpha}\cdot(\hat{p}+\hat{r}) + L'\boldsymbol{\alpha}\cdot\hat{p}\boldsymbol{\alpha}\cdot(\hat{p}+\hat{r})], \tag{42}$$

where

$$N' = 2\sum_{k=1}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}k\,{}_1F_1(\alpha;\beta;x) \times(P_{k-1}'+P_k') \tag{43}$$

$$M' = \sum_{k=1}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2}\,{}_1F_1(\alpha;\beta;x) \times(P_{k-1}'-P_k') \tag{44}$$

$$L' = \sum_{k=1}^{\infty}\frac{\Gamma(\alpha)}{\Gamma(\beta)}e^{\nu\pi/2}x^{\gamma-1}e^{-x/2} \times\{[(\gamma-k)\,{}_1F_1(\alpha;\beta;x) - 2\gamma\,{}_1F_1(\alpha;\beta-1;x)]P_{k-1}' - [(\gamma+k)\,{}_1F_1(\alpha;\beta;x) - 2\gamma\,{}_1F_1(\alpha;\beta-1;x)]P_k'\} \tag{45}$$

$P_k = P_k(\hat{p}\cdot\hat{r})$.

For large r, ψ' has the form of a plane wave plus incoming spherical waves.

II. SOMMERFELD-MAUE APPROXIMATION

In the following, the wave function ψ in the form (39) is used to derive the Sommerfeld-Maue approximation for a Dirac electron in a Coulomb potential.[6] It was shown by Bethe and Maximon that the latter approximation is obtained from the Darwin series by neglecting terms of order α^2Z^2/l^2 as compared to unity.[7] We here verify this result. We show in fact that the Sommerfeld-Maue function follows directly from ψ as the result of approximating γ by k in the latter function. This approximation is precisely stated by the equation

$$\gamma = (k^2-\alpha^2Z^2)^{\frac{1}{2}} = k(1-\alpha^2Z^2/k^2)^{\frac{1}{2}} \approx k. \tag{46}$$

We first show that the coefficient L vanishes in this approximation. Substitution of (46) in (33) gives

$$\begin{aligned} L &\approx 2\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k-i\nu)}{\Gamma(2k+1)}e^{\nu\pi/2}x^{k-1}e^{-x/2}k \\ &\quad\times\{{}_1F_1(k-i\nu;2k+1;x)P_k' - {}_1F_1(k-i\nu;2k;x)P_k' \\ &\qquad - {}_1F_1(k-i\nu;2k;x)P_{k-1}'\} \\ &\approx 2\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k-i\nu)}{\Gamma(2k+1)}e^{\nu\pi/2}x^{k-1}e^{-x/2}k \\ &\quad\times[{}_1F_1(k-i\nu;2k+1;x) - {}_1F_1(k-i\nu;2k;x)]P_k' \\ &\quad+\sum_{k=0}^{\infty}(-1)^k\frac{\Gamma(k+1-i\nu)}{\Gamma(2k+3)}e^{\nu\pi/2}x^ke^{-x/2}(2k+2) \\ &\qquad\times{}_1F_1(k+1-i\nu;2k+2;x)P_k'. \end{aligned} \tag{47}$$

Since P_0' vanishes, the above can be rewritten in the form

$$\begin{aligned} L &\approx \sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k-i\nu)}{\Gamma(2k+1)}e^{\nu\pi/2}x^ke^{-x/2} \\ &\quad\times\{(2k/x)\,{}_1F_1(k-i\nu;2k+1;x) \\ &\quad-(2k/x)\,{}_1F_1(k-i\nu;2k;x) \\ &\quad+\frac{(k-i\nu)}{2k+1}\,{}_1F_1(k+1-i\nu;2k+2;x)\}P_k'. \end{aligned} \tag{47'}$$

By use of the recursion relation[5]

$$[(\beta-1)/x][{}_1F_1(\alpha;\beta;x) - {}_1F_1(\alpha;\beta-1;x)] = -(\alpha/\beta)\,{}_1F_1(\alpha+1;\beta+1;x) \tag{48}$$

the coefficient of P_k' in Eq. (47') is shown to be identically zero. Whence

$$L \approx 0. \tag{49}$$

Subject to the approximation (46), the coefficients N and M given by (31) and (32) can be written

$$N \approx 2\sum_{k=1}^{\infty}(-1)^k k\frac{\Gamma(k-i\nu)}{\Gamma(2k+1)}e^{\nu\pi/2}x^{k-1}e^{-x/2} \times{}_1F_1(k-i\nu;2k+1;x)(P_{k-1}'-P_k') \tag{50}$$

$$M \approx \sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k-i\nu)}{\Gamma(2k+1)}e^{\nu\pi/2}x^{k-1}e^{-x/2} \times{}_1F_1(k-i\nu;2k+1;x)(P_{k-1}'+P_k'). \tag{51}$$

The latter sums can be simplified by eliminating the complex parameters of the hypergeometric functions. To this end we introduce, after Gordon,[8] the following

[6] A. Sommerfeld and A. W. Maue, Ann. Physik **22**, 629 (1935).
[7] H. A. Bethe and L. C. Maximon, Phys. Rev. **93**, 768 (1954).
[8] W. Gordon, Z. Physik, **48**, 187 (1928).

contour integral representation[9] for ${}_1F_1$

$$
\begin{aligned}
&{}_1F_1(k-i\nu;\,2k+1;\,x)\\
&\quad=-\frac{\Gamma(1-i\nu)\Gamma(k)}{\Gamma(k-i\nu)}\frac{1}{2\pi i}\int_1^{0+}(-s)^{i\nu-1}(1-s)^{k-i\nu-1}\\
&\qquad\times{}_1F_1(k;\,2k+1;\,x[1-s])ds. \qquad (52)
\end{aligned}
$$

From (50) and (51)

$$
\begin{aligned}
N\approx&-\frac{2\Gamma(1-i\nu)}{2\pi i}e^{\frac{1}{2}(\nu\pi-x)}\int_1^{0+}(-s)^{i\nu-1}(1-s)^{-i\nu}\\
&\times N(s)ds \qquad (53)
\end{aligned}
$$

$$
\begin{aligned}
M\approx&-\frac{\Gamma(1-i\nu)}{2\pi i}e^{\frac{1}{2}(\nu\pi-x)}\int_1^{0+}(-s)^{i\nu-1}(1-s)^{-i\nu}\\
&\times M(s)ds, \qquad (54)
\end{aligned}
$$

where

$$
\begin{aligned}
N(s)=&\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k)}{\Gamma(2k+1)}(x[1-s])^{k-1}\\
&\times{}_1F_1(k;\,2k+1;\,x[1-s])k(P_{k-1}'-P_k') \qquad (55)
\end{aligned}
$$

$$
\begin{aligned}
M(s)=&\sum_{k=1}^{\infty}(-1)^k\frac{\Gamma(k)}{\Gamma(2k+1)}(x[1-s])^{k-1}\\
&\times{}_1F_1(k;\,2k+1;\,x[1-s])(P_{k-1}'+P_k'). \qquad (56)
\end{aligned}
$$

The function $N(s)$ can be written

$$
\begin{aligned}
N(s)=&\sum_{k=0}^{\infty}(-1)^{k-1}\frac{\Gamma(k+1)}{\Gamma(2k+1)}(x[1-s])^{k-1}\\
&\times\left\{\frac{x[1-s]}{2(2k+1)}{}_1F_1(k+1;\,2k+3;\,x[1-s])\right.\\
&\left.+{}_1F_1(k;\,2k+1;\,x[1-s])\right\}P_k' \qquad (57)
\end{aligned}
$$

$$
\begin{aligned}
N(s)=&\sum_{k=0}^{\infty}\frac{\Gamma(k+1)}{\Gamma(2k+1)}(-x[1-s])^{k-1}\\
&\times{}_1F_1(k+1;\,2k+2;\,x[1-s])P_k'. \qquad (58)
\end{aligned}
$$

Equation (58) is obtained from (57) with the aid of the recursion relations[5]

$$
\begin{aligned}
&{}_1F_1(\alpha;\beta-1;\,z)-{}_1F_1(\alpha-1;\beta-1;\,z)\\
&\qquad=[z/(\beta-1)]{}_1F_1(\alpha;\beta;\,z) \qquad (59.1)
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{z}{\beta-1}\left[\frac{\beta-\alpha}{\beta}{}_1F_1(\alpha;\beta+1;\,z)-{}_1F_1(\alpha;\beta;\,z)\right]\\
&\qquad={}_1F_1(\alpha;\beta;\,z)-{}_1F_1(\alpha;\beta-1;\,z). \qquad (59.2)
\end{aligned}
$$

[9] The path of integration passes through the point $s=1$ and closes around the origin in a positive direction.

As an explicit function of r and θ, $N(s)$ is written

$$
\begin{aligned}
N(s)=&\frac{1}{2ipr(1-s)}\frac{\partial}{\partial\cos\theta}\sum_{k=0}^{\infty}\frac{\Gamma(k+1)}{\Gamma(2k+1)}(2ipr[1-s])^k\\
&\times{}_1F_1(k+1;\,2k+2;\,-2ipr[1-s])P_k(\cos\theta). \qquad (60)
\end{aligned}
$$

The hypergeometric function here can be expressed as a spherical Bessel function via the relation[10]

$$
\begin{aligned}
&{}_1F_1(k+1;\,2k+2;\,-2iy)\\
&\quad=e^{-2iy}{}_1F_1(k+1;\,2k+2;\,2iy)\\
&\quad=(2y/\pi)^{\frac{1}{2}}(\tfrac{1}{2}y)^{-k-\frac{1}{2}}e^{-iy}\Gamma(k+\tfrac{1}{2}+1)j_k(y), \qquad (61)
\end{aligned}
$$

whereupon the series in Eq. (60) is directly summable;

$$
\begin{aligned}
&\sum_{k=0}^{\infty}\frac{\Gamma(k+1)}{\Gamma(2k+1)}(2ipr[1-s])^k\\
&\qquad\times{}_1F_1(k+1;\,2k+2;\,-2ipr[1-s])P_k(\cos\theta)\\
&\qquad=\exp\{-ipr[1-s](1-\cos\theta)\}. \qquad (62)
\end{aligned}
$$

It follows that

$$
N(s)=\tfrac{1}{2}\exp\{-ipr[1-s](1-\cos\theta)\}. \qquad (63)
$$

Using relation (26.2) and the result of (62), we obtain for $M(s)$

$$
\begin{aligned}
M(s)=&[2ipr[1-s](1-\cos\theta)]^{-1}\\
&\times\exp[-ipr[1-s](1-\cos\theta)\}. \qquad (64)
\end{aligned}
$$

Substitution of (63) and (64) in Eqs. (53) and (54) gives

$$
\begin{aligned}
N\approx&-\frac{\Gamma(1-i\nu)}{2\pi i}e^{\nu\pi/2}e^{ipr\cos\theta}\\
&\times\int_1^{0+}(-s)^{i\nu-1}(1-s)^{-i\nu}e^{ipr s(1-\cos\theta)}ds \qquad (65)
\end{aligned}
$$

$$
\begin{aligned}
M\approx&-\frac{\Gamma(1-i\nu)}{2\pi i}e^{\nu\pi/2}\frac{e^{ipr\cos\theta}}{2ipr(1-\cos\theta)}\\
&\times\int_1^{0+}(-s)^{i\nu-1}(1-s)^{-i\nu-1}e^{ipr s(1-\cos\theta)}ds. \qquad (66)
\end{aligned}
$$

The integrals involved in the latter relations are expressible in terms of hypergeometric functions by use of the standard integral representation[11]

$$
\begin{aligned}
{}_1F_1(\alpha;\beta;\,z)=&-\frac{1}{2\pi i}\frac{\Gamma(\beta)\Gamma(1-\alpha)}{\Gamma(\beta-\alpha)}\\
&\times\int_1^{0+}e^{zt}(1-t)^{\beta-\alpha-1}(-t)^{\alpha-1}dt\\
&\qquad\mathrm{Re}(\beta-\alpha)>0 \qquad (67)
\end{aligned}
$$

[10] See, for instance, W. Magnus and F. Oberhettinger, *Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1954).

[11] Reference 4, p. 272.

and the recursion relation[5]

$$\beta[{}_1F_1(\alpha+1;\beta;z)-{}_1F_1(\alpha;\beta;z)] = z\,{}_1F_1(\alpha+1;\beta+1;z). \quad (68)$$

One obtains

$$N\approx\Gamma(1-i\nu)e^{\nu\pi/2}e^{ipr\cos\theta}\,{}_1F_1(i\nu;1;ipr[1-\cos\theta]) \quad (69)$$

$$M\approx-\tfrac{1}{2}\Gamma(1-i\nu)e^{\nu\pi/2}e^{ipr\cos\theta} \times{}_1F_1(1+i\nu;2;ipr[1-\cos\theta]). \quad (70)$$

The wave function ψ in the approximation of Eq. (46) is obtained after substitution in (39) of the result (49) and the above values of N and M.

$$\psi\approx\Gamma(1-i\nu)e^{\nu\pi/2}\exp(i\mathbf{p}\cdot\mathbf{r})[{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r}) -\tfrac{1}{2}i\alpha Z\boldsymbol{\alpha}\cdot(\hat{p}-\hat{r})\,{}_1F_1(1+i\nu;2;ipr-i\mathbf{p}\cdot\mathbf{r})]u. \quad (71)$$

The result can be rewritten in the more familiar form of the Sommerfeld-Maue wave function;

$$\psi\approx N_1\exp(i\mathbf{p}\cdot\mathbf{r})[1-(i/2W)\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_r] \times{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r})u, \quad (72)$$

with

$$N_1=\Gamma(1-i\nu)e^{\nu\pi/2}$$

$$\boldsymbol{\nabla}_r\,{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r}) =\alpha ZW(\hat{p}-\hat{r})\,{}_1F_1(1+i\nu;2;ipr-i\mathbf{p}\cdot\mathbf{r}). \quad (73)$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

Superselection Principle and Pure States of n-Identical Particles

A. GALINDO,* A. MORALES, AND R. NUÑEZ-LAGOS
Junta de Energía Nuclear, Madrid, Spain

(Received April 28, 1961)

Wightman's assumption of commutative superselection rules is proved to be equivalent to Jauch's assumption of existence of a complete set of commuting observables. The analysis of coherence is carried out by using the Gel'fand representation theory and the general structure of the complete sets of compatible observables is given when the Hilbert space is separable. The above equivalence provides the physical grounds to incorporate both assumptions into the axiomatic foundations of quantum theory as a single "superselection principle." To illustrate how this principle works, the problem of determining the physically realizable states of an assembly of n-identical particles is analyzed and solved.

INTRODUCTION

THE possibility that the set $\mathcal{Q}$ of the bounded Hermitian operators representing observables of a quantum system were actually reducible was first raised by Wick, Wightman, and Wigner.[1] Later, in a natural attempt to provide the simplest substitute for irreducibility, Wightman[2] put forward the "assumption of commutative superselection rules," i.e., the commutor $\mathcal{Q}'$ is Abelian.[3] More recently, a careful analysis of the theory of measurement compels Jauch to state the following assumption[4]: $\mathcal{Q}''$ has a maximal Abelian subalgebra, or equivalently, $\mathcal{Q}$ contains a complete set of compatible observables. Needless to say, this last hypothesis is profoundly rooted in the quantum philosophy, and that as such, its proper place has to be looked for in the very postulation frame of quantum theory.

In this paper we present (Sec. I) a simple proof of the equivalence of the Wightman (W) and Jauch (J) assumptions in a Hilbert space of arbitrary dimension; this equivalence strongly supports the existence of a "superselection principle," according to which the von Neumann algebra generated by the (bounded) observables of any specific quantum situation must be hyperreducible. In carrying out the general discussion of (in)coherent subspaces, extensive use is made of the Gel'fand representation theory. Whenever the underlying Hilbert space is separable, the complete sets of compatible observables appear as direct integrals of maximal Abelian real W^* algebras in the elementary coherent subspaces.

Section II presents an interesting application of the superselection principle. After some preparatory results on the structure of the group algebras of finite unitary groups and their commutors, a specific problem is undertaken: to determine those subspaces $\mathcal{T}_n\subset\mathcal{H}_n \equiv\mathcal{H}_1\otimes\mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_1$ (n factors) satisfying the following two conditions: (a) $\mathcal{T}_n$ reduces both the symmetric group $\mathcal{G}$ and its commutor $\mathcal{G}'$; (b) $\mathcal{G}'$ is hyper-reducible in $\mathcal{T}_n$. From the foregoing considerations it follows that $\mathcal{T}_n$ are exactly those subspaces compatible with the superselection principle if one requires from the observables to commute with permutations. This is obviously the case if the physical system under consideration is the quantum assembly of n-identical particles (indiscernability principle). As a result, we prove that

* Present address: Department of Mathematics, New York University, New York.
[1] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. 88, 101 (1952).
[2] A. S. Wightman, Suppl. Nuovo cimento 14, 81 (1959).
[3] For symbols and definitions see following material.
[4] J. M. Jauch, Helv. Phys. Acta 33, 711 (1960).

and the recursion relation[5]

$$\beta[{}_1F_1(\alpha+1;\beta;z)-{}_1F_1(\alpha;\beta;z)] = z\,{}_1F_1(\alpha+1;\beta+1;z). \quad (68)$$

One obtains

$$N\approx\Gamma(1-i\nu)e^{\nu\pi/2}e^{ipr\cos\theta}\,{}_1F_1(i\nu;1;ipr[1-\cos\theta]) \quad (69)$$

$$M\approx-\tfrac{1}{2}\Gamma(1-i\nu)e^{\nu\pi/2}e^{ipr\cos\theta} \times{}_1F_1(1+i\nu;2;ipr[1-\cos\theta]). \quad (70)$$

The wave function ψ in the approximation of Eq. (46) is obtained after substitution in (39) of the result (49) and the above values of N and M.

$$\psi\approx\Gamma(1-i\nu)e^{\nu\pi/2}\exp(i\mathbf{p}\cdot\mathbf{r})[{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r}) -\tfrac{1}{2}i\alpha Z\boldsymbol{\alpha}\cdot(\hat{p}-\hat{r})\,{}_1F_1(1+i\nu;2;ipr-i\mathbf{p}\cdot\mathbf{r})]u. \quad (71)$$

The result can be rewritten in the more familiar form of the Sommerfeld-Maue wave function;

$$\psi\approx N_1\exp(i\mathbf{p}\cdot\mathbf{r})[1-(i/2W)\boldsymbol{\alpha}\cdot\boldsymbol{\nabla}_r] \times{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r})u, \quad (72)$$

with

$$N_1=\Gamma(1-i\nu)e^{\nu\pi/2}$$

$$\boldsymbol{\nabla}_r\,{}_1F_1(i\nu;1;ipr-i\mathbf{p}\cdot\mathbf{r}) =\alpha ZW(\hat{p}-\hat{r})\,{}_1F_1(1+i\nu;2;ipr-i\mathbf{p}\cdot\mathbf{r}). \quad (73)$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

Superselection Principle and Pure States of n-Identical Particles

A. GALINDO,* A. MORALES, AND R. NUÑEZ-LAGOS
Junta de Energía Nuclear, Madrid, Spain

(Received April 28, 1961)

Wightman's assumption of commutative superselection rules is proved to be equivalent to Jauch's assumption of existence of a complete set of commuting observables. The analysis of coherence is carried out by using the Gel'fand representation theory and the general structure of the complete sets of compatible observables is given when the Hilbert space is separable. The above equivalence provides the physical grounds to incorporate both assumptions into the axiomatic foundations of quantum theory as a single "superselection principle." To illustrate how this principle works, the problem of determining the physically realizable states of an assembly of n-identical particles is analyzed and solved.

INTRODUCTION

THE possibility that the set $\mathcal{Q}$ of the bounded Hermitian operators representing observables of a quantum system were actually reducible was first raised by Wick, Wightman, and Wigner.[1] Later, in a natural attempt to provide the simplest substitute for irreducibility, Wightman[2] put forward the "assumption of commutative superselection rules," i.e., the commutor $\mathcal{Q}'$ is Abelian.[3] More recently, a careful analysis of the theory of measurement compels Jauch to state the following assumption[4]: $\mathcal{Q}''$ has a maximal Abelian subalgebra, or equivalently, $\mathcal{Q}$ contains a complete set of compatible observables. Needless to say, this last hypothesis is profoundly rooted in the quantum philosophy, and that as such, its proper place has to be looked for in the very postulation frame of quantum theory.

In this paper we present (Sec. I) a simple proof of the equivalence of the Wightman (W) and Jauch (J) assumptions in a Hilbert space of arbitrary dimension; this equivalence strongly supports the existence of a "superselection principle," according to which the von Neumann algebra generated by the (bounded) observables of any specific quantum situation must be hyperreducible. In carrying out the general discussion of (in)coherent subspaces, extensive use is made of the Gel'fand representation theory. Whenever the underlying Hilbert space is separable, the complete sets of compatible observables appear as direct integrals of maximal Abelian real W^* algebras in the elementary coherent subspaces.

Section II presents an interesting application of the superselection principle. After some preparatory results on the structure of the group algebras of finite unitary groups and their commutors, a specific problem is undertaken: to determine those subspaces $\mathcal{T}_n\subset\mathcal{H}_n \equiv\mathcal{H}_1\otimes\mathcal{H}_1\otimes\cdots\otimes\mathcal{H}_1$ (n factors) satisfying the following two conditions: (a) $\mathcal{T}_n$ reduces both the symmetric group $\mathcal{G}$ and its commutor $\mathcal{G}'$; (b) $\mathcal{G}'$ is hyper-reducible in $\mathcal{T}_n$. From the foregoing considerations it follows that $\mathcal{T}_n$ are exactly those subspaces compatible with the superselection principle if one requires from the observables to commute with permutations. This is obviously the case if the physical system under consideration is the quantum assembly of n-identical particles (indiscernability principle). As a result, we prove that

* Present address: Department of Mathematics, New York University, New York.
[1] G. C. Wick, A. S. Wightman, and E. P. Wigner, Phys. Rev. 88, 101 (1952).
[2] A. S. Wightman, Suppl. Nuovo cimento 14, 81 (1959).
[3] For symbols and definitions see following material.
[4] J. M. Jauch, Helv. Phys. Acta 33, 711 (1960).

no pure states other than symmetric or antisymmetric ones are compatible with the superselection principle. At this point, we should emphasize that the "proofs" one usually finds in many textbooks concerning the symmetry or antisymmetry of wave functions are quite unsatisfactory from an operational viewpoint. Indeed, if ψ and P are arbitrary elements of $\mathfrak{H}_n$ and $\mathcal{G}$, respectively, it is plain that the expectation values $(\psi,Q\psi)$ and $(P\psi,QP\psi)$ are equal for every observable Q, as it follows from the relations $PQ=QP$ (mathematically equivalent to indiscernability) and $P^*=P^{-1}$. Therefore, ψ and $P\psi$ are physically identical (even if $\psi\neq\pm P\psi$!). In order for the physical equivalence criterium to imply $\psi=\pm P\psi$, it would be necessary that the set of observables were irreducible in $\mathfrak{H}_n$, a fact which would violate the indiscernability principle.

Finally, an Appendix is devoted to discuss the nature of the set $\mathcal{Q}$ of observables, from a physical and mathematical viewpoint.

I. EQUIVALENCE OF (W) AND (J) ASSUMPTIONS: SUPERSELECTION PRINCIPLE

Let $\mathfrak{H}$ be an arbitrary complex Hilbert space, $\mathfrak{L}(\mathfrak{H})$ the W^* algebra of all the bounded linear operators in $\mathfrak{H}$, and $\mathcal{O}$ a W^* subalgebra of $\mathfrak{L}(\mathfrak{H})$. (A self-adjoint algebra of operators is said to be a von Neumann or W^* algebra if it is closed relative to the weak neighborhood topology).[5,6] The commutor $\mathfrak{A}'$ of a subset $\mathfrak{A}\subset\mathfrak{L}(\mathfrak{H})$ is defined as the subset of $\mathfrak{L}(\mathfrak{H})$ consisting of those operators which commute with every element of $\mathfrak{A}$. It can be proved that $\mathfrak{A}'$ is a W^* algebra and that the bicommutor $\mathfrak{A}''$ is the minimal W^* algebra containing $\mathfrak{A}$ (i.e., the weak self-adjoint closure of $\mathfrak{A}$). Therefore $\mathcal{O}''=\mathcal{O}$. A W^* algebra $\mathcal{C}$ will be called maximal Abelian if $\mathcal{C}=\mathcal{C}'$. Finally, we shall say that the W^* algebra $\mathcal{O}$ is hyper-reducible if $\mathcal{O}'$ is Abelian, i.e., $\mathcal{O}'\subset\mathcal{O}$ (these algebras are in particular discrete or of type I).[5,7]

Theorem 1.

Let $\mathcal{O}$ be a W^* algebra; then the following two statements are equivalent:

a_1: $\mathcal{O}$ is hyper-reducible,

a_2: $\mathcal{O}$ contains a W^* subalgebra $\mathcal{C}$ that is maximal Abelian.

Proof.[8]

$a_2\Rightarrow a_1$: In fact, if $\mathcal{C}\subset\mathcal{O}$, and $\mathcal{C}=\mathcal{C}'$, then $\mathcal{O}'\subset\mathcal{C}'=\mathcal{C}\subset\mathcal{O}$; therefore $\mathcal{O}'\subset\mathcal{O}$.

$a_1\Rightarrow a_2$: If the Abelian W^* subalgebras of $\mathcal{O}$ are partially ordered by inclusion, Zorn's lemma states that there exists some maximal element, say $\mathcal{C}$, containing $\mathcal{O}'$; the maximal character of $\mathcal{C}$ in the above ordering implies that $\mathcal{C}'\cap\mathcal{O}\subset\mathcal{C}$; this relation, together with $\mathcal{O}'\subset\mathcal{C}\subset\mathcal{C}'$ and its consequent $\mathcal{C}'\subset\mathcal{O}$, prove that $\mathcal{C}=\mathcal{C}'$; so $a_1\Rightarrow a_2$.

Therefore, $a_1(=)a_2$, Q.E.D.

It is not difficult to realize that from this rather simple theorem the equivalence of (W) and (J) assumptions is easily shown.[9] It suffices to take $\mathcal{O}$ as the W^* algebra generated by the (bounded) observables, and to note that any Abelian W^* algebra can be considered as generated by a set of commuting self-adjoint operators. As mentioned in the introduction, this equivalence is basic to state a physically grounded superselection principle.

In the rest of this paper we shall deal exclusively with quantum theories in which the superselection principle holds. A well-established fact is that whenever the Hilbert space of state vectors is separable, it can be decomposed into a direct integral of coherent subspaces (coherent=minimal invariant with respect to the observables).[10] Although in the general case a similar reduction theory is not available, it is yet possible, nevertheless, to display the main features concerning (in)coherence. Since the Gel'fand representation theory will play a relevant role in this analysis, we summarize its essential results concerning the structure of Abelian W^* algebras[11]:

Let $\mathfrak{d}$ be an Abelian W^* subalgebra of $\mathfrak{L}(\mathfrak{H})$. The set Z of all nonzero homomorphisms λ of $\mathfrak{d}$ into the field of complex numbers, with the weakest topology which makes $A(\lambda)\equiv\lambda(A)$ continuous ($A\in\mathfrak{d}$ fixed, λ variable), is called the carrier space or spectrum of $\mathfrak{d}$. The following statements are true:

a) Z is a Stonian compact Haussdorf space.[12]

b) There exists an involution and norm-preserving one-to-one linear mapping τ of $\mathfrak{d}$ onto the B^* algebra[13] $C(Z)$ of continuous complex-valued functions on Z. (Gel'fand isomorphism.)

c) Let $A\in\mathfrak{d}$, and $f_A\equiv\tau A$. The range of $f_A(\lambda)$ as λ runs over Z coincides with the spectrum of A.

It is now an easy task to go through the problem of (in)coherence. Let us take $\mathfrak{d}$ as the commutor $\mathcal{Q}'$ of the W^* algebra $\mathcal{Q}''$ generated by the system $\mathcal{Q}$ of observables of a quantum system. From the aforementioned properties it follows in a straightforward

---

[5] J. Dixmier, *Les Algèbres d'opérateurs dans l'espace Hilbertien* (Gauthier-Villars, Paris, France, 1957).

[6] C. E. Rickart, *General Theory of Banach Algebras* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1960).

[7] I. E. Segal, Memoirs Am. Math. Soc. No. 9 (1951).

[8] $A\Rightarrow B$ means "if A, then B." Similarly, $A(=)B$ expresses that $A\Rightarrow B$ and $B\Rightarrow A$.

[9] We are tacitly assuming that $\mathcal{Q}=(\mathcal{Q}'')_{sa}$, i.e., that every self-adjoint element of $\mathcal{Q}''$ belongs to $\mathcal{Q}$. See the Appendix for a more detailed discussion of this point. We are indebted to the referee for his suggestion that this sort of discussion would be desirable.

[10] J. von Neumann, Ann. Math. **50**, 401 (1949). Cf. also references 4 through 7.

[11] I. M. Gel'fand, Mat. Sbornik **9**, 3 (1941); Uspekhi. Mat. Nauk **12**, 249 (1957). See especially reference 6.

[12] A. Haussdorf space is said to be extremely disconnected or Stonian if the closure of every open subset is also open. See M. H. Stone, Can. J. Math. **1**, 176 (1949).

[13] See reference 6, Chapters III and IV.

way that the three following statements are equivalent:

1) $E\mathfrak{H}$ reduces $\mathfrak{L}$, $(E^*=E=F^2)$.
2) E is a self-adjoint idempotent in $\mathfrak{d}$.
3) τE is the characteristic function φ_X of a clopen (= open and closed) subset X of Z.

Consequently, if X_1 and X_2 are disjoint clopen subsets of Z, then $(\tau^{-1}\varphi_{X_1})\mathfrak{H}$ and $(\tau^{-1}\varphi_{X_2})\mathfrak{H}$ are orthogonal subspaces of $\mathfrak{H}$ which are mutually incoherent relative to $\mathfrak{L}$, and conversely. Therefore, in order for $E\mathfrak{H}$ to be coherent, it is necessary and sufficient that X be relatively connected. But Z is Stonian; so X must consist of a single isolated point. Consequently, the coherent subspaces are in one-to-one correspondence with the isolated points of Z, and so, $\mathfrak{H}$ will break up into the direct sum of the coherent subspaces if and only if the set of isolated points is dense in Z; from property (c) it easily follows that this condition is equivalent to requiring that all the operators in $\mathfrak{d}$ (which are clearly normal) have pure point spectra (i.e., the eigenstates of one such operator span the whole $\mathfrak{H}$). If this were not the case, such a decomposition would not be possible any more. However, there exists a particularly relevant case, namely, when $\mathfrak{H}$ is separable, in which a generalized decomposition can be carried out.[14]

In fact, under this assumption of separability, the following propositions can be proved:

b_1: There exist a non-negative Radon basic measure ν on Z, a ν-measurable field of separable Hilbert spaces $\mathfrak{H}(\lambda)$, and an isometric linear mapping σ of $\mathfrak{H}$ onto $\mathfrak{H}\equiv\int_Z\oplus\mathfrak{H}(\lambda)d\nu(\lambda)$, such that $\sigma\mathfrak{d}\sigma^{-1}$ is the W^* algebra $\mathfrak{D}$ of diagonalizable operators in $\mathfrak{H}$. [ν as well as the field $\mathfrak{H}(\cdot)$ are essentially unique].

b_2: If $\mathfrak{A}$ is any W^* algebra such that $\mathfrak{d}\subset\mathfrak{A}\subset\mathfrak{d}'$, then there exists a ν-measurable and ν-unique field of W^* algebras $\mathfrak{A}(\lambda)\subset\mathfrak{L}[\mathfrak{H}(\lambda)]$ such that

$$\sigma\mathfrak{A}\sigma^{-1}=\int_Z\oplus\mathfrak{A}(\lambda)d\nu(\lambda) \quad\text{and}\quad \sigma\mathfrak{A}'\sigma^{-1}=\int_Z\oplus\mathfrak{A}'(\lambda)d\nu(\lambda).$$

In particular

$$\sigma\mathfrak{d}'\sigma^{-1}=\int_Z\oplus\mathfrak{L}[\mathfrak{H}(\lambda)]d\nu(\lambda).$$

From b_1, b_2, one easily realizes that the ν-measurable subsets of Z now play the role that the clopen subsets did before with respect to (in)coherence. However, the topological properties of Z imply, in particular, that any ν-measurable subset of Z is equivalent (in the ν-measure sense) to a clopen subset; in other words, any bounded ν-measurable function on Z coincides (ν-almost everywhere) with a continuous function. Consequently, we meet again the parallelism displayed before between topology and (in)coherence, as it was to be expected for simple reasons. The Hilbert spaces $\mathfrak{H}(\lambda)$ are called elementary coherent subspaces, although they cannot be properly considered as subspaces of $\mathfrak{H}$ except in the case that ν be discrete at the point λ, or equivalently that λ be an isolated point of Z. We must admit, therefore, that the theory of generalized decompositions does not lead to any new result as regards the analysis of (in)coherence. Nevertheless, it will be most useful to render explicit the structure of the complete sets of commuting observables. In fact, if $\mathfrak{R}$ is such a set, then

$$\mathfrak{d}\equiv\mathfrak{L}'\subset\mathfrak{R}''\equiv\mathfrak{C}=\mathfrak{C}'\subset\mathfrak{L}''=\mathfrak{d}',$$

where $\mathfrak{L}''$ is the W^* algebra generated by the observables. Therefore, there exists (proposition b_2) a ν-measurable field of W^* algebras $\mathfrak{C}(\lambda)=\mathfrak{C}'(\lambda)$ such that

$$\sigma\mathfrak{C}\sigma^{-1}=\int_Z\oplus\mathfrak{C}(\lambda)d\nu(\lambda).$$

Conversely, if $\mathfrak{C}(\cdot)$ is a ν-measurable field of maximal Abelian W^* algebras, then

$$\mathfrak{C}\equiv\sigma^{-1}\left\{\int_Z\oplus\mathfrak{C}(\lambda)d\nu(\lambda)\right\}\sigma$$

is a maximal Abelian W^* algebra contained in $\mathfrak{L}''$.

Finally, it is plain that $\sigma\mathfrak{R}\sigma^{-1}=\int_Z\oplus\mathfrak{R}(\lambda)d\nu(\lambda)$, where $\mathfrak{R}(\lambda)$ are sets of commuting self-adjoint operators determining $\mathfrak{C}(\lambda)$. From the mathematical viewpoint, however, a greater simplicity is reached if we take the complete sets of commuting observables as being maximal Abelian real W^* algebras, i.e., if we assume that any self-adjoint operator in the weak closure of $\mathfrak{R}$ belongs to $\mathfrak{R}$. In that case, $\mathfrak{R}(\lambda)$ are also maximal Abelian real W^* algebras in $\mathfrak{H}(\lambda)$, and conversely.

II. PURE STATES OF n IDENTICAL PARTICLES

Theorem 2. Let $\mathfrak{H}$ be a complex Hilbert space and $\mathfrak{A}$ the group algebra of a finite group $\mathfrak{G}$ of unitary operators in $\mathfrak{H}$, i.e., the subalgebra of $\mathfrak{L}(\mathfrak{H})$ consisting of all the linear combinations of elements of $\mathfrak{G}$. Then the following statements are true:

c_1: $\mathfrak{A}$ is a W^* algebra.

c_2: It is possible to endow $\mathfrak{A}$ with a structure of H^* algebra,[15] i.e., a scalar product (A,B) can be defined for any A, $B\in\mathfrak{A}$ such that $\mathfrak{A}$ becomes a Hilbert space and $(A,BC)=(B^*A,C)$.

c_3: As an H^* algebra, $\mathfrak{A}$ is the direct sum of its (mutually orthogonal) minimal two-sided ideals $\mathfrak{A}_i$ $(i=1, 2, \cdots, s)$. The identity E_i of $\mathfrak{A}_i$ is a self-adjoint idempotent and belongs to the center of $\mathfrak{A}$. Each $\mathfrak{A}_i$

[14] See references 4 and 10. For a detailed analysis of b_1, b_2, the reader may profitably consult reference 5, especially Chap. II.

[15] This structure is, of course, compatible with the structure quoted in c_1 in the sense that the adjoint of any $A\in\mathfrak{A}$ when considered as an operator acting on the left in $\mathfrak{A}$ coincides with its adjoint as an operator in $\mathfrak{H}$.

admits a decomposition as a direct sum of mutually orthogonal minimal left-ideals $I_{i1}, I_{i2}, \cdots, I_{ir_i}$. (Although I_{ij} are not uniquely determined, unless $r_i=1$, the length r_i of the decomposition is fixed for a given $\mathfrak{A}_i$.) Each I_{ij} is of the form $\mathfrak{A}E_{ij}$, where E_{ij} is an irreducible self-adjoint idempotent, and $E_{ij}\mathfrak{A}E_{ik}$ is one dimensional; finally, there exist some elements $T_{ijk}\in I_{ik}$ such that $T_{ijk}=T_{ikj}^*$, $T_{ijk}T_{ikj}=E_{ij}$.

c_4: $\mathfrak{A}'$ is completely reducible; the subspaces $E_{ij}\mathfrak{H}$ are minimal invariant relative to $\mathfrak{A}'$.

c_5: A subspace $\mathfrak{H}'\subset\mathfrak{H}$ reduces $\mathfrak{A}$ and $\mathfrak{A}'$ if and only if

$$\mathfrak{H}'=\Big(\sum_{i\in J'} E_i\Big)\mathfrak{H},$$

where J' is any subset of $J\equiv(1,2,\cdots,s)$.

c_6: A subspace $\mathfrak{H}''\subset\mathfrak{H}$ reduces $\mathfrak{A}$ and $\mathfrak{A}'$, and $\mathfrak{A}$ is Abelian in $\mathfrak{H}''$ if and only if

$$\mathfrak{H}''=\Big(\sum_{i\in J''} E_i\Big)\mathfrak{H},$$

where J'' is any subset of $\{i\in J: r_i=1\}$.

Before proving this theorem, let us present an important consequence of it: Let $\mathfrak{H}\equiv\mathfrak{H}_n=\mathfrak{H}_1\otimes\mathfrak{H}_1\otimes\cdots\otimes\mathfrak{H}_1$, and $\mathcal{G}$ the symmetric group: $P\{\psi_{i_1}\otimes\psi_{i_2}\otimes\cdots\otimes\psi_{i_n}\}\equiv\psi_{i_1'}\otimes\psi_{i_2'}\otimes\cdots\otimes\psi_{i_n'}$, where $\psi_i\in\mathfrak{H}_1$ and $P\in\mathcal{G}$ is the permutation $(1,2,\cdots,n)\to(1',2',\cdots,n')$. But $\mathfrak{A}$ is a faithful representation of a two-sided ideal of the abstract group algebra of $\mathcal{G}$. (If $\dim\mathfrak{H}_1\geqslant n$, $\mathfrak{A}$ is actually isomorphic to the entire abstract group algebra.) So, if we combine c_6 with those classical results[16,17] on the structure of this abstract algebra which state that $r_i=1$ if and only if E_i is either the symmetrizer S or the antisymmetrizer A, we reach the following fundamental result: In order for a subspace $\mathcal{T}_n\subset\mathfrak{H}_n$ to reduce both $\mathcal{G}$ and $\mathcal{G}'$ in such a way that $\mathcal{G}'$ be hyper-reducible in $\mathcal{T}_n$ it is necessary and sufficient that $\mathcal{T}_n$ be any one of these three subspaces: $S\mathfrak{H}_n$, $A\mathfrak{H}_n$, and $(S+A)\mathfrak{H}_n$. Both $S\mathfrak{H}_n$ and $A\mathfrak{H}_n$ are coherent with respect to $\mathfrak{A}'$, while $(S+A)\mathfrak{H}_n$ is obviously incoherent as a direct orthogonal sum of two subspaces ($S\mathfrak{H}_n$ and $A\mathfrak{H}_n$) which are mutually incoherent. Therefore, since $\mathcal{G}'$ can be considered as the W^* algebra generated by the observables of a quantum assembly of n-identical particles, this result can be interpreted as saying that the unique pure states of the assembly which are compatible with the superselection principle are the symmetric or the antisymmetric ones. As said in the introduction, that constitutes (in our view) a satisfactory proof of an old feature of quantum mechanics.

Proof of Theorem 2.

c_1: Let $\{A_1,A_2,\cdots,A_n\}$ be a linear basis of $\mathfrak{A}$, so that any $A\in\mathfrak{A}$ can be uniquely written down as

$$\sum_{i=1}^{n}\lambda_i A_i,$$

and let $\{\mathfrak{H}_\gamma\}(\gamma\in K)$ denote the collection of all the finite subspaces $\mathfrak{H}_\gamma\subset\mathfrak{H}$ reducing $\mathfrak{A}$. Let $B\in\mathfrak{A}''$; it is clear that every $\mathfrak{H}_\gamma$ reduces B. If B_γ, $\mathfrak{A}_\gamma$ denote, respectively, the restrictions of B, $\mathfrak{A}$ to $\mathfrak{H}_\gamma$, then $B_\gamma\in\mathfrak{A}_\gamma''$; but $\dim\mathfrak{H}_\gamma<\infty$, so $B_\gamma\in\mathfrak{A}_\gamma$, i.e.,

$$B_\gamma=\sum_{i=1}^{n}\lambda_{i,\gamma}A_{i,\gamma}.$$

As $\dim\mathfrak{A}_\gamma$ may likely be less than n, we need to prove that a unique consistent choice of the coefficients $\lambda_{i,\gamma}$ is actually possible: In fact, there always exists a subset $K'\subset K$ such that

$$\mathfrak{H}=\sum_{\alpha\in K'}\oplus\mathfrak{H}_\alpha$$

[note that any cyclic (relative to $\mathfrak{A}$) subspace of $\mathfrak{H}$ is finite dimensional]; therefore, and because of the finite dimensionality of $\mathfrak{A}$, we can assert the existence of some $\beta\in K'$ such that $\dim\mathfrak{A}_\beta=n$. It is clear that $\lambda_{i,\beta}$ $(i=1,2,\cdots,n)$ are unique; but $\mathfrak{H}_\beta\oplus\mathfrak{H}_\alpha(\alpha\neq\beta,\alpha\in K')$ belongs to $\{\mathfrak{H}_\gamma\}$; therefore $\lambda_{i,\alpha}=\lambda_{i,\beta}$ is a consistent choice for all $\alpha\in K'$, and consequently,

$$B=\sum_{i=1}^{n}\lambda_{i,\beta}A_i,\quad \text{Q.E.D.}$$

c_2: Let us consider $\mathcal{G}$ as an abstract group with elements $g_1,g_2,\cdots,g_m$, and let $\tilde{\mathfrak{A}}$ be its abstract group algebra. Since $\mathcal{G}$ is, in particular, a compact group, and $\tilde{\mathfrak{A}}=L_2(\mathcal{G})$we can readily apply that well-known general result which states that $\tilde{\mathfrak{A}}$ is an H^* algebra[18]; because of the finiteness of $\mathcal{G}$, we can further assert that $\mathcal{G}$ is a unitary group in $\tilde{\mathfrak{A}}$. Consequently, $\mathfrak{A}$ is *-isomorphic to a two-sided ideal of $\tilde{\mathfrak{A}}$, and hence, $\mathfrak{A}$ is an H^* algebra, Q.E.D.

c_3: These properties are essentially the Wederburn classical structure theorems and their proofs can be found, for instance, in reference 18.

c_4: It suffices to prove that $E_{ij}\mathfrak{H}$ are minimal invariant with respect to $\mathfrak{A}'$. But this conclusion is a simple consequence of both the irreducibility of E_{ij} and the proposition c_1. [In fact, let us assume $E_{ij}\mathfrak{H}$ to be reducible under $\mathfrak{A}'$; then there would exist two subspaces $E_{ij}'\mathfrak{H}$ and $E_{ij}''\mathfrak{H}$ (E_{ij}', E_{ij}'' being mutually orthogonal self-adjoint idempotents) with $E_{ij}'+E_{ij}''=E_{ij}$, such that they would reduce $\mathfrak{A}'$. This last property would imply E_{ij}', $E_{ij}''\in\mathfrak{A}''(=\mathfrak{A})$, and therefore, E_{ij} would not be an irreducible idempotent.]

c_5: Because of c_3 and c_4, we have just to prove that

[16] B. L. van der Waerden, *Modern Algebra* (Frederick Ungar Publishing Company, New York, 1950).

[17] H. Boerner, *Darstellungen von Gruppen* (Springer Verlag, Berlin, Germany, 1955).

[18] L. H. Loomis, *An Introduction to Abstract Harmonic Analysis* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1953). See also reference 6.

if $E_{ij}\mathcal{H}\subset\mathcal{H}'$, and $\mathcal{H}'$ reduces $\mathfrak{A}$ and $\mathfrak{A}'$, then $E_{ik}\mathcal{H}\subset\mathcal{H}'$ for any $k\neq j$. But $T_{ikj}E_{ij}\mathcal{H}\subset E_{ik}\mathcal{H}$, and $T_{ikj}{}^{*}T_{ikj}E_{ij}\mathcal{H}=E_{ij}\mathcal{H}$; therefore $T_{ikj}E_{ij}\mathcal{H}$ is a non-null subspace in $\mathcal{H}'\bigcap E_{ik}\mathcal{H}$, and hence $E_{ik}\mathcal{H}\subset\mathcal{H}'$.

c_6: Since $T_{ijk}T_{ikj}\neq T_{ikj}T_{ijk}$ if $k\neq j$, and $E_l\mathfrak{A}F_l$ is isomorphic to the field of complex numbers whenever $r_l=1$ (as it easily follows from c_3), c_6 is obviously true.

ACKNOWLEDGMENTS

We are especially indebted to Professor Sanchez del Rio for his valuable discussions as well as to the Junta de Energía Nuclear for its financial support.

APPENDIX

That the equivalence (W)⟨=⟩(J) cannot be logically inferred from Theorem 1 without any further knowledge about $\mathcal{Q}$ is revealed by this example:

Let $\mathcal{H}$ be 3 dimensional, with $\{\varphi_1,\varphi_2,\varphi_3\}$ as an orthonormal basis. Define A and B as the self-adjoint operators given by

$$A\varphi_1=\varphi_1,\quad A\varphi_{2,3}=-\varphi_{2,3},\quad B\varphi_1=\varphi_1+2\varphi_3,$$
$$B\varphi_2=2\varphi_2,\quad B\varphi_3=2\varphi_1-2\varphi_3.$$

It is easy to verify that if $\mathcal{Q}\equiv\{A,B\}$, then $\mathcal{Q}'$ is Abelian; but $AB\neq BA$; therefore $\mathcal{Q}$ has no other Abelian subsets than $\{A\}$ and $\{B\}$. Since both A and B have no simple spectra, $\{A\}''\neq\{A\}'$ and $\{B\}''\neq\{B\}'$. Consequently (W) is logically weaker than (J).

However, the set $\mathcal{Q}$ of all the bounded observables of a quantum system is by no means arbitrary; endowing it, in a natural way, with a definite mathematical structure has actually been the subject of many sound investigations, mainly by Segal[19] and Mackey,[20] whose results we hurriedly summarize:

Mackey's analysis leads reasonably to the assumption that every projection in any coherent space must represent an experimental "question," and so, that $\mathcal{Q}=(\mathcal{Q}'')_{sa}$; therefore (W)⟨=⟩(J) whenever Mackey's postulational approach is adopted.

However, Segal's method gives $\mathcal{Q}$ the following more general structure; $\mathcal{Q}=(\bar{\mathcal{Q}})_{sa}$, where $\bar{\mathcal{Q}}$ is a C^* algebra. Fortunately, a remark of a physical nature will enable us to consider $\bar{\mathcal{Q}}$ as being a W^* algebra, retrieving thereby the equivalence (W)⟨=⟩(J): let $\{A_i\}$ be an arbitrary nondecreasing bounded directed set of elements of $\mathcal{Q}$. If $A\equiv\lim A_i$, it is plain that the expectation value of A in any physically realizable state can be experimentally measured with unlimited accuracy by means of A_i's. On the other hand, these expectation values uniquely define A, since this is true for A_i's. So, it is perfectly natural to assume that A is an observable, i.e. $A\in\mathcal{Q}$, and as a consequence, it follows that $\bar{\mathcal{Q}}$ is a W^* algebra.[21]

[19] I. E. Segal, Ann. Math. 48, 930 (1947). I. E. Segal, Lectures, Summer seminar on applied mathematics, Boulder, Colorado (1960).

[20] G. Mackey, *The Mathematical Foundations of Quantum Mechanics* (Harvard University Press, Cambridge, Massachusetts, 1960).

[21] R. V. Kadison, Ann. Math. 64, 175 (1956).

Ergodicity Conditions in Quantum Mechanics and Van Hove's Hamiltonians

G. M. PROSPERI*
Instituto di Scienze Fisiche dell'Università di Milano,
Istituto Nazionale di Fisica Nucleare—Sezione di Milano, Milan, Italy
(Received March 7, 1961)

In this work we start from a formulation of the quantum-ergodic problem developed in a preceding paper. We prove that the ergodicity conditions which we established there in an abstract and very general form are in particular satisfied for the class of Hamiltonians for which Van Hove was able to prove a master equation. Some restrictive assumptions have been introduced to Van Hove's theory which are necessary to identify the unperturbed states of Van Hove with the states $\omega_{\nu i}$ of the ergodic theory.

INTRODUCTION

AS is well known, statistical mechanics can be founded on two different approaches, the ergodic method and the master equation method.

The two approaches are not in contradiction, but according to modern views they correspond to two complementary aspects of the problem. Ergodic theories have essentially the aim of establishing existence theorems for an equilibrium state of a system. The master equation has the aim of describing how the approach of the system towards the abovementioned equilibrium state takes place.

In principle, the ergodic theorems should be more general, the master equation should hold under more particular conditions, but should be richer in physical contents.

The situation has recently, in a certain sense, been reversed, while notable improvements were made in the field of the master equation mainly by Van Hove,[1-2] only minor improvements were made in the field of the ergodic theories.

More specifically Van Hove has developed, in a series of papers,[1-5] a perturbation theory for continuous spectra (owing to the great number of degrees of freedom of a macroscopic system, the spectrum of its Hamiltonian can be practically considered as continuous) and he succeeded in deducing a master equation which governs the time evolution of the occupation probability of the unperturbed states of the system for a certain class of Hamiltonians and a certain class of initial conditions.

The aim of the present work is to show that certain ergodicity conditions of a rather abstract type which we gave in a preceding paper[6] (which we shall hereafter indicate with the letter A), starting from a new formulation of the quantum-ergodic problem, are satisfied for the class of Hamiltonians dealt with by Van Hove. The formulation which we have adopted of the quantum ergodic problem can be summarized in this way:

As usual, we assume the energy spectrum of the system to be decomposed in intervals, we consider the subspaces (energy shells) spanned by the eigenvectors corresponding to those intervals and we assume they characterize a "macroenergy" (in the sense that by a macroscopic experiment can only be established to what interval the energy of the system belongs and hence in what subspace its state vector is). Further, we suppose every energy shell V to be decomposed in new subspaces V_ν (cells) orthogonal to each other which are spanned by a system of basis vector $\{\omega_{\nu i}\}$ $(\nu=1,2,\cdots,N;$ $i=1,2,\cdots s_\nu;$ $\sum_\nu s_\nu=S)$ and we assume that all state vectors belonging to the same V_ν, correspond to the same macroscopic properties of the system.

We assume that a macroscopic observation has been made on the system at the time $t=0$, and thus, that its state vector ψ_0 at that instant has been assigned to a determined subspace V_μ.

We define as ergodic those systems for which the relation

$$\sum_{\nu=1}^{N} \frac{\mathfrak{G}[Mu_\nu(t)-s_\nu/S]^2}{s_\nu^2/S^2} \ll 1 \qquad \text{(I1)}$$

holds. Here M is the time average and $\mathfrak{G}$ is an average on all possible initial states ψ_0's belonging to the cell μ. More precisely the criterium by which $\mathfrak{G}$ is calculated is the following: one observes that all possible ψ_0's belonging to the cell V_μ, constitute a complex s_μ-dimensional hypersphere of radius=1 and one attributes to every set of ψ_0's a weight proportional to the area of the region which they intercept on the surface of the hypersphere.

If we put

$$H=\sum_\rho E_\rho P_\rho,$$

where E_ρ are the eigenvalues of the energy and P_ρ the projectors on the corresponding eigensubspaces, the principal result of A is that relation (I1) is satisfied if

---

* Present address: Lawrence Radiation Laboratory, University of California, Berkeley, California.

[1] L. Van Hove, Physica 21, 517 (1955).
[2] L. Van Hove, Physica 23, 441 (1957); 25, 268 (1959).
[3] L. Van Hove, Physica 21, 901 (1955).
[4] L. Van Hove, Physica 22, 343 (1956).
[5] N. M. Hugentoltz, Physica 23, 481 (1957).
[6] G. M. Prosperi and A. Scotti, J. Math. Phys. 1, 218 (1960). For another formulation of the ergodic theorem see G. Ludwig Z. Physik 150, 346 (1958), and G. Ludwig "Axiomatic Quantum Statistics of Macroscopic Systems," Reports of the Scuola Internazionale di Fisica "Enrico Fermi," Corso XIV, Italy, 1960 (to be published).

the other relations

$$|(1/s_\mu)\sum_{1}^{s_\mu}{}_j\sum_{1}^{s_\nu}{}_i\sum_\rho|(\omega_{\nu i},P_\rho\omega_{\mu j})|^2-(s_\nu/S)| \ll N^{-\frac{1}{2}}s_\nu/S, \quad (I2')$$

$$(1/s_\mu^2)\sum_{1}^{s_\mu}{}_j\sum_{1}^{s_\mu}{}_{j'}|\sum_{1}^{s_\nu}{}_i\sum_\rho(\omega_{\nu i},P_\rho\omega_{\mu j})^*(\omega_{\nu i},P_\rho\omega_{\mu j'}) -(s_\nu/S)\delta_{jj'}|^2\ll(s_\nu^2/NS^2) \quad (I2'')$$

are satisfied. These are the ergodicity conditions which we have mentioned above.

The exposition plan of our work is the following:

The proof of the relations (I.2'), (I.2'') is essentially contained in the fourth section.

In the first section we recall the assumptions which Van Hove makes on the Hamiltonian and some results which are necessary for our work.

In the second section we discuss the conditions under which it is possible to construct, throughout Van Hove's unperturbed states, an energy shell and we make some hypothesis on the way in which the cells V_ν must be chosen. We also establish some preliminary relation which we need in the fourth section.

In the third section we establish another important preliminary relation. Finally in the last two sections we discuss the results.

1. PRINCIPAL ASSUMPTIONS AND RESULTS OF VAN HOVE'S THEORY

As we stated in the introduction, in this section we shall recall the hypothesis which Van Hove put at the basis of his perturbation theory for continuous spectrum and of his theory of the master equation and we shall summarize some results which will be necessary later on.

We assume the Hamiltonian of the system to be of the form

$$H=H_0+\lambda V \quad (1.1)$$

we denote by $\alpha\equiv(\alpha^{(1)},\alpha^{(f)},\cdots,\alpha^{(f)})$ the ensemble of the variables which characterize an eigenstate $|\alpha\rangle_0$ of the unperturbed Hamiltonian

$$H_0|\alpha\rangle_0=\epsilon(\alpha)|\alpha\rangle_0.$$

We assume the eigenstates of H_0 to be discrete and hence the values of α which characterize them to be also so, but distributed with a very large density ρ_0 (which, for simplicity, we shall assume as constant).

We assume the states $|\alpha\rangle_0$ to be normalized in the ordinary sense

$${}_0\langle\alpha|\alpha'\rangle_0=\delta_{\alpha\alpha'}.$$

We then introduce the new states $|\alpha\rangle\equiv(\rho_0)^{\frac{1}{2}}|\alpha\rangle_0$; they result normalized in this way

$$\langle\alpha|\alpha'\rangle=\rho_0\delta_{\alpha\alpha'}.$$

If we accept the approximation in which ρ_0 is very large, we can write

$$\langle\alpha|\alpha'\rangle=\delta(\alpha-\alpha').$$

We call O any operator; we have then in general

$$\langle\alpha|O|\alpha'\rangle=O_{\rm d}(\alpha)\delta(\alpha-\alpha')+Q(\alpha,\alpha'),$$

where $Q(\alpha,\alpha')$ is an expression which can contain delta singularities in the components of α and α' but these must be of a smaller order than

$$\delta(\alpha-\alpha')\equiv\prod_{k=1}^{b}\delta(\alpha^{(k)}-\alpha'^{(k)}).$$

We define the two operators $O_{\rm d}$ and $O_{\rm nd}$ as the diagonal part and nondiagonal part of O:

$$O_{\rm d}|\alpha\rangle\equiv O_{\rm d}(\alpha)|\alpha\rangle,$$
$$\langle\alpha|O_{\rm nd}|\alpha'\rangle\equiv Q(\alpha,\alpha').$$

If $f(\alpha)$ and $g(\alpha,\alpha')$ are functions which change very little with increments of α of the order of $1/\rho_0$, we can put

$$(1/\rho_0)\sum_\alpha f(\alpha)\cong\int d\alpha f(\alpha),$$

from which in particular we have

$$\sum_\alpha|\alpha\rangle_0\,{}_0\langle\alpha|\cong\int d\alpha|\alpha\rangle\langle\alpha|,$$

$$\sum_{\alpha\alpha'}{}_0\langle\alpha|O|\alpha'\rangle_0 g(\alpha,\alpha')\cong\int d\alpha\int d\alpha'\langle\alpha|O|\alpha'\rangle g(\alpha,\alpha'), \quad (1.2')$$

$$\frac{1}{\rho_0}\sum_\alpha f(\alpha)\,{}_0\langle\alpha|O|\alpha\rangle_0\cong\int d\alpha f(\alpha)O_{\rm d}(\alpha).$$

As it is clear from these relations, $O_{\rm d}(\alpha)$ and $Q(\alpha,\alpha')$ are essentially mean values of ${}_0\langle\alpha|O|\alpha\rangle_0$ and $\rho_0\,{}_0\langle\alpha|O|\alpha'\rangle_0$ calculated in a neighborhood of α and α'.

On the operator V we make the following assumptions:

(i) The diagonal part of V is zero.

(ii) If $A_1, A_2, \cdots, A_n$ are diagonal operators in the α representation, the expression

$$VA_1\cdots A_nV \quad (1.3)$$

has in general a nonvanishing diagonal part, and the matrix element of its nondiagonal part

$$\langle\alpha|(VA_1V\cdots A_nV)_{\rm nd}|\alpha'\rangle$$

does not contain any factor of the form $\delta[\epsilon(\alpha)-\epsilon(\alpha')]$ (in particular $\langle\alpha|V|\alpha'\rangle$ does not contain any factor of that type).

(iii) Let us consider the expression

$$\langle\alpha|VA_1V\cdots A_nV|\alpha'\rangle$$
$$=\int d\alpha_1\int d\alpha_2\cdots\int d\alpha_n\langle\alpha|V|\alpha_1\rangle A_1(\alpha_1)$$
$$\times\langle\alpha_1|V|\alpha_2\rangle\cdots A_n(\alpha_n)\langle\alpha_n|V|a'\rangle, \quad (1.4)$$

and in this the submatrix

$$\langle\alpha_j|VA_{j+1}V\cdots A_{l-1}V|\alpha_l\rangle$$
$$=\int d\alpha_{j+1}\int d\alpha_{j+2}\cdots\int d\alpha_{l-1}\langle\alpha_j|V|\alpha_{j+1}\rangle A_{j+1}(\alpha_{j+1})$$
$$\times\langle\alpha_{j+1}|V|\alpha_{j+2}\rangle\cdots A_{l-1}(\alpha_{l-1})\langle\alpha_{l-1}|V|\alpha_l\rangle,$$

where

$$1<j<l<n.$$

We suppose that this has a singularity of the type $\delta(\alpha_j-\alpha_l)$ (i.e., that $VA_{j+1}V\cdots A_{l-1}V$ has a nonvanishing diagonal part) then the expressions

$$\langle\alpha_j|(VA_{j+1}V\cdots A_{k-1}V|\alpha_k\rangle A_k(\alpha_k)$$
$$\times\langle\alpha_k|VA_{k+1}V\cdots A_{l-1}V)_{\rm d}A_lVA_{l+1}V\cdots A_{m-1}V|\alpha_m\rangle,$$

and

$$\langle\alpha_i|VA_{i+1}V\cdots A_{j-1}VA_j(VA_{j+1}V\cdots A_k|\alpha_k\rangle A_k(\alpha_k)$$
$$\times\langle\alpha_k|VA_{k+1}V\cdots A_{l-1}V)_{\rm d}|\alpha_l\rangle,$$

where

$$1\leq i<j<l<m\leq n,$$

have no singularities of the type $\delta(\alpha_k-\alpha_m)$ and $\delta(\alpha_i-\alpha_k)$, respectively.

We define as the irreducible diagonal part

$$(VA_1V\cdots A_nV)_{\rm id}$$

and as the irreducible nondiagonal part

$$(VA,V\cdots A_nV)_{\rm ind}$$

of the expression (1.3) the contribution to the respective diagonal and nondiagonal parts which we obtain if, in performing the integrations in expression (1.4), we maintain the variables $\alpha_1, \alpha_2, \cdots \alpha_n$ each outside of a small neighborhood of any others and out of a small neighborhood of α and α' (this corresponds to eliminating the contribution derived from the diagonal parts of all possible submatrices).

Then, as a consequence of the postulate (iii), the nondiagonal part of an expression of the type (1.3) can be expressed as a sum of its irreducible nondiagonal part and of the terms which we obtain if we replace in all possible ways in (1.4) noncontiguous submatrices with their diagonal parts, and if we calculate the irreducible nondiagonal part of the complete expression. The diagonal part of (1.3) can be expressed in a strictly similar way, the only difference consists in the fact that, in this case, groups of terms which are placed on the two extremities of the expression must be considered as contiguous and the replacment of extreme submatrices must be made only on a single side chosen once and for all. We denote the aforementioned decomposition as decomposition of the diagonal (nondiagonal) part of the expression (1.3) in its "irreducible components." It is widely used in Van Hove's theory.

Let us introduce the resolvent operator[3]

$$R_l\equiv(H-l)^{-1}. \quad (1.5)$$

The projector on the subspace corresponding to a value of the energy between E' and E'' can be written[4]

$$P(E',E'')=\frac{i}{2\pi}\oint_{\gamma(E',E'')}dlR_l$$
$$=\frac{i}{2\pi}\int_{E'}^{E''}dE(R_{E-io}-R_{E+io}) \quad (1.6)$$

where $\gamma(E',E'')$ is a path which encircles just the portion of the real axis between E' and E''.

The following formal properties of R_l are important[3]:

$$(R_l)^\dagger=R_{l^*}, \quad (1.7)$$
$$(l-l')R_lR_{l'}=R_l-R_{l'}. \quad (1.8)$$

If we replace in (1.5) the (1.1) and expand in series of λ, we obtain[3]

$$R_l=(H_0-l)^{-1}-\lambda(H_0-l)^{-1}V(H_0-l)^{-1}$$
$$+\lambda^2(H_0-l)^{-1}V(H_0-l)^{-1}V(H_0-l)^{-1}-\cdots. \quad (1.9)$$

Let us put[3]

$$D_l\equiv(R_l)_{\rm d}. \quad (1.10)$$

By decomposing every term of the expression of $(R_l)_{\rm nd}$ into irreducible components and by making appropriate partial summations, it can be proved[3,5] that

$$R_l=D_l-\lambda D_l(V-\lambda VD_lV+\cdots)_{\rm ind}D_l. \quad (1.11)$$

By introducing the auxiliary quantity

$$G_l\equiv(VD_lV-\lambda VD_lVD_lV+\cdots)_{\rm id}, \quad (1.12)$$

it can be proved[3,5] that D_l can be expressed in a closed form:

$$D_l(\alpha)=[\epsilon(\alpha)-l-\lambda^2G_l(\alpha)]^{-1}. \quad (1.13)$$

From the definitions (1.10) and (1.12), and from (1.13), one can obtain the following important properties for $D_l(\alpha)$ and $G_l(\alpha)$ regarded as functions of l.[3] $D_l(\alpha)$ and $G_l(\alpha)$ are analytic functions out of the real axis; the asymptotic behavior of $D_l(\alpha)$ is given by the relation

$$D_l(\alpha)\underset{|l|\to\infty}{\to}1/|l|.$$

One has

$$[D_l(\alpha)]^*=D_{l^*}(\alpha), \quad [G_l(\alpha)]^*=G_{l^*}(\alpha);$$

it can be proved that in the half-plane $\mathrm{Im}l>0$ one has $\mathrm{Im}G_l(\alpha)>0$. Across the real axis $G_l(\alpha)$ has finite discontinuities (generally only in a certain interval); $D_l(\alpha)$ too has finite discontinuities of the same kind and in addition it can have poles.

For the above properties we can put[3]

$$G_{E+i0}(\alpha)=K_E(\alpha)+iJ_E(\alpha),$$
$$G_{E-i0}(\alpha)=K_E(\alpha)-iJ_E(\alpha),$$

being

$$J_E(\alpha)\geq 0.$$

We have then

$$D_{E\pm i0}(\alpha)=[\epsilon(\alpha)-E-\lambda^2K_E(\alpha)\mp i\lambda^2J_E(\alpha)]^{-1}. \quad (1.14)$$

Let us assume that there are certain intervals of values of E in which

$$J_E(\alpha)=0;$$

$D_l(\alpha)$ will or will not have poles depending on whether the equation

$$\epsilon(\alpha)-E-\lambda^2K_E(\alpha)=0 \quad (1.15)$$

has roots or does not in those intervals.

In the former case we call the state α nondissipative, in the latter dissipative.[2] This classification exhibits the profoundly different behavior in the time evolution of the vector $\int_{I_\alpha}d\alpha'|\alpha'\rangle$ in the two cases (I_α is a small neighborhood of α).

The preceding results, in particular relations (1.11), (1.13), (1.14), and Eq. (1.15), are the basis of all developments of Van Hove's perturbation theory for continuous spectra.

Now we conclude these considerations of a general character and we intend to recall definitions and properties of some other quantities more strictly connected with the theory of the master equation and particularly interesting for us.

We suppose A to be a diagonal operator in α representation and its eigenvalue $A(\alpha)$ to be a smooth enough function of α, and we make the position[2]

$$(R_lAR_{l'})_{\mathrm{d}}(\alpha)\equiv\int d\alpha' A(\alpha')X_{ll'}(\alpha',\alpha), \quad (1.16)$$

$$[(V-\lambda VD_lV+\cdots)A(V-\lambda VD_{l'}V+\cdots)]_{\mathrm{id}}(\alpha)$$
$$\equiv\int d\alpha' A(\alpha')W_{ll'}(\alpha',\alpha). \quad (1.17)$$

If we substitute (1.11) into (1.16), decompose each term of the expansion obtained in irreducible components, and sum the terms in a suitable way, the following relations can be proved[5]:

$$X_{ll'}(\alpha',\alpha)$$
$$=D_l(\alpha')D_{l'}(\alpha')\delta(\alpha'-\alpha)+\lambda^2D_l(\alpha')D_{l'}(\alpha')\times\int d\alpha_1W_{ll'}(\alpha',\alpha_1)X_{ll'}(\alpha_1,\alpha)$$
$$=D_l(\alpha')D_{l'}(\alpha')\delta(\alpha'-\alpha)+\lambda^2D_l(\alpha')D_{l'}(\alpha')\times[W_{ll'}(\alpha',\alpha)+\lambda^2\int d\alpha_1W_{ll'}(\alpha',\alpha_1)D_l(\alpha_1)D_{l'}(\alpha_1)\times W_{ll'}(\alpha_1,\alpha)+\cdots]D_l(\alpha)D_{l'}(\alpha). \quad (1.18)$$

$X_{ll'}(\alpha',\alpha)$ and $W_{ll'}(\alpha',\alpha)$ have the following properties: Both are analytic for l and l' out of the real axis; they satisfy the relations

$$[X_{ll'}(\alpha',\alpha)]^*=X_{l'^*l^*}(\alpha',\alpha),$$
$$[W_{ll'}(\alpha',\alpha)]^*=W_{l'^*l^*}(\alpha',\alpha); \quad (1.19)$$

for l or l' across the real axis $W_{ll'}(\alpha',\alpha)$ has finite discontinuities, whereas $X_{ll'}(\alpha',\alpha)$ has finite discontinuities and pseudopoles (i.e., polar singularities operlapping finite discontinuities) the position of the latter is dependent on the reciprocal position of l and l'.

The two functions

$$X_{E+l\ E+l}(\alpha',\alpha) \quad \text{and} \quad X_{E+l\ E-l}(\alpha',\alpha)$$

are particularly interesting for us with regard to the study of their singularity.

As we can deduce from the relation[2]

$$D_l(\alpha)-D_{l'}(\alpha)=(l-l')\int d\alpha' X_{ll'}(\alpha',\alpha), \quad (1.20)$$

of the two, the function $X_{E+l\ E-l}(\alpha',\alpha)$ has always a pseudopole of the first order in the point $l=0$, whereas the function $X_{E+l\ E+l}(\alpha',\alpha)$ has no pseudopole of this kind. If the states α and α' are both dissipative, the pseudopole $l=0$ is the only pseudopole for $X_{E+lE-l}(\alpha',\alpha)$, whereas $X_{E+l\ E+l}(\alpha',\alpha)$ has no pseudopoles and it has in this case only finite discontinuities. The origin of the pseudopole in $l=0$ for $X_{E+l\ E-l}(\alpha',\alpha)$ is different according to the nature of the states; in particular if all the states are dissipative, its origin is in the lack of convergence of the series (1.18). It is useful to note that asymptotically we have

$$\left.\begin{matrix}X_{E+l\ E-l}(\alpha',\alpha)\\ X_{E+l\ E-l}(\alpha',\alpha)\end{matrix}\right\}\xrightarrow[|l|\to\infty]{}\propto\frac{1}{|l|^2}. \quad (1.21)$$

The two expressions

$$q_E^+(\alpha',\alpha)\equiv(1/\pi)\lim_{\eta\to+0}\eta X_{E-i\eta\ E+i\eta}(\alpha',\alpha)$$

and

$$q_E^-(\alpha',\alpha)\equiv(1/\pi)\lim_{\eta\to+0}\eta X_{E+i\eta\ E-i\eta}(\alpha',\alpha)$$

are particularly interesting for us; in the theory of the master equation they are connected with the limit distribution for $t \to \pm\infty$ of the probability. An explicit evaluation of these is possible only when we make further assumptions on the physical system.

We shall make the following assumptions:

(iv) The following symmetry relation holds

$$W_{ll'}(\alpha',\alpha) = W_{l'l}(\alpha,\alpha'); \tag{1.22}$$

for (1.18); this also implies

$$X_{ll'}(\alpha',\alpha) = X_{l'l}(\alpha,\alpha'). \tag{1.23}$$

(v) If we consider two states α and α', a succession of states $\alpha_1, \alpha_2, \cdots, \alpha_n$ exists for which

$$\epsilon(\alpha) = \epsilon(\alpha_1), \quad \epsilon(\alpha_1) = \epsilon(\alpha_2), \quad \cdots \epsilon(\alpha_n) = \epsilon(\alpha)$$

and

$$W^{(0)}(\alpha',\alpha_1) \neq 0, \quad W^{(0)}(\alpha_1,\alpha_2) \neq 0, \cdots, \quad W^{(0)}(\alpha_n,\alpha) \neq 0,$$

where $W^{(0)}(\alpha',\alpha)$ represents the zero approximation of

$$W_{ll'}(\alpha',\alpha) \left[\text{i.e., } \int d\alpha' A(\alpha') W^{(0)}(\alpha',\alpha) \equiv (VAV)_d(\alpha) \right].$$

We can prove that this assumption implies, moreover, that all unperturbed states are dissipative.

On these assumptions we can prove $q_E^+(\alpha',\alpha)$ and $q_E^-(\alpha',\alpha)$ to be equal and, for λ smaller than a certain value λ_c, the relation[2]

$$q_E(\alpha',\alpha) \equiv (1/\pi) \lim_{\eta \to \pm 0} \eta X_{E-i\eta\ E+i\eta}(\alpha',\alpha) = \frac{\Delta_E(\alpha')\Delta_E(\alpha)}{\int d\alpha'' \Delta_E(\alpha'')} \tag{1.24}$$

holds, where

$$\Delta_E(\alpha) \equiv (i/2\pi)(D_{E-i\sigma}(\alpha) - D_{E+i\sigma}(\alpha)). \tag{1.25}$$

2. ENERGY SHELL

In Van Hove's theory we have no explicit reference to the concept of energy shell. Such a reference, instead, is for us essential if we want to identify the states $|\alpha\rangle_0$ with the vectors $\omega_{\nu i}$ which we considered in the introduction and we think that it would be important also for a physical interpretation of the results of Van Hove. Hence we wish to know under what conditions it is possible to construct a set Ω of values of α such that the subspace spanned by the vectors $|\alpha\rangle_0$ with $\alpha \in \Omega$ practically coincides with the subspace spanned by the energy eigenvectors, corresponding to an eigenvalue between E_1 and E_2.

From the mathematical point of view, the problem is to establish under what conditions a set Ω can be constructed which has the following properties:

$$P(E_1,E_2)|\alpha\rangle_0 \cong \begin{cases} |\alpha\rangle_0 & \text{for } \alpha \in \Omega, \\ 0 & \text{for } \alpha \notin \Omega, \end{cases} \tag{2.1}$$

we have[7]

$$\begin{aligned} \|P(E_1,E_2)|\alpha\rangle_0\|^2 &= {}_0\langle\alpha|P(E_1,E_2)|\alpha\rangle_0 \\ &= \frac{i}{2\pi} \int_{E_1}^{E_2} dE\ {}_0\langle\alpha|(R_{E-i\sigma} - R_{E+i\sigma})|\alpha\rangle_0 \\ &\cong \frac{i}{2\pi} \int_{E_1}^{E_2} dE [D_{E-i\sigma}(\alpha) - D_{E+i\sigma}(\alpha)]. \end{aligned} \tag{2.2}$$

The problem therefore consists in the construction of a set Ω for which

$$\int_{E_1}^{E_2} dE \Delta_E(\alpha) \cong I_\Omega(\alpha), \tag{2.3}$$

where

$$I_\Omega(\alpha) \equiv \begin{cases} 1 & \text{for } \alpha \in \Omega, \\ 0 & \text{for } \alpha \notin \Omega. \end{cases}$$

We shall now make a number of reasonable assumptions under which the construction of the aforementioned set Ω is possible.

In the first place we shall assume that Eq. (1.15) has only one root $E(\alpha)$ (it always has at least one root), for the assumption (v), then, in particular $J_{E(\alpha)}(\alpha) > 0$ holds; we shall also assume that it is possible to take $\Delta E = E_2 - E_1$ so large that

$$\Delta E \gg \lambda^2 J_E(\alpha) \tag{2.4}$$

(the sense of the symbol $\gg$ will be examined in Sec. 5), and so that the relation

$$\lambda^2[K_{E2}(\alpha) - K_{E1}(\alpha)] \ll \Delta E, \text{ hence } \lambda^2 \partial K_E/\partial E \ll 1 \tag{2.5}$$

holds.

We call Ω the set of the values of α for which the condition

$$E_1 \leq E(\alpha) \leq E_2,$$

is satisfied. For $\alpha \in \Omega$, for (2.5) we can then write

$$\begin{aligned} D_{E\pm i\sigma}(\alpha) &\cong [\epsilon(\alpha) - E - \lambda^2 K_{E(\alpha)}(\alpha) \mp i\lambda^2 J_E(\alpha)]^{-1} \\ &= [E(\alpha) - E \mp i\lambda^2 J_E(\alpha)]^{-1}, \end{aligned} \tag{2.6}$$

[7] Strictly speaking the final equality (as we deduce from what is said in Sec. 1) holds, not for one single value of α of the discrete spectrum, but when an average has been made in a neighborhood of this value. We note, however, that this equality may be considered without doubt exact when it is used, as we shall always do in the future, within integral expressions. The same statement holds for analogous relations established in the present section.

and hence

$$\Delta_E(\alpha)\cong(1/2\pi i)\{[E(\alpha)-E-i\lambda^2 J_E(\alpha)]^{-1} -[E(\alpha)-E+i\lambda^2 J_E(\alpha)]^{-1}\} =\frac{1}{\pi}\frac{\lambda^2 J_E(\alpha)}{[E(\alpha)-E]^2+[\lambda^2 J_E(\alpha)]^2}, \tag{2.7}$$

under the condition (2.4), (2.3) is therefore satisfied.

Let us now consider some important consequences of the relation (2.3). We have

$$P(E_1,E_2)|\alpha\rangle_0=\frac{i}{2\pi}\int_{E_1}^{E_2}dE[D_{E-io}(\alpha)-D_{E+io}(\alpha)]|\alpha\rangle_0 +\frac{i}{2\pi}\int_{E_1}^{E_2}dE[(R_{E-io})_{\rm nd}-(R_{E+io})_{\rm nd}]|\alpha\rangle_0. \tag{2.8}$$

We notice that the two terms in which $P(E_1,E_2)|\alpha\rangle_0$ results decomposed in this way are orthogonal:

$$\begin{aligned}&{}_0\langle\alpha|\frac{i}{2\pi}\int_{E_1}^{E_2}dE[D_{E-io}(\alpha)-D_{E+io}(\alpha)]\\ &\times\frac{i}{2\pi}\int_{E_1}^{E_2}dE[(R_{E-io})_{\rm nd}-(R_{E+io})_{\rm nd}]|\alpha\rangle_0\\ &\cong-\frac{1}{4\pi^2}\int_{E_1}^{E_2}dE\int_{E_1}^{E_2}dE'\{(D_{E-io}-D_{E+io})\\ &\qquad\times[(R_{E'-io})_{\rm nd}-(R_{E'+io})_{\rm nd}]\}_{\rm d}(\alpha)=0.\end{aligned}$$

We then have

$$\begin{aligned}&\left|\left|\frac{i}{2\pi}\int_{E_1}^{E_2}dE[(R_{E-io})_{\rm nd}-(R_{E+io})_{\rm nd}]|\alpha\rangle_0\right|\right|^2\\ &\cong\|P(E_1,E_2)|\alpha\rangle_0\|^2\\ &\quad-\left|\left|\frac{i}{2\pi}\int_{E_1}^{E_2}dE[D_{E-io}(\alpha)-D_{E+io}(\alpha)]|\alpha\rangle_0\right|\right|^2\\ &=\int_{E_1}^{E_2}dE\Delta_E(\alpha)-\left(\int_{E_1}^{E_2}dE\Delta_E(\alpha)\right)^2\ll 1,\end{aligned} \tag{2.9}$$

where the last inequality is a consequence of (2.3). From (1.8) we have

$$\lim_{\eta\to\pm 0}(2i\eta R_{E-i\eta}R_{E+i\eta})=R_{E+io}-R_{E-io},$$

therefore (2.9) can be rewritten

$$\lim_{\substack{\eta\to\pm 0\\ \eta''\to\pm 0}}\frac{1}{\pi^2}\eta\eta''\int_{E_1}^{E_2}dE\int_{E_1}^{E_2}dE''[(R_{E-i\eta}R_{E+i\eta})_{\rm nd} \times(R_{E''-i\eta''}R_{E''+i\eta''})_{\rm nd}]_{\rm d}(\alpha)\ll 1. \tag{2.10}$$

This important relation will be useful subsequently. Let us now introduce a new system of variables $E(\alpha)$, $\beta^{(1)},\cdots,\beta^{(f-1)}$ for characterizing states instead of $\alpha^{(1)}$, $\alpha^{(2)},\cdots,\alpha^{(f)}$, and let us assume that these can be chosen in such a way that the set Ω is described when $E(\alpha)$ varies in the interval (E_1,E_2) and β in a suitable set Γ. Let us further assume that we decompose Ω in subsets Ω_ν characterized by the relations

$$\alpha\in\Omega_\nu \text{ is equivalent to } E(\alpha)\in(E_1,E_2),\quad \beta\in\Gamma_\nu,$$

where the Γ_ν's are appropriate subsets of Γ, and that the subspaces spanned by the $|\alpha\rangle_0$'s with $\alpha\in\Omega_\nu$ can be identified with the subspaces V_ν spanned by the $\omega_{\nu i}$'s, which we discussed in the introduction. We further assume that the Jacobian determinant

$$\frac{\partial(\alpha^{(1)},\alpha^{(2)},\cdots,\alpha^{(f)})}{\partial[E(\alpha),\beta^{(1)},\cdots,\beta^{(f-1)}]}$$

is practically constant for variations of $E(\alpha)$ of the order of ΔE.

We notice now that, owing to the substantial symmetry of (2.7) in $E(\alpha)$ and in E, we have

$$\int_{E_1}^{E_2}dE(\alpha)\Delta_E(\alpha)=\begin{cases}1 & \text{for } E\in(E_1,E_2),\\ 0 & \text{for } E\notin(E_1,E_2).\end{cases} \tag{2.11}$$

From this relation it follows that

$$\begin{aligned}&\int_{\Gamma_\nu}d\alpha\Delta_E(\alpha)\\ &\cong\int_{\Gamma_\nu}d\beta\left|\frac{\partial(\alpha^{(1)},\alpha^{(2)},\cdots,\alpha^{(b)})}{\partial[E(\alpha),\beta^{(1)},\cdots,\beta^{(b-1)}]}\right|\int_{E_1}^{E_2}dE(\alpha)\Delta_E(\alpha)\\ &\cong\begin{cases}\displaystyle\int_{\Gamma_\nu}d\beta\left|\frac{\partial(\alpha^{(1)},\alpha^{(2)},\cdots,\alpha^{(b)})}{\partial[E(\alpha),\beta^{(1)},\cdots,\beta^{(b-1)}]}\right|\cong\frac{1}{\Delta E}\mu(\Omega_\nu) & \text{for } E\in(E_1,E_2),\\ 0 & \text{for } E\notin(E_1,E_2),\end{cases}\end{aligned} \tag{2.12}$$

and, in the same way,

$$\int_\Omega d\alpha\Delta_E(\alpha)\cong\begin{cases}(1/\Delta E)\mu(\Omega) & \text{for } E\in(E_1,E_2),\\ 0 & \text{for } E\notin(E_1,E_2),\end{cases} \tag{2.13}$$

where $\mu(\Omega_\nu)$ and $\mu(\Omega)$ represent Lebesque's measures of Ω_ν and Ω in α space.

From the relation (2.11) we can further deduce

$$\int d\alpha\Delta_E(\alpha)\cong\int_\Omega d\alpha\Delta_E(\alpha) \quad \text{if} \quad E\in(E_1,E_2). \tag{2.14}$$

From (2.12), (2.13), and (2.14) we have further,

$$\int_{\Omega_\nu} d\alpha \int_{-\infty}^{+\infty} dE q_E(\alpha,\alpha_0)$$

$$= \int_{-\infty}^{+\infty} dE \frac{\Delta_E(\alpha_0)}{\int d\alpha'' \Delta_E(\alpha'')} \int_{\Omega_\nu} d\alpha \Delta_E(\alpha)$$

$$\cong \frac{\mu(\Omega_\nu)}{\Delta E} \int_{E_1}^{E_2} dE \frac{\Delta_E(\alpha_0)}{\int_\Omega d\alpha'' \Delta_E(\alpha'')} \cong \frac{\mu(\Omega_\nu)}{\mu(\Omega)} \int_{E_1}^{E_2} dE \Delta_E(\alpha_0)$$

$$\cong \frac{\mu(\Omega_\nu)}{\Omega(\Omega)} I_\Omega(\alpha_0). \quad (2.15)$$

To illustrate the meaning of (2.15) we may apply it to the results of Van Hove. Then this relation shows that, under our supplementary assumptions, the probability distribution of Van Hove coincides with the microcanonic probability distributions as t goes to $\pm\infty$, not only in a formal way, but also in the strict sense which we used in our formulation of the ergodic problem.[8]

3. A PRELIMINARY RELATION

Equation (1.24) essentially gives us the limit for $\eta \to 0$ of the expression

$$\eta(R_{E-i\eta} A R_{E+i\eta})_{\mathrm{d}};$$

we will subsequently need an analogous limit for the expression

$$\eta(R_{E-i\eta} A R_{E+i\eta})_{\mathrm{nd}}$$

also. This can be calculated immediately if we bear in mind the relation

$$\begin{aligned}(R_l A R_{l'})_{\mathrm{nd}} = {} & (R_l A R_{l'})_{\mathrm{d}} (-\lambda V + \lambda^2 V D_{l'} V - \cdots)_{\mathrm{ind}} D_{l'} \\ & + D_l (-\lambda V + \lambda^2 V D_l V - \cdots)_{\mathrm{ind}} (R_l A R_{l'})_{\mathrm{d}} \\ & + D_l [(-\lambda V + \lambda^2 V D_l V - \cdots)(R_l A R_{l'})_{\mathrm{d}} \\ & \times (-\lambda V + \lambda^2 V D_{l'} V - \cdots)]_{\mathrm{ind}} D_{l'}. \quad (3.1)\end{aligned}$$

This relation can be proved by considerations similar to those with which Van Hove proves (1.18).

From (3.1), taking into account (1.24), one obtains

$$\begin{aligned}\lim_{\eta \to \pm 0} & [\eta \langle \alpha | (R_{E-i\eta} A R_{E+i\eta})_{\mathrm{nd}} | \alpha' \rangle] \\ = {} & \int d\alpha_1 A(\alpha_1) q_E(\alpha_1,\alpha) \\ & \times \langle \alpha | (-\lambda V + \lambda^2 V D_{E \pm io} V - \cdots)_{\mathrm{ind}} | \alpha' \rangle D_{E \pm io}(\alpha') \\ & + D_{E \mp io}(\alpha) \langle \alpha | (-\lambda V + \lambda^2 V D_{E \mp io} V - \cdots)_{\mathrm{ind}} | \alpha' \rangle \\ & \times \int d\alpha_1 A(\alpha_1) q_E(\alpha_1,\alpha') + D_{E \mp io}(\alpha) \\ & \times \int d\alpha_1 \int d\alpha_2 A(\alpha_1) q_E(\alpha_1,\alpha_2) \\ & \times \langle \alpha | [(-\lambda V + \lambda^2 V D_{E \mp io} V - \cdots) | \alpha_2 \rangle \\ & \times \langle \alpha_2 | (-\lambda V + \lambda^2 V D_{E \pm io} V - \cdots)]_{\mathrm{ind}} | \alpha' \rangle D_{E \pm io}(\alpha'),\end{aligned}$$

where the upper and lower signs must be taken always together.

4. ERGODICITY CONDITIONS

Let us go back now to ergodicity conditions (I2′) and (I2″). We wish to examine whether they are satisfied for a system which satisfies the postulates of Secs. 1 and 2 when the energy shell is identified with the subspace spanned by vectors $|\alpha\rangle_0$ with $\alpha \in \Omega$ and the subspaces V_ν (cells) with the subspaces spanned by vectors $|\alpha\rangle_0$ with $\alpha \in \Omega_\nu$. In fact, instead of proving (I2′) and (I2″) directly, we shall prove certain other relations, more restrictive and formally simpler. If we assume $s_\nu \gg N^{\frac{1}{2}}$, the left-hand side of (I2″) can be expanded and simplified and (I2′) and (I2″) can be replaced by the other conditions

$$p_{\mu\nu} \equiv (1/s_\mu) \sum_1^{s_\mu} {}_j \sum_1^{s_\nu} {}_i (\omega_{\mu j}, P_\rho \omega_{\nu i})(\omega_{\nu i}, P_\rho \omega_{\mu j}) = s_\nu / S[1 + O(\chi/N^{\frac{1}{2}})], \quad (4.1')$$

$$q_{\mu\nu} \equiv (1/s_\mu) \sum_1^{s_\mu} {}_j \sum_1^{s_\nu} {}_i (\omega_{\mu j}, P_\rho \omega_{\nu i})(\omega_{\nu i}, P_\rho \omega_{\mu j'}) \cdot (\omega_{\mu j'}, P_{\rho'} \omega_{\nu i'})(\omega_{\nu i'}, P_{\rho'} \omega_{\mu j}) = s_\nu^2 / S^2[1 + O(\chi/N^{\frac{1}{2}})], \quad (4.1'')$$

where χ is a number ($\ll 1$) of the order required for the left-hand side of (I1). If we replace the $\omega_{\nu i}$'s by the $|\alpha_0\rangle$'s, the left-hand sides of (4.1′) and (4.1″) can be

[8] Van Hove's results are deduced under the hypothesis that in the initial state $|\psi_0\rangle = \int d\alpha c(\alpha) |\alpha\rangle$, the phases of the $c(\alpha)$ are randomly distributed. In comparing the point of view of Van Hove with ours, the question arises whether this hypothesis is consistent with our assumption that $|\psi_0\rangle$ belongs to the energy shell or to a particular cell V_μ; this is clearly true in our case, but this would not be true if the relations (2.4) and (2.5) did not hold.

written

$$p_{\mu\nu}=\frac{1}{\rho_0\mu(\Omega_\mu)}\sum_{\alpha\in\Omega_\mu}\sum_{\alpha'\in\Omega_\nu}\sum_\rho {}_0\langle\alpha|P_\rho|\alpha'\rangle_0\,{}_0\langle\alpha'|P_\rho|\alpha\rangle_0, \quad (4.2')$$

$$q_{\mu\nu}=\frac{1}{\rho_0\mu(\Omega_\mu)}\sum_{\alpha\in\Omega_\mu}\sum_{\alpha'\in\Omega_\nu}\sum_{\alpha''\in\Omega_\mu}\sum_{\alpha'''\in\Omega_\nu}\sum_\rho\sum_{\rho'} {}_0\langle\alpha|P_\rho|\alpha'\rangle_0$$
$$\times{}_0\langle\alpha'|P_\rho|\alpha''\rangle_0\,{}_0\langle\alpha''|P_{\rho'}|\alpha'''\rangle_0\,{}_0\langle\alpha'''|P_{\rho'}|\alpha\rangle_0. \quad (4.2'')$$

If we now denote by δE the distance between two energy levels in a neighborhood of E_ρ, we have for (1.6),

$$P_\rho=\frac{i}{2\pi}\int_{E_\rho-\delta E/2}^{E_\rho+\delta E/2}dE'(R_{E'-io}-R_{E'+io})$$
$$=\frac{i}{2\pi}\int_{E-\delta E/2}^{E+\delta E/2}dE'(R_{E'-io}-R_{E'+io}), \quad (4.3)$$

where in the last expression E is a generic internal point in the interval $(E_\rho\delta E-/2, E_\rho+\delta E/2)$. In (4.2') and (4.2'') we therefore get

$$\sum_\rho {}_0\langle\alpha|P_\rho|\alpha'\rangle_0\,{}_0\langle\alpha'|P_\rho|\alpha''\rangle_0=-\frac{1}{4\pi^2}\sum_\rho\int_{E_\rho-\delta E/2}^{E_\rho+\delta E/2}dE\int_{E-\delta E/2}^{E+\delta E/2}dE'\,{}_0\langle\alpha|(R_{E-io}-R_{E+io})|\alpha'\rangle_0\,{}_0\langle\alpha'|(R_{E'-io}-R_{E'+io})|\alpha''\rangle_0$$
$$=-\frac{1}{4\pi^2}\int_{-\infty}^{+\infty}dE\int_{E-\delta E/2}^{E+\delta E/2}dE'\,{}_0\langle\alpha|(R_{E-io}-R_{E+io})|\alpha'\rangle_0\,{}_0\langle\alpha'|(R_{E'-io}-R_{E'+io})|\alpha''\rangle_0. \quad (4.4)$$

For the above considerations, by (1.2) and (1.2') we can write

$$\frac{1}{\rho_0}\sum_{\alpha\in\Omega_\mu}\sum_{\alpha\in\Omega_\nu}{}_0\langle\alpha|(R_{E-io}-R_{E+io})|\alpha'\rangle_0\,{}_0\langle\alpha'|(R_{E'-io}-R_{E'+io})|\alpha\rangle_0$$
$$\cong\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'[(R_{E-io}-R_{E+io})|\alpha'\rangle\langle\alpha'|(R_{E'-io}-R_{E'+io})]_{\mathrm{d}}(\alpha)$$
$$=\int_{\Omega_\mu}d\alpha[(R_{E-io}I_{\Omega_\nu}R_{E-io})_{\mathrm{d}}(\alpha)-(R_{E-io}I_{\Omega_\nu}R_{E'+io})_{\mathrm{d}}(\alpha)-(R_{E+io}I_{\Omega_\nu}R_{E'-io})_{\mathrm{d}}(\alpha)+(R_{E+io}I_{\Omega_\nu}R_{E'+io})_{\mathrm{d}}(\alpha)],$$

where I_{Ω_ν} is a diagonal operator which is defined by the relation $I_{\Omega_\nu}|\alpha\rangle=I_{\Omega_\nu}(\alpha)|\alpha\rangle$. Then for (1.16) we have

$$p_{\mu\nu}=-\frac{1}{2\pi^2}\frac{1}{\mu(\Omega_\mu)}\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'\int_{-\infty}^{+\infty}dE\int_{E-\delta E/2}^{E+\delta E/2}dE'[X_{E-ioE'-io}(\alpha',\alpha)-X_{E-ioE'+io}(\alpha',\alpha)$$
$$-X_{E+ioE'-io}(\alpha',\alpha)+X_{E+ioE'+io}(\alpha',\alpha)]$$
$$=-\frac{1}{2\pi^2}\frac{1}{\mu(\Omega_\mu)}\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'\int_{-\infty}^{+\infty}dW\int_{-\delta E}^{\delta E}dW'[X_{W+W'-io\;W-W'-io}(\alpha',\alpha)-X_{W+W'-io\;W-W'+io}(\alpha',\alpha)$$
$$-X_{W+W'+io\;W-W'-io}(\alpha',\alpha)+X_{W+W'+io\;W-W'+io}(\alpha',\alpha)].$$

Since δE is very small (see Sec. 5), in the preceding relation we can replace the right-hand side by its limit for $\delta E\to 0$. Taking into account then (see Sec. 1) that $X_{E+l\;E+l}(\alpha',\alpha)$ has on the real axis only finite discontinuities, that $X_{E+l\;E-l}(\alpha',\alpha)$ has a pseudopole in $l=0$, and taking into account the relations (1.24) and (2.15), we have

$$p_{\mu\nu}\cong\frac{1}{2\pi}\frac{1}{\mu(\Omega_\mu)}\lim_{\eta\to 0}\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'\int_{-\infty}^{+\infty}dW[\eta X_{W-i\eta W+i\eta}(\alpha',\alpha)+\eta X_{W+i\eta W-i\eta}(\alpha',\alpha)]$$
$$=\frac{1}{\mu(\Omega_\mu)}\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'\int_{-\infty}^{+\infty}dW q_W(\alpha',\alpha)\cong\frac{\mu(\Omega_\nu)}{\mu(\Omega)}=\frac{s_\nu}{S}. \quad (4.5)$$

Reasoning in a similar way as we did for $p_{\mu\nu}$, we can write

$$q_{\mu\nu}\cong\frac{1}{16\pi^4\mu(\Omega_\mu)}\int_{\Omega_\mu}d\alpha\int_{\Omega_\nu}d\alpha'\int_{\Omega_\mu}d\alpha''\int_{\Omega_\nu}d\alpha'''\int_{-\infty}^{+\infty}dE\int_{E-\delta E/2}^{E+\delta E/2}dE'\int_{-\infty}^{+\infty}dE''\int_{E''-\delta E/2}^{E''+\delta E/2}dE'''$$
$$\times[(R_{E-io}-R_{E+io})|\alpha'\rangle\langle\alpha'|(R_{E'-io}-R_{E'+io})|\alpha''\rangle\langle\alpha''|(R_{E''-io}-R_{E''+io})|\alpha'''\rangle\;\langle\alpha'''|(R_{E'''-io}-R_{E'''+io})]_{\mathrm{d}}(\alpha)$$

$$= \frac{1}{16\pi^4\mu(\Omega_\mu)} \int_{\Omega_\mu} d\alpha [(R_{E-io} I_{\Omega_\nu} R_{E'-io} I_{\Omega_\mu} R_{E''-io} I_{\Omega_\nu} R_{E'''-io})_{\mathrm{d}}(\alpha) - (R_{E+io} I_{\Omega_\nu} R_{E'-io} I_{\Omega_\mu} R_{E''-io} I_{\Omega_\nu} R_{E'''-io})_{\mathrm{d}}(\alpha) - \cdots$$
$$+ (R_{E+io} I_{\Omega_\nu} R_{E'+io} I_{\Omega_\mu} R_{E''+io} I_{\Omega_\nu} R_{E'''+io})_{\mathrm{d}}(\alpha)]$$
$$\cong \frac{1}{4\pi^2\mu(\Omega_\mu)} \lim_{\substack{\eta\to+0\\ \eta''\to+0}} \eta\eta'' \int_{\Omega_\mu} d\alpha \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' [(R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta} I_{\Omega_\mu} R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{d}}(\alpha)$$
$$+ (R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta} I_{\Omega_\mu} R_{E''+i\eta''} I_{\Omega_\nu} R_{E''-i\eta''})_{\mathrm{d}}(\alpha) + (R_{E+i\eta} I_{\Omega_\nu} R_{E-i\eta} I_{\Omega_\mu} R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{d}}(\alpha)$$
$$+ (R_{E+i\eta} I_{\Omega_\nu} R_{E-i\eta} I_{\Omega_\mu} R_{E''+i\eta''} I_{\Omega_\nu} R_{E''-i\eta''})_{\mathrm{d}}(\alpha)],$$

where we have taken into account the fact that the matrix element $\langle\alpha| R_{E+l} A R_{E+l} |\alpha'\rangle$ has only finite discontinuities on the real axis as can be deduced from (1.8).

The calculation of $q_{\mu\nu}$ is therefore brought back to the evaluation of limits of the kind

$$\lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' (R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta} I_{\Omega_\mu} R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{d}}(\alpha). \tag{4.7}$$

Since obviously the relation

$$(R_l A_1 R_{l'} A_2 R_{l''} A_3 R_{l'''})_{\mathrm{d}} = (R_l A_1 R_{l'})_{\mathrm{d}} A_2 (R_{l''} A_3 R_{l'''})_{\mathrm{d}} + [(R_l A_1 R_{l'})_{\mathrm{nd}} A_2 (R_{l''} A_3 R_{l'''})_{\mathrm{nd}}]_{\mathrm{d}}$$

holds, we see that (4.7) is the sum of the other expressions

$$\lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' (R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{d}}(\alpha) I_{\Omega_\mu}(\alpha) \cdot (R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{d}}(\alpha) \tag{4.8}$$

and

$$\lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' [(R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{nd}} I_{\Omega_\mu} \cdot (R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{nd}}]_{\mathrm{d}}(\alpha). \tag{4.9}$$

Equation (4.8) can be immediately calculated by means of the preceding considerations; we get

$$\lim_{\substack{\eta\to0\\ \eta''\to0}} \eta\eta'' \frac{1}{\pi^2} \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' (R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{d}}(\alpha) I_{\Omega_\mu}(\alpha) \cdot (R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{d}}(\alpha)$$
$$= \int_{\Omega_\nu} d\alpha' \int_{-\infty}^{+\infty} dE q_E(\alpha',\alpha) \int_{\Omega_\nu} d\alpha''' \int_{-\infty}^{+\infty} dE'' q_{E''}(\alpha''',\alpha) I_{\Omega_\mu}(\alpha) \cong \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2 I_{\Omega_\mu}(\alpha). \tag{4.10}$$

For the calculation of (4.9) we notice that for (2.12), (2.13), and (2.14) we have

$$\int_{\Omega_\nu} d\alpha_1 q_E(\alpha_1,\alpha) \cong \frac{\mu(\Omega_\nu)}{\mu(\Omega)} \int_{\Omega} d\alpha_1 q_E(\alpha_1,\alpha) \cong \begin{cases} \mu(\Omega_\nu)/\mu(\Omega) \int d\alpha_1 q_E(\alpha_1,\alpha) & \text{for } E\in(E_1,E_2), \\ 0 & \text{for } E\in(E_1,E_2). \end{cases} \tag{4.11}$$

Then for (3.2) we have

$$\lim_{\eta\to\pm0} \eta \int_{-\infty}^{+\infty} dE (R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{nd}} \cong \frac{\mu(\Omega_\nu)}{\mu(\Omega)} \lim_{\eta\to\pm0} \eta \int_{E_1}^{E_2} dE (R_{E-i\eta} R_{E+i\eta})_{\mathrm{nd}} \tag{4.12}$$

and taking into account (2.10),

$$\lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' [(E_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{nd}} I_{\Omega_\mu} (R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{nd}}]_{\mathrm{d}}(\alpha)$$
$$\cong \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2 \lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' [(R_{E-i\eta} R_{E+i\eta})_{\mathrm{nd}} I_{\Omega_\mu} (R_{E''-i\eta''} R_{E''+i\eta''})_{\mathrm{nd}}]_{\mathrm{d}}(\alpha)$$
$$\leq \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2 \lim_{\substack{\eta\to\pm0\\ \eta''\to\pm0}} \eta\eta'' \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE'' [(R_{E-i\eta} R_{E+i\eta})_{\mathrm{nd}} (R_{E''-i\eta''} R_{E''+i\eta''})_{\mathrm{nd}}]_{\mathrm{d}}(\alpha) \ll \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2. \tag{4.13}$$

Equations (4.10) and (4.13) allow us to write

$$q_{\mu\nu}\cong[\mu(\Omega_\nu)/\mu(\Omega)]^2=s_\nu^2/S^2. \tag{4.14}$$

5. EVALUATION OF ERRORS INVOLVED IN THE APPROXIMATION MADE IN PRECEDING SECTIONS

If we want to identify the relations (4.5) and (4.14) with (4.1′) and (4.1″), it is necessary and sufficient to prove that the relative errors which were made as a consequence of the different approximation, are of the order of $\chi/N^{\frac{1}{4}}$.

With this in view we notice that the approximations which we made in the preceding section are essentially of the three following types:

(1) The relations (1.2) and (1.2′) have been used. The relative error which the replacing of an expression of the type $(1/\rho_0)\sum_\alpha f(\alpha)$ with the other $\int d\alpha f(\alpha)$ introduces, is of the order of $\delta\alpha/\Delta\alpha$ where $\delta\alpha$ represents the distance between two points of the discrete spectrum and $\Delta\alpha$ a variation of α to which a variation of $f(\alpha)$, comparable with $f(\alpha)$ itself, corresponds. In our case $D_{E\pm io}(\alpha)$ may be considered the expression in the integrals that varies faster. We have a relevant variation of this quantity for variations of $E(\alpha)$ of the order of $\lambda^2 J$, where J is a number of the order of the typical $J_E(\alpha)$. Therefore $\Delta\alpha$ results equal to $\lambda^2 J/(\partial E(\alpha)/\partial\alpha)$ and hence [obviously $\delta\alpha\partial E(\alpha)/\partial\alpha\sim\delta E$] the error which is a consequence of the approximation we have made is of the order of

$$\delta E/\lambda^2 J. \tag{5.1}$$

(2) The expressions

$$\int_{-\delta E}^{\delta E} dW' X_{W+W'\pm io W-W'\pm io}(\alpha',\alpha),$$

$$\int_{-\delta E}^{\delta E} dW' \langle\alpha|\,(R_{W+'W\pm io} A R_{W-W'\pm io})_{\mathrm{nd}}\,|\alpha'\rangle$$

have been replaced by their limit for $\delta E\to 0$. The absolute error that we make when we replace an expression of the kind

$$\frac{1}{i\pi}\int_{-\epsilon}^{\epsilon}\frac{g(w)}{w-io}dw$$

with ϵ very small by its limit for $\epsilon\to 0$ $g(0)$, is of the order of the variations of $g(w)$ in the interval $(-\epsilon,\epsilon)$. Then the corresponding relative error results of the order of ϵ/d, where d represents a variation of w (in a neighborhood of $w=0$) for which $g(w)$ has a variation of the order of $g(w)$ itself. In our case $\epsilon=\delta E$ and, as we may deduce for example from (1.20), d can be identified by $\lambda^2 J$. Therefore the relative error which one makes in consequence of this approximation is expressed again by the quantity (5.1).

(3) The relations (2.3) and (2.11) and the others that have been deduced in Secs. 2 and 3 by means of (2.3) (2.11) have been used.

To evaluate the error which we make if we use the relations (2.3) and (2.11), we notice in the first place that they are always used in expressions of the kind

$$\int_{\Omega_\lambda} d\alpha f(\alpha)\int_{E_1}^{E_2} dE\Delta_E(\alpha), \tag{5.2}$$

$$\int_{-\infty}^{+\infty} dE g(E)\int_{E_1}^{E_2} dE(\alpha)\Delta_E(\alpha), \tag{5.3}$$

where $f(\alpha)$ and $g(E)$ are very smooth functions, respectively, of α and E. We note again, therefore, that the approximations made are in effect two distinct ones. The first consists of the use of the relation (2.7) and shows a relative error of the order of

$$\Delta E/\lambda^2\Delta K, \tag{5.4}$$

where ΔK is a number of the order of $K_{E_2}(\alpha)-K_{E_1}(\alpha)$. The second appears in identifying the values of the integrals

$$\frac{1}{\pi^2}\int_{E_1}^{E_2} dE\frac{\lambda^2 J_E(\alpha)}{[E-E(\alpha)]^2+[\lambda^2 J_E(\alpha)]^2}, \tag{5.5}$$

$$\frac{1}{\pi^2}\int_{E_1}^{E_2} dE(\alpha)\frac{\lambda^2 J_E(\alpha)}{[E-E(\alpha)]^2+[\lambda^2 J_E(\alpha)]^2}, \tag{5.6}$$

with the right-hand sides of the (2.3) and (2.11). This shows, in expressions of the type (5.2) and (5.3), a relative error of the order

$$\lambda^2 J/\Delta E. \tag{5.7}$$

To convince ourselves of this, it is sufficient to note that the expression (5.7) represents the relative error made in the evaluation of the integrals (5.5) and (5.6) for the greatest part of the values of α and, respectively, of E. It also represents the relation between the measure of the "exceptional" ensemble, for which the said error is of the order of one [this is the ensemble of the values of α for which $E(\alpha)-E_1\sim\lambda^2 J$ or $E(\alpha)-E_2\sim\lambda^2 J$ and respectively $E-E_1\sim\lambda^2 J$ or $E-E_2\sim\lambda^2 J$] and the measure of the set to which the integral (5.2) is extended [or which contributes to the integral (5.3)]. Summing up, it is possible to identify (4.5) and (4.14) with (4.1′) and (4.1″) if we determine the qualitative hypothesis of the foregoing sections on the density of levels and the largeness of ΔE, ΔK, and J by requiring that the expressions (5.1), (5.4), and (5.7) be of the order of $\chi/N^{\frac{1}{4}}$. In symbols

$$\left.\begin{matrix}\delta E/\lambda^2 J\\ \Delta E/\lambda^2\Delta K\\ \lambda^2 J/\Delta K\end{matrix}\right\}\sim\chi/N^{\frac{1}{4}}. \tag{5.8}$$

The only cases which need particular consideration because they are not altogether trivial are (2.10) and (4.13).

Equation (2.10), when (5.8) is presumed to be

verified, must be determined on the basis that the left-hand side is of the order $\chi/N^{\frac{1}{2}}$ when $E(\alpha)-E_1$ and $E(\alpha)-E_2$ are sufficiently large with respect to $\lambda^2 J$ and is of the order of one in the region in which these expressions are of the order of $\lambda^2 J$. We may substitute for (4.13) the more exact relation

$$\lim_{\substack{\eta\to\pm 0\\ \eta''\to\pm 0}} \eta\eta''\frac{1}{\mu(\Omega_\mu)}\int_{\Omega_\mu} d\alpha \int_{-\infty}^{+\infty} dE \int_{-\infty}^{+\infty} dE''\left[(R_{E-i\eta} I_{\Omega_\nu} R_{E+i\eta})_{\mathrm{nd}} I_{\Omega_\mu}\cdot (R_{E''-i\eta''} I_{\Omega_\nu} R_{E''+i\eta''})_{\mathrm{nd}}\right]_{\mathrm{d}}(\alpha)$$

$$\leq \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2 \lim_{\substack{\eta\to\pm 0\\ \eta''\to\pm 0}} \eta\eta''\frac{1}{\mu(\Omega_\mu)}\int_{E_1}^{E_2} dE\cdot\int_{E_1}^{E_2} dE''\left[(R_{E-i\eta} R_{E+i\eta})_{\mathrm{nd}} (R_{E''-i\eta''} R_{E''+i\eta''})_{\mathrm{nd}}\right]_{\mathrm{d}}(\alpha)\cdot\left[1+\mathrm{O}\left(\frac{\chi}{N^{\frac{1}{2}}}\right)\right]$$

$$= \left(\frac{\mu(\Omega_\nu)}{\mu(\Omega)}\right)^2 \mathrm{O}\left(\frac{\chi}{N^{\frac{1}{2}}}\right)$$

and what we have said above about (4.5) and (4.14) is confirmed.

6. CONCLUDING CONSIDERATIONS

In this section we will make some observations on the results of the preceding section and, successively, compare briefly some aspects of the ergodic method and of the master equation method, and, in consequence, some aspects of Van Hove's deduction and of our own.

Bearing in mind all of the hypotheses, successively introduced on the analytical properties of the Hamiltonian of the physical system discussed, we saw in the preceding section that if the relative errors with which $p_{\mu\nu}$ and $q_{\mu\nu}$ were evaluated were to be of the order $\chi/N^{\frac{1}{2}}$ as required by (4.1), then the conditions (5.8) would have to be satisfied. These conditions are rather restrictive, because, although the value of ΔE and the dimensions of the cells are largely arbitrary, N must, in all cases, be a very large number. We wish to note, however, that they are excessively restrictive and that at least, on principle, it is possible to weaken them.

The appearance of the factor $1/N^{\frac{1}{2}}$ in the right-hand side of (5.8) is essentially due to the demand that the following relation should be satisfied

$$\mathfrak{B}\sum_{\nu=1}^{N}\frac{(Mu_\nu(t)-s_\nu/S)^2}{s_\nu^2/S^2}\sim\chi. \tag{6.1}$$

This would not have occurred, had we required the relation

$$\mathfrak{B}\frac{(Mu_\nu(t)-s_\nu/S)^2}{s_\nu^2/S^2}\sim\chi(\nu=1,2,\cdots,N). \tag{6.2}$$

The reason why we chose relation (6.1) instead of (6.2) is the following. Relation (6.1) gives us the assurance that with the exception of an ensemble of initial states of very small weight, the expression

$$\sum_{\nu=1}^{N}\bar{\Delta}_\nu,$$

where

$$\bar{\Delta}_\nu\equiv\frac{(Mu_\nu(t)-s_\nu/S)^2}{s_\nu^2/S^2},$$

is very small and, consequently, the quantities $\bar{\Delta}_\nu$ are simultaneously small for all the values of ν. Instead, the relation (6.2) assures us that, with exception made for an ensemble of initial states of very small weight, each single $\bar{\Delta}_\nu$ is small, but tells us nothing with regard to the weight of the ensemble of initial states for which all the $\bar{\Delta}_\nu$'s are simultaneously very small. On the other hand, if relation (6.2) is too weak, (6.1) is certainly too strong because, in order for us to prove the desired result (the existence of a class of states, the weight of which very closely approaches unity, for which the $\bar{\Delta}_\nu$'s are simultaneously small), the demand that (6.1) be satisfied implies the selection of the most unfavorable circumstances, i.e., to assume as empty the intersections of the ensembles of exceptional initial states (for which the single $\bar{\Delta}_\nu$'s are of the order of one) when taken two by two.

We shall now pass to the already mentioned comparison between particular aspects of the ergodic method and the master equation. We must at first remark that the systems to which statistical mechanics are applied are essentially spatially-limited systems and composed of a very large, but finite number of particles. The energy spectrum, effective or unperturbed, of these systems is therefore always discrete, even if the density of the levels is very high owing to the large number of degrees of freedom and to the large spatial extension of the system. The general solution of the Schrödinger equation is therefore, a Fourier series in time and is consequently an almost periodic function of the time. Thus a recurrence theorem holds[9] which is analogous to Poincaré's classical one and the recurrence time T is of the order of $1/\delta E$.

We now note that the master equation describes essentially the temporal evolution of the system from a certain initial situation towards a final situation, a

[9] P. Bocchieri and A. Loinger, Phys. Rev. **107**, 337 (1957).

state of macroscopic equilibrium in which the system itself remains indefinitely. Here the character of recurrences of the physical system is in no manner present and therefore it can describe the temporal evolution of the system solely in an interval of time which is small in respect to T.

The state of equilibrium is defined in classical statistics as the state, presumed existent, in which the system remains the greater part of the time. In quantum statistics, equilibrium is defined as a macrostate which possesses at a time randomly taken an occupational probability very close to one. [$Mu_\nu(t)$ has precisely the meaning of occupation probability of the macrostate which corresponds to the variety V_ν at a randomly taken time very close to one.]

To these differences of conceptual character between the ergodic method and the master equation method there correspond important differences even in the analytical developments which are necessary in order to prove the relative theorems. A fundamental approximation introduced by Van Hove in his deduction of the master equation consists in considering the spectrum of the system as continuous, when, as we have seen, is essentially discrete. By this means Fourier's series which expresses, for instance, the occupation probability of an unperturbed state is replaced by the corresponding integral of Fourier, which does not present the character of quasi-periodicity of the original series. In the demonstration which we have given herewith of our ergodic theorem, we too have considered the spectrum of the system as continuous, but we have done so only to evaluate expressions such as $p_{\mu\nu}$ and $q_{\mu\nu}$ which are already the result of time averages which were calculated bearing in mind the discrete character of the spectrum itself. It would have been pointless if instead of having started from the ergodicity conditions (I.2′) and (I.2″), [i.e., from (12′) and (12″) of paper A], we had started from the conditions (10′) and (10″) of the same paper and deals with the left-hand sides of the latter in the above cited approximation before calculating he time average.

A final observation which we wish to make is that the coarse grained probability used by us is, in effect, very much coarser grained than the one used by Van Hove. The Van Hove probability density $p_t(\alpha)$ in a certain point α is essentially an average in a small neighborhood of α of the microscopic probability; the coarse grained probability $u_\nu(t)$ used by us is the sum of the microscopic probabilities relative to all the values of α which belong to the set Ω_ν or (which amounts to the same thing) the integral extended to Ω_ν of $p_t(\alpha)$.

ACKNOWLEDGMENTS

I would like to thank Professor P. Caldirola, Professor Loinger, and Dr. A. Scott, for the kind interest they have shown in this work and for their stimulating discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

On the Calculation of Analytic Functions of Cyclic Matrices*

PHILLIP B. ABRAHAM† AND GEORGE WEISS
The Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland

(Received August 17, 1961)

Recently Löwdin, Pauncz, and de Heer, have discussed the calculation of functions of cyclic matrices. They presented three exact methods for doing this. We show that the last of their methods can be generalized to a method which is convenient for approximate calculations and which in addition can be extended to the calculation of functions of higher order cyclic matrices. A number of examples are presented. It is also shown that by the same techniques it is possible to evaluate functions of a skew circulant matrix.

IN a recent paper Löwdin, Pauncz, and de Heer[1] have presented several methods for the calculation of the inverse, and inverse square root, of a simple symmetric cyclic matrix. These matrices arise in many problems which involve the use of cyclic boundary conditions, such as in the theory of lattice vibrations or the tight binding approximation in the theory of metals. The treatment of these problems in higher dimensions requires the definition of generalized cyclic mattices. While these matrices were discussed by Lewis and Keller[2] the problem of calculating various functions of generalized cyclic matrices was only barely touched on in their paper.

It is the purpose of this paper to discuss a method for the calculation of functions of cyclic matrices, both for

* This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command.

† The material in this article is part of a thesis to be submitted by P. B. Abraham in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Maryland.

[1] P.-O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. 1, 461 (1960).

[2] R. M. Lewis and J. B. Keller, Phys. Rev. 121, 1022 (1961).

state of macroscopic equilibrium in which the system itself remains indefinitely. Here the character of recurrences of the physical system is in no manner present and therefore it can describe the temporal evolution of the system solely in an interval of time which is small in respect to T.

The state of equilibrium is defined in classical statistics as the state, presumed existent, in which the system remains the greater part of the time. In quantum statistics, equilibrium is defined as a macrostate which possesses at a time randomly taken an occupational probability very close to one. [$Mu_\nu(t)$ has precisely the meaning of occupation probability of the macrostate which corresponds to the variety V_ν at a randomly taken time very close to one.]

To these differences of conceptual character between the ergodic method and the master equation method there correspond important differences even in the analytical developments which are necessary in order to prove the relative theorems. A fundamental approximation introduced by Van Hove in his deduction of the master equation consists in considering the spectrum of the system as continuous, when, as we have seen, is essentially discrete. By this means Fourier's series which expresses, for instance, the occupation probability of an unperturbed state is replaced by the corresponding integral of Fourier, which does not present the character of quasi-periodicity of the original series. In the demonstration which we have given herewith of our ergodic theorem, we too have considered the spectrum of the system as continuous, but we have done so only to evaluate expressions such as $p_{\mu\nu}$ and $q_{\mu\nu}$ which are already the result of time averages which were calculated bearing in mind the discrete character of the spectrum itself. It would have been pointless if instead of having started from the ergodicity conditions (I.2′) and (I.2″), [i.e., from (12′) and (12″) of paper A], we had started from the conditions (10′) and (10″) of the same paper and deals with the left-hand sides of the latter in the above cited approximation before calculating he time average.

A final observation which we wish to make is that the coarse grained probability used by us is, in effect, very much coarser grained than the one used by Van Hove. The Van Hove probability density $p_t(\alpha)$ in a certain point α is essentially an average in a small neighborhood of α of the microscopic probability; the coarse grained probability $u_\nu(t)$ used by us is the sum of the microscopic probabilities relative to all the values of α which belong to the set Ω_ν or (which amounts to the same thing) the integral extended to Ω_ν of $p_t(\alpha)$.

ACKNOWLEDGMENTS

I would like to thank Professor P. Caldirola, Professor Loinger, and Dr. A. Scott, for the kind interest they have shown in this work and for their stimulating discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH–APRIL, 1962

On the Calculation of Analytic Functions of Cyclic Matrices*

PHILLIP B. ABRAHAM† AND GEORGE WEISS
The Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland

(Received August 17, 1961)

Recently Löwdin, Pauncz, and de Heer, have discussed the calculation of functions of cyclic matrices. They presented three exact methods for doing this. We show that the last of their methods can be generalized to a method which is convenient for approximate calculations and which in addition can be extended to the calculation of functions of higher order cyclic matrices. A number of examples are presented. It is also shown that by the same techniques it is possible to evaluate functions of a skew circulant matrix.

IN a recent paper Löwdin, Pauncz, and de Heer[1] have presented several methods for the calculation of the inverse, and inverse square root, of a simple symmetric cyclic matrix. These matrices arise in many problems which involve the use of cyclic boundary conditions, such as in the theory of lattice vibrations or the tight binding approximation in the theory of metals. The treatment of these problems in higher dimensions requires the definition of generalized cyclic mattices. While these matrices were discussed by Lewis and Keller[2] the problem of calculating various functions of generalized cyclic matrices was only barely touched on in their paper.

It is the purpose of this paper to discuss a method for the calculation of functions of cyclic matrices, both for

* This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command.

† The material in this article is part of a thesis to be submitted by P. B. Abraham in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Maryland.

[1] P.-O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. 1, 461 (1960).

[2] R. M. Lewis and J. B. Keller, Phys. Rev. 121, 1022 (1961).

those arising from one-dimensional problems and those arising from higher-dimensional problems. Our techniques is a generalization of the method quoted in reference 1 for the calculation of the inverse square root of a cyclic matrix. It is probably the only one of the methods mentioned by Löwdin, Pauncz, and de Heer which permits of a simple generalization to higher dimensions. Furthermore, the results are in such a form that approximate results are easily obtained when the conditions are such that these exist. We will also discuss the evaluation of functions of asymmetric cyclic matrices, since these present no more difficulty in theory than do symmetric cyclic matrices.

A cyclic matrix of order n is one in which there are n independent elements. They appear in the first row and the elements in successive rows are the successive cyclic permutations of these.

We begin by considering the cyclic, not necessarily symmetric, matrix $\mathbf{\Delta}$ which we specify in the form

$$\mathbf{\Delta}=(1,S_1,S_2,\cdots,S_N)_{\text{cyc}}, \tag{1}$$

with eigenvalues given by

$$\lambda_k=\sum_{j=0}^{N} S_j \exp\left(\frac{2\pi ikj}{N+1}\right), \quad S_0=1. \tag{2}$$

The problem with which we shall be concerned, is to find an expression for the elements of functions $F(\mathbf{\Delta})$. We restrict ourselves to functions $F(x)$ which are analytic in some neighborhood of the origin and which are defined on the spectrum of the matrix $\mathbf{\Delta}$. We also require that the circle of convergence of $F(x)$ contain all of the eigenvalues of $\mathbf{\Delta}$. For these functions it is easily shown that

$$[F(\mathbf{\Delta})]_{mn}=\frac{1}{N+1}\sum_{k=0}^{N} F(\lambda_k)\exp\left[\frac{2\pi i(m-n)k}{N+1}\right] \tag{3}$$

since the result is true for positive powers $F(x)=x^k$. As it stands, this formula, while correct, is not useful for computation unless N is small, or unless $F(\lambda_k)$ has special features which enable the sum to be evaluated in closed form. We shall convert this sum into a form which is more convenient for accurate calculation or for approximate evaluation.

Let us define a function $\lambda(\theta)$ as

$$\lambda(\theta)=\sum_{j=0}^{N} S_j \exp(ij\theta), \tag{4}$$

from which it is obvious that

$$\lambda_k=\lambda\left(\frac{2\pi k}{N+1}\right). \tag{5}$$

We shall assume that a Fourier series expansion for $F\{\lambda(\theta)\}$ can be written

$$F\{\lambda(\theta)\}=\sum_{j=0}^{\infty} A_j \exp(ij\theta), \tag{6}$$

where

$$A_j=\frac{1}{2\pi}\int_0^{2\pi} F\{\lambda(\theta)\}\exp(-ij\theta)d\theta. \tag{7}$$

Only positive j are required in Eq. (6) since $\lambda(\theta)$ contains only positive powers of $\exp(i\theta)$ and $F(x)$ contains positive powers of x. We will only be interested in $F(\lambda_k)$ where λ_k is given in Eq. (2). This, however, can be written as a finite sum in terms of quantities which we shall denote by A^*;

$$F(\lambda_k)=\sum_{j=0}^{N} A_j^* \exp\left(\frac{2\pi ijk}{N+1}\right), \tag{8}$$

where the A^* are expressible in terms of the A's:

$$A_j^*=\sum_{r=0}^{\infty} A_{j+r(N+1)}. \tag{9}$$

The rearrangement of the series is legitimate since the analyticity of $F(x)$ guarantees uniform convergence within the radius of convergence. Rather than work with the general matrix element $[F(\mathbf{\Delta})]_{mn}$, we will calculate only the $[F(\mathbf{\Delta})]_{0n}$ which we denote by U_n. All of the remaining matrix elements can be calculated by using the cyclic properties. Substituting the expression for $F(\lambda_k)$ given in Eq. (8) into Eq. (3), we see that U_n can be written

$$U_n=\frac{1}{N+1}\sum_{j=0}^{N}\sum_{k=0}^{N} A_j^* \exp\left[\frac{2\pi i(j-n)k}{N+1}\right]$$

$$=A_n^*, \quad n=0,1,2,\cdots,N. \tag{10}$$

Thus we see that the evaluation of the elements of $F(\mathbf{\Delta})$ is reduced to the evaluation of the Fourier coefficients and the summation of Eq. (9). Very often the approximation

$$U_n\sim A_n \tag{11}$$

will be sufficient for computational purposes since when $F(x)$ is analytic, the Fourier coefficients will fall off exponentially with the index. This will be a particularly effective approximation when N is large.

Similar procedures to those given above suffice to discuss the important special case when $\mathbf{\Delta}$ is a symmetric matrix. In this case we can write

$$\mathbf{\Delta}=(1,S_1,S_2,\cdots,S_M,\cdots,S_2,S_1)_{\text{cyc}}, \tag{12}$$

where $M+1$ is the number of generating elements in $\mathbf{\Delta}$. If the order of the matrix $N+1$, is even, then the element S_M occurs only once, while if $N+1$ is odd S_M occurs twice. The eigenvalues are again given by Eq. (5) but the exponentials can now be combined,

resulting in the expression

$$\lambda_k=1+2\sum_{j=1}^{M-1} S_j\cos\left(\frac{2\pi kj}{N+1}\right)+2\epsilon_M S_M\cos\left(\frac{2\pi Mk}{N+1}\right), \quad (13)$$

where $\epsilon_M=\frac{1}{2}$ if N is odd and $\epsilon_M=1$ if N is even. In the present case we replace the exponentials which appear in the definition of $\lambda(\theta)$ in Eq. (4) by cosines. The coefficients which appear in the expansion

$$F\{\lambda(\theta)\}=\tfrac{1}{2}A_0+\sum_{n=1}^{\infty} A_n\cos n\theta \quad (14)$$

are now

$$A_n=(1/\pi)\int_0^{2\pi} F\{\lambda(\theta)\}\cos n\theta d\theta. \quad (15)$$

As for the case of asymmetric cyclic matrices we can write

$$F(\lambda_k)=\sum_{j=0}^{N} A_j^*\cos\left(\frac{2\pi jk}{N+1}\right), \quad (16)$$

where

$$A_0^*=\tfrac{1}{2}A_0+\sum_{j=1}^{\infty} A_{j(N+1)} \quad (17)$$

and the remaining A^* are defined as in Eq. (9). Following the steps leading to Eq. (10) and replacing the exponential in Eq. (3) by a cosine, we find, in the symmetric case,

$$U_n=\tfrac{1}{2}(A_n^*+A_{N+1-n}^*). \quad (18)$$

These results are essentially the generalization of Sec. IV of reference 1.

There are a number of exact results which can be derived through the use of the formulas given so far. Let us first consider the inverse matrix, or the matrix of Green's functions of the linear difference equation characterized by Δ. We first discuss the general, not necessarily symmetric, matrix. Let us assume for Δ the expression shown in Eq. (1). Then the A_j's are given by

$$A_j=\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{-ij\theta}d\theta}{1+S_1e^{i\theta}+S_2e^{2i\theta}+\cdots+S_Ne^{iN\theta}}$$

$$=\frac{1}{2\pi i}\int_c\frac{dz}{(1+S_1z+S_2z^2+\cdots+S_Nz^N)z^{j+1}}, \quad (19)$$

where C is the unit circle. Thus we see that A_j is the coefficient of z^j in

$$[P(z)]^{-1}=[1+S_1z+S_2z^2+\cdots+S_Nz^N]^{-1}, \quad (20)$$

provided that $P(z)$ has no roots in or on the unit circle. If we now assume that

$$S_1>S_2>\cdots>S_N>0, \quad (21)$$

then it can be shown[3] that $P(z)$ has no roots in the unit circle. Other simple sufficient conditions for $P(z)$ to have all of its roots outside of the unit circle can be obtained by using Rouché's theorem.[4] For example if

$$|S_1|+|S_2|+\cdots+|S_k|<1, \quad (22)$$

then the roots of $P(z)$ will lie outside of the unit circle provided

$$(1+S_1+S_2+\cdots+S_k)>(S_{k+1}+S_{k+2}\cdots+S_N). \quad (23)$$

In particular, if

$$\sum_{j=1}^{N}|S_j|<1 \quad (24)$$

then the roots of $P(z)$ are greater than one in absolute value.

Let us assume that $P(z)$ has no multiple roots (although the extension to the more general case is trivial). Then we can write

$$P(z)=\prod_{i=1}^{N}(1+\sigma_iz), \quad (25)$$

where $|\sigma_i|<1$. In terms of the σ_i, the value of A_n is calculated to be

$$A_n=(-1)^n\sum_{k=1}^{N}\frac{\sigma_k{}^{n+1}}{P'(-1/\sigma_k)}. \quad (26)$$

Since the only dependence on n is through the term $\sigma_k{}^{n+1}$ we may find an explicit expression for the $A_n{}^*$:

$$A_n{}^*=(-1)^n\sum_{k=1}^{N}\frac{1}{P'(-1/\sigma_k)}\frac{\sigma_k{}^{n+1}}{1-(-\sigma_k)^{N+1}}. \quad (27)$$

An expression similar to that of Eq. (27) can be obtained for the $A_n{}^*$ in the symmetric case. Let us start from the expression for A_n:

$$A_n=\frac{1}{\pi}\int_0^{2\pi}\frac{\cos n\theta d\theta}{\prod_{i=1}^{N}(1+\sigma_i\cos\theta)}=\frac{1}{\pi}\int_0^{2\pi}\frac{\cos n\theta d\theta}{P(\cos\theta)}, \quad (28)$$

where it is assumed that the S_i are such that the σ_i are all distinct and satisfy $|\sigma_i|<1$ for all i. This will certainly be the case if

$$\sum_{i=1}^{N}|S_i|<1. \quad (29)$$

Equation (28) can be simplified by means of partial

[3] E. Landau, *Darstellung und Begündung einiger neuerer Ergebnisse der Funktionentheorie* (Chelsea Publishing Company, New York, 1946).

[4] Rouché's theorem states that if $f(z)$ and $g(z)$ are analytic within and on a closed curve C, and if $|f(z)|>|g(z)|$ on C, then $f(z)$ and $f(z)+g(z)$ have the same number of zeros in the region bounded by C.

fractions and the resulting integrals evaluated

$$A_n=\frac{1}{\pi}\int_0^{2\pi}\cos n\theta\sum_{i=1}^{N}\frac{\sigma_i}{P'(-1/\sigma_i)}\frac{d\theta}{1+\sigma_i\cos\theta}$$
$$=2\sum_{i=1}^{N}\frac{\sigma_i}{P'(-1/\sigma_i)}\frac{1}{(1-\sigma_i^2)^{\frac{1}{2}}}\left(\frac{(1-\sigma_i^2)^{\frac{1}{2}}-1}{\sigma_i}\right)^n. \quad (30)$$

The series for A_n* can be summed explicitly, leading to a final result:

$$U_n=\sum_{i=1}^{N}\frac{\sigma_i}{(1-\sigma_i^2)^{\frac{1}{2}}}\frac{1}{P'(-1/\sigma_i)}$$
$$\times\frac{1}{1-\{[(1-\sigma_i^2)^{\frac{1}{2}}-1]/\sigma_i\}^{N+1}}\left[\left(\frac{(1-\sigma_i^2)^{\frac{1}{2}}-1}{\sigma_i}\right)^n\right.$$
$$\left.+\left(\frac{(1-\sigma_i^2)^{\frac{1}{2}}-1}{\sigma_i}\right)^{N+1-n}\right],\quad n=0,1,2,\cdots. \quad (31)$$

A number of exact results can be obtained for the important special case

$$\Delta=(1,S,0,0,\cdots 0,S). \quad (32)$$

For example the elements of Δ^ν can be found, where $\nu\neq 0, 1, 2, \cdots$. Assuming that $0\leq|2S|\leq 1$, and letting

$$\beta=\frac{-1+(1-4S^2)^{\frac{1}{2}}}{2S}, \quad (33)$$

where $-1\leq\beta\leq 0$, we find

$$A_j=\frac{1}{\pi}\int_0^{2\pi}(1+2S\cos\theta)^\nu\cos j\theta d\theta$$
$$=\frac{1}{\pi}\frac{1}{(1+\beta^2)^\nu}\int_0^{2\pi}(1-2\beta\cos\theta+\beta^2)^\nu\cos j\theta d\theta$$
$$=\frac{2\beta^j\Gamma(\nu+j)}{(1+\beta^2)^\nu\Gamma(\nu)\Gamma(j+1)}F(\nu,\nu+j,j+1;\beta^2), \quad (34)$$

where the last result is taken from reference 5, p. 383. Hence

$$U_n=\frac{1}{(1+\beta^2)^\nu\Gamma(\nu)}\sum_{j=0}^{\infty}\left\{\frac{\Gamma[\nu+n+j(N+1)]}{\Gamma[n+j(N+1)+1]}F[\nu,\nu+n+j(N+1),n+j(N+1);\beta^2]\right.$$
$$\left.+\frac{\Gamma[\nu+(j+1)(N+1)-n]}{\Gamma[(j+1)(N+1)-n+1]}F[\nu,\nu+(j+1)(N+1)-n,(j+1)(N+1)-n;\beta^2]\right\}. \quad (35)$$

In the particular case $\nu=\frac{1}{2}$, the result can be obtained in terms of Legendre functions[5] and we find

$$U_0=\frac{1}{\sqrt{\beta}}\left\{P_{\frac{1}{2}}(\beta)+2\Gamma(\tfrac{3}{2})\sum_{j=1}^{\infty}\frac{P_{\frac{1}{2}}^{j(N+1)}(\beta)}{\Gamma[j(N+1)+\frac{3}{2}]}\right\}$$
$$U_n=\frac{\Gamma(\frac{3}{2})}{\sqrt{\beta}}\sum_{j=0}^{\infty}\left\{\frac{P_{\frac{1}{2}}^{n+j(N+1)}(\beta)}{\Gamma[n+j(N+1)+\frac{3}{2}]}\right.$$
$$\left.+\frac{P_{\frac{1}{2}}^{(j+1)(N+1)-n}(\beta)}{\Gamma[(j+1)(N+1)-n+\frac{3}{2}]}\right\} \quad (36)$$

somewhat analogous to the result obtained by Löwdin, Pauncz, and de Heer. All of these series are rapidly convergent since the successive terms decrease exponentially.

Another result that can be obtained fairly simply for Δ given by Eq. (32) is $\exp(\Delta)$. The integrals for the A's are easily found in terms of Bessel's functions of imaginary argument and the final results are

$$U_n=e\sum_{j=0}^{\infty}[I_{n+j(N+1)}(2S)+I_{(j+1)(N+1)-n}(2S)],$$
$$n\geq 0. \quad (37)$$

[5] I. M. Ryshik and I. S. Gradstein, *Tables of Series, Products, and Integrals* (Deutscher Verlag der Wissenschaften, Berlin, Germany, 1957).

Similar results can be obtained for the sine and cosine functions.

Approximate formulas for functions of cyclic matrices can be obtained by expanding $F(\lambda_k)$ or $F\{\lambda(\theta)\}$ in powers of the S's, and then retaining terms only up to the desired order. The complete formal expansion of U_n is

$$U_n=\sum_{\nu=0}^{\infty}F^{(\nu)}(0)$$
$$\times\sum_{\substack{j_1+j_2+\cdots+j_N=\nu\\ j_1+2j_2+\cdots+Nj_N=n+j(N+1)\\ j=0,1,2,\cdots}}\cdots\sum_{j_N}\frac{S_1^{j_1}S_2^{j_2}\cdots S_N^{j_N}}{j_1!j_2!\cdots j_N!}. \quad (38)$$

The lowest order terms in this expansion are explicitly

$$U_0\sim F(0)+(S_1S_N+S_2S_{N-1}+\cdots)F^{(2)}(0)+\cdots$$
$$U_1\sim S_1F^{(1)}(0)+\cdots$$
$$U_2\sim\tfrac{1}{2}S_1^2F^{(2)}(0)+S_2F^{(1)}(0)+\cdots \quad (39)$$
$$U_3\sim S_3F^{(1)}(0)+\tfrac{1}{6}S_1^3F^{(3)}(0)+\tfrac{1}{2}S_1S_1^2F^{(2)}(0)+\cdots.$$
$$\vdots\qquad\qquad\vdots$$

Thus far we have been dealing with what we might call first-order circulant matrices. These arise in a natural way from one-dimensional problems in which

periodic boundary conditions are used. When periodic boundary conditions are used in higher dimensions, they lead to the notion of generalized circulant matrices, which were also discussed by Lewis and Keller.[2] The first-order circulant matrix can be defined as having elements Δ_{ij} such that

$$\Delta_{ij}=\begin{cases}S_{i-j} & i>j\\ S_{N+1-(j-i)} & j>i\end{cases} \tag{40}$$

or if we define the S_k over all of the integers by the prescription $S_k=S_{k+j(N+1)}$, $j=\cdots,-2,-1,0,1,2,\cdots$,

$$\Delta_{ij}=S_{i-j}. \tag{41}$$

In the following we shall assume that there is an $N\times N$ matric function $\mathbf{S}(\mathbf{k})$ defined over all d-dimensional vectors $\mathbf{k}=(k_1,k_2,k_3,\cdots,k_d)$, such that

$$\mathbf{S}[\mathbf{k}+\mathbf{n}(N+1)]=\mathbf{S}(\mathbf{k}), \tag{42}$$

where $\mathbf{n}$ is any d-dimensional vector whose components are integers. The generalized circulant matrix will now be assumed to be made up of $N\times N$ cyclic submatrices as blocks (rather than just numbers as is the case for ordinary circulants). The submatrix elements of the generalized circulant matrix will now be defined by $\mathbf{\Delta}(\mathbf{I},\mathbf{J})$ where $\mathbf{I}$ and $\mathbf{J}$ are d-dimensional vectors. In analogy with Eq. (41) we define a generalized circulant matrix by

$$\mathbf{\Delta}(\mathbf{I},\mathbf{J})=\mathbf{S}(\mathbf{I}-\mathbf{J}). \tag{43}$$

For purposes of explicitly representing the matrix $\mathbf{\Delta}(\mathbf{I},\mathbf{J})$ we assign an ordering to the d-dimensional vectors $\mathbf{k}=(k_1,k_2,\cdots k_d)$ such that $\mathbf{k}^{(1)}>\mathbf{k}^{(2)}$ if and only if

$$\begin{gathered}k_1^{(1)}=k_1^{(2)},\quad k_2^{(1)}=k_2^{(2)},\quad\cdots\quad k_r^{(1)}=k_r^{(2)},\\ k_{r+1}^{(1)}>k_{r+1}^{(2)}\quad r=1,2,\cdots,d-1,\end{gathered} \tag{44}$$

where $k_j^{(i)}$ is the j component of $\mathbf{k}^{(i)}$. With this definition the component blocks can be arranged unambiguously. For example, in the theory of lattice dynamics the following equations represent the steady state amplitudes of a two dimensional simple cubic lattice:

$$M\omega^2u_{mn}=\gamma_1\Delta_m{}^2u_{mn}+\gamma_2\Delta_n{}^2u_{mn} \tag{45}$$

where, for example,

$$\Delta_m{}^2u_{mn}=u_{m+1n}-2u_{mn}+u_{m-1n}.$$

Under the assumption of cyclic boundary conditions,

$$u_{m+kN,n}=u_{m,n+jN}=u_{m,n};\ j,k=\cdots,-1,0,1,\cdots \tag{46}$$

we see that the $u_{m,n}$ are the eigenvectors and $M\omega^2$ are the eigenvalues, of a generalized circulant matrix $\mathbf{\Delta}$,

$$\mathbf{\Delta}=\begin{pmatrix}\mathbf{S}_0 & \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{S}_N\\ \mathbf{S}_N & \mathbf{S}_0 & \mathbf{S}_1 & \cdots & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \vdots & & \vdots & \vdots\end{pmatrix} \tag{47}$$

where

$$\begin{aligned}\mathbf{S}_0&=(-2\gamma_1-2\gamma_2,\gamma_1,0,\cdots,0,\gamma_1)_{\text{cyc}},\\ \mathbf{S}_1&=(\gamma_2,0,0,\cdots,0)_{\text{cyc}},\\ \mathbf{S}_N&=(\gamma_2,0,0,\cdots,0)_{\text{cyc}}.\end{aligned} \tag{48}$$

The α matrices which appear in the Dirac equations can also be regarded as generalized circulant matrices.

The elements of the r-dimensional generalized circulant matrix can be represented in the form of a multiple trigonometric expansion:

$$\mathbf{\Delta}(\mathbf{I},\mathbf{J})=(N+1)^{-r}\sum_{\mathbf{K}}\mathbf{\Lambda}(\mathbf{K})\times\exp\left[-\frac{2\pi i}{(N+1)^r}(\mathbf{I}-\mathbf{J})\cdot\mathbf{K}\right], \tag{49}$$

where the matrices $\mathbf{\Lambda}(\mathbf{K})$ are the generalized eigenvalues of $\mathbf{\Delta}$ and can be written

$$\mathbf{\Lambda}(\mathbf{K})=\sum_{\mathbf{J}}\mathbf{S}(\mathbf{J})\exp\left(\frac{2\pi i}{(N+1)^r}\mathbf{J}\cdot\mathbf{K}\right), \tag{50}$$

in which $\mathbf{S}(\mathbf{J})$ is the matrix of Eq. (43). The sums run over all r-dimensional vectors with positive integral or zero elements, in which the elements range from 0 to N. It is also readily verified that for any analytic function $F(x)$, one can write for the elements of $F(\mathbf{\Delta})$

$$[F(\mathbf{\Delta})]_{\mathbf{IJ}}=(N+1)^{-r}\sum_{\mathbf{K}}F\{\mathbf{\Lambda}(\mathbf{K})\}\times\exp\left[-\frac{2\pi i}{(N+1)^r}(\mathbf{I}-\mathbf{J})\cdot\mathbf{K}\right], \tag{51}$$

provided that the $\mathbf{\Lambda}(\mathbf{K})$ are such that the Taylor series for $F\{\mathbf{\Lambda}(\mathbf{K})\}$ converges, or equivalently that the eigenvalues of $\mathbf{\Delta}$ lie in the circle of convergence of $F(x)$. Now the theory for the representation of $F(\mathbf{\Delta})$ follows exactly as in the one dimensional case. Let us define $\mathbf{\Lambda}(\boldsymbol{\theta})$ to be the generalization of $\mathbf{\Lambda}(\mathbf{K})$:

$$\mathbf{\Lambda}(\boldsymbol{\theta})=\sum_{\mathbf{J}}\mathbf{S}(\mathbf{J})\exp\left(\frac{2\pi i}{(N+1)^r}\mathbf{J}\cdot\boldsymbol{\theta}\right). \tag{52}$$

Assuming now the Fourier expansion

$$F\{\mathbf{\Lambda}(\boldsymbol{\theta})\}=\sum_{J_1=0}^{\infty}\cdots\sum_{J_r=0}^{\infty}\mathbf{A}_{\mathbf{J}}\exp(i\mathbf{J}\cdot\boldsymbol{\theta}), \tag{53}$$

with

$$\mathbf{A}_{\mathbf{J}}=(2\pi)^{-r}\int_0^{2\pi}\cdots\int F\{\mathbf{\Lambda}(\boldsymbol{\theta})\}\exp(-i\mathbf{J}\cdot\boldsymbol{\theta})d^r\boldsymbol{\theta}, \tag{54}$$

we see that

$$F\{\mathbf{\Lambda}(\mathbf{K})\}=\sum_{\mathbf{J}}\mathbf{A}_{\mathbf{J}}{}^*\exp\left(\frac{2\pi i}{(N+1)^r}\mathbf{J}\cdot\mathbf{K}\right), \tag{55}$$

where

$$\mathbf{A}_{\mathbf{J}}{}^*=\sum_{L_1=0}^{\infty}\cdots\sum_{L_r=0}^{\infty}A_{\mathbf{J}+\mathbf{L}(N+1)}. \tag{56}$$

The remaining analysis follows exactly as in the one-dimensional case, that is to say, the submatrices of $\mathbf{\Delta}$ are just the $\mathbf{A}_J^*$. In similar fashion, when $\mathbf{\Delta}$ is symmetric, the submatrix blocks of $\mathbf{F}(\mathbf{\Delta})$ are

$$\mathbf{U}_n=\tfrac{1}{2}(\mathbf{A}_J^*+A^*_{(N+1)\mathbf{i}-n}) \tag{57}$$

where $\mathbf{i}=(1,1,\cdots 1)$.

The results of the preceding analysis are easily generalized to the calculation of analytic functions of matrices which have the form

$$\mathbf{\Delta}^+=\begin{bmatrix} 1 & S_1 & S_2 & S_3 & \cdots & S_N \\ -S_N & 1 & S_1 & S_2 & \cdots & S_{N-1} \\ -S_{N-1} & -S_N & 1 & S_1 & \cdots & S_{N-2} \\ \vdots & \vdots & \vdots & & & \vdots \\ -S_1 & -S_2 & \cdots & & & -S_N \end{bmatrix}, \tag{58}$$

which might be called skew-circulant matrices. The eigenvalues of such a matrix are found to be

$$\lambda_k=\sum_{j=0}^{N} S_j \exp\left\{\frac{(2k+1)\pi ij}{N+1}\right\}, \quad S_0=1 \tag{59}$$

and the unitary matrix $\mathbf{U}$ which diagonilizes $\mathbf{\Delta}^+$ has elements

$$U_{kj}=(N+1)^{-\frac{1}{2}} \exp\left\{\frac{(2j+1)\pi ik}{N+1}\right\}. \tag{60}$$

If $F(x)$ is analytic in some neighborhood of the origin, then the elements of $F(\mathbf{\Delta}^+)$ are given by

$$[F(\mathbf{\Delta}^+)]_{mn}=(N+1)^{-1}\sum_{k=0}^{N} F(\lambda_k) \times\exp\left\{\frac{(2k+1)(m-n)i\pi}{N+1}\right\}. \tag{61}$$

The argument from Eq. (4) to Eq. (9) can be repeated with just a slight modification to yield the result

$$[F(\mathbf{\Delta}^+)]_{0n}=U_n=\sum_{j=0}^{\infty}(-1)^j A_{n+j(N+1)}, \tag{62}$$

where now A_m is defined as the Fourier coefficient

$$A_m=\frac{1}{2\pi}\int_0^{2\pi} F\{\lambda(\theta)\}e^{-im\theta}d\theta \tag{63}$$

and

$$\lambda(\theta)=\sum_{j=0}^{N} S_j \exp(ij\theta) \tag{64}$$

as before. If now $\mathbf{\Delta}^+$ is expressed as

$$\mathbf{\Delta}^+=\mathbf{I}+\mathbf{S} \tag{65}$$

where $\mathbf{S}$ is a skew symmetric circulant (notice that $\mathbf{\Delta}^+$ itself is not) then

$$U_n=[F(\mathbf{\Delta})]_{0n}=\tfrac{1}{2}(A_n{}^*+A_{N+1-n}{}^*), \tag{66}$$

where

$$\begin{aligned} A_0{}^*&=\tfrac{1}{2}A_0+\sum_{j=1}^{\infty}(-1)^j A_{j(N+1)} \\ A_n{}^*&=\sum_{j=0}^{\infty}(-1)^j A_{n+j(N+1)} \end{aligned} \tag{67}$$

and A_m is now to be defined as

$$A_m=\frac{1}{\pi}\int_0^{2\pi} F\{\lambda(\theta)\}\cos m\theta d\theta. \tag{68}$$

The proofs of these results parallel those for functions of ordinary circulant matrices.

Unusual Eigenvalues of the Energy Operator

ADEL DA SILVEIRA
Colégio Pedro II and Núcleo de Estudos e Pesquisas Científicas do Rio de Janeiro, Estado de Guanabara, Brazil
(Received June 23, 1961)

A generalization of quantum mechanics different from the theory proposed recently by Phipps is considered. Difficulties connected with the commutation relations of the angular momentum components defined with the help of the operators p_i of the fundamental equations of Phipps are presented as new arguments in favor of the choice of Phipps of the Hermitian operators to be used in his generalized theory. The application of the equations proposed in the present paper to the motion of a particle in a central Coulombic field of forces is considered. In the Dirac theory the spectrum is different from the traditional one. As the Heisenberg postulate is now obeyed, this result shows that an unusual spectrum is not a consequence of the violation of the postulate.

I. INTRODUCTION

A GENERALIZATION of quantum mechanics which contains the Hamilton-Jacobi theory as a particular case was analyzed recently by Phipps.[1] The application of the new equations to the problem of the stationary motion of a Dirac particle bound by Coulombic central forces to a massive center, results in a highly unusual energy spectrum that was attributed by Phipps to the local violation of the Heisenberg postulate. Actually, as we shall show, we can obtain an unusual energy spectrum without violating this postulate, if we conveniently modify the initial equations.

The classical equations of the Hamilton-Jacobi theory and the Schrödinger equation are particular cases of the generalized equations of Phipps. However, the quantities which are the energy and the linear momentum in the classical equations and in the ordinary quantum equations do not have the properties of these quantities in the generalized theory. Besides, as we shall show (in Sec. II), if we define the angular momentum operator by means of the classical expression with the **p** interpreted as the operator of the Phipps equations, we obtain a quantity with commutation properties different from the usual ones. The definition of the spin operator is in this case impossible. A further assumption was made by Phipps in order to overcome these difficulties. It was supposed that in the general theory, the energy and the linear momentum operators are quantities connected with the operators H and p defined by the fundamental equations (1). If this is done, in fact, the Heisenberg postulate is not violated for the new operators; the commutation relations of the angular momentum operator are the usual ones and the energy is the conjugate of time. However, the operators p_k of the initial equations are not Hermitian.

In the present paper another way of solving these problems is considered. A set of fundamental equations is postulated so that the classical and quantum equations are obtained as particular cases of the proposed equations (in Sec. III). The Heisenberg postulate is not violated at all, and the angular momentum and the energy operators are obtained from the classical expressions with the help of the operators p_k of the fundamental equations. The introduction of a further assumption to define the energy and the linear moment is not necessary. With respect to the reduction of the equations to the classical ones we must observe that the P_k given by the relations (13) and (20) are not constants (in apparent contradiction with the Hamilton-Jacobi theory). If we compare, however, the magnitudes of the two terms of P given by (20), we see that they are comparable for distances r of the order of 10^{-8} cm if the particle in motion is an electron, or 10^{-11} cm if it is a proton. For distances greater than these values (i.e., 10^{-2} cm), P is reduced to a constant term. The dependence of P on r is, therefore, associated with small distances, far beyond the limits of validity of classical physics. Therefore, to get classical results as the constant values of P_k, we must consider classical distances between interacting centers. This is indicated in the condition (i) in Sec. III.

The application of the new equations to the stationary case analyzed by Phipps gives rise to an unusual energy spectrum which is different from the spectrum encountered by Phipps; the new spectrum being Bohr-like in first approximation (in Sec. IV). The values of the energy lie in the region between $-m_0c^2$ and $+m_0c^2$ and contain only one quantum number. The spectrum obtained in the present paper depends on the charge Ze and on the Planck constant. It is worth mentioning that the spectrum of Phipps depends only on the rest energy of the particle in motion and, therefore, cannot be reduced to the spectrum of the Bohr theory.

II. EQUATIONS OF PHIPPS

The equations proposed by Phipps are

$$H(x_k,p_k,t)\Psi_f=-\frac{\partial S\Psi_f}{\partial t}, \tag{1a}$$

$$p_k\Psi_f=\frac{\partial S\Psi_f}{\partial x_k}, \tag{1b}$$

$$P_k\Psi_f=-\frac{\partial S\Psi_f}{\partial X_k}, \tag{1c}$$

[1] T. E. Phipps, Jr., Phys. Rev. 118, 1653 (1960).

where $k=1, 2, \cdots, 3n$. S may depend on x_k, X_k, and t.

These equations are reduced to the equations of the Hamilton-Jacobi theory if Ψ_f is taken to be a nonzero c number different from zero. Equations (1a) and (1b) are the ones of the Schrödinger theory if we take $S=\hbar/i$. If, however, neither S nor Ψ_f are constant, we obtain a set of equations which constitute the basis of a new theory. It is easy to show that, in this case, we obtain

$$[p_k,x_m]\Psi_f=S\delta_{km}\Psi_f. \tag{2}$$

Hence, if $\mathbf{p}$ is the linear momentum operator, the Heisenberg postulate is violated because of the factor S in the right-hand side of the relation (2). It can be shown that different commutators are obtained for other operators. If, for instance, we define the angular momentum operator by the relation

$$L_k\Psi_f=\epsilon_{klm}x_lp_m\Psi_f, \tag{3}$$

where ϵ_{klm} is the alternating symbol of tensor analysis, we have, in view of (1b),

$$L_k\Psi_f=\epsilon_{klm}x_l\frac{\partial S\Psi_f}{\partial x_m}. \tag{4}$$

The commutation relations of the operators L_k are, because of (4):

$$[L_k,L_r]\Psi_f=\epsilon_{klm}\epsilon_{rpq}\Big[x_lS\delta_{pm}\frac{\partial S\Psi_f}{\partial x_q}-x_pS\delta_{lq}\frac{\partial S\Psi_f}{\partial x_m}+x_lx_p\Big(\frac{\partial S}{\partial x_m}\frac{\partial S\Psi_f}{\partial x_q}-\frac{\partial S}{\partial x_q}\frac{\partial S\Psi_f}{\partial x_m}\Big)\Big],$$

where δ_{ik} is the Kronecker delta.

In view of the relation

$$\epsilon_{klr}\epsilon_{jrs}-\epsilon_{krs}\epsilon_{jlr}=-\epsilon_{kjr}\epsilon_{rls}$$

with sum in r, and of the definition (4), we obtain

$$[L_k,L_r]\Psi_f=-S\epsilon_{krp}L_p\Psi_f+x_l(\partial S/\partial x_m)\times(\epsilon_{klm}L_r-\epsilon_{rlm}L_k)\Psi_f. \tag{5}$$

If we choose the value $\hbar/i$ for S, the relation (5) will turn into the usual commutation relations of the components of the angular momentum operator. It is easy to show that L_k is not the only operator to satisfy (5). We can show that N_k, M_k, and U_k defined by

$$N_k\Psi_f=S\epsilon_{klm}x_l\partial\Psi_f/\partial x_m,$$
$$M_k\Psi_f=(iS\sigma_k/2+N_k)\Psi_f,$$
$$U_k\Psi_f=(iS\sigma_k/2+L_k)\Psi_f,$$

satisfy (5). In these definitions σ_k are the Pauli matrices. We can show, however, that (5) is not satisfied by the operator R_k defined by the relation

$$R_k\Psi_f=[\hbar\sigma_k/2+\epsilon_{klm}x_l(\partial/\partial x^m)S]\Psi_f.$$

The operators M_k and U_k contain a term with the Pauli matrices so that we are tempted to interpret $iS\sigma/2$ as the spin operator of the theory and U_k as the total angular momentum. However, if (5) is admitted for the angular momentum components, this interpretation is not a correct one. Besides, $iS\sigma/2$ would be a spin operator dependent on the position x^l. These are, of course, new arguments against the interpretation of p_i as the linear momentum operator.

In order to solve these difficulties, Phipps has introduced the further assumption that H and p are not the energy and the linear momentum operators of the generalized theory. These operators are, according to Phipps, the quantities defined by

$$\mathcal{H}\equiv Hs^{-1}$$

and

$$\pi\equiv ps^{-1},$$

where s is a real quantity connected with S by

$$S\equiv(\hbar/i)s.$$

If we introduce the wave function

$$\Psi\equiv s\Psi_f$$

we obtain, from (1b), the equation

$$\mathcal{H}\Psi=-(\hbar/i)\partial\Psi/\partial t.$$

Therefore, $\mathcal{H}$ is the conjugate of time. We can obtain from the definitions introduced above the usual commutation relations of the linear momentum. The angular momentum operator must be defined with the help of the quantity π.

In Sec. III we shall examine another way of solving the problems considered above by means of a new set of fundamental equations.

III. GENERAL EQUATIONS

Among the forms of classical mechanics that could form the basis of the quantum theory, the Hamilton-Jacobi and Poisson bracket ones have been used extensively. We can ask, however, whether some generalized form of the Hamilton-Jacobi theory can be used as, for instance, the one that takes as the Hamilton generating function not the usual S but some function $G(S)$ of it. If this is admitted, we can write down the classical equations of motion in the form

$$H(x_k,p_k,t)=-\partial G(S)/\partial t, \tag{6a}$$
$$p_k=\partial G(S)/\partial x_k, \tag{6b}$$
$$P_k=-\partial G(S)/\partial X_k, \tag{6c}$$

$(k=1,2,3,\cdots,3n)$.

It is easy to show that the equations of Hamilton

$$\dot{p}_k=-\partial H/\partial x_k,$$
$$\dot{x}_k=\partial H/\partial p_k,$$

are not altered. According to (6b) and (6c), P_k and X_k are the canonical transforms of p_k and x_k by means of

$$p_k\delta x_k - P_k\delta X_k = \delta G(S).$$

Now, if classical equations like these are to be used as a guide to a quantum generalization, the form of $G(S)$ as well as the ordering of factors which appear in it and in its derivatives must be chosen before the transition to the quantum equations. Although this order is unimportant in the classical theory, it is fundamental in quantum theory because the quantum operators in general do not commute. Once these choices are made, the quantum equations are obtained when a wave function operand is adjoined to the right in Eqs. (6). In this way it would be possible to obtain an infinite number of equations, one for each $G(S)$ and a chosen order of the operators. In order to reduce the infinite number of possibilities one should impose some reasonable conditions to the equations obtained in this way, as follows:

(i) When the wave function is taken to be an arbitrary nonzero c number, the postulated equations should reduce to some form of the Hamilton-Jacobi equations, for distances between interacting centers for which these equations are known to hold.

(ii) Another type of reduction should bring the equations into predictive agreement with ordinary quantum mechanics.

(iii) The postulated equations, to constitute a nontrivial generalization of quantum mechanics, should possess new solutions not present either in classical or quantum mechanics.

In the work of Phipps, the Eqs. (6) are satisfied as well as the assumed conditions by means of the choice

$$G(S) = S.$$

In the present paper we shall consider another choice

$$G(S) = a \ln S,$$

where a is the imaginary constant $-i\hbar$. The derivatives indicated in (6) are now calculated with this $G(S)$ and are written in the form

$$H(x_k,p_k,t) = -aS^{-1}(\partial/\partial t)S,$$
$$p_k = aS^{-1}(\partial/\partial x_k)S,$$
$$P_k = -aS^{-1}(\partial/\partial X_k)S.$$

The quantum equations are obtained, as explained above, by means of the wave function operand Ψ and are:

$$H(x_k,p_k,t)\Psi = -aS^{-1}(\partial/\partial t)S\Psi, \tag{7a}$$
$$p_k\Psi = aS^{-1}(\partial/\partial x_k)S\Psi, \tag{7b}$$
$$P_k\Psi = -aS^{-1}(\partial/\partial X_k)S\Psi. \tag{7c}$$

The Hamiltonian operator H of these equations is obtained from the classical total mechanical energy with the help of (7b). The operator $\mathbf{p}$ defined in (7b) is obtained from the usual linear momentum by means of a similarity transformation and obeys the usual commutation relations

$$[p_k,x_l]\Psi = -i\hbar\delta_{kl}\Psi.$$

If S is a constant, Eqs. (7a) and (7b) are those of ordinary quantum mechanics. If, besides, the P_k are constants, the relation (7c) is associated with a physically meaningless phase factor of the wave function. To show this, consider Eq. (7c) which is satisfied by

$$\Psi = e^{-P_kX_k}\Phi(x_k,t).$$

Equations (7a) and (7b) reduce, after cancellation of the constant exponential factor, to the equations of the Schrödinger theory

$$H\Phi = -(\hbar/i)\partial\Phi/\partial t$$
$$p_k\Phi = -i\hbar\partial\Phi/\partial x_k.$$

If S is a constant but P_k is not a constant, Eq. (7c) is a condition imposed on the wave function which is the origin of a solution different from the traditional one.

In the relativistic theory, Eq. (7a) must take the form of the Dirac equation[2]

$$(E + c\boldsymbol{\alpha}\cdot\mathbf{p} + \beta m_0c^2)\Psi = 0, \tag{8}$$

where the operators E and $\mathbf{p}$ are defined in the manner of the relation (7b). If we substitute for E and $\mathbf{p}$ the corresponding expressions in terms of derivatives and use the hypothesis that S commutes with the Dirac matrices, we obtain

$$(i\hbar\partial/\partial t - i\hbar c\boldsymbol{\alpha}\cdot\mathbf{grad} + \beta m_0c^2)S\Psi = 0; \tag{9}$$

the new wave function being, therefore, $S\Psi$. The form of the P_k of the relation (7c) is not altered and is, in the special case analyzed in the next section, given by the relations (13) and (20).

IV. SPECTRUM OF THE DIRAC EQUATION

In order to apply the equations to a concrete case, let us consider the motion of a Dirac particle of charge e and rest mass m_0 bound by central Coulombic forces to an infinitely massive center. The Dirac particle is assumed to be in a stationary state and the point center is in the origin of coordinates in the laboratory system. We can use a two-component radial-wave function and can write the two equations for the radial part[3]

$$(S\Psi)_2' + \frac{l}{r}(S\Psi)_2 + \frac{1}{\hbar c}\left(-m_0c^2 + \frac{Ze^2}{r} - E\right)(S\Psi)_1 = 0, \tag{10a}$$

$$-(S\Psi)_1' + \frac{l}{r}(S\Psi)_1 + \frac{1}{\hbar c}\left(m_0c^2 + \frac{Ze^2}{r} - E\right)(S\Psi)_2 = 0. \tag{10b}$$

[2] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1949), Chap. XII, p. 311.

[3] Reference 2, p. 323. Schiff uses F and G instead of $(S\Psi)_1$ and $(S\Psi)_2$.

Now, let us consider the relation (7c). Since we are considering the stationary state, S and Ψ can be assumed to be independent of time. Then we can write

$$-P_k S\Psi=(\hbar/i)\partial S\Psi/\partial X_k \tag{11}$$

for (7c). The use of polar coordinates R, θ, and φ

$$X_1=R\sin\theta\cos\varphi,$$
$$X_2=R\sin\theta\sin\varphi,$$
$$X^3=R\cos\theta,$$

and the hypothesis that S and Ψ do not depend on θ and φ affords the following relation

$$\partial/\partial X_k=\xi_k\partial/\partial R, \tag{12}$$

where ξ_k are direction cosines. We now consider an arbitrarily fixed direction specified by θ and φ so that

$$P_k=P\xi_k, \tag{13}$$

where P can be a function of r. In this case Eq. (11) is reduced with the help of (12) and (13) to the form

$$P(r)S\Psi=-(\hbar/i)\partial S\Psi/\partial R. \tag{14}$$

As in the paper of Phipps, in order to interpret the action of the operator $\partial/\partial R$ on $S\Psi$ we must assume temporarily that $S\Psi$ depends on R. If we admit that the dependence on $\mathbf{r}$ and $\mathbf{R}$ occurs only in the combination $|\mathbf{r}-\mathbf{R}|$, we shall have

$$P(r)S\Psi=(\hbar/i)\partial S\Psi/\partial r, \tag{15}$$

with the initial condition

$$\mathbf{R}=(0,0,0).$$

We shall not analyze the meaning of this condition in the present paper as it was sufficiently discussed in reference 1. The separated equations for the two components of $S\Psi$ are implied by (15) so we have

$$(S\Psi)_2=K(S\Psi)_1,$$

where K is a constant. If we introduce this value of $(S\Psi)_2$ in (10a) and (10b), we obtain the two equations

$$K(S\Psi)_1'+\left[\frac{Kl}{r}+\frac{1}{\hbar c}\left(-m_0c^2+\frac{Ze^2}{r}-E\right)\right](S\Psi)_1=0,$$

$$-(S\Psi)_1'+\left[\frac{l}{r}+\frac{K}{\hbar c}\left(m_0c^2+\frac{Ze^2}{r}-E\right)\right](S\Psi)_1=0,$$

which are compatible if

$$K^2=\frac{m_0c^2+E}{m_0c^2-E}, \tag{16}$$

and

$$2l=-\gamma(K+K^{-1}) \tag{17}$$

where

$$\gamma=Ze^2/\hbar c.$$

In order to obtain the energy eigenvalues E, we substitute in (17) the value of K^2 given by (16). We obtain

$$K=\frac{\gamma}{l}\frac{m_0c^2}{(E-m_0c^2)}. \tag{18}$$

Now, if we introduce this value of K back in (16), we obtain

$$E=\pm m_0c^2(1-\gamma^2/l^2)^{\frac{1}{2}}, \tag{19}$$

which is different from the energy spectrum found by Phipps. These values of E depend on the quantum number l and correspond to the region between $-m_0c^2$ and $+m_0c^2$. If we develop E in series, retaining only terms up to the first power in γ^2/l^2, we obtain

$$E\sim\pm m_0c^2\mp chR/l^2,$$

where R is the Rydberg constant. The second term is the energy which appears in the theory of Bohr. We can obtain the form of $P(r)$ if we substitute in (10b) the derivative of $S\Psi$ given by (15) and use (18) and (19). We obtain

$$P(r)=\frac{\hbar}{i}\left(\frac{s}{r}-\frac{Zm_0e^2}{\hbar^2l}\right), \tag{20}$$

where

$$s=\pm(l^2-\gamma^2)^{\frac{1}{2}}.$$

The two terms of P are comparable for distances of the order of 10^{-8} cm if m_0 is the electronic rest mass while the last is predominant for values of r much greater than 10^{-8} cm. If m_0 is the protonic rest mass, the two terms are comparable for distances, of the order of 10^{-11} cm. The solutions $(S\Psi)_1$ and $(S\Psi)_2$ of the wave equation are obtained with the use of Eq. (15). We obtain

$$K(S\Psi)_1=(S\Psi)_2=Ar^s\exp(-Zm_0e^2r/\hbar^2l)$$

where A is the constant of integration. It is curious that the expression

$$Zm_0e^2/\hbar^2l$$

is the coefficient (α) of the formula (44.16) of reference 2 if (E) is given by relation (19).

ACKNOWLEDGMENTS

The writer's grateful thanks are due to the members of the Núcleo de Estudos e Pesquisas Científicas, especially Dr. Armando Dias Tavares, Dr. Edson Rodrigues, Dr. Aldizio F. Costa, and Dr. Henrique A. G. de Morais for many helpful discussions.

Theory of Transport Coefficients. I. General Theory and Electrical Conductivity of Electron-Phonon System*

S. FUJITA† AND R. ABE‡
Department of Physics, Northwestern University, Evanston, Illinois
(Received May 11, 1961)

A general method is presented for the calculation of static transport coefficients based on Kubo's formulas. Techniques of perturbation expansion, diagram representation, and linked-cluster expansion are used. As an example, the electrical conductivity of an electron-phonon system is calculated with the natural introduction of irreversibility, following the ideas of Van Hove, and Prigogine and his collaborators. Under certain conditions the present method is shown to be equivalent to the conventional method by means of the Boltzmann equation. These conditions are examined and the improvement of the approximation is discussed.

I. INTRODUCTION

In the past few years remarkable progress has been made in several branches of statistical mechanical theory of irreversible processes. Van Hove,[1] Prigogine and his collaborators,[2,3] and Bogoliubov[4] were successful in deriving the master equation from the Liouville equation for a certain type of many-particle system, by using some special property of its asymptotic time behavior. In every case their derivations are based on the assumption that the system under consideration is infinitely large. In mathematical terms this assumption is expressed by the following limit:

$$N\to\infty, \quad V\to\infty, \quad (N/V=\text{finite}), \tag{1.1}$$

where N is the number of particles and V the volume of the container. Physically this ensures the infinite recurrence time of Poincaré's cycle. In addition to this basic assumption on the size of system, some particular limiting processes are introduced, relating the quantities such as coupling constant g, the number density or the concentration c, and the relaxation time of the system t. In the case of weakly coupled systems, e.g., electron-phonon system, this is stated as

$$g\to 0, \quad t\to\infty, \quad (g^2t=\text{finite}). \tag{1.2}$$

For dilute gases consisting of particles interacting with strong short-range forces such as those usually treated by kinetic theory of gases, the behavior of the systems is considered in the limit:

$$c=N/V\to 0, \quad t\to\infty, \quad (ct=\text{finite}). \tag{1.3}$$

These conditions (1.2) and (1.3) are closely related to the existence of well-separated relaxation time and collision time of the systems. In fact they are effectively equivalent to the assumption of instantaneous collisions. The recent theory of irreversible processes due to Prigogine and his collaborators[5] does not depend on these restrictive conditions.

On the other hand, Kubo and others[6] have established the rigorous statistical mechanical expressions for transport coefficients. Solving the equation of motion for the density matrix of the system in the first order of the external perturbing field, they obtained transport coefficients in a form of integrals of time relaxed current correlation functions. For example, the $\mu-\nu$ component of static electrical conductivity tensor ($\omega=0$, ω being the frequency of the external field) is expressed as

$$\sigma_{\mu\nu}=\lim_{T'\to\infty}\beta V^{-1}\,\mathrm{Re}\left\{\int_0^{T'}du\langle J_\mu J_\nu(u)\rangle\right\}, \tag{1.4}$$

where Re means that the real part of the bracketed quantity is to be taken, $\beta=(kT)^{-1}$, k being the Boltzmann constant, T the absolute temperature, and the symbol $\langle\ \rangle$ means that the thermal average is taken over the canonical or grand canonical ensemble. Furthermore, $\mathbf{J}(u)$ is the Heisenberg operator of the total current $\mathbf{J}$ given by

$$\mathbf{J}(u)=e^{iuH}\mathbf{J}e^{-iuH}, \tag{1.5}$$

where H is the total Hamiltonian of the system. The formula (1.4) makes it possible to calculate the transport coefficients without the help of the approximate Boltzmann equation.

Several attempts have been made to explicitly calculate transport coefficients starting from Kubo's formula (1.4). Nakano[7] evaluated the electrical conductivity in this way. However, his argument about the irreversibil-

* Work supported by the Office of Naval Research and by the Air Research and Development Command, U. S. Air Force, through its European office.

† Present Address: Université Libre de Bruxelles, Bruxelles, Belgique.

‡ On leave of absence from Tokyo Institute of Technology, Tokyo, Japan.

[1] L. Van Hove, Physica **21**, 517 (1955); **23**, 441 (1957).

[2] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1961). The book contains major contributions on this subject.

[3] R. Brout and I. Prigogine, Physica **22**, 621 (1956); I. Prigogine and R. Balescu, *ibid.* **25**, 281 (1959); **25**, 302 (1959); **26**, 145 (1960).

[4] N. N. Bogoliubov, J. Phys. (U.S.S.R.) **10**, 265 (1946).

[5] See reference 2.

[6] R. Kubo, J. Phys. Soc. Japan **12**, 570 (1957); H. Mori, Phys. Rev. **112**, 1829 (1958); R. Kubo, M. Yokota, and S. Nakajima, J. Phys. Soc. Japan **12**, 1203 (1957); W. Kohn and J. M. Luttinger, Phys. Rev. **108**, 590 (1957); **109**, 1829 (1958).

[7] H. Nakano, Progr. Theoret. Phys. (Kyoto) **15**, 77 (1956); **17**, 145 (1957).

ity and the approximate procedure of calculation are not very convincing. Montroll and Ward[8] developed a linked-cluster expansion method for these coefficients employing toron diagrams. Although their method is applicable to all kinds of classical and quantum statistical transport coefficients, their prescription for the calculation seems to be incomplete for practical purposes, because of the following reason. If one uses the perturbation method for $\mathbf{J}(u)$, the zeroth order term in g of the integrand in (1.4) becomes independent of u, because the unperturbed Hamiltonian H_0 usually commutes with $\mathbf{J}$. Therefore, the u integral of this term diverges as $T' \to \infty$. The above-mentioned method of Montroll and Ward contains the same kind of divergence. Without resolving such divergence difficulty the calculation of nontrivial transport coefficients is impossible. Chester and Thellung[9] overcame this difficulty by reordering the perturbation series in a certain way and summing up partial series and then performing the u integration. Their techniques rely heavily on the conditions (1.1) and (1.2). However, their theory is restricted to the case in which the Hamiltonian consists of single-particle energies. Furthermore, it is not clear how the thermodynamical averages of the phonon occupation numbers are brought into their Eq. (6.10).

In this paper we shall present a general perturbation method for the calculation of quantum statistical transport coefficients starting from Kubo's formulas. The theory is formulated by taking a simple example of the electrical conductivity of an electron-phonon system.

The perturbation expansion and the diagrammatic representation for the current correlation function are given in Sec. II, using the theorem due to Bloch and de Dominicis.[10] The linked-cluster expansion is established in Sec. III by the use of Feynman diagrams. In Sec. IV various linked diagrams are examined for their dependency on coupling constant, time, and volume. Then a special class of diagrams are picked out under the conditions (1.1) and (1.2), and it is shown that if the electrons obey Boltzmann statistics the u integral of the sum of these selected diagrams leads to an integral equation which is closely related to the linearized form of the usual Boltzmann equation. In the present stage of approximation and in the present example, our method yields the identical conductivity as the one we would obtain from the Boltzmann equation approach. This result is obtained under only assumptions (1.1) and (1.2). The physical meaning of these assumptions is discussed in Sec. V. Throughout in the text the following units are chosen: $2M=\hbar=1$, where M is the mass of a particle.

[8] E. W. Montroll and J. C. Ward, Physica **25**, 423 (1959).
[9] G. R. Chester and A. Thellung, Proc. Phys. Soc. (London) **73**, 745 (1959).
[10] C. Bloch and C. de Dominicis, Nuclear Phys. **7**, 459 (1958).

II. PERTURBATION THEORY (DIAGRAM REPRESENTATION)

In this section we shall develop a perturbation expansion method for the time relaxed-current correlation function and represent each term of the expansion by a set of Feynman diagrams.

Let us consider a system characterized by a Hamiltonian H, which consists of two parts:

$$H=H_0+gH_I, \tag{2.1}$$

where H_0 is the unperturbed Hamiltonian and H_I is the Hamiltonian describing the interactions between the unperturbed fields.

Define an operator $S(t)$ by

$$e^{-itH}=e^{-itH_0}S(t). \tag{2.2}$$

It is formally solved in an infinite series:

$$S(t)=1+\sum_{n=1}^{\infty}(-ig)^n\int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n \times H_I(t_1)H_I(t_2)\cdots H_I(t_n), \tag{2.3}$$

where $H_I(t')$ is the interaction energy operator in the interaction representation defined by

$$H_I(t')=e^{it'H_0}H_Ie^{-it'H_0}. \tag{2.4}$$

Taking the Hermitian conjugates to (2.2) and (2.3), we have

$$e^{itH}=S^\dagger(t)e^{itH_0} \tag{2.2a}$$

and

$$S^\dagger(t)=1+\sum_{n=1}^{\infty}(ig)^n\int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n \times H_I(t_n)\cdots H_I(t_2)H_I(t_1). \tag{2.3a}$$

It is to be noted that products of time dependent operators are ordered in mutually opposite directions with respect to time in the integrands on the right-hand sides of (2.3) and (2.3a).

Substituting (2.2) and (2.2a) into the expression (1.4), we obtain

$$\sigma_{\mu\nu}=\beta V^{-1}\,\mathrm{Re}\left\{\int_0^\infty du X(u)\right\}, \tag{2.5}$$

where $X(u)$ is the current correlation function given by

$$X(u)=\langle J_\mu S^\dagger(u)e^{iuH_0}J_\nu e^{-iuH_0}S(u)\rangle. \tag{2.6}$$

In order to proceed further we shall use the explicit form of the current operator $\mathbf{J}$ as well as the Hamiltonian H. We shall develop the theory by taking a simple example of the electrical conductivity of a weakly coupled electron-phonon system (neglecting Coulomb interaction and spin coordinates). We assume the Hamiltonian H of this system has the form:

$$H=\sum_{\mathbf{p}}p^2a_{\mathbf{p}}^\dagger a_{\mathbf{p}}+\sum_{\mathbf{q}}\omega_q b_{\mathbf{q}}^\dagger b_{\mathbf{q}}+g\sum_{\mathbf{p},\mathbf{q}}V^{-\frac{1}{2}}\gamma_q^{\frac{1}{2}} \times(a_{\mathbf{p}+\mathbf{q}}^\dagger a_{\mathbf{p}}b_{\mathbf{q}}+a_{\mathbf{p}}^\dagger a_{\mathbf{p}+\mathbf{q}}b_{\mathbf{q}}^\dagger). \tag{2.7}$$

Here $\mathbf{p}$ and $\mathbf{q}$ are momenta of an electron and a phonon, respectively; ω_q is the energy of a phonon with momentum $\mathbf{q}$; $a^\dagger$ or a is the creation or annihilation operator for electron satisfying anticommutation relations characteristic of Fermi statistics, and $b^\dagger$ or b is the corresponding operator for phonon. In the case of electron conduction in metals, γ_q is given by $\gamma_q=q^2/\omega_q$, but the following procedure is independent of a particular choice of γ_q. (We assume that γ_q is a function of $|\mathbf{q}|$.)

The electric current $\mathbf{J}$ is given by

$$\mathbf{J}=2e\sum_{\mathbf{p}}\mathbf{p}a_{\mathbf{p}}^\dagger a_{\mathbf{p}}, \tag{2.8}$$

where e is the electronic charge. Substituting this into (2.6), one has

$$X(u)=4e^2\sum_{\mathbf{p},\mathbf{p}'}p_\mu p_\nu' \times\langle a_{\mathbf{p}}^\dagger(0)a_{\mathbf{p}}(0)S^\dagger(u)a_{\mathbf{p}'}^\dagger(u)a_{\mathbf{p}'}(u)S(u)\rangle, \tag{2.9}$$

where

$$a_{\mathbf{p}}(t)=e^{itH_0}a_{\mathbf{p}}e^{-itH_0}, \qquad a_{\mathbf{p}}^\dagger(t)=e^{itH_0}a_{\mathbf{p}}^\dagger e^{-itH_0}. \tag{2.10}$$

In our example the unperturbed Hamiltonian H_0 commutes with the electric current operator as well as the number operator $n_{\mathbf{p}}=a_{\mathbf{p}}^\dagger a_{\mathbf{p}}$, i.e.,

$$[H_0,\mathbf{J}]=0, \quad [H_0,n_{\mathbf{p}}]=0, \tag{2.11}$$

whence

$$a_{\mathbf{p}}^\dagger(u)a_{\mathbf{p}}(u)=a_{\mathbf{p}}^\dagger a_{\mathbf{p}}. \tag{2.12}$$

If the infinite series (2.3) and (2.3a) are substituted in (2.9), the very first term of $X(u)$ is independent of u, so that the subsequent integration on u from zero to infinity leads to an apparent divergence as mentioned in the introduction. Also, there will be many other terms which become divergent upon the u integration. However, we hope that we can remove this difficulty by appropriately reordering the infinite series (2.9) and performing partial summation before the integration. In this rearrangement of terms, the diagrammatic representation will be found very useful. In the remainder of this section we shall give a short account of this diagrammatic method.

In calculating the statistical mechanical averages such as the one in (2.9), it is usually much simpler to deal with the grand canonical average rather than with the canonical average, although both types of averaging should yield the same result as the number of particles involved tends to infinity. The grand canonical average $\langle A\rangle$ of a quantity A is defined by

$$\langle A\rangle=\mathrm{Tr}(e^{\alpha N-\beta H}A)/\mathrm{Tr}(e^{\alpha N-\beta H}). \tag{2.13}$$

Here e^α is fugacity and N is given by

$$N=\sum_{\mathbf{p}}a_{\mathbf{p}}^\dagger a_{\mathbf{p}}. \tag{2.14}$$

The symbol Tr stands for a trace taken over a complete set of states with all integral numbers of particles. The fugacity e^α is to be determined from the relation:

$$\langle N\rangle/V=c. \tag{2.15}$$

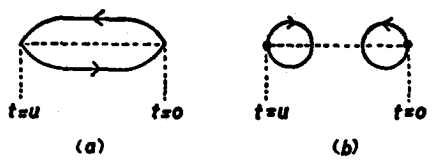


FIG. 1. Diagrams of the zeroth-order terms [Eq. (2.27)].

Under the approximation we shall adopt in the later sections, the grand canonical average referring to the true Hamiltonian H can be replaced by the one referring to H_0, i.e.,

$$\langle A\rangle\approx\langle A\rangle_0=\mathrm{Tr}(e^{\alpha N-\beta H_0}A)/\mathrm{Tr}(e^{\alpha N-\beta H_0}). \tag{2.16}$$

This point will be discussed in full detail when the approximation scheme is actually introduced. Equation (2.16) leads to

$$X(u)\approx X'(u)=4e^2\sum_{\mathbf{p},\mathbf{p}'}p_\mu p_\nu'\langle a_{\mathbf{p}}(0)S^\dagger(u)a_{\mathbf{p}'}^\dagger(u) \times a_{\mathbf{p}'}(u)S(u)a_{\mathbf{p}}^\dagger(0)\rangle_0\exp(\alpha-\beta p^2) \tag{2.17}$$

in place of (2.9).

The explicit form of the operator $H_I(t)$ defined in (2.4) is written as

$$H_I(t)=\sum_{\mathbf{p},\mathbf{q}}V^{-\frac{1}{2}}\gamma_q^{\frac{1}{2}}[a_{\mathbf{p}+\mathbf{q}}^\dagger(t)a_{\mathbf{p}}(t)b_{\mathbf{q}}(t) +a_{\mathbf{p}}^\dagger(t)a_{\mathbf{p}+\mathbf{q}}(t)b_{\mathbf{q}}^\dagger(t)], \tag{2.18}$$

where $b_{\mathbf{q}}(t)$ and $b_{\mathbf{q}}^\dagger(t)$ are defined by equation similar to (2.10) It is easily proved that[11]

$$a_{\mathbf{p}}(t)=a_{\mathbf{p}}\exp(-ip^2t), \quad a_{\mathbf{p}}^\dagger(t)=a_{\mathbf{p}}^\dagger\exp(ip^2t), \tag{2.19}$$

$$b_{\mathbf{q}}(t)=b_{\mathbf{q}}\exp(-i\omega_q t), \quad b_{\mathbf{q}}^\dagger(t)=b_{\mathbf{q}}^\dagger\exp(i\omega_q t). \tag{2.20}$$

When we substitute infinite series $S(u)$ and $S^\dagger(u)$ from (2.3), (2.3a), and (2.18) into the grand canonical average in (2.17), we have to calculate the average of the type $\langle UVWX\cdots YZ\rangle_0$, where the product $UVWX\cdots YZ$ consists of various factors a, $a^\dagger$, b, and $b^\dagger$. According to Bloch and de Dominicis,[10] this can be decomposed into a set of complete contractions of the product. Namely, it is proved that

$$\langle UVW\cdots Z\rangle_0=\sum(\pm1)^{\delta(P)}\langle UV\rangle_0\langle WZ\rangle_0\cdots, \tag{2.21}$$

where $+$ corresponds to Bose operators, $-$ to Fermi operators, $\delta(P)$ is the number of pair exchanges which transpose the product of factors on the left-hand side into the one on the right-hand side. In (2.21) every factor in the product must be paired by another and the summation extends over all possible ways of pairings.

[11] See, for the proof, e.g., S. S. Schweber, H. A. Bethe, and A. de Hoffmann, *Mesons and Fields* (Row, Peterson & Company, Evanston, Illinois, 1955), Vol. 1, p. 170.

If the product consists of an odd number of factors, its average vanishes indentically. A contracted pair such as $\langle UV\rangle_0$ vanishes except for the following four cases,

$$\left.\begin{aligned}\langle a_{\mathbf{p}}(t)a_{\mathbf{p}'}{}^{\dagger}(t')\rangle_0&=\delta(\mathbf{p},\mathbf{p}')(1-f_p)\\&\quad\times\exp[-i(t-t')p^2],\\ \langle a_{\mathbf{p}'}{}^{\dagger}(t')a_{\mathbf{p}}(t)\rangle_0&=\delta(\mathbf{p},\mathbf{p}')f_p\exp[-i(t-t')p^2],\end{aligned}\right\}\quad(2.22)$$

$$\left.\begin{aligned}\langle b_{\mathbf{q}}(t)b_{\mathbf{q}'}{}^{\dagger}(t')\rangle_0&=\delta(\mathbf{q},\mathbf{q}')(1+n_q)\\&\quad\times\exp[-i(t-t')\omega_q],\\ \langle b_{\mathbf{q}'}{}^{\dagger}(t')b_{\mathbf{q}}(t)\rangle_0&=\delta(\mathbf{q},\mathbf{q}')n_q\exp[-i(t-t')\omega_q],\end{aligned}\right\}\quad(2.23)$$

where $\delta(\mathbf{p},\mathbf{p}')$ is Kronecker's delta symbol,[12] f_p and n_q are the Fermi and the Planck distribution functions, respectively:

$$f_p=\exp(\alpha-\beta p^2)/[1+\exp(\alpha-\beta p^2)],\quad(2.24)$$

$$n_q=\exp(-\beta\omega_q)/[1-\exp(-\beta\omega_q)].\quad(2.25)$$

Equations (2.22) and (2.23) are easily proved by direct calculation using expressions (2.19) and (2.20).

Following the standard procedures we can represent each of the complete contractions by a Feynman diagram.

Consider the average

$$U(\mathbf{p}',\mathbf{p};u)=\langle a\mathbf{p}(0)S^{\dagger}(u)a_{\mathbf{p}'}{}^{\dagger}(u)a_{\mathbf{p}'}(u)S(u)a_{\mathbf{p}}{}^{\dagger}(0)\rangle_0,\quad(2.26)$$

which appeared in (2.17). Its zeroth order term is given by

$$\begin{aligned}&\langle a_{\mathbf{p}}(0)a_{\mathbf{p}'}{}^{\dagger}(u)a_{\mathbf{p}'}(u)a_{\mathbf{p}}{}^{\dagger}(0)\rangle_0\\&\quad=\langle a_{\mathbf{p}}(0)a_{\mathbf{p}'}{}^{\dagger}(u)\rangle_0\langle a_{\mathbf{p}'}{}^{\dagger}(u)a_{\mathbf{p}}(0)\rangle_0\\&\qquad+\langle a_{\mathbf{p}}(0)a_{\mathbf{p}}{}^{\dagger}(0)\rangle_0\langle a_{\mathbf{p}'}{}^{\dagger}(u)a_{\mathbf{p}'}(u)\rangle_0.\end{aligned}\quad(2.27)$$

These two nontrivial contractions may be represented by two diagrams shown in Fig. 1, where the time is measured from the right to the left and the contracted pair is represented by a solid line directed from the point of $a^{\dagger}$ to that of a. The first-order term in g of (2.26) vanishes, because it contains either one $b^{\dagger}$ or b. The second-order term can arise from the expansion of $S(u)$ or from the expansion of $S^{\dagger}(u)$ or from both. A most general term of the nth order contains the kth iterates of $H_I(t)$ from $S(u)$ and the $(n-k)$th iterates of $H_I(t)$ from $S^{\dagger}(u)$. The contraction procedure for each term may be performed by making complete contractions in all possible ways. However, all nontrivial contractions can be obtained most simply by drawing diagrams.

We shall give the prescription for drawing diagrams without further discussions, for the correspondence between contractions and diagrams could easily be traced.

[12] In order to avoid a possible confusion, we shall write Kronecker's delta symbol as $\delta(\mathbf{p},\mathbf{p}')$ and Dirac's delta function as $\delta(\mathbf{p}-\mathbf{p}')$. We have therefore an equation $\delta(\mathbf{p},\mathbf{p}')\to(2\pi)^3V^{-1}\times\delta(\mathbf{p}-\mathbf{p}')$ in the limit $V\to\infty$.

(i) Mark two points at $t=0$ and $t=u$ on a horizontal boundary (dashed) line.

(ii) Mark k points above the boundary line at t_1, t_2, $\cdots$, t_k, $(t_1>t_2>\cdots>t_k)$, also mark $n-k$ points below the boundary at t_1', t_2', $\cdots$, t_{n-k}', $(t_1'>t_2'>\cdots>t_{n-k}')$.

(iii) Draw $n/2$ (n : even) directed wavy lines between the points marked in the process (ii); every point should be connected with another by only one wavy line. Draw directed particle (solid) lines, one entering and one leaving each point marked in (i) and (ii); these lines may run between points or from a point to itself. The different diagrams are obtained by drawing the phonon (wavy) and the particle (solid) lines in all possible ways.

It is seen from (2.22) and (2.23) that the contracted pair $\langle a_{\mathbf{p}}(t)a_{\mathbf{p}'}{}^{\dagger}(t')\rangle_0$, which is denoted by a particle line running from a creation point at t' to an annihilation point at t, can be associated with a single momentum vector $\mathbf{p}$, not with two momenta $\mathbf{p}$ and $\mathbf{p}'$, and similarly the contracted pair $\langle b_{\mathbf{q}}(t)b_{\mathbf{q}'}{}^{\dagger}(t')\rangle_0$ is denoted by a directed phonon line with a single momentum $\mathbf{q}$. If we call each of the points marked in (ii) a corner, each corner has by construction two particle lines, one entering and one leaving, and one phonon line, either entering or leaving. It is noted that the momentum flows associated with directed lines are conserved at each corner in view of a particular choice of the interaction Hamiltonian (2.18).

Now it is convenient to introduce the following functions which are closely related to the contractions given by (2.22) and (2.23):

$$\left.\begin{aligned}G_+(\tau',\tau;\mathbf{p})&=(1-f_p)\exp[-i(\tau'-\tau)p^2],\\ G_-(\tau',\tau;\mathbf{p})&=f_p\exp[-i(\tau'-\tau)p^2],\end{aligned}\right\}\quad(2.28)$$

$$\left.\begin{aligned}P_+(\tau',\tau;\mathbf{q})&=(1+n_q)\exp[-i(\tau'-\tau)\omega_q],\\ P_-(\tau',\tau;\mathbf{q})&=n_q\exp[-i(\tau'-\tau)\omega_q].\end{aligned}\right\}\quad(2.29)$$

Then the contribution to (2.26) corresponding to a given Feynman diagram can be written down from the following prescriptions:

(i) For every particle line with momentum $\mathbf{p}$ or phonon line with momentum $\mathbf{q}$ running from a point at τ to another point at τ', introduce a factor which is seen from Table I. For example, if a particle line with

TABLE I. Correspondence between parricle or phonon line (from τ to τ') and introduced factor.[a]

τ' \ τ	Above	Below
Above	G_+ or $P_+(\tau'>\tau)$ G_- or $P_-(\tau'<\tau)$	G_- or $P_-(\tau'\gtrless\tau)$
Below	G_+ or $P_+(\tau'\gtrless\tau)$	G_- or $P_-(\tau'>\tau)$ G_+ or $P_+(\tau'<\tau)$

[a] If a particle line runs from a point at $\tau=0$ to a point at τ' below the boundary introduce a factor $G_+(\tau',0;\mathbf{p})$. Similarly, if a particle line enters into a point $\tau=0$ from τ' above the boundary introduce $G_+(0,\tau';\mathbf{p})$.

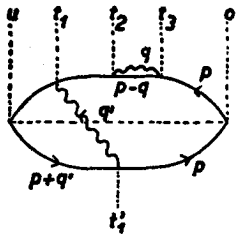


FIG. 2. Example of a diagram which corresponds to the expression (2.30).

momentum $\mathbf{p}$ runs from a point above the boundary at τ to another point above the boundary at τ' introduce a factor $G_+(\tau',\tau;\mathbf{p})$ if $\tau'>\tau$, and $G_-(\tau',\tau;\mathbf{p})$ if $\tau'<\tau$. Similarly, if a phonon line with momentum $\mathbf{q}$ runs crossing the boundary upwards introduce a factor $P_-(\tau',\tau;\mathbf{q})$, and so on.

(ii) For every particle line with momentum $\mathbf{p}$ running from a point to itself introduce $f_\mathbf{p}$, except for the point at $t=0$ where $1-f_p$ should be introduced.

(iii) Make a product of all factors given by (i) and (ii), integrate with respect to $t_1, t_2, \cdots, t_k$ $(u\geqslant t_1\geqslant t_2, \cdots \geqslant t_k\geqslant 0)$ and $t_1', t_2', \cdots, t_{n-k}'$ $(u\geqslant t_1'\geqslant t_2'\cdots\geqslant t_{n-k}'\geqslant 0)$.

(iv) Multiply the integral by $(-1)^{l+k-1}(ig)^n V^{-n/2}\times\delta(\mathbf{p}',\mathbf{p}_0)\prod\gamma_{q_i}$ and perform the summations with respect to all momentum variables of particles and phonons in the intermediate states. Here l is the number of closed particle loops, $\mathbf{p}_0$ is the momentum associated with a particle line entering the end point at $t=u$. As an example of the above prescriptions we give the contribution corresponding to the diagram shown in Fig. 2.

$$-(ig)^4V^{-2}\int_0^u dt_1\int_0^{t_1}dt_2\int_0^{t_2}dt_3\int_0^u dt_1'\sum_{\mathbf{q},\mathbf{q}'}\gamma_q\gamma_{q'}$$
$$\times\delta(\mathbf{p}',\mathbf{p}+\mathbf{q}')G_+(u,t_1;\mathbf{p}+\mathbf{q}')G_+(t_1,t_2;\mathbf{p})$$
$$\times G_+(t_2,t_3;\mathbf{p}-\mathbf{q})G_+(t_3,0;\mathbf{p})G_+(0,t_1';\mathbf{p})$$
$$\times G_+(t_1',u;\mathbf{p}+\mathbf{q}')P_+(t_2,t_3;\mathbf{q})P_-(t_1,t_1';\mathbf{q}'). \quad (2.30)$$

III. LINKED-CLUSTER EXPANSION

The linked-cluster expansion for the ground-state energy or the thermodynamic functions has been discussed by many authors.[13] Also its generalization to the current correlation function was established by Montroll and Ward[8] in terms of toron diagrams. However, our diagrams being different from theirs, it is not immediately clear how their proof is applied to the present case. In this section we shall briefly discuss the linked-cluster expansion using our diagrams.

A diagram constructed from prescriptions in Sec. II may or may not consist of two or more connected parts. In the former case it is called *unlinked* and in the latter case *linked*. For example the diagram of Fig. 1(a) is linked and the one of Fig. 1(b) is unlinked. We shall show through the linked-cluster expansion that $U(\mathbf{p}',\mathbf{p};u)$ given by (2.26) is expressed as a sum of terms corresponding to diagrams which have parts, linked or unlinked, comprising the fixed points at $t=0$ and $t=u$. In other words, it is not necessary to take into account vacuum diagrams which have no fixed points in themselves. In the following, we shall discuss the problem along the line conjectured by Bloch.[13]

Let us first consider a trivial case; the calculation of the quantity

$$\langle S^\dagger(u)S(u)\rangle_0=1 \quad (3.1)$$

in terms of our diagrams. It is clear that the diagrams contributing to $\langle S^\dagger(u)S(u)\rangle_0$ are all vacuum diagrams. A given diagram of this type is in general composed of a certain number of linked vacuum diagrams. We shall now say that two linked diagrams have the same structure if they have the same arrangement of corners, phonon and particle lines and the same location of corners relative to the boundary line and differ only in the labeling of summation variables. It is noted that in our definition two structures are to be distinguished from each other when their corners are located differently. For example two diagrams (a) and (b) in Fig. 3 have the same structure but (a) and (c) are different in structure. If the structure Γ of a general diagram for (3.1) is made up of k_1 linked structures Γ_1, k_2 linked structures Γ_2, and so on, we shall write it as

$$\Gamma=k_1\Gamma_1+k_2\Gamma_2+\cdots. \quad (3.2)$$

Now let us denote the contribution of the structure Γ to $\langle S^\dagger(u)S(u)\rangle_0$ by $U(\Gamma)$. If the integrations with respect to time variables extend independently from 0 to u, $U(\Gamma)$ may be expressed as a product of individual $U(\Gamma_i)$, but this is not the case, since the integration is restricted by the prescription (iii) in Sec. II. However, one may remedy the situation in the following way.

Suppose for simplicity that a given diagram of structure Γ is composed of two connected parts of structures Γ_1 and Γ_2. Let the diagram of Γ_2 be fixed and let that of Γ_1 move from right to left along the boundary line, keeping the relative locations of corners on Γ_1 unchanged. In doing so, the relative positions of corners on Γ_1 with respect to those on Γ_2 will change, thus the labeling of time variables on Γ_1 (and Γ_2) will change. Therefore, by this movement one may have many Feynman diagrams which have the same structure

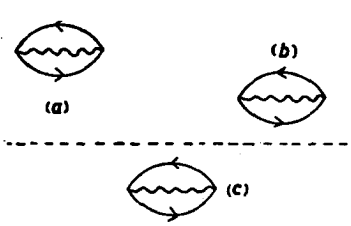


FIG. 3. a and b are identical in structure but a and c are different.

[13] T. Matsubara, Progr. Theoret. Phys. (Kyoto) 14, 351 (1955); K. A. Brueckner, Phys. Rev. 100, 36 (1955); J. Goldstone, Proc. Roy. Soc. (London) A239, 267 (1957); N. Hugenholtz, Physica 23, 481 (1957); J. Hubbard, Proc. Roy. Soc. (London) A240, 539 (1957); D. J. Thouless, Phys. Rev. 107, 1162 (1957); C. Bloch, Nuclear Phys. 7, 451 (1958); E. Montroll and J. Ward, Phys. Fluids 1, 55 (1958); T. D. Lee and C. N. Yang, Phys. Rev. 113, 1165 (1959); A. E. Glassgold, W. Heckrotte, and K. M. Watson, *ibid.* 115, 1374 (1959).

but differ only in the labeling of time variables. If one sums up the contributions of these diagrams, introducing an appropriate change of variables, it may be easily seen that

$$U(\Gamma)=U(\Gamma_1)U(\Gamma_2). \tag{3.3}$$

Similarly, if the structure Γ has the form given by (3.2), it is proved that

$$U(\Gamma)=[U(\Gamma_1)]^{k_1}[U(\Gamma_2)]^{k_2}\cdots/k_1!k_2!\cdots. \tag{3.4}$$

The factors $k_1!$, $k_2!$, etc., come from the fact that the permutations of k_1 connected diagrams among themselves do not lead to a new diagram. (The detail will be seen in Bloch's paper.)

From (3.4) it follows that

$$\langle S^\dagger(u)S(u)\rangle_0=\sum_{k_1,k_2,\cdots=0}^{\infty}\frac{[U(\Gamma_1)]^{k_1}[U(\Gamma_2)]^{k_2}\cdots}{k_1!k_2!\cdots}$$

$$=\exp[\sum_{\Gamma_j}U(\Gamma_j)], \tag{3.5}$$

where the summation on Γ_j is to be taken over all linked structures. It is noted here that the contribution 1 from the expansion of $S^\dagger(u)S(u)$ is automatically included in (3.4) as a term with all $k_j=0$. Equation (3.1) requires that all the contributions summed over linked structures should cancel out among themselves in every order of coupling constant g. This can be checked by direct calculations for the lowest order terms.

Let us now consider the average $U(\mathbf{p}',\mathbf{p};u)$. If the contributions are represented by diagrams, a general diagram is found to be a sum of subdiagrams, linked or unlinked, comprising fixed points at $t=0$ and $t=u$ and a certain number of vacuum diagrams. Summing up all the contributions in the same way as we did for $\langle S^\dagger(u)S(u)\rangle_0$, we obtain the following expression

$$U(\mathbf{p}',\mathbf{p};u)=\exp[\sum_j U(\Gamma_j)]\sum_{\Gamma_i}a(\mathbf{p}'u;\mathbf{p}0;\Gamma_i)$$

$$=\sum_{\Gamma_i}a(\mathbf{p}'u;\mathbf{p}0;\Gamma_i). \tag{3.6}$$

Here the quantity $a(\mathbf{p}'u;\mathbf{p}0;\Gamma_i)$ is the contribution associated with a structure Γ_i, free from vacuum structure, which has parts, linked or unlinked, containing the fixed points at $t=0$ and $t=u$. Equation (3.6) is the required linked-cluster expansion in our case.

In closing this section, we would like to mention some remarks on the calculation of $\sigma_{\mu\nu}$. From (2.5), (2.17), and (2.26), it follows that

$$\sigma_{\mu\nu}=\frac{4e^2\beta}{V}\mathrm{Re}\left\{\int_0^\infty du\sum_{\mathbf{p},\mathbf{p}'}p_\mu p_\nu' U(\mathbf{p}',\mathbf{p};u)\right.$$

$$\left.\times\exp(\alpha-\beta p^2)\right\}. \tag{3.7}$$

If one substitutes (3.6) into this equation, the contributions can be classified into two types according to whether their representative diagrams are linked or unlinked. It will be easily seen that the latter type will not contribute to $\sigma_{\mu\nu}$. In fact, a term of this type should have the form: $a(\mathbf{p}'u;\mathbf{p}0;\Gamma_i)=c_1(\mathbf{p})c_2(\mathbf{p}')$, where one factor depends on the momentum $\mathbf{p}$ and the other on $\mathbf{p}'$, corresponding to its unlinked structure. There is no preferred direction in the system under consideration, so that each factor should depend on the magnitude of momentum vector, i.e.,

$$c_1(\mathbf{p})c_2(\mathbf{p}')=c_1(|\mathbf{p}|)c_2(|\mathbf{p}'|). \tag{3.8}$$

One sees then by the symmetry argument that the property (3.8) makes the sums over $\mathbf{p}$ and $\mathbf{p}'$ in (3.7) vanishing, which proves the above statement.

IV. ELECTRICAL CONDUCTIVITY OF ELECTRON-PHONON SYSTEM

We have discussed in the last two sections how the current correlation function is expanded in terms of linked-cluster structures. In this section these diagrams are analyzed for their dependencies on coupling constant and time. The analysis will be subsequently used to select, for the lowest order calculation of electrical conductivity, a special class of diagrams according to the rule identical to that used in statistical mechanical theory of Van Hove, and Prigogine and his collaborators on irreversible processes. It is then shown that if the electron states are not degenerate, the sum of the selected diagrams satisfies an integral equation, which is closely related to the solution of the Boltzmann equation.

We have shown in Sec. III that only the linked diagrams having fixed points at $t=0$ and $t=u$ contribute to the electrical conductivity. In order to facilitate the analysis of these diagrams, let us assume that the electrons obey the Boltzmann statistics. This situation is physically realized in a semiconductor where the number density of conduction electron is so low that the electronic states are no longer degenerate. This assumption reduces the number of diagrams significantly. It is equivalent to dropping all diagrams whose corresponding expressions contain the factor f and replacing the factor $(1-f)$ by unity in the integrand obtained from the prescriptions in Sec. II. Symbolically

$$f\to 0,\quad 1-f\to 1. \tag{4.1}$$

One may easily see that under this assumption the particle lines always run from right to left above the boundary and from left to right below the boundary.

We enumerate various contributions in the following.

(1) The zeroth order. There is only one diagram in the zeroth order. It was given in Fig. 1(a). Its contribution is

$$\delta(\mathbf{p},\mathbf{p}'). \tag{4.2}$$

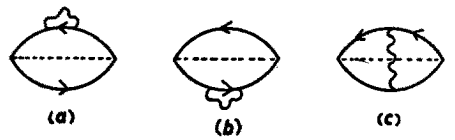


FIG. 4. The second-order diagrams.

(2) The second order. Three types of nonvanishing diagram are found and drawn in Fig. 4, where undirected phonon line is purposely drawn with the understanding that an undirected phonon line corresponds to two lines directed in the mutually opposite senses.

Let us calculate the contribution corresponding to Fig. 4(a). It is easily shown to be

$$-\delta(\mathbf{p}',\mathbf{p})g^2\int_0^u dt_1\int_0^{t_1} dt_2 V^{-1}\sum_{\mathbf{q}}\gamma_q\{(1+n_q)\times\exp[i\epsilon_1(t_1-t_2)]+n_q\exp[i\epsilon_2(t_1-t_2)]\},\quad(4.3)$$

where

$$\epsilon_1=\mathbf{p}^2-(\mathbf{p}-\mathbf{q})^2-\omega_q,\quad \epsilon_2=\mathbf{p}^2-(\mathbf{p}-\mathbf{q})^2+\omega_q.\quad(4.4)$$

In evaluating the time integral, we shall employ the frequently used asymptotic formulas:

$$\int_0^u dt_1\int_0^{t_1} dt_2\frac{\exp[\pm i\epsilon(t_1-t_2)]}{(p-1)!}t_2^{p-1}=\frac{\pi u^p}{p!}\delta_\pm(\epsilon)+O(u^{p-1}),\quad(4.5)$$

$$\delta_\pm(\epsilon)=\delta(\epsilon)\pm(i/\pi)P(1/\epsilon).\quad(4.6)$$

Here the symbol P means to take Cauchy's principal value.

It is now convenient to introduce functions $\Gamma_\pm$ in the followlng way

$$\Gamma_\pm(\mathbf{p})=\pi g^2V^{-1}\textstyle\sum_{\mathbf{q}}\gamma_q[(1+n_q)\delta_\pm(\epsilon_1)+n_q\delta_\pm(\epsilon_2)].\quad(4.7)$$

Then (4.3) is reduced to

$$-u\delta(\mathbf{p}',\mathbf{p})\Gamma_+(\mathbf{p}).\quad(4.8)$$

Similarly, the contributions of Fig. 4(b) and (c) are given by, respectively,

$$-u\delta(\mathbf{p}',\mathbf{p})\Gamma_-(\mathbf{p}),\quad(4.9)$$

$$2g^2u\pi V^{-1}\textstyle\sum_{\mathbf{q}}\gamma_q\delta(\mathbf{p}',\mathbf{p}-\mathbf{q})[(1+n_q)\delta(\epsilon_1)+n_q\delta(\epsilon_2)].\quad(4.10)$$

It is noted that these three expressions (4.8), (4.9), and (4.10) are all proportional to g^2u.

FIG. 5. Examples of diagram. a is proportional to g^4u and b to $g^{10}u^6$. The contribution of c vanishes in the limit (4.12).

(3) Higher orders. In the similar manner we can evaluate various integrals of higher orders. For example, two diagrams (a) and (b) in Fig. 5 are shown to have contributions proportional to g^4u and $g^{10}u^4$, respectively. In general it is shown that every diagram in the present case has a contribution proportional to

$$g^{2m}(g^2u)^n,\quad m,n=0,1,2,\cdots.\quad(4.11)$$

We shall now state the rule for selecting diagrams for the purpose of calculating the conductivity systematically. We shall adopt the same selection rule which Van Hove, Prigogine, and others have used to derive the master equation for the weakly coupled systems. It is given by [cf. (1.1) and (1.2)]

1. $N\to\infty$, $V\to\infty$ such that $N/V=$finite, (4.12)

2. $g\to 0$, $t\to\infty$ such that $g^2t=$finite. (4.13)

In the present case we may expect that the current correlation function $X(u)$ would decay exponentially as $\exp(-u/t)$ for large u. The limit is then equivalent to retain all the terms of order of $(g^2u)^n$, $n=0,1,2,\cdots$, in the series. Symbolically,

$$g^2u=\text{finite},\quad g^m(g^2u)^n\to 0,\quad m>0,\quad n=0,1,2,\cdots.\quad(4.14)$$

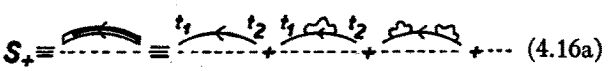
We immediately see that in the limit (4.14) the approximation (2.17) is justified because the correction terms are of the orders of $g^k(k\geqslant 1)$, and therefore can be neglected in the limit (4.14). Such corrections would be important in higher order calculations, for example in the calculation in which one keeps all contributions of the order $g^m(g^2u)^n$ $(m>0, n=0,1,\cdots)$.

In the limit (4.12) the diagram shown in Fig. 5(c) does not contribute to the current correlation function. It is seen to have extra restrictions on the summation over momenta in the intermediate states. This property is characteristic of those diagrams which vanish in the limit (4.12).

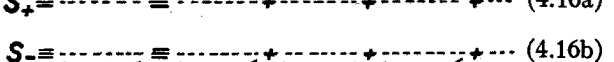
In the same limit the momentum eigenvalues become continuous so that summations over momenta will be replaced by the corresponding integrals

$$V^{-1}\sum_{\mathbf{q}}\to(2\pi)^{-3}\int d\mathbf{q}.\quad(4.15)$$

We now wish to sum up all terms of the form $(g^2u)^n$, $n=0,1,2,\cdots$. Examples of this type of diagrams are shown in Fig. 1(a), 4(a), 4(b), 4(c), and 5(b). In order to sum these diagrams we shall introduce modified propagators $S_\pm$ defined symbolically by a series of diagrams



(4.16a)



(4.16b)

where two ends of each diagram have the same time coordinates t_1 and t_2 ($u \geqslant t_1 \geqslant t_2 \geqslant 0$). By repeatedly using (4.5), the corresponding mathematical expressions for $S_\pm$ are shown to be

$$S_\pm(t_1,t_2;\mathbf{p}) = \exp[\mp i(t_1-t_2)p^2] \sum_{k=0}^{\infty} \frac{[-(t_1-t_2)\Gamma_\pm(\mathbf{p})]^k}{k!}$$

$$= \exp[\mp i(t_1-t_2)p^2 - \Gamma_\pm(\mathbf{p})(t_1-t_2)]. \quad (4.17)$$

If we approximate $U(\mathbf{p}',\mathbf{p};u)$ by $Y(\mathbf{p}',\mathbf{p};u)$, i.e.,

$$U(\mathbf{p}',\mathbf{p};u) \approx Y(\mathbf{p}',\mathbf{p};u), \quad (4.18)$$

which is defined by a series:

$Y(\mathbf{p}',\mathbf{p};u) \equiv$  $+ \cdots,$ (4.19)

where every part of the particle lines represents a modified propagator (4.16a) or (4.16b). Evidently this series contains all contributions proportional to $(g^2u)^n$, $n=0, 1, 2, \cdots$. Simple calculation yields for $Y(\mathbf{p}',\mathbf{p};u)$:

$$Y(\mathbf{p}',\mathbf{p};u) = \delta(\mathbf{p}',\mathbf{p})e^{-\Gamma(\mathbf{p})u} + \sum_{n=1}^{\infty} \left(\frac{g}{2\pi}\right)^{2n} \times \int_0^u dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} dt_n \int d\mathbf{q}_1 \cdots d\mathbf{q}_n \times \delta(\mathbf{p}',\mathbf{p}_1)e^{-\Gamma(\mathbf{p})t_n} \prod_{i=1}^{n} \{W(\mathbf{p}_{i+1},\mathbf{q}_i) \times \exp[-\Gamma(\mathbf{p}_i)(t_{i-1}-t_i)]\}, \quad (4.20)$$

where

$$\mathbf{p}_i = \mathbf{p} - \sum_{j=i}^{n} \mathbf{q}_j, \quad \mathbf{p}_{n+1} = \mathbf{p}, \quad t_0 = u, \quad (4.21)$$

$$W(\mathbf{p},\mathbf{q}) = \gamma_q[(1+n_q)\delta(\epsilon_1) + n_q\delta(\epsilon_2)],$$

and

$$\Gamma(\mathbf{p}) = \Gamma_+(\mathbf{p}) + \Gamma_-(\mathbf{p}) = \frac{g^2}{(2\pi)^2} \int d\mathbf{q} W(\mathbf{p},\mathbf{q}) > 0. \quad (4.22)$$

Let us define a function $\psi(\mathbf{p}',\mathbf{p})$ by

$$\psi(\mathbf{p}',\mathbf{p}) = \int_0^\infty du Y(\mathbf{p}',\mathbf{p};u). \quad (4.23)$$

Then it follows from (4.20) that

$$\psi(\mathbf{p}',\mathbf{p}) = \frac{\delta(\mathbf{p}',\mathbf{p})}{\Gamma(\mathbf{p})} + \sum_{n=1}^{\infty} \left(\frac{g}{2\pi}\right)^{2n} \int \frac{\delta(\mathbf{p}',\mathbf{p}_1)}{\Gamma(\mathbf{p})} \times \prod_{i=1}^{n} \left\{\frac{W(\mathbf{p}_{i+1},\mathbf{q}_i)}{\Gamma(\mathbf{p}_i)} d\mathbf{q}_i\right\}. \quad (4.24)$$

One may easily see that $\psi(\mathbf{p}',\mathbf{p})$ satisfies the following integral equation:

$$\psi(\mathbf{p}',\mathbf{p}) = \frac{\delta(\mathbf{p}',\mathbf{p})}{\Gamma(\mathbf{p})} + \frac{g^2}{(2\pi)^2} \int \frac{W(\mathbf{p},\mathbf{q})}{\Gamma(\mathbf{p})} \psi(\mathbf{p}', \mathbf{p}-\mathbf{q}) d\mathbf{q}. \quad (4.25)$$

If we further define a function $\varphi_\nu(\mathbf{p})$ by

$$\varphi_\nu(\mathbf{p}) = \frac{V}{(2\pi)^3} \int p_\nu' \psi(\mathbf{p}',\mathbf{p}) d\mathbf{p}', \quad (4.26)$$

then from (4.25) and the footnote 12, we have

$$\varphi_\nu(\mathbf{p}) = \frac{p_\nu}{\Gamma(\mathbf{p})} + \frac{g^2}{(2\pi)^2} \int \frac{W(\mathbf{p},\mathbf{q})}{\Gamma(\mathbf{p})} \varphi_\nu(\mathbf{p}-\mathbf{q}) d\mathbf{q}, \quad (4.27)$$

or from (4.22) it follows that

$$p_\nu = \frac{g^2}{(2\pi)^2} \int W(\mathbf{p},\mathbf{q})[\varphi_\nu(\mathbf{p}) - \varphi_\nu(\mathbf{p}-\mathbf{q})] d\mathbf{q}. \quad (4.28)$$

Since the system under consideration is isotropic, the conductivity tensor will be diagonal and

$$\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = \sigma. \quad (4.29)$$

From (3.7), (4.23), and (4.26), we have

$$\sigma = \frac{4\beta e^2}{(2\pi)^3} \int d\mathbf{p} p_x \varphi_x(\mathbf{p}) \exp(\alpha - \beta p^2). \quad (4.30)$$

Unfortunately the integral equation (4.28) is not easy to solve. Instead of solving this we shall be content with the fact that it is exactly the same type of integral equation which we would encounter when we calculate the conductivity starting from the Boltzmann equation. We shall briefly discuss this point in what follows.

The Boltzmann equation for the present case is written as[14]

$$e\mathcal{E}\frac{\partial F}{\partial p_x} = \frac{g^2}{(2\pi)^2} \int d\mathbf{q}\gamma_q\{[n_qF(\mathbf{p}-\mathbf{q}) - (1+n_q)F(\mathbf{p})]\delta(\epsilon_1) + [(1+n_q)F(\mathbf{p}-\mathbf{q}) - n_qF(\mathbf{p})]\delta(\epsilon_2)\}. \quad (4.31)$$

Here $F(\mathbf{p})$ is the distribution function for the electron and n_q that for the phonon, which is assumed to have the same form in the equilibrium state, i.e., Planck

[14] A. H. Wilson, *The Theory of Metals* (Cambridge University Press, New York, 1954), 2nd ed., p. 259.

distribution function, and $\mathcal{E}$ is the magnitude of th stationary electric field which is applied to the system along the x axis.

If we assume that $\mathcal{E}$ is small, we may put $F(\mathbf{p})$ into the form:

$$F(\mathbf{p})=f(\mathbf{p})+\mathcal{E}\phi_x(\mathbf{p})\partial f(\mathbf{p})/\partial(p^2), \tag{4.32}$$

where $f(\mathbf{p})=\exp(\alpha-\beta p^2)$, $(2\pi)^{-3}f(\mathbf{p})d\mathbf{p}$ being the number of electrons per unit volume in the range $\mathbf{p}\sim\mathbf{p}+d\mathbf{p}$. Substituting (4.32) into (4.31) and neglecting higher order terms of $\mathcal{E}$, we get

$$2ep_x=\frac{g^2}{(2\pi)^2}\int d\mathbf{q}W(\mathbf{p},\mathbf{q})[\phi_x(\mathbf{p}-\mathbf{q})-\phi_x(\mathbf{p})]. \tag{4.33}$$

On the other hand σ is given by from Ohm's law:

$$\sigma=\frac{\langle j_x\rangle_{\mathrm{Av}}}{\mathcal{E}}=-\frac{2e\beta}{(2\pi)^3}\int d\mathbf{p}p_x\phi_x(\mathbf{p})\exp(\alpha-\beta p^2). \tag{4.34}$$

Comparison of (4.28) and (4.33) shows that

$$\phi_x(\mathbf{p})=-2e\varphi_x(\mathbf{p}), \tag{4.35}$$

which yields the identical conductivity for both methods in view of (4.30) and (4.34).

In conclusion, our result is stated as follows. After summing up the sawtooth diagrams and introducing the modified particle propagator, if one considers all the diagrams of ladder type, one is led to the electrical conductivity calculated by means of Boltzmann equation. We would like to emphasize that our derivation of this result is based only on the assumptions (4.12) and (4.13).

V. DISCUSSIONS

We would like to discuss in this section the physical meaning of the assumptions (4.12) and (4.13) on which our selection of diagrams is based. One may not raise any objection to the first assumption, stating that the system under consideration is infinitely large, for if this were not satisfied, the starting formula (Kubo's formula) itself would not be justified. The second assumption (4.13) is, as mentioned in the introduction, closely related to the existence of two well separated time scales. We shall examine this point in some detail.

In our theory we can find two distinct time scales. These may be represented for example by T' in (1.4) and $t_c=[\mathbf{p}^2-(\mathbf{p}-\mathbf{q})^2-\omega_q]^{-1}\approx(2\mathbf{p}\cdot\mathbf{q})^{-1}$ in the exponential that appeared in (4.3). Let us estimate the order of magnitude of $(2pq)^{-1}$. In the conventional units it is $(2pq/2M)^{-1}=(vq)^{-1}$, where v is the velocity of an electron. The range of values of q is limited from 0 to q_D (q_D: Debye's characteristic wave number). Therefore the average value of q^{-1} will be of the order $q_D^{-1}\approx d$ (lattice spacing) above Debye's characteristic temperature. The time $t_c=(d/v)$, which may be called duration of collision or collision time, will be very small compared with the relaxation time. Below the Debye temperature, scattering of electrons due to phonons of low momenta will become predominant, for the preferred presence of these phonons is expected from the form of n_q. Therefore, the average time $(vq)^{-1}$ becomes larger and larger as the temperature is lowered. However, the same preferred presence of low momentum phonons also makes the relaxation time t_r large in a proportional manner. In fact it may be shown from elementary physical analysis that any average contribution to t_r due to phonons of momentum $\mathbf{q}$ is proportional to n_q^{-1}. [This is in fact borne out in the expression (4.28).] Therefore it is highly unlikely that t_c exceeds t_r, i.e.,

$$t_r>t_c. \tag{5.1}$$

In many cases we would expect that

$$t_r\gg t_c, \tag{5.2}$$

which is certainly true at high temperatures. This strong inequality is considered to be essential for the discussion of the scattering as well as for that of transport phenomena. If the coupling constant g is small, we may assume that t_r is roughly proportional to g^{-2}. Therefore one of the surest ways of insuring (5.2) is to take the weak coupling limit in the form of (4.13). Another way is to assume the instantaneous collision time. It is very interesting to note that both of these limits lead effectively to the same selection rule (4.14). The time T' is assumed to be considerably larger than t_r.

It is noted that in our discussion the assumption of repeated random phase and the restrictive assumption about relaxation time, both of which assumptions are inherent in the Boltzmann equation, are completely avoided. This advantage was also demonstrated by Chester and Thellung.[9]

In order to improve the present calculation, we must take account of the finiteness of collision time t_c. This may be done by replacing the rule (4.13) by a new condition.[15] Then we shall have to consider the following corrections, perhaps simultaneously:

(1) Selection of a larger class of diagrams than those given by (4.14).

(2) Appropriate calculation of the grand canonical average. The approximation (2.16) was justified by assuming the condition (4.14).

(3) Appropriate evaluation of the time integrals.

In the present example the calculation of electrical conductivity is performed with the assumption that the electrons obey the Boltzmann statistics. In many cases of physical interest, for example in the case of a metal, the electrons obey the Fermi-Dirac statistics.

[15] *Note added in proof.* A calculation in this direction has been published by E. Verboven, Physica **26**, 1091 (1960).

Our general formalism can be applied to this case, though some complication arises from the statistics effect on the intermediate states of electrons. The analysis of this case will be reported in a separate paper.

The viscosity coefficient of a classical dilute gas can be treated essentially by replacing the assumption (1.2) by (1.3). This will be reported in a subsequent paper.

ACKNOWLEDGMENTS

The authors wish to thank Professor A. J. F. Siegert, Dr. K. Aizu, and Dr. R. Hirota of Northwestern University, for valuable discussions. One of the authors (S.F.) wishes to acknowledge the illuminating discussions with Professor I. Prigogine and Dr. R. Balescu of Universitié Libre de Bruxelles.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH-APRIL, 1962

Theory of Transport Coefficients. II Viscosity Coefficients of Quantum Gases Obeying the Boltzmann Statistics*

S. FUJITA
Université Libre de Bruxelles, Brussels, Belgium
(Received May 11, 1961)

The viscosity coefficients of a weakly coupled and a dilute quantum gas, both obeying the classical statistics, are calculated using the method developed by Abe and the present author. In this method transport coefficients are calculated from exact formulas due to Kubo with the natural introduction of the irreversibility following the ideas of Van Hove, Prigogine, and their collaborators. It is shown explicitly that the low-density limit as well as the weak-coupling limit of the viscosity are those which are expected from the conventional calculation via the Boltzmann equation. The nature of these limits is closely related to the two well separated time scales—average duration of collision and average time between successive collisions.

I. INTRODUCTION

THE general idea and method developed in the previous paper[1] is applied to the treatment of the viscosity coefficients of classical statistical gases.

The object of the present paper is not to present a new method for the explicit evaluation of the viscosity coefficients but to show how these coefficients are calculated from the exact formulas without the aid of the Boltzmann equation and by doing so to clarify the nature of the approximations inherent in the calculation.

It is hoped that the present treatment will form a prototype of the future calculation of viscosity coefficients in the case where the traditional treatment via the Boltzmann equation is no longer justified.

Let us consider a gas characterized by the Hamiltonian H

$$H\equiv\sum_{i=1}^{N}p_i^2+\tfrac{1}{2}g\sum_{i\neq j=1}^{N}\sum\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j), \tag{1.1}$$

where the vectors $\mathbf{p}_i$ and $\mathbf{r}_i$ are the momentum and the position of the ith particle and $\mathcal{U}$ is a pair-force potential with a *short range*, N the number of particles, and g the coupling constant.

The static viscosity coefficient η of the gas is, in analogy with the formula (I.1.4), given by[2]

$$\eta=\beta V^{-1}\,\mathrm{Re}\left[\lim_{T'\to\infty}\int_0^{T'}\frac{du\,\mathrm{Tr}\{e^{-\beta H}J_{xy}J_{xy}[u]\}}{\mathrm{Tr}\{e^{-\beta H}\}}\right], \tag{1.2}$$

where the current operator J_{xy} is defined by[3]

$$J_{xy}\equiv 2\sum_{i=1}^{N}p_{ix}p_{iy}-\tfrac{1}{4}g\sum_{i\neq j}^{N}\sum\left[(x_i-x_j)\frac{\partial\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j)}{\partial(y_i-y_j)}+(y_i-y_j)\frac{\partial\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j)}{\partial(x_i-x_j)}\right] \tag{1.3}$$

with $\mathbf{r}\equiv(x,y,z)$. The operator $J_{xy}[u]$ is the current in the Heisenberg picture defined by

$$J_{xy}[u]\equiv e^{iuH}J_{xy}e^{-iuH}. \tag{1.4}$$

The volume of container is denoted by V.

We wish to calculate the viscosity coefficient from (1.2) in the two limiting cases— a weakly-coupled gas and a dilute gas, both having short-range pair interactions and obeying the classical statistics. There is no real gas in nature which can be treated in the category of a weakly coupled gas. However, it turns out that the

* This work was supported by the Air Research and Development Command, U. S. Air Force, through its European office.

[1] S. Fujita and R. Abe, J. Math. Phys. 3, 350 (1962). This paper is referred to as I.

[2] H. Mori, Phys. Rev. 112, 1829 (1958). More detailed references may be found in reference 1.

[3] The units are chosen in the present paper such that $\hbar=2M=1$, M being the mass of a particle, unless otherwise specified.

Our general formalism can be applied to this case, though some complication arises from the statistics effect on the intermediate states of electrons. The analysis of this case will be reported in a separate paper.

The viscosity coefficient of a classical dilute gas can be treated essentially by replacing the assumption (1.2) by (1.3). This will be reported in a subsequent paper.

ACKNOWLEDGMENTS

The authors wish to thank Professor A. J. F. Siegert, Dr. K. Aizu, and Dr. R. Hirota of Northwestern University, for valuable discussions. One of the authors (S.F.) wishes to acknowledge the illuminating discussions with Professor I. Prigogine and Dr. R. Balescu of Universitié Libre de Bruxelles.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH-APRIL, 1962

Theory of Transport Coefficients. II Viscosity Coefficients of Quantum Gases Obeying the Boltzmann Statistics*

S. FUJITA
Université Libre de Bruxelles, Brussels, Belgium
(Received May 11, 1961)

The viscosity coefficients of a weakly coupled and a dilute quantum gas, both obeying the classical statistics, are calculated using the method developed by Abe and the present author. In this method transport coefficients are calculated from exact formulas due to Kubo with the natural introduction of the irreversibility following the ideas of Van Hove, Prigogine, and their collaborators. It is shown explicitly that the low-density limit as well as the weak-coupling limit of the viscosity are those which are expected from the conventional calculation via the Boltzmann equation. The nature of these limits is closely related to the two well separated time scales—average duration of collision and average time between successive collisions.

I. INTRODUCTION

THE general idea and method developed in the previous paper[1] is applied to the treatment of the viscosity coefficients of classical statistical gases.

The object of the present paper is not to present a new method for the explicit evaluation of the viscosity coefficients but to show how these coefficients are calculated from the exact formulas without the aid of the Boltzmann equation and by doing so to clarify the nature of the approximations inherent in the calculation.

It is hoped that the present treatment will form a prototype of the future calculation of viscosity coefficients in the case where the traditional treatment via the Boltzmann equation is no longer justified.

Let us consider a gas characterized by the Hamiltonian H

$$H\equiv\sum_{i=1}^{N}p_i^2+\tfrac{1}{2}g\sum_{i\neq j=1}^{N}\sum\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j), \tag{1.1}$$

where the vectors $\mathbf{p}_i$ and $\mathbf{r}_i$ are the momentum and the position of the ith particle and $\mathcal{U}$ is a pair-force potential with a *short range*, N the number of particles, and g the coupling constant.

The static viscosity coefficient η of the gas is, in analogy with the formula (I.1.4), given by[2]

$$\eta=\beta V^{-1}\,\mathrm{Re}\left[\lim_{T'\to\infty}\int_0^{T'}\frac{du\,\mathrm{Tr}\{e^{-\beta H}J_{xy}J_{xy}[u]\}}{\mathrm{Tr}\{e^{-\beta H}\}}\right], \tag{1.2}$$

where the current operator J_{xy} is defined by[3]

$$J_{xy}\equiv 2\sum_{i=1}^{N}p_{ix}p_{iy}-\tfrac{1}{4}g\sum_{i\neq j}^{N}\sum\left[(x_i-x_j)\frac{\partial\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j)}{\partial(y_i-y_j)}+(y_i-y_j)\frac{\partial\mathcal{U}(\mathbf{r}_i-\mathbf{r}_j)}{\partial(x_i-x_j)}\right] \tag{1.3}$$

with $\mathbf{r}\equiv(x,y,z)$. The operator $J_{xy}[u]$ is the current in the Heisenberg picture defined by

$$J_{xy}[u]\equiv e^{iuH}J_{xy}e^{-iuH}. \tag{1.4}$$

The volume of container is denoted by V.

We wish to calculate the viscosity coefficient from (1.2) in the two limiting cases— a weakly-coupled gas and a dilute gas, both having short-range pair interactions and obeying the classical statistics. There is no real gas in nature which can be treated in the category of a weakly coupled gas. However, it turns out that the

* This work was supported by the Air Research and Development Command, U. S. Air Force, through its European office.

[1] S. Fujita and R. Abe, J. Math. Phys. 3, 350 (1962). This paper is referred to as I.

[2] H. Mori, Phys. Rev. 112, 1829 (1958). More detailed references may be found in reference 1.

[3] The units are chosen in the present paper such that $\hbar=2M=1$, M being the mass of a particle, unless otherwise specified.

dilute gas, which interests us, can be treated in a manner very much similar to the weakly-coupled gas. Since the mathematical treatment of the latter is considerably simpler than that of the former, it will serve as a useful mathematical exercise for problems of the former. Using the techniques of linked-cluster expansion and assuming the weak coupling, the viscosity of a classical statistical gas is formally calculated in Sec. II. In Sec. III the viscosity coefficient of a dilute gas is treated. Effectively, the result for this case can be obtained from the formulas derived in Sec. II through a redefinition of differential cross section. It is then shown that the formal result thus obtained is completely equivalent to that which one would expect from the calculation via the Boltzmann equation. In the derivation of these results we have assumed only the following conditions:

1. The particles obey the Boltzmann statistics,
2. The system is infinitely large: $N\to\infty$, $V\to\infty$ such that $c\equiv N/V=$finite;
3. The particles are weakly coupled (for the treatment in Sec. II): $g\to 0$, $t\to\infty$ such that $g^2t=$finite, or 3′. The density of the gas is low (for the treatment in Sec. III): $c\to 0$, $t\to\infty$ such that $ct=$finite, where t is a time of the order of the relaxation time.

The conditions 2 and 3 or 2 and 3′ are those under which the Markoffian master equations were derived by Van Hove, Prigogine, and their collaborators.[4] The condition 3 or 3′ is closely related to the well-separated time scales—average duration of collision and average time between successive collisions. This point is briefly discussed in Sec. IV where a few comments on the present calculation are given.

II. VISCOSITY COEFFICIENT OF A WEAKLY COUPLED QUANTUM GAS OBEYING THE BOLTZMANN STATISTICS

We would like to treat the problem in a similar manner to the one in which we calculated the electrical conductivity in the preceding paper I. For this reason we shall formulate the problem in second quantization although we are primarily interested in the classical statistical case.

The second term of the expression in (1.3) for the current is proportional to the coupling constant g. This term will become negligible in the limit of weak coupling, the precise definition of the limit being given later in (2.14). For simplicity we shall drop it from the beginning[5]:

$$J_{xy}\to J_{xy}'=2\textstyle\sum_i p_{ix}p_{iy}. \quad \text{(approximation.)} \tag{2.1}$$

In second quantization one has for (1.1) and (2.1)

$$H=\textstyle\sum_{\mathbf{p}} p^2a^\dagger(\mathbf{p})a(\mathbf{p})+\frac{1}{2}gV^{-1}\sum_{\mathbf{p}_1\mathbf{p}_2\mathbf{q}}\mathfrak{V}(q)a^\dagger(\mathbf{p}_2+\mathbf{q})\times a^\dagger(\mathbf{p}_1-\mathbf{q})a(\mathbf{p}_1)a(\mathbf{p}_2), \tag{2.2}$$

$$J_{xy}=2\textstyle\sum_{\mathbf{p}} p_xp_ya^\dagger(\mathbf{p})a(\mathbf{p}), \tag{2.3}$$

where the scalar $\mathfrak{V}(q)$ is the Fourier transform of $\mathfrak{V}(r)$ defined by

$$\mathfrak{V}(q)\equiv\int d^3r\mathfrak{V}(r)e^{i\mathbf{q}\cdot\mathbf{r}}; \tag{2.4}$$

the creation and annihilation operator $\{a^\dagger,a\}$ satisfy the following commutation or anticommutation relations:

$$\left.\begin{aligned}[a(\mathbf{p}),a^\dagger(\mathbf{p}')]_\mp&\equiv a(\mathbf{p})a^\dagger(\mathbf{p}')\mp a^\dagger(\mathbf{p}')a(\mathbf{p})\\&=\delta^{(3)}(\mathbf{p},\mathbf{p}')\\ [a(\mathbf{p}),a(\mathbf{p}')]_\mp&=[a^\dagger(\mathbf{p}),a^\dagger(\mathbf{p}')]_\mp=0\end{aligned}\right\} \tag{2.5}$$

where the upper signs correspond to the case of bosons and the lower signs to that of fermions.

Techniques of perturbation expansion, diagrammatic representation, and linked-cluster expansion can be introduced closely following the procedure taken in I. Without duplicating the old procedure we shall note the main points, emphasizing the differences:

The current correlation function $X(u)$, which is the integrand in the expansion (1.2) has the following form:

$$\begin{aligned}X(u)&=4\sum_{\mathbf{p},\mathbf{p}_0}p_xp_yp_{0x}p_{0y}\\&\quad\times\frac{\mathrm{Tr}\{e^{\alpha N-\beta H}a^\dagger(\mathbf{p})a(\mathbf{p})S^\dagger(u)a^\dagger(\mathbf{p}_0)a(\mathbf{p}_0)S(u)\}}{\mathrm{Tr}\{e^{\alpha N-\beta H}\}}\\&\to 4\sum_{\mathbf{p},\mathbf{p}_0}p_xp_yp_{0x}p_{0y}\exp[\alpha-\beta p^2]\\&\quad\times\langle a(\mathbf{p},0)S^\dagger(u)a^\dagger(\mathbf{p}_0,u)a(\mathbf{p}_0,u)S(u)a^\dagger(\mathbf{p},0)\rangle_0,\end{aligned}$$

(approximation). (2.6)

Here e^α is the fugacity of the system and the operator $S(u)$ and its Hermitian conjugate $S^\dagger(u)$ are defined by

$$\begin{aligned}S(u)&\equiv e^{iuH_0}e^{-iuH}\\&=\sum_{n=0}^{\infty}(-ig)^n\int_0^u dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n\\&\quad\times H_I(t_1)H_I(t_2)\cdots H_I(t_n),\end{aligned} \tag{2.7a}$$

$$\begin{aligned}S^\dagger(u)&\equiv e^{iuH}e^{-iuH_0}\\&=\sum_{n=0}^{\infty}(ig)^n\int_0^u dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n\\&\quad\times H_I(t_n)\cdots H_I(t_n)H_I(t_1),\end{aligned} \tag{2.7b}$$

where $H_I(t)$ is the perturbation in the intermediate picture:

$$H_I(t)\equiv V^{-1}\textstyle\sum_{\mathbf{p}_1,\mathbf{p}_2,\mathbf{q}}\mathfrak{V}(q)a^\dagger(\mathbf{p}_2+\mathbf{q},t)a^\dagger(\mathbf{p}_1-\mathbf{q},t)\times a(\mathbf{p}_1,t)a(\mathbf{p}_2,t). \tag{2.8}$$

The time dependent operators $\{a(\mathbf{p},t),a^\dagger(\mathbf{p},t)\}$ are defined by

$$\left.\begin{aligned}a(\mathbf{p},t)&\equiv e^{itH_0}a(\mathbf{p})e^{-itH_0}\\ a^\dagger(\mathbf{p},t)&\equiv e^{itH_0}a^\dagger(\mathbf{p})e^{-itH_0}\end{aligned}\right\}. \tag{2.9}$$

[4] The references for the derivation of the master equations and the related topics may be found in reference 1.

[5] The prime is suppressed hereafter.

The average symbol $\langle\cdots\rangle_0$ means the average with the grand canonical ensemble of free particles. The approximation is the replacement of the average with respect to the ensemble of interacting particles with the average with respect to the free particle ensemble. This approximation is justified in the weak coupling limit (2.14).

The average

$$\langle a(\mathbf{p},0)S^\dagger(u)a^\dagger(\mathbf{p}_0,u)a(\mathbf{p}_0,u)S(u)a^\dagger(\mathbf{p},0))\rangle_0 \quad (2.10)$$

can be represented by Feynman diagrams as before.[1] We must, however, modify the prescription for drawing diagrams because we have now the different form (2.8) of the perturbation. A general diagram of order n will be drawn in the following way (cf. I, Sec. II):

(i) Mark two points at $t=0$ and $t=u$ on a horizontal boundary (dashed) line (the same as before).

(ii) Mark k pairs of points above the boundary line at $t=t_1, t_2, \cdots, t_k, u>t_1>t_2>\cdots>t_k>0$, exactly one pair being on the same time; also mark $n-k$ pairs of points below the boundary at $t=t_1', t_2', \cdots, t_{n-k}', u>t_1'>t_2'>\cdots>t_{n-k}'>0$.

(iii) Draw n *undirected* (wavy) lines between the pairs of points on the same times.

Draw *directed* particle (solid) lines, one entering and one leaving each point marked in (i) and (ii); these lines may run between points or from a point to itself.

A momentum variable can be attached to each particle and to each quantum line as before. The momentum value will be conserved at each corner drawn in the process (ii) because of our form (2.8) of the perturbation.

The contribution to the average (2.10) corresponding to a given diagram can be written down from the prescription similar to that given in I, Sec. II. The complete prescription for the present case will not be given here. It reproduces all the old rules except the rule (i), where the new rule replaces the regulation about the phonon lines with:

(i') For every quantum line with a momentum $\mathbf{q}$, introduce a factor

$$V^{-1}v(q). \quad (2.11)$$

The linked-cluster expansion (I.3.6) can be established in the same manner. All those unlinked diagrams which can be decomposed into two parts, one part depending on $\mathbf{p}$ and the other on $\mathbf{p}_0$, do not contribute to the current correlation function $X(u)$.

Now we can go on to analyze the various linked diagrams and to select an important set of diagrams.

First we assume that the particles obey the Boltzmann statistics. This is equivalent (i) to pick out the terms of lowest order in f, (ii) to replace the factor $1\pm f$ by 1 and (iii) to drop "exchange" diagrams.

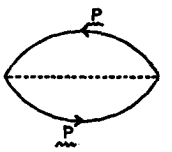


FIG. 1. The zeroth order diagram of C.

Symbolically

$$\left.\begin{array}{l} f\ll 1 \\ 1\pm f\to 1 \\ \text{exchange diagram}\to 0 \end{array}\right\}. \quad (2.12)$$

Secondly we introduce the condition that the system under consideration is infinitely large

$$N\to\infty, \quad V\to\infty \quad \text{such that} \quad c\equiv N/V=\text{finite}. \quad (2.13)$$

Finally we assume that the gas is very weakly coupled:

$$g\to 0, \quad t\to\infty \quad \text{such that} \quad g^2t=\text{finite}$$

or equivalently

$$g^2u=\text{finite}, \quad g^m(g^2u)^n\to 0 \quad m\geqq 1, \quad n=0,1,2\cdots. \quad (2.14)$$

These three conditions restrict the type of the important diagrams to a special class, say, class C. In the zeroth order there is only one diagram in C which is drawn in Fig. 1. The contribution is, of course,

$$\delta^{(3)}(\mathbf{p}-\mathbf{p}_0). \quad (2.15)$$

There exists no first-order diagram which remains finite. In the second order we have six diagrams belonging to the class C. They are shown in Figs. 2(a)-(f). The corresponding contributions can be written down

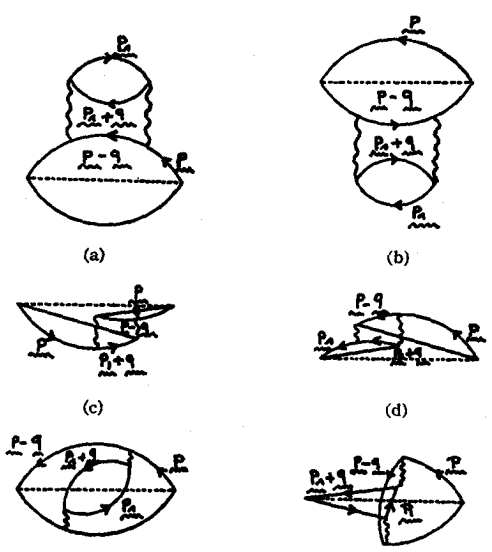


FIG. 2. The second-order diagrams of C.

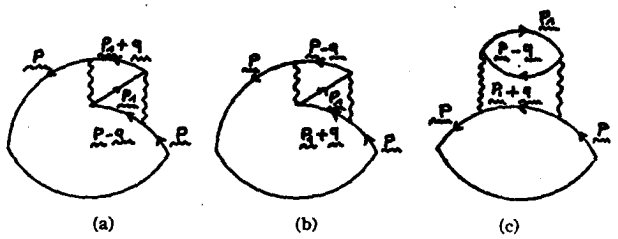


FIG. 3. Exchange diagrams.

from the new prescription. Using the asymptotic formulas

$$\int_0^u dt_1 \int_0^{t_1} dt_2 \frac{e^{\pm i\epsilon(t_1-t_2)} t_2^{p-1}}{(p-1)!} = \frac{u^p}{p!}[\pi\delta_\pm(\epsilon)] + O(u^{p-1}), \quad (2.16)$$

$$\delta_\pm(\epsilon) = \delta(\epsilon) \pm (i/\pi)P(1/\epsilon), \quad (2.17)$$

$$\int_0^u dt_1 \int_0^u dt_2 e^{\pm i\epsilon(t_1-t_2)} = u 2\pi\delta(\epsilon) + O(1), \quad (2.18)$$

we obtain the following expressions:

$$\text{a}: \quad -\delta^{(3)}(\mathbf{p}-\mathbf{p}_0)(2\pi)^{-6} f(\mathbf{p}_1) \times W_+(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \quad (2.19a)$$

$$\text{b}: \quad -\delta^{(3)}(\mathbf{p}-\mathbf{p}_0)(2\pi)^{-6} f(\mathbf{p}_1) \times W_-(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \quad (2.19b)$$

$$\text{c}: \quad +\delta^{(3)}(\mathbf{p}-\mathbf{q}-\mathbf{p}_0) 2(2\pi)^{-6} f(\mathbf{p}_1) \times W(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}, \mathbf{p}, \mathbf{p}_1) \quad (2.19c)$$

$$\text{d}: \quad -\delta^{(3)}(\mathbf{p}_1-\mathbf{p}_0)(2\pi)^{-6} f(\mathbf{p}_1) \times W_+(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \quad (2.19d)$$

$$\text{e}: \quad -\delta^{(3)}(\mathbf{p}_1-\mathbf{p}_0)(2\pi)^{-6} f(\mathbf{p}_1) \times W_-(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \quad (2.19e)$$

$$\text{f}: \quad +\delta^{(3)}(\mathbf{p}_1+\mathbf{q}-\mathbf{p}_0) 2(2\pi)^{-6} f(\mathbf{p}_1) \times W(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \quad (2.19f)$$

all integrated with respect to $\mathbf{p}_1$ and $\mathbf{q}$, and multiplied by a factor $\pi g^2 u$, where the functions $W_\pm$ and W are defined as

$$W_\pm(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \equiv \mathcal{U}(q)^2 \delta_\pm[p^2+p_1^2-(\mathbf{p}-\mathbf{q})^2-(\mathbf{p}_1+\mathbf{q})^2]. \quad (2.20)$$

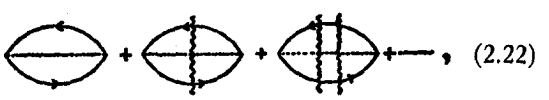
$$W \equiv \tfrac{1}{2}(W_+ + W_-) \quad (2.21)$$

All these contributions a, b, $\cdots$, f are proportional to g^2uf. Also in the second order there exist eighteen diagrams whose contributions are proportional to g^2uf. The three typical diagrams are shown in Fig. 3. They are diagrams of *exchange-scattering type* and are ruled out from C by the assumption (2.12). It can be easily verified that if the contributions corresponding to these three are added to the expression (2.19a), it would amount simply to replacing the function W_+ defined in (2.20) with

$$W_+'(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1) \equiv |\mathcal{U}(q) \pm \mathcal{U}(|\mathbf{p}_1-\mathbf{p}+\mathbf{q}|)|^2 \times \delta_+[p^2+p_1^2-(\mathbf{p}-\mathbf{q})^2-(\mathbf{p}_1+\mathbf{q})^2]. \quad (2.20a)$$

We shall postpone the discussion of these exchange scattering terms to Sec. IV.

The higher order diagrams can be analyzed in a similar manner using the asymptotic formulas (2.16) and (2.18). It is found that a general diagram of the class C has a very simple structure. The totality of these diagrams can be written down in a series:



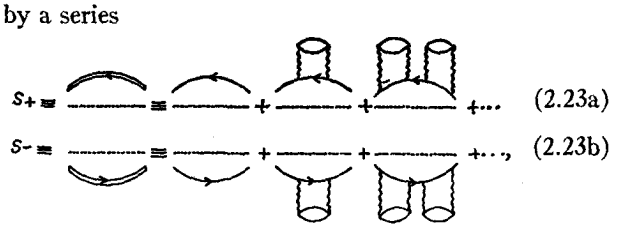
(2.22)

where each part cut by a vertical wavy line means the sum of six parts shown in Fig. 4. The first and the second term of (2.22) correspond to the diagram in Fig. 1 and the diagrams in Fig. 2, respectively.

For the purpose of calculation it is more convenient to reorder the series (2.22) in the following way.

Let us call "collision" a process in which two particle lines exchange one or more quanta between them. We notice that in the first two processes shown in Fig. 4 the particle line running from the right to the left as well as the one running in the opposite direction carries the same momentum before and after "collision" while in the next four the particle lines have different momenta before and after collision. For the latter case we shall denote the different momenta by $\mathbf{p}$ and $\mathbf{p}_0$ and designate the sum of these four processes by $w(\mathbf{p}_0, \mathbf{p})$. Mathematically this function $w(\mathbf{p}_0, \mathbf{p})$ is given by the algebraic sum of four terms (2.19c), (2.19d), (2.19e), and (2.19f) integrated with respect to $\mathbf{p}_1$ and $\mathbf{q}$.

Let us then introduce modified propagators defined by a series



(2.23a)

(2.23b)

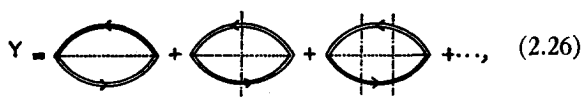
where the two ends of each diagram have the same time coordinates t_1 and t_2, $(t_1 > t_2)$. The corresponding mathematical expressions for $S_\pm$ are given by

$$S_\pm(\mathbf{p}; t_1, t_2) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!}[\pi g^2 \Gamma_\pm(\mathbf{p})(t_1-t_2)]^k \times \exp[\mp i(t_1-t_2)p^2]$$
$$= \exp[\mp i(t_1-t_2)p^2 - \pi g^2 \Gamma_\pm(\mathbf{p})(t_1-t_2)], \quad (2.24)$$

with

$$\Gamma_\pm(\mathbf{p}) \equiv (2\pi)^{-6} \int\int d^3q d^3p_1 f(\mathbf{p}_1) \times W_\pm(\mathbf{p}-\mathbf{q}, \mathbf{p}_1+\mathbf{q}; \mathbf{p}, \mathbf{p}_1). \quad (2.25)$$

If we now form an infinite series



(2.26)

where every part of the particle lines represents a modified propagator S_+ or S_- and each vertical broken line the sum of four collision processes w, this series evidently contain all terms of the series in (2.22) and vice versa. In mathematical terms, the sum $Y(\mathbf{p}_0,\mathbf{p};u)$ is given by

$$
\begin{aligned}
&Y(\mathbf{p}_0,\mathbf{p};u)\\
&=\delta^{(3)}(\mathbf{p}-\mathbf{p}_0)\exp\{-2\pi g^2 u\Gamma(\mathbf{p})\}\\
&\quad+2\pi g^2\int_0^u dt_1\int d^3p'\delta^{(3)}(\mathbf{p}'-\mathbf{p}_0)\\
&\quad\times\exp[-2\pi g^2\Gamma(\mathbf{p}')(u-t_1)]w(\mathbf{p}';\mathbf{p})\\
&\quad\times\exp[-2\pi g^2\Gamma(\mathbf{p})t_1]+(2\pi g^2)^2\\
&\quad\times\int_0^u dt_1\int_0^{t_1}dt_2\int\int d^3p'd^3p''\delta^{(3)}(\mathbf{p}'-\mathbf{p}_0)\\
&\quad\times\exp[-2\pi g^2\Gamma(\mathbf{p}')(u-t_1)]w(\mathbf{p}',\mathbf{p}'')\\
&\quad\times\exp[-2\pi g^2\Gamma(\mathbf{p}'')(t_1-t_2)]w(\mathbf{p}'',\mathbf{p})\\
&\qquad\times\exp[-2\pi g^2\Gamma(\mathbf{p})t_2]+\cdots. \qquad (2.27)
\end{aligned}
$$

If we integrate the series term by term with respect to u from 0 to infinity and denote the result by

$$\psi(\mathbf{p}_0,\mathbf{p})\equiv\int_0^\infty du Y(\mathbf{p}_0,\mathbf{p};u), \qquad (2.28)$$

we can easily see that the ψ satisfy the following integral equation:

$$
\begin{aligned}
\psi(\mathbf{p}_0,\mathbf{p})&=\delta^{(3)}(\mathbf{p}-\mathbf{p}_0)[2\pi g^2\Gamma(\mathbf{p})]^{-1}\\
&\quad+\int d^3p'\psi(\mathbf{p}_0,\mathbf{p}')w(\mathbf{p},\mathbf{p}')\Gamma(\mathbf{p})^{-1}. \qquad (2.29)
\end{aligned}
$$

Multiplying the both sides by $p_{0x}p_{0y}$ and subsequently integrating on $\mathbf{p}_0$, one obtains

$$
\begin{aligned}
\varphi(\mathbf{p})&\equiv\int d^3p_0p_{0x}p_{0y}\psi(\mathbf{p}_0;\mathbf{p})\\
&=p_xp_y[2\pi g^2\Gamma(\mathbf{p})]^{-1}\\
&\quad+\int d^3p'\varphi(\mathbf{p}')w(\mathbf{p}',\mathbf{p})\Gamma(\mathbf{p})^{-1}. \qquad (2.30)
\end{aligned}
$$

From the definition of $\Gamma(\mathbf{p})$ and $w(\mathbf{p}',\mathbf{p})$ and after the appropriate change of integration variables, this

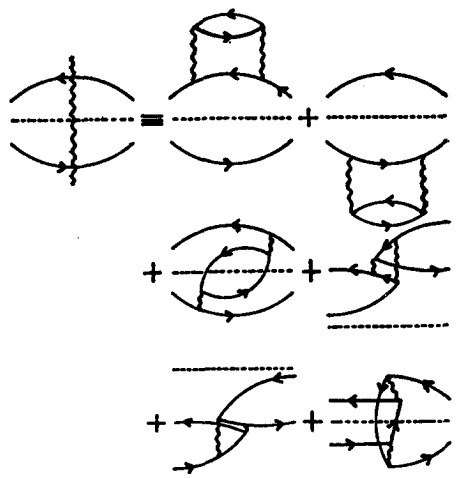


FIG. 4. Diagram components important to the present calculation.

equation may be written as

$$
\begin{aligned}
p_xp_y&=(2\pi)^{-5}g^2\int\int d^3p_1d^3q|\mathcal{U}(q)|^2f(\mathbf{p}_1)\\
&\quad\times\delta[p^2+p_1^2-(\mathbf{p}-\mathbf{q})^2-(\mathbf{p}_1+\mathbf{q})^2]\\
&\quad\times[\varphi(\mathbf{p})+\varphi(\mathbf{p}_1)-\varphi(\mathbf{p}-\mathbf{q})-\varphi(\mathbf{p}_1+\mathbf{q})]. \qquad (2.31)
\end{aligned}
$$

If we introduce the coordinates describing the motion of center of mass and the relative motion, we can rewrite the equation in the following form:

$$
\begin{aligned}
p_xp_y&=\tfrac{1}{4}(2\pi)^{-5}g^2\int d^3p_1d\Omega p_i|\mathcal{U}(\mathbf{p}_f-\mathbf{p}_i)|^2f(\mathbf{p}_1)\\
&\quad\times[\varphi(\mathbf{p})+\varphi(\mathbf{p}_1)-\varphi(\mathbf{p}-\mathbf{p}_f+\mathbf{p}_i)\\
&\qquad-\varphi(\mathbf{p}_1+\mathbf{p}_f-\mathbf{p}_i)]. \qquad (2.32)
\end{aligned}
$$

Here the momenta ($\mathbf{p}_i\equiv\frac{1}{2}(\mathbf{p}_1-\mathbf{p})$, $\mathbf{p}_f\equiv\mathbf{p}_i+\mathbf{q}$) of the relative motion before and after collision have the same magnitude and the differential element of the angle between $\mathbf{p}_i$ and $\mathbf{p}_f$ is denoted by $d\Omega$.

The viscosity coefficient η is now given from (1.2), (2.6), (2.28), and (2.30) by

$$\eta=4\beta(2\pi)^{-3}\int d^3pp_xp_yf(\mathbf{p})\varphi(\mathbf{p}). \qquad (2.33)$$

The integral equation (2.31) or (2.32) is closely related to the so-called linearized Boltzmann equation. This point will be discussed at the last part of Sec. III as well as in Sec. IV.

III. VISCOSITY COEFFICIENT OF A DILUTE GAS OBEYING THE BOLTZMANN STATISTICS

The low-density limit of the viscosity coefficient of a classical gas can be calculated almost in the same manner as the weak-coupling limit of the same.

It is interesting to note that the two approximations (2.1) and (2.6) are justified in the low-density limit as well as in the weak-coupling limit. The second term of the current J in (1.3) is of higher order in density than the first term and therefore may be neglected in the low-density limit[5]:

$$J_{xy} \to J_{xy}' = 2\sum_{\mathbf{p}} p_x p_y a^\dagger(\mathbf{p})a(\mathbf{p}) \quad \text{(approximation)}. \quad (3.1)$$

The current correlation function $X(u)$ may be written as

$$\begin{aligned} X(u) &= 4\sum_{\mathbf{p},\mathbf{p}_0} p_x p_y p_{0x} p_{0y} \\ &\times \langle S(-i\beta)a^\dagger(\mathbf{p})a(\mathbf{p})S^\dagger(u)a^\dagger(\mathbf{p}_0)a(\mathbf{p}_0)S(u)\rangle_0 / \\ &\times \langle S(-i\beta)\rangle_0 \to 4\sum_{\mathbf{p},\mathbf{p}_0} p_x p_y p_{0x} p_{0y} \exp[\alpha - \beta p^2] \\ &\times \langle a(\mathbf{p},0)S^\dagger(u)a^\dagger(\mathbf{p}_0,u)a(\mathbf{p}_0,u)S(u)a^\dagger(\mathbf{p},0)\rangle_0 \\ &\qquad \text{(approximation)}. \quad (3.2) \end{aligned}$$

The justification of this will be discussed after the introduction of the low-density limit. We then see that we have to deal with exactly the same average

$$\langle a(\mathbf{p},0)S^\dagger(u)a^\dagger(\mathbf{p}_0,u)a(\mathbf{p}_0,u)S(u)a^\dagger(\mathbf{p},0)\rangle_0$$

as before. Therefore all earlier remarks about the diagrammatic expansion are unchanged.

We now introduce the following three assumptions:

1. Boltzmann statistics

$$f \ll 1, \quad 1 \pm f \to 1 \quad (3.3)$$

Exchange term $\to 0$.

2. Infinite system

$$N \to \infty, \quad N \to \infty \quad \text{such that} \quad c \equiv N/V = \text{finite}. \quad (3.4)$$

3. Low-density limit

$$c \to 0, \quad t \to \infty \quad \text{such that} \quad ct = \text{finite}. \quad (3.5)$$

In the present case we may expect that the current correlation function $X(u)$ will decay exponentially $\exp(-u/t)$ for large u. Then the limit (3.5) is equivalent to the limit

$$\begin{aligned} c^m(cu)^n \to 0 \quad & N = 0,1,2,\cdots \quad (3.5a) \\ & m \geqq 1 \\ & cu = \text{finite}. \end{aligned}$$

The limit (3.5) or (3.5a) is to be taken *after* the limit (3.4).

It is easy to see that the approximation (3.2) is justified in the low-density limit (3.5a). Consider the difference $e^{\beta H_0} e^{-\beta H} - 1$. Every term of its expansion contains at least one potential which depends on a pair of particles. Therefore any correction which would arise from more precise treatment of the grand canonical average will contain an extra power of c and hence will not contribute in the present approximation.

Analysis of diagrams concerning their dependency on density c, time n, and volume V can be carried out in a usual manner. The analysis shows that the contribution corresponding to an arbitrary linked diagram is proportional to $V^{-k}(1-f)^l f^m (fu)^n$ with k, l, m, and n being certain non-negative integers 0, 1, 2, $\cdots$. In the limit of classical statistics the (Boltzmann) distribution function f is proportional to the density c.

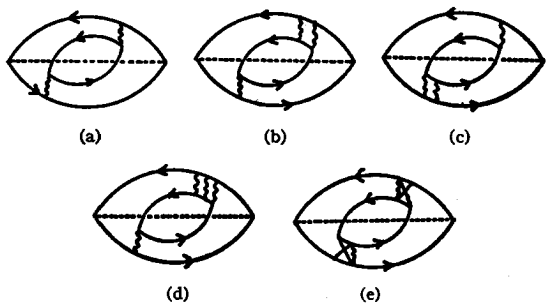
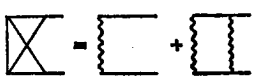


FIG. 5. Diagrams of C'.

$$f \propto c. \quad (3.6)$$

It is found then that the above mentioned three assumptions (3.3), (3.4), and (3.5a) select a special class of diagrams whose contributions are proportional to $(1-f)^m(fu)^n$ or $(cu)^n$ with m and n being non-negative integers. We shall denote this class of diagrams by C'.

Some of simple diagrams in C' are drawn in Fig. 5. Their dominant contributions are all proportional to cu. In the diagram (e) a special symbol is introduced:


(3.7)

It turns out to correspond to a sum of successive Born binary collisions. Figure 5(e) represents a double series, i.e., the sum of a, b, c, $\cdots$. We shall use this abbreviation wherever possible. Now it is a remarkable fact that if we replace every complete binary collision series in a given diagram of C' with the lowest Born approximation, i.e., the first term of the series (3.7), we would see that the resulting diagram is an element of C defined in the Sec. II. Conversely if we replace every lowest Born approximation in an arbitrary diagram of C with a complete series (3.7), we find an element of C'. Furthermore, if the pair potential $\mathcal{U}(r)$ is of short range such that

$$r\mathcal{U}(r) \to 0 \quad \text{as} \quad r \to \infty, \quad (3.8)$$

then the theory of scattering[6] shows that the above mentioned replacement of a lowest Born approximation with a complete binary collision series in a diagram, simply amounts to the replacement of transformed potential $\mathcal{U}$ with the so-called T matrix in the corresponding mathematical expressions. This matrix T may be defined as the solution of the following integral

[6] See e.g., B. A. Lipmann and J. Schwinger, Phys. Rev. 79, 469 (1950); I. Prigogine and P. Resibois, Physica 24, 795 (1958).

equation

$$T_{kl}=\mathcal{U}_{kl}+\frac{\pi i^{-1}}{(2\pi)^3}\int d^3m\,\mathcal{U}_{km}\delta_+(E_l-E_m)T_{ml}, \quad (3.9)$$

where the subscripts k, l, and m denote momenta of relative motion in the binary scattering; and $\mathcal{U}_{kl}$ and E_l are defined by

$$\mathcal{U}_{kl}\equiv\int d^3r\,\mathcal{U}(r)e^{i(\mathbf{k}-\mathbf{l})\cdot\mathbf{r}} \quad (3.10)$$

$$E_l\equiv 2l^2. \quad (3.11)$$

Equation (3.9) can be derived from the definition (3.7) of the binary collision series, using the asymptotic time integration (2.16).

Now the resulting formula for the viscosity of a dilute gas can be immediately written down from the formula for the viscosity of a weakly coupled gas. Replacing $g\mathcal{U}$ in (2.32) with T

$$g\mathcal{U}\rightarrow T, \quad (3.12)$$

one has

$$p_xp_y=(2\pi)^{-5}\tfrac{1}{4}\int d^3p_1\int d\Omega p_i|T(\mathbf{p}_f-\mathbf{p}_i)|^2f(\mathbf{p}_1)$$
$$\times[\varphi(\mathbf{p})+\varphi(\mathbf{p}_1)-\varphi(\mathbf{p}-\mathbf{p}_f+\mathbf{p}_i)$$
$$-\varphi(\mathbf{p}_1+\mathbf{p}_f-\mathbf{p}_i)] \quad (3.13)$$

with $p_f=p_i$.

If we use the well-known formula[7] which connects the matrix T with the differential cross section σ (quantum mechanical)

$$\sigma(p_i,\theta_{fi})\delta(E_{\mathbf{p}_f}-E_{\mathbf{p}_i})=|T(\mathbf{p}_f,\mathbf{p}_i)|^2(64\pi^2)^{-1}$$
$$\times\delta(E_{\mathbf{p}_f}-E_{\mathbf{p}_i}) \quad (3.14)$$

with θ_{fi} being the angle between $\mathbf{p}_i$ and $\mathbf{p}_f$, we can rewrite (3.13) as

$$p_xp_y=4(2\pi)^{-3}\int d^3p_1\int d\Omega p_i\sigma(p_i,\theta_{if})f(\mathbf{p}_1)$$
$$\times[\varphi(\mathbf{p})+\varphi(\mathbf{p}_1)-\varphi(\mathbf{p}-\mathbf{p}_f+\mathbf{p}_i)$$
$$-\varphi(\mathbf{p}_1+\mathbf{p}_f-\mathbf{p}_i)]. \quad (3.15)$$

The viscosity coefficient η is given in terms of φ as before [cf. (2.33)]

$$\eta=4\beta(2\pi)^{-3}\int d^3p\,p_xp_y f(\mathbf{p})\varphi(\mathbf{p}). \quad (3.16)$$

We shall show in the following that the results (3.15) and (3.16) can be obtained simply from the calculation via the Boltzmann equation.

[7] See, e.g., S. S. Schweber, H. A. Bethe, and F. de Hoffmann, *Mesons and Fields* (Row, Peterson & Company, Evanston, Illinois, 1955), Vol. I, p. 46.

The Boltzmann equation for the distribution function $F(\mathbf{r},\mathbf{p},t)$ of a dilute gas subject to no external force is written as

$$\frac{\partial F}{\partial t}+2\mathbf{p}\cdot\frac{\partial F}{\partial\mathbf{r}}=4\int\int d^3p_1d\Omega(F'F_1'-FF_1)b_i\sigma(p_i,\theta) \quad (3.17)$$

with the abreviations

$$F\equiv F(\mathbf{r},\mathbf{p},t) \quad F'\equiv F(\mathbf{r},\mathbf{p}',t)$$
$$F_1\equiv F(\mathbf{r},\mathbf{p}_1,t) \quad F_1'\equiv F(\mathbf{r},\mathbf{p}_1',t), \quad (3.18)$$

where the pairs $(\mathbf{p},\mathbf{p}_1)$ and $(\mathbf{p}',\mathbf{p}_1')$ are, respectively, the momenta before and after collision.

For simplicity let us consider a stationary situation in which a gas moves along the x axis with an average velocity $\langle v_x\rangle_{\rm av}=2\langle p_x\rangle_{\rm av}$ depending only on y and in which the gas has the same temperature everywhere. Here the average of a momentum dependent function $A(\mathbf{p})$ is defined by

$$\langle A\rangle_{\rm av}\equiv c^{-1}\int d^3pF(\mathbf{r},\mathbf{p},t)A(\mathbf{p}). \quad (3.19)$$

When the velocity gradient $d\langle v_x\rangle_{\rm av}/dy$ is assumed to be sufficiently small, the viscosity coefficient η can be defined as a proportionality factor in the following equation:

$$cM^{-1}\langle p_xp_y\rangle_{\rm av}=-\eta\frac{d\langle v_x\rangle_{\rm av}}{dy}=2c\langle p_xp_y\rangle_{\rm av} \quad (3.20)$$

with M being mass of a particle. In the same condition the drift term, i.e., the right-hand side of (3.17) will become

$$-2f\beta p_xp_yd\langle v_x\rangle_{\rm av}/dy. \quad (3.21)$$

If the velocity gradient does not exist, the function F agrees with the Boltzmann distribution function f. Therefore the difference $|F-f|$ will be small in comparison with f as long as the gradient is small. Hence we may put

$$F(\mathbf{p},\mathbf{r})\equiv f(\mathbf{p})\left[1+\phi(p)\frac{d\langle v_x\rangle_{\rm av}(y)}{dy}\right]. \quad (3.22)$$

Substituting this into the right-hand side of (3.17) and neglecting the second order $(d\langle v_x\rangle_{\rm av}/dy)^2$ in the result, one has

$$4\int\int dp_1d\Omega f(\mathbf{p})f(\mathbf{p}_1)[\phi(\mathbf{p}')+\phi(\mathbf{p}_1')-\phi(\mathbf{p})-\phi(\mathbf{p}'_1)]$$
$$\times p_i\sigma(p_i,\theta)\frac{d\langle v_x\rangle_{\rm av}}{dy}. \quad (3.23)$$

Equating this with (3.21) and dropping the common

factor, one obtains

$$\beta p_i p_\nu = 2\int\int dp_1 d\Omega f(\mathbf{p}_1) p_i \sigma(p_i,\theta) \times[\phi(\mathbf{p})+\phi(\mathbf{p}_1)-\phi(\mathbf{p}')-\phi(\mathbf{p}_1')]. \quad (3.24)$$

This is sometimes referred to as the linearized Boltzmann equation. From (3.20) and (3.22) one sees that the viscosity coefficient η is given by

$$\eta = -2\int d^3p\, p_x p_\nu f(\mathbf{p})\phi(\mathbf{p}). \quad (3.25)$$

Comparing (3.24) and (3.15) one obtains

$$\phi(\mathbf{p}) = (4\pi^3)^{-1}\beta\varphi(\mathbf{p}). \quad (3.26)$$

With this relation we can immediately see that the calculation via (3.24) and (3.25) gives the viscosity identical to that given by (3.15) and (3.16).

In the Sec. II we made an attempt to calculate the viscosity of a weakly coupled gas obeying the Boltzmann statistics. The result is given by the expressions (2.32) and (2.33). It is quite clear that, following the same procedure just taken above, we can discuss this result in connection with the solution of the linearized Boltzmann equation.

IV. COMMENTS AND DISCUSSIONS

We shall make several comments and discussions on our calculation of the viscosity coefficients in the preceding two sections.

1. We have assumed that the particles obey the Boltzmann statistics. This does not simply mean that we have treated a completely classical gas. In fact, the differential cross section introduced in (3.14) is a quantum-mechanical one. The viscosity coefficient given by (2.32) and (2.33) is also quantum mechanical.

It is well known[8] that the quantum-mechanical diffraction effect is not negligible in binary collision theory if the de Broglie wavelength

$$\lambda_\beta \equiv 2\pi\beta^{\frac{1}{2}} \quad (4.1)$$

associated with temperature $(k\beta)^{-1}$ is comparable to, or greater than, the molecular dimension characterized, for example, by the range of force R:

$$\lambda_\beta \gtrsim R. \quad (4.2)$$

The de Broglie wavelength λ_β can be considered as the measure of uncertainty associated with the linear position of each particle of the gas with temperature $(k\beta)^{-1}$. In the temperature region specified by (4.2) we must calculate the cross section quantum mechanically. There arises another sort of modification necessitated in the same region. If a pair of colliding particles are identical as we have assumed in our calculation, they would not in general be distinguishable from each other around the closest approach ($\sim R$), which is comparable to or even smaller than the de Broglie wavelength λ_β. This modifies the collision theory. In our theory the modification corresponds to the retention of all the diagrams of exchange type such as shown in Fig. 3. The result amounts simply to redefine the differential cross section σ [cf. (2.20a)]

$$\sigma(p_i,\theta) = (64\pi^2)^{-1}|T(p_i,\theta) \pm T(p_i,\pi-\theta)|^2 \quad (4.3)$$

for the case of a dilute gas where θ is the angle between $\mathbf{p}_i$ and $\mathbf{p}_f$ [c.f. (3.14)]. For the case of a weakly-coupled gas the T is to be replaced by a transformed potential $\mathcal{U}$ in the above formula.

2. As mentioned before the viscosity given by (2.32) and (2.33) is quantum mechanical. In the classical mechanics an interaction potential can introduce only an infinitesimal amount of momentum transfer[9] between the colliding particles in the lowest order in g. Therefore the Boltzmann collision term $\partial F/\partial t)_{\text{coll.}}$ for a weakly-coupled gas has, instead of the usual difference form, a differential form[10]

$$\left(\frac{\partial F}{\partial t}\right)_{\text{coll.}} = (8\pi^2)^{-1}\int d^3p_1 \int d^3q\, |\mathcal{U}(\mathbf{q})|^2 \times \mathbf{q}\cdot\frac{\partial}{\partial \mathbf{p}}\delta[2\mathbf{q}\cdot(\mathbf{p}-\mathbf{p}_1)] \times \mathbf{q}\cdot\left(\frac{\partial}{\partial \mathbf{p}}-\frac{\partial}{\partial \mathbf{p}_1}\right)F(\mathbf{r},\mathbf{p},t)F(\mathbf{r},\mathbf{p}_1,t). \quad (4.4)$$

On the other hand, in quantum mechanics a finite amount of momentum transfer is allowed between the particles even in the lowest Born approximation. The collision term then has a usual difference form

$$\left(\frac{\partial F}{\partial t}\right)_{\text{coll.}} = (4\pi^2)^{-1}\int d^3p_1 \int d^3q\, |\mathcal{U}(\mathbf{q})|^2 \times\delta[p^2+p_1^2-(\mathbf{p}-\mathbf{q})^2-(\mathbf{p}_1+\mathbf{q})^2] \times[F(\mathbf{r},\mathbf{p}-\mathbf{q},t)F(\mathbf{r},\mathbf{p}_1+\mathbf{q},t) - F(\mathbf{r},\mathbf{p},t)F(\mathbf{r},\mathbf{p}_1,t)]. \quad (4.5)$$

The Boltzmann drift term is conventionally assumed to have the same form for both the classical and the quantum-mechanical case. It is clear that the integral equation (2.32) is essentially equivalent to the linearized Boltzmann equation with the quantum-mechanical collision term (4.5) separate from the classical collision term (4.4).

3. It is also well known that if the temperature is further lowered from the temperature region discussed

[8] See, e.g., E. A. Uehling, Phys. Rev. 46, 917 (1934).

[9] For interesting discussions on this point, see I. Prigogine and S. Ono, Physica 25, 171 (1959).

[10] I. Prigogine and R. Balescu, Physica 26, 145 (1960).

in 1 so that the de Broglie wavelength λ_β is comparable to, or greater than, the average molecular separation characterized for example by $c^{-\frac{1}{3}}$:

$$\lambda_\beta \gtrsim c^{-\frac{1}{3}} \gg R, \tag{4.6}$$

the statistics effect (or exchange effect) becomes important. In this temperature region, we must treat the problem with the assumption that the particles obey the Bose or the Fermi statistics. The usual Uehling-Uhlenbeck equation is not fully justified, for the statistics effect of the intermediate states is not dully taken into account. The treatment for this quantum statistical region will be reported in a subsequent paper.

4. The main result (3.15) and (3.16) in Sec. III was obtained on the four assumptions (3.3), (3.4), (3.5), and (3.8). The assumption (3.3) about the statistics was just discussed in 1 and 3. We shall not discuss here the assumptions (3.4) and (3.8) as they are clear and clean physically. The assumption (3.5) concerns with the presupposed form of the relaxation of the system. The elementary dynamic theory of gases tells us that if a gas of many particles is coupled with strong but short range forces and is very dilute, its relaxation time is proportional to c^{-1}. This knowledge made us use the assumption (3.5). From looking at the chosen diagrams consistent with (3.5) we can see that the rule (3.5a) [equivalent to (3.5)] is completely equivalent to the assumption of instantaneous collision. In the actual dilute gas we know that

$$t_c \ll t, \tag{4.7}$$

t_c being the average collision time. It is very interesting that the assumption (3.5) as well as the assumption of instantaneous collision are the limiting case of the above strong inequality.

ACKNOWLEDGMENTS

The author wishes to thank Professor I. Prigogine, Professor J. McLennan, Jr., and Dr. P. Resibois for valuable discussions for the material of this paper.

The author wishes to express his sincere thanks to Dr. R. Balescu for enlightening discussions and his kind reading of the manuscripts.

Non-Rectilinear Current Flow in a Straight Conductor

A. C. PIPKIN AND R. S. RIVLIN
Brown University, Providence, Rhode Island
(Received August 22, 1961)

The flow of current in a long, straight, isotropic, conducting rod of uniform, noncircular cross section, due to a potential difference between its ends, is considered. The electric field in the material is assumed to depend linearly on the current density and to be quadratic in the magnetic field, while the magnetic permeability of the material is assumed constant. It is shown in the case of a rod of elliptical cross section that the current has a circulatory component superposed on the rectilinear flow.

1. INTRODUCTION

IT has been pointed out[1] that in isotropic materials with general galvanomagnetic properties, rectilinear flow of current is possible only in exceptional circumstances. In this paper we will consider the steady flow of electrical current in infinitely long rods of uniform cross section, composed of materials which obey certain particularly simple constitutive equations. We will show that in rods of elliptical cross section, in particular, the current has a circulatory component superimposed on the expected rectilinear flow.

If the electric field strength $\mathbf{E}$ and the magnetic induction $\mathbf{B}$ in a holohedral isotropic material are functions of the electric current density $\mathbf{J}$ and the magnetic field strength $\mathbf{H}$, then the functional relations must have the forms

$$\mathbf{E}=a_1\mathbf{J}+a_2(\mathbf{J}\cdot\mathbf{H})\mathbf{H}+a_3\mathbf{J}\times\mathbf{H} \tag{1.1}$$

and

$$\mathbf{B}=c_1\mathbf{H}+(\mathbf{J}\cdot\mathbf{H})(c_2\mathbf{J}+c_3\mathbf{J}\times\mathbf{H}), \tag{1.2}$$

where the scalar coefficients $a_1, a_2, \cdots, c_3$ are functions of

$$\mathbf{J}\cdot\mathbf{J},\ \mathbf{H}\cdot\mathbf{H}, \text{ and } (\mathbf{J}\cdot\mathbf{H})^2. \tag{1.3}$$

In circular rods of such materials, the electric current can flow along the rod in straight lines, no matter what the coefficients $a_1, a_2, \cdots, c_3$ may be. In rods of arbitrary cross section, however, rectilinear flow of current is possible only if these coefficients satisfy the relation

$$c_1=0 \quad \text{or} \quad \frac{\partial(\mathbf{J}\cdot\mathbf{J}a_1^2, \mathbf{J}\cdot\mathbf{J}a_3^2/c_1^2)}{\partial(\mathbf{J}\cdot\mathbf{J}, \mathbf{H}\cdot\mathbf{H})}=0 \tag{1.4}$$

whenever $\mathbf{J}\cdot\mathbf{H}=0$ (see Appendix).

To illustrate these facts and to show the nature of the flow when rectilinear flow is not possible, we will consider materials for which the following simplified forms of (1.1) and (1.2) are valid:

$$\mathbf{E}=(a_{10}+a_{11}\mathbf{H}\cdot\mathbf{H})\mathbf{J}+a_2(\mathbf{J}\cdot\mathbf{H})\mathbf{H}+a_3\mathbf{J}\times\mathbf{H}, \tag{1.5}$$

$$\mathbf{B}=\mu\mathbf{H}. \tag{1.6}$$

Here a_{10}, a_{11}, a_2, a_3, and μ are constants. These forms are obtained from (1.1) and (1.2) by neglecting terms nonlinear in $\mathbf{J}$ and terms of higher than the second degree in $\mathbf{H}$. A relation of the form (1.4) holds among the coefficients in (1.5) and (1.6) only if either a_{11}, a_3, or μ is equal to zero.

[1] A. C. Pipkin and R. S. Rivlin, J. Math. Phys. 2, 636 (1961); 2, 865 (1961).

2. STEADY CURRENTS IN LONG CYLINDRICAL RODS

We consider an infinitely long cylindrical rod lying with its axis in the z^* direction of a system of rectangular Cartesian coordinates x^*, y^*, z^*. The cross section of the rod is bounded by a curve C which is not necessarily circular. The rod is composed of a homogeneous conducting material for which the constitutive equations (1.5) and (1.6) are valid, and the medium surrounding the rod is an insulator of infinite extent, for which the relevant constitutive equations are

$$\mathbf{J}=0 \quad \text{and} \quad \mathbf{B}=\mu_0\mathbf{H}. \tag{2.1}$$

For steady flow of current in the rod, Maxwell's equations take the form

$$\nabla^*\times\mathbf{H}=\mathbf{J}, \tag{2.2}$$

$$\nabla^*\cdot\mathbf{B}=0, \tag{2.3}$$

and

$$\nabla^*\times\mathbf{E}=0, \tag{2.4}$$

where ∇^* denotes the operator $(\partial/\partial x^*, \partial/\partial y^*, \partial/\partial z^*)$. The tangential component of $\mathbf{H}$ and the normal component of $\mathbf{B}$ are continuous across the boundary of the conductor. $\mathbf{H}$ is assumed to approach zero suitably fast as distance from the conductor increases.

If rectilinear flow of current ($J_z=J_y=0$) is not possible, the simplest alternative possibility is that the current distribution is independent of position along the rod, i.e., $\partial\mathbf{J}/\partial z^*=0$. Equation (2.2) can then be satisfied by a magnetic field which is also independent of z^*. With these conditions on $\mathbf{J}$ and $\mathbf{H}$, it follows from the constitutive equations that $\partial\mathbf{B}/\partial z^*=0$, and that $\partial\mathbf{E}/\partial z^*=0$ in the conductor.

Let L be a characteristic linear dimension of the conductor cross section, and let x, y, and z be dimensionless coordinates:

$$x=x^*/L, \quad y=y^*/L, \quad z=z^*/L, \quad \nabla=L\nabla^*. \tag{2.5}$$

With $\partial\mathbf{E}/\partial z=0$, it follows from (2.4) that $\mathbf{E}$ is of the

form

$$\mathbf{E}=E[\mathbf{k}-\nabla V(x,y)], \tag{2.6}$$

where $\mathbf{k}$ is a unit vector in the z direction. The axial component of the field is uniform, with a strength E which is regarded as given.

From $\partial\mathbf{B}/\partial z=0$ and Eq. (2.3), it follows that

$$\mathbf{B}=(\mu LE/a_{10})[\mathbf{k}B(x,y)+\mathbf{k}\times\nabla\psi(x,y)]. \tag{2.7}$$

The factor $\mu LE/a_{10}$ is introduced in order to make B and ψ dimensionless; we assume that neither μ nor a_{10} is zero. The curves $\psi=$ constant are the projections on the xy plane of the lines of magnetic flux. The condition that the normal component of $\mathbf{B}$ is continuous across C can be satisfied with no loss of generality by requiring that ψ be continuous.

From (1.6), (2.1), and (2.7) we obtain

$$\begin{aligned}\mathbf{H}&=(LE/a_{10})[\mathbf{k}B+\mathbf{k}\times\nabla\psi] \text{ inside } C,\\ &\quad(\mu/\mu_0)(LE/a_{10})[\mathbf{k}B+\mathbf{k}\times\nabla\psi] \text{ outside } C.\end{aligned} \tag{2.8}$$

Since $\mathbf{J}=0$ outside C, it follows from (2.2) and (2.8) that $B=0$ outside C, provided that $B=0$ at infinity. Continuity of the tangential component of $\mathbf{H}$ across C then requires that

$$B|_{\text{inside}}=0 \text{ on } C, \tag{2.9}$$

and

$$\left.\frac{\partial\psi}{\partial n}\right|_{\text{inside}}=\frac{\mu}{\mu_0}\left.\frac{\partial\psi}{\partial n}\right|_{\text{outside}} \text{ on } C. \tag{2.10}$$

Here $\partial/\partial n$ denotes the derivative in the direction normal to C.

With (2.2) and (2.8), the condition that $\mathbf{J}=0$ outside C implies not only that $B=0$, but also that

$$\nabla^2\psi=0 \text{ outside } C. \tag{2.11}$$

Since $\mathbf{H}=O(1/r)$ for r large, where r denotes distance from the conductor, it follows that $\psi=O(\ln r)$ for r large.

Inside the conductor, it follows from (2.2), (2.5), and (2.8) that

$$\mathbf{J}=(E/a_{10})(\nabla B\times\mathbf{k}+\mathbf{k}\nabla^2\psi). \tag{2.12}$$

The portion of the magnetic field accounted for by ψ is associated with the axial component of the current. The axial field B is associated with the circulatory component of $\mathbf{J}$. The curves $B=$ constant are the projections on the xy plane of the flow lines of the electrical current. Since $B=0$ on the conductor boundary, the current automatically satisfies the condition that it be tangential to the boundary.

By using Eqs. (2.6), (2.8), and (2.12) in Eq. (1.5), we obtain

$$\begin{aligned}1=(1+\alpha\epsilon^2\nabla\psi\cdot\nabla\psi+\alpha\epsilon^2B^2)\nabla^2\psi+\beta\epsilon^2(B\nabla^2\psi-\nabla B\cdot\nabla\psi)B\\ -\epsilon\mathbf{k}\cdot(\nabla B\times\nabla\psi)\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}-\nabla V=(1+\alpha\epsilon^2\nabla\psi\cdot\nabla\psi+\alpha\epsilon^2B^2)\nabla B\times\mathbf{k}\\ +\beta\epsilon^2(B\nabla^2\psi-\nabla B\cdot\nabla\psi)\mathbf{k}\times\nabla\psi\\ -\epsilon(\nabla^2\psi\nabla\psi+B\nabla B),\end{aligned} \tag{2.14}$$

where ϵ, α, and β are dimensionless parameters defined by

$$\epsilon=a_3EL/a_{10}^2,\quad \alpha=a_{10}a_{11}/a_3^2,\quad \beta=a_{10}a_2/a_3^2. \tag{2.15}$$

The potential V can be eliminated from (2.14) by taking the curl of both sides.

Equations (2.13) and (2.14) are the governing equations for ψ and B inside the conductor. Outside the conductor, $B=0$ and $\nabla^2\psi=0$. ψ is continuous across the boundary C, and (2.9) and (2.10) are satisfied on C. At infinity, $\psi=O(\ln r)$. It appears that the problem of finding ψ and B is well-set.

3. SERIES EXPANSION

The parameters α and β defined in (2.15) depend only on properties of the material. The parameter ϵ, however, is proportional to EL, and can be made as small as desired by using a weak electric field or a conductor of small cross section. To solve Eqs. (2.13) and (2.14), we assume that $|\epsilon|\ll 1$, and expand ψ, B, and V in powers of ϵ. Some preliminary analysis indicates that these expansions are of the forms

$$\psi=\sum_{n=0}^{\infty}\epsilon^{2n}\psi_n,\quad V=\epsilon\sum_{n=0}^{\infty}\epsilon^{2n}V_n,\quad B=\epsilon^3\sum_{n=0}^{\infty}\epsilon^{2n}B_n. \tag{3.1}$$

When the expansions (3.1) are used in Eq. (2.13), we obtain

$$\nabla^2\psi_0=1, \tag{3.2}$$

$$\nabla^2\psi_1=-\alpha(\nabla\psi_0\cdot\nabla\psi_0)\nabla^2\psi_0, \tag{3.3}$$

$$\begin{aligned}\nabla^2\psi_2=-\alpha(2\nabla\psi_1\cdot\nabla\psi_0\nabla^2\psi_0+\nabla\psi_0\cdot\nabla\psi_0\nabla^2\psi_1)\\ +\mathbf{k}\cdot(\nabla B_0\times\nabla\psi_0),\end{aligned} \tag{3.4}$$

and so on. $\nabla^2\psi_n$ is given in terms of ψ_i, $i\leq n-1$, and B_j, $j\leq n-2$.

The equations obtained by using (3.1) in (2.14) are

$$\nabla V_0=\nabla^2\psi_0\nabla\psi_0, \tag{3.5}$$

$$\nabla V_1=\mathbf{k}\times\nabla B_0+\nabla^2\psi_0\nabla\psi_1+\nabla^2\psi_1\nabla\psi_0, \tag{3.6}$$

$$\begin{aligned}\nabla V_2=\mathbf{k}\times\nabla B_1+\nabla^2\psi_0\nabla\psi_2\\ +\nabla^2\psi_1\nabla\psi_1+\nabla^2\psi_2\nabla\psi_0+\alpha(\nabla\psi_0\cdot\nabla\psi_0)\mathbf{k}\times\nabla B_0\\ +\beta(B_0\nabla^2\psi_0-\nabla B_0\cdot\nabla\psi_0)\mathbf{k}\times\nabla\psi_0,\end{aligned} \tag{3.7}$$

and so forth. Taking the curl of Eqs. (3.6) and (3.7), and using (3.2) and the fact that $\partial B_0/\partial z=0$, we obtain

$$\nabla^2B_0=\mathbf{k}\cdot[\nabla\psi_0\times\nabla(\nabla^2\psi_1)] \tag{3.8}$$

$$\begin{aligned}\nabla^2B_1=\mathbf{k}\cdot[\nabla\psi_0\times\nabla(\nabla^2\psi_2)]\\ +\mathbf{k}\cdot[\nabla\psi_1\times\nabla(\nabla^2\psi_1)]-\alpha\nabla\cdot[(\nabla\psi_0\cdot\nabla\psi_0)\nabla B_0]\\ -\beta\nabla\cdot[(B_0\nabla^2\psi_0-\nabla B_0\cdot\nabla\psi_0)\nabla\psi_0].\end{aligned} \tag{3.9}$$

In general, $\nabla^2 B_n$ is expressed in terms of ψ_i, $i \leq n+1$ and B_j, $j \leq n-1$. However, $\nabla^2 B_n$ depends on ψ_{n+1} only through $\nabla^2 \psi_{n+1}$, which in turn can be expressed in terms of ψ_i, $i \leq n$ and B_j, $j \leq n-1$. Hence $\nabla^2 B_n$ can be expressed in terms of ψ_i, $i \leq n$ and B_j, $j \leq n-1$.

Outside the conductor, ψ_n satisfies $\nabla^2 \psi_n = 0$. The boundary conditions on ψ_n and B_n are the same as those on ψ and B, respectively.

The functions ψ_n and B_n can be found in the order ψ_0, B_0, ψ_1, B_1, etc. At each stage ψ_n, or B_n, satisfies a Poisson equation whose inhomogeneous part is given in terms of functions already determined. V_n can be determined, if it is desired to do so, when ψ_n and B_{n-1} are known.

We summarize, in particular, the conditions satisfied by ψ_0, V_0, and B_0: ψ_0 satisfies the equation

$$\nabla^2 \psi_0 = \begin{matrix} 1 \text{ inside } C, \\ 0 \text{ outside } C. \end{matrix} \tag{3.10}$$

ψ_0 is continuous across C, and its normal derivative satisfies the condition

$$\left.\frac{\partial \psi_0}{\partial n}\right|_{\text{inside}} = \frac{\mu}{\mu_0} \left.\frac{\partial \psi_0}{\partial n}\right|_{\text{outside}} \quad \text{on } C. \tag{3.11}$$

At infinity, $\psi_0 = O(\ln r)$. V_0 is given directly by Eq. (3.5), if (3.2) is used:

$$V_0 = \psi_0 + \text{constant inside } C. \tag{3.12}$$

By using (3.2) in (3.3), and the latter in turn in Eq. (3.8), we obtain

$$\nabla^2 B_0 = -\alpha \mathbf{k} \cdot [\nabla \psi_0 \times \nabla (\nabla \psi_0 \cdot \nabla \psi_0)]. \tag{3.13}$$

On and outside C, $B_0 = 0$.

4. APPARENT AXIAL CONDUCTIVITY

According to Eq. (2.12), the total flow of current in the axial direction is given by

$$J = (E/a_{10}) \int\int \nabla^2 \psi \, dx^* dy^* = (EL^2/a_{10}) \int\int \nabla^2 \psi \, dx dy, \tag{4.1}$$

where the integral extends over the cross section of the conductor. The cross-sectional area is $L^2 A$, where

$$A = \int\int dx dy. \tag{4.2}$$

The apparent axial conductivity, which would be obtained by measuring the total current J as a function of E, is J/EL^2A.

Retaining only the first two terms of the expansion of ψ in powers of ϵ, we obtain from (4.1) the following expression for the apparent conductivity:

$$\begin{aligned} J/EL^2A &= (1/a_{10}A) \int\int [\nabla^2 \psi_0 + \epsilon^2 \nabla^2 \psi_1] dx dy \\ &= (1/a_{10}A) \int\int [1 - \alpha \epsilon^2 \nabla \psi_0 \cdot \nabla \psi_0] dx dy \\ &= (1/a_{10}) [1 - (\alpha \epsilon^2/A) \int\int \nabla \psi_0 \cdot \nabla \psi_0 dx dy]. \end{aligned} \tag{4.3}$$

The second line of (4.3) is obtained by using (3.2) and (3.3), and the third by using (4.2).

The apparent conductivity deviates from $1/a_{10}$ by a term proportional to $\alpha \epsilon^2$, which, from (2.15), is equal to $(EL)^2 a_{11}/a_{10}^3$. Because of the factor E^2, the apparent conductivity varies as a function of field strength even though $\mathbf{E}$ and $\mathbf{J}$ are linearly related in the constitutive equation (1.5). The apparent conductivity also depends on the area of the conductor cross section, through the factor L^2, and on its shape.

5. EXAMPLE: ELLIPTICAL CROSS SECTION

Let the curve C bounding the conductor be an ellipse:

$$(x/R)^2 + (y/S)^2 = 1, \quad R > S. \tag{5.1}$$

The solution of Eq. (3.10) under the specified boundary conditions is then given inside C by

$$\psi_0 = \tfrac{1}{2} \frac{R^2 S^2}{R^2 + S^2} \left(\frac{x^2}{R^2} + \frac{y^2}{S^2} \right) + a(x^2 - y^2), \tag{5.2}$$

and outside C by

$$\begin{aligned} \psi_0 = \tfrac{1}{2} \frac{R^2 S^2}{R^2 + S^2} + \tfrac{1}{2} \frac{\mu_0}{\mu} RS(\xi - \xi_0) \\ + \tfrac{1}{2} a [R^2 - S^2 + (R^2 + S^2) e^{-2(\xi - \xi_0)} \cos 2\eta], \end{aligned} \tag{5.3}$$

where ξ and η are elliptical coordinates such that

$$\begin{aligned} x &= (R^2 - S^2)^{\frac{1}{2}} \cosh \xi \cos \eta, \\ y &= (R^2 - S^2)^{\frac{1}{2}} \sinh \xi \sin \eta. \end{aligned} \tag{5.4}$$

$\xi = \xi_0 = \tanh^{-1}(S/R)$ is the curve C. The constant a in Eqs. (5.2) and (5.3) has the value

$$a = \tfrac{1}{2} RS \frac{R^2 - S^2}{R^2 + S^2} [2RS + (\mu/\mu_0)(R^2 + S^2)]^{-1}. \tag{5.5}$$

When the expression (5.2) is used in Eq. (3.13), the equation for B_0 is found to be

$$\nabla^2 B_0 = Dxy, \tag{5.6}$$

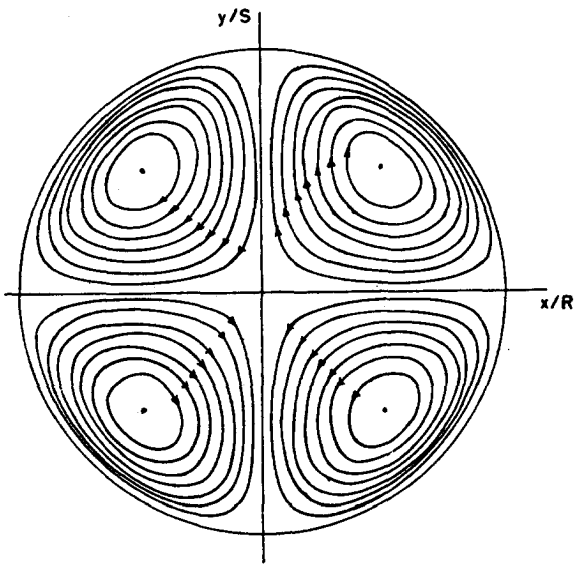


FIG. 1. Streamlines for circulatory current in elliptical cylinder of semiaxes R and S.

where

$$D=2\alpha\left(4a-\frac{R^2-S^2}{R^2+S^2}\right)\left(\frac{R^2}{R^2+S^2}-2a\right)\left(\frac{S^2}{R^2+S^2}+2a\right). \quad (5.7)$$

The solution of Eq. (5.6) with $B_0=0$ on the boundary (5.1) is

$$B_0=\tfrac{1}{6}D\frac{R^2S^2}{R^2+S^2}xy\left(\frac{x^2}{R^2}+\frac{y^2}{S^2}-1\right). \quad (5.8)$$

If B is not identically zero, then from (2.12) it is seen that a circulatory component of current is superimposed on the axial flow. If $\epsilon=0$ in (3.1), B vanishes and rectilinear flow is obtained. Assuming that $EL\neq 0$, then from (2.15), ϵ vanishes only if $a_3=0$. From (5.8), B_0 vanishes if $D=0$. From (5.7) and (5.5) it is found that $D=0$ only if $\alpha=0$, $\mu=0$, $R=S$, $\mu/\mu_0=-S/R$, or $\mu/\mu_0=-R/S$. Assuming $a_{10}\neq 0$, then from (2.15), $\alpha=0$ implies that $a_{11}=0$. The case $R=S$ is that of a circular cross section. The cases $R=S$, $a_{11}=0$, $a_3=0$, and $\mu=0$ are the cases mentioned in Sec. 1 for which rectilinear flow is possible. The two remaining cases, in which μ/μ_0 is negative, are of no physical interest.

As was mentioned earlier, the curves $B=$constant are the streamlines for the circulatory component of the electrical current. The curves $B_0=$constant obtained from Eq. (5.8) are shown in Fig. 1, with arrows to indicate the direction of flow when D is positive.

ACKNOWLEDGMENT

The work described in this paper was carried out under a grant from the National Science Foundation. We gratefully acknowledge their support.

APPENDIX

It has been shown[1] that rectilinear flow of current is possible in rods of arbitrary cross section only if the scalar coefficients in (1.1) and (1.2) satisfy the following relation whenever $\mathbf{J}\cdot\mathbf{H}=0$:

$$c_1\frac{d(Ja_3)}{d(H^2)}=Ja_3\frac{dc_1}{d(H^2)}. \quad (A.1)$$

Here $J=(\mathbf{J}\cdot\mathbf{J})^{\frac{1}{2}}$, $H^2=\mathbf{H}\cdot\mathbf{H}$, and

$$\frac{d}{d(H^2)}=\frac{\partial}{\partial(H^2)}-\left[\frac{\partial(J^2a_1^2)}{\partial(H^2)}\Big/\frac{\partial(J^2a_1^2)}{\partial(J^2)}\right]\frac{\partial}{\partial(J^2)}. \quad (A.2)$$

The condition (A.1) can be written in the more convenient form (1.4) in the following way. First, if $c_1=0$, then (A.1) is satisfied. If $c_1\neq 0$, then (A.1) may be written as

$$\frac{d}{d(H^2)}\left[\frac{J^2a_3^2}{c_1^2}\right]=0. \quad (A.3)$$

By using (A.2) in (A.3), and multiplying by $\partial(J^2a_1^2)/\partial(J^2)$, we obtain the condition that the Jacobian in (1.4) vanishes.

Traveling-Wave Technique of Cyclotron Radiation

P. PARZEN AND G. NOMICOS
Republic Aviation Corporation, Farmingdale, Long Island, New York
(Received April 17, 1961)

A traveling-wave technique of computing the radiation from accelerated electrons is described. In contrast to the usual method which expands the field due to the moving electrons in spherical waves, an expansion in traveling waves along a preferred direction is utilized.

This method is especially suitable in those cases where the radiation is confined by cylindrical conducting boundaries or plasmas. The technique is illustrated by considering the problem of the radiation from an electron in circular motion both in free space and within a coaxial conducting cylinder, and from an electron in a variable-pitch helical orbit.

I. INTRODUCTION

THE usual method[1] of computing the radiation from an accelerated electron leads to a spherical-wave representation of the electromagnetic field of the moving electron. This method is hardly applicable in the presence of conducting surfaces, dielectrics, and plasmas. An alternative method suggested here is a traveling-wave representation in a preferred direction (z direction) of the electromagnetic field. This enables the use of the methods of the theory of cylindrical waveguides. The method is especially suited to handle the presence of cylindrical conducting boundaries, dielectrics, and plasmas.

Basically the traveling-wave method is as follows. We expand the current density due to the moving electron in a Fourier integral in time and in the z direction (traveling waves). The fields due to each traveling wave is determined by solving Maxwell's equations, as in the theory of waveguides, with the currents as forcing terms. Boundary conditions on cylindrical boundaries are easily satisfied. Usually it is a fairly simple matter to transfer from the traveling-wave to the spherical-wave representation with the use of the steepest descent method.

We illustrate the traveling-wave method in this paper with three simple problems.

1. The radiation of an electron moving in a circle.
2. The radiation from an electron moving in a circle which is coaxial with a circular conducting cylinder.
3. The radiation due to an electron moving in a variable-pitch helical orbit.

II. RADIATION OF AN ELECTRON IN A CIRCULAR ORBIT

We consider the radiation from an electron moving in a circle and in free space with a uniform constant magnetic field. Its velocity v is arbitrary and the cyclic frequency of the motion ω is expressed by

$$\omega=\frac{e\mathbf{B}}{m}\left(1-\frac{v^2}{c^2}\right)^{\frac{1}{2}}. \tag{II1}$$

The center of the circle is taken as the origin of the cylindrical coordinate system with the axis Z in the direction of the magnetic field $\mathbf{B}$.

The current density of the electron $\mathbf{J}$ is a product of delta functions, i.e.,

$$\mathbf{J}=ev\frac{\delta(\rho-a)}{a}\delta(z)\delta(\varphi-\omega t), \tag{II2}$$

where a is the radius of the orbit. Fourier analyzing the last two delta functions, Eq. (II2) is written as

$$\mathbf{J}=\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\mathbf{J}_n dh \tag{II3}$$

where the traveling-wave component $\mathbf{J}_n$ is given by

$$J_n=\frac{e\omega}{(2\pi)^2}\delta(\rho-a)e^{-jn\varphi}e^{jhz}e^{jn\omega t}. \tag{II4}$$

From the above equations, it is obvious that the current $\mathbf{J}_n$ has only one component in the φ direction, i.e.,

$$J_{n\varphi}=\frac{e\omega}{(2\pi)^2}\delta(\rho-a)e^{j(n\omega t+hz-n\varphi)} \tag{II5}$$

$$J_{nz}\equiv 0, \quad J_{n\rho}\equiv 0.$$

The electromagnetic fields are obtained from Maxwell's equations due to a traveling-wave current of the form of Eq. (II4) which in mks units are,

[1] L. D. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Press, Inc., Reading, Massachusetts, 1951), p. 189.

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial E_z}{\partial\rho}\right)+\left(k_n^2-h^2-\frac{n^2}{\rho^2}\right)E_z=\frac{j}{n\omega\epsilon_0}\left\{(k_n^2-h^2)J_z+jh\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho J_\rho)-\frac{n}{\rho}J_\varphi\right]\right\} \tag{II6}$$

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\frac{\partial H_z}{\partial\rho}\right)+\left(k_n^2-h^2-\frac{n^2}{\rho^2}\right)H_z=-\frac{1}{\rho}\left[\frac{\partial}{\partial\rho}(\rho J_\varphi)+jnJ_\rho\right] \tag{II7}$$

$$(k_n^2-h^2)E_\rho=jh\frac{\partial E_z}{\partial\rho}-\frac{jn\omega\mu_0}{\rho}(-jn)H_z+jn\omega\mu_0 J_\rho \tag{II8}$$

$$(k_n^2-h^2)E_\varphi=\frac{jh}{\rho}(-jn)E_z+jn\omega\mu_0\frac{\partial H_z}{\partial\rho}+jn\omega\mu_0 J_\varphi \tag{II9}$$

$$(k_n^2-h^2)H_\rho=\frac{jn\omega\epsilon_0(-jn)}{\rho}E_z+jh\frac{\partial H_z}{\partial\rho}+jhJ_\varphi \tag{II10}$$

$$(k_n^2-h^2)H_\varphi=-jn\omega\epsilon_0\frac{\partial E_z}{\partial\rho}+jh\frac{(-jn)}{\rho}H_z-jhJ_\rho, \tag{II11}$$

where

$$k_n^2=n^2\omega^2\mu_0\epsilon_0. \tag{II12}$$

Since $J_\rho=0$, $J_z=0$, and J_φ is a delta function, the solution of Maxwell's equations are very simple for the regions $\rho<a$ and $\rho>a$.

1. Region $\rho<a$

For this region, the solution of Eqs. (II6) to (II11) is given by

$$E_{zn}=AJ_n(\gamma_n\rho),\quad H_{zn}=BJ(\gamma_n\rho) \tag{II13}$$

$$E_{\rho n}=\frac{1}{\gamma_n^2}\left\{jh\gamma_nAJ_n'(\gamma_n\rho)-\frac{n^2\omega\mu_0}{\rho}BJ_n(\gamma_n\rho)\right\} \tag{II14}$$

$$E_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{hn}{\rho}AJ_n(\gamma_n\rho)+jn\omega\mu_0\gamma_nBJ_n'(\gamma_n\rho)\right\} \tag{II15}$$

$$H_{\rho n}=\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega\epsilon_0}{\rho}AJ_n(\gamma_n\rho)+jh\gamma_nBJ_n'(\gamma_n\rho)\right\} \tag{II16}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{-jn\omega\epsilon_0\gamma_nAJ_n'(\gamma_n\rho)+\frac{hn}{\rho}BJ(\gamma_n\rho)\right\} \tag{II17}$$

where J_n is the Bessel function of order n and

$$\gamma_n^2=k_n^2-h^2=n^2\omega^2\mu_0\epsilon_0-h^2. \tag{II18}$$

2. Region $\rho>a$

Similarly, for the region $\rho>a$, the solution of Eqs. (II6) to (II11) gives

$$E_{zn}=CH_n^{(2)}(\gamma_n\rho),\quad H_{zn}=DH_n^{(2)}(\gamma_n\rho) \tag{II19}$$

$$E_{\rho n}=\frac{1}{\gamma_n^2}\left\{jh\gamma_nCH_n^{(2)\prime}(\gamma_n\rho)-\frac{n^2\omega\mu_0}{\rho}DH_n^{(2)}(\gamma_n\rho)\right\} \tag{II20}$$

$$E_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{hn}{\rho}CH_n^{(2)}(\gamma_n\rho)+jn\omega\mu_0\gamma_nDH_n^{(2)\prime}(\gamma_n\rho)\right\} \tag{II21}$$

$$H_{\rho n}=\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega\epsilon_0}{\rho}CH_n^{(2)}(\gamma_n\rho)+jh\gamma_nDH_n^{(2)\prime}(\gamma_n\rho)\right\} \tag{II22}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{-jn\omega\epsilon_0\gamma_nCH_n^{(2)\prime}(\gamma_n\rho)+\frac{hn}{\rho}DH_n^{(2)}(\gamma_n\rho)\right\} \tag{II23}$$

where $H_n^{(2)}$ is the Hankel function of order n.

The constants A, B, C, and D are computed from the matching of the tangential fields E_z, E_φ, H_z, and H_φ at the radius $\rho=a$ with the proper jump conditions, which are obtained by integrating across the point $\rho=a$ with the properties of the delta functions.

$$\begin{aligned}A&=-\frac{he}{8\pi\epsilon_0}H_n^{(2)}(\gamma_na),\quad B=\frac{je\omega a\gamma_n}{8\pi}H_n^{(2)\prime}(\gamma_na)\\ C&=-\frac{he}{8\pi\epsilon_0}J_n(\gamma_na),\quad D=\frac{je\omega a\gamma_n}{8\pi}J_n'(\gamma_na).\end{aligned} \tag{II24}$$

3. Region $\rho\gg a$

Substituting Eqs. (II24) into Eqs. (II21) and (II23), we find that the nth Fourier component of the electric

and magnetic field in the φ direction becomes

$$E_{\varphi n}=\frac{-ea}{8\pi\gamma_n}\left\{n\omega^2\mu_0\gamma_n J_n'(\gamma_n a)H_n^{(2)\prime}(\gamma_n\rho)+\frac{nh^2}{\epsilon_0 a\gamma_n\rho}J_n(\gamma_n a)H_n^{(2)}(\gamma_n\rho)\right\} \quad \text{(II25)}$$

$$H_{\varphi n}=\frac{jn\omega hea}{8\pi\gamma_n}\left\{\frac{1}{a}J_n(\gamma_n a)H_n^{(2)\prime}(\gamma_n\rho)+\frac{1}{\rho}J_n'(\gamma_n a)H_n^{(2)}(\gamma_n\rho)\right\}. \quad \text{(II26)}$$

The total electric and magnetic field in the φ direction is expressed as:

$$\mathfrak{A}_T=\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\mathfrak{A}_n e^{j(n\omega t+hz-n\varphi)}dh, \quad \text{(II27)}$$

where $\mathfrak{A}$ stands for E_φ and H_φ.

Since the integral of Eq. (II27) with respect to h will give a finite contribution to the radiation field only from the first terms of Eqs. (II25) and (II26) for $E_{\varphi n}$ and $H_{\varphi n}$, respectively, we neglect the second terms from both equations. Making use of the asymptotic formula of the Hankel function for large argument and interchanging summation and integration in Eq. (II27), we get the following for the magnetic field $H_{\varphi T}$.

$$H_{\varphi T}=\left(\frac{2}{\pi}\right)^{\frac{1}{2}}\frac{\omega e}{8\pi}e^{j\pi/4}\sum_{n=-\infty}^{\infty}ne^{-jn(\varphi-\omega t)}e^{jn\pi/2}\times\int_{-\infty}^{\infty}\frac{h}{\gamma_n(\gamma_n\rho)^{\frac{1}{2}}}J_n(\gamma_n a)e^{-j\gamma_n\rho}e^{jhz}dh. \quad \text{(II28)}$$

Integrating the above equation with the method of steepest descent, we have

$$\int_{-\infty}^{\infty}\phi(h)e^{jf(h)}dh\simeq\frac{\phi(h_0)(2\pi)^{\frac{1}{2}}}{(|f''(h_0)|)^{\frac{1}{2}}}\exp\{j[f(h_0)+\tfrac{1}{4}\pi\,\mathrm{sign}f''(h_0)]\}, \quad \text{(II29)}$$

where

$$f(h)=hz-\gamma_n\rho,\qquad f'(h_0)=0$$
$$h_0=-k_n\cos\theta,\qquad \gamma_{n0}\rho-h_0 z=k_n\zeta$$
$$\phi(h_0)=\frac{h_0}{\gamma_{n0}(\gamma_{n0}\rho)^{\frac{1}{2}}}J_n(\gamma_{n0}a),\quad |f''(h_0)|^{\frac{1}{2}}=\frac{\sqrt{\zeta}}{(k_n)^{\frac{1}{2}}\sin\theta}. \quad \text{(II30)}$$

Thus, the integral of Eq. (II28) becomes

$$I=-\frac{(2\pi)^{\frac{1}{2}}}{\gamma_{n0}(\gamma_{n0}\zeta\sin\theta)^{\frac{1}{2}}}k_n\cos\theta J_n(\gamma_{n0}a)e^{-jk_n\zeta+j\pi/4}\frac{(k_n)^{\frac{1}{2}}\sin\theta}{\sqrt{\zeta}}. \quad \text{(II31)}$$

Substituting Eq. (II31) into Eq. (II28), we get

$$H_{\varphi T}=-\frac{j\omega e}{4\pi}\sum_{n=-\infty}^{\infty}ne^{-jn(\varphi-\omega t)}e^{jn\pi/2}\cos\theta J_n(\gamma_{n0}a)\frac{e^{-jk_n\zeta}}{\zeta}. \quad \text{(II32)}$$

Following the same procedure for the total electric field in the φ direction and making use of Eqs. (II27), (II29), and (II30), we find

$$E_{\varphi T}=-\frac{e\omega^2\mu_0 a}{4\pi}\sum_{n=-\infty}^{\infty}ne^{-jn(\varphi-\omega t)}e^{jn\pi/2}J_n'(\gamma_{n0}a)\frac{e^{-jk_n\zeta}}{\zeta}. \quad \text{(II33)}$$

The total radiated energy per one revolution of the electron is found by

$$\int_0^T\int_s(\mathbf{E}\times\mathbf{H})\cdot d\mathbf{S}=\int_0^T\int_s(E_\theta H_\varphi-E_\varphi H_\theta)dsdt, \quad \text{(II34)}$$

where

$$T=2\pi/\omega.$$

From Maxwell's equations in spherical coordinates, it follows that for the radiation field at infinity from outgoing waves we have

$$H_\theta=-(\epsilon_0/\mu_0)^{\frac{1}{2}}E_\varphi,\quad E_\theta=(\mu_0/\epsilon_0)^{\frac{1}{2}}H_\varphi. \quad \text{(II35)}$$

From the Fourier analysis for the contribution to the radiation intensity of the nth component per unit solid angle, we have that

$$S_n=\frac{1}{T}\int_0^T(E_{\theta n}H_{\varphi n}-E_{\varphi n}H_{\theta n})\zeta^2dt=2\zeta^2[E_{\theta n}H_{\varphi n}^*-E_{\varphi n}H_{\theta n}^*]. \quad \text{(II36)}$$

Making use of Eqs. (II35), Eq. (II36) becomes

$$S_n=2\zeta^2\left\{\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}E_{\varphi n}E_{\varphi n}^*+\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}H_{\varphi n}H_{\varphi n}^*\right\}. \quad \text{(II37)}$$

Substituting Eqs. (II32) and (II33) into Eq. (II37), we get the following for the nth Fourier component of

the intensity of the radiation energy per unit solid angle

$$S_n=\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}\frac{e^2\omega^2\mu_0^2}{8\pi^2}n^2\omega^2a^2J_n'^2(\gamma_{n0}a)+\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{e^2\omega^2}{8\pi^2}n^2\cot^2\theta J_n^2(\gamma_{n0}a). \quad \text{(II38)}$$

Making use of Eqs. (II12), (II16), and (II30), the above equation becomes

$$S_n=\frac{n^2\omega^2e^2}{8\pi^2}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\left\{\cot^2\theta J_n^2\left(\frac{nv}{c}\sin\theta\right)+\frac{v^2}{c^2}J_n'^2\left(\frac{nv}{c}\sin\theta\right)\right\}, \quad \text{(II39)}$$

which agrees with Landau and Lifschitz.[1]

III. RADIATION OF AN ELECTRON IN A CIRCULAR ORBIT IN A WAVEGUIDE

We consider the radiation from an electron moving in a circle bounded by a coaxial cylindrical conducting wall with radius **b** and in a uniform constant magnetic field.

In this case both the current density of the electron and the electromagnetic fields are given by the same equations as those in Sec. II.

1. Region $\rho<a$

For this region, the solution of Eqs. (II6) to (II11) is given by

$$E_{zn}=A_1J_n(\gamma_n\rho), \quad H_{zn}=B_1J_n(\gamma_n\rho) \quad \text{(III1)}$$

$$E_{\rho n}=\frac{1}{\gamma_n^2}\left\{jh\gamma_nA_1J_n'(\gamma_n\rho)-\frac{n^2\omega\mu_0}{\rho}B_1J_n(\gamma_n\rho)\right\} \quad \text{(III2)}$$

$$E_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{hn}{\rho}A_1J_n(\gamma_n\rho)+jn\omega\mu_0\gamma_nB_1J_n'(\gamma_n\rho)\right\} \quad \text{(III3)}$$

$$H_{\rho n}=\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega\epsilon_0}{\rho}A_1J_n(\gamma_n\rho)+jh\gamma_nB_1J_n'(\gamma_n\rho)\right\} \quad \text{(III4)}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{-jn\omega\epsilon_0\gamma_nA_1J_n'(\gamma_n\rho)+\frac{hn}{\rho}B_1J_n(\gamma_n\rho)\right\}. \quad \text{(III5)}$$

2. Region $a<\rho<b$

Since the walls of the cylinder with $r=b$ are conducting, the solution of Eqs. (II6) to (II11) is given by

$$E_{zn}=C_1\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_nb)}{J_n(\gamma_nb)}J_n(\gamma_n\rho)\right] \quad \text{(III6)}$$

$$H_{zn}=D_1H_n^{(2)}(\gamma_n\rho)+E_1J_n(\gamma_n\rho) \quad \text{(III7)}$$

$$E_{\rho n}=\frac{1}{\gamma_n^2}\left\{jh\gamma_nC_1\left[H_n^{(2)'}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_nb)}{J_n(\gamma_nb)}J_n'(\gamma_n\rho)\right]-\frac{n^2\omega\mu_0}{\rho}[D_1H_n^{(2)}(\gamma_n\rho)+E_1J_n(\gamma_n\rho)]\right\} \quad \text{(III8)}$$

$$E_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{hn}{\rho}C_1\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_nb)}{J_n(\gamma_nb)}J_n(\gamma_n\rho)\right]+jn\omega\mu_0\gamma_n[D_1H_n^{(2)'}(\gamma_n\rho)+E_1J_n'(\gamma_n\rho)]\right\} \quad \text{(III9)}$$

$$H_{\rho n}=\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega\epsilon_0}{\rho}C_1\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_nb)}{J_n(\gamma_nb)}J_n(\gamma_n\rho)\right]+jh\gamma_n[D_1H_n^{(2)'}(\gamma_n\rho)+E_1J_n'(\gamma_n\rho)]\right\} \quad \text{(III10)}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{-jn\omega\epsilon_0\gamma_nC_1\left[H_n^{(2)'}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_nb)}{J_n(\gamma_nb)}J_n'(\gamma_n\rho)\right]+\frac{hn}{\rho}[D_1H_n^{(2)}(\gamma_n\rho)+E_1J_n(\gamma_n\rho)]\right\}. \quad \text{(III11)}$$

The constants A_1, B_1, C_1, D_1, and E_1 are computed from the matching of the tangential fields E_z, E_φ, H_z, and H_φ at the radius $\rho=a$ with the proper jump conditions and from the boundary conditions at $\rho=b$.

$$A_1=-\frac{he}{8\pi\epsilon_0}\cdot\frac{H_n^{(2)}(\gamma_na)J_n(\gamma_nb)-H_n^{(2)}(\gamma_nb)J_n(\gamma_na)}{J_n(\gamma_nb)} \quad \text{(III12)}$$

$$B_1=\frac{j\gamma_nae}{8\pi}\cdot\frac{H_n^{(2)'}(\gamma_na)J_n'(\gamma_nb)-H_n^{(2)'}(\gamma_nb)J_n'(\gamma_na)}{J_n'(\gamma_nb)} \quad \text{(III13)}$$

$$C_1=-\frac{he}{8\pi\epsilon_0}J_n(\gamma_na) \quad \text{(III14)}$$

$$D_1=\frac{je\omega a\gamma_n}{8\pi}J_n'(\gamma_na) \quad \text{(III15)}$$

$$E_1=-\frac{je\omega a\gamma_n}{8\pi}\cdot\frac{H_n^{(2)'}(\gamma_nb)J_n'(\gamma_na)}{J_n'(\gamma_nb)}. \quad \text{(III16)}$$

Substituting Eqs. (III12) to (III16) into Eqs. (III2) to (III5), we find the following for the nth Fourier component of the electromagnetic field in the region $\rho<a$:

$$E_{\rho n}=-\frac{1}{\gamma_n^2}\left\{\frac{jeh^2\gamma_n}{8\pi\epsilon_0}\cdot\frac{H_n^{(2)}(\gamma_n a)J_n(\gamma_n b)-H_n^{(2)}(\gamma_n b)J_n(\gamma_n a)}{J_n(\gamma_n b)}J_n'(\gamma_n\rho)\right.$$
$$\left.+\frac{jen^2\omega^2\mu_0 a\gamma_n}{8\pi\rho}\cdot\frac{H_n^{(2)\prime}(\gamma_n a)J_n'(\gamma_n b)-H_n^{(2)\prime}(\gamma_n b)J_n'(\gamma_n a)}{J_n'(\gamma_n b)}J_n(\gamma_n\rho)\right\} \quad \text{(III17)}$$

$$E_{\varphi n}=-\frac{1}{\gamma_n^2}\left\{\frac{neh^2}{8\pi\epsilon_0\rho}\cdot\frac{H_n^{(2)}(\gamma_n a)J_n(\gamma_n b)-H_n^{(2)}(\gamma_n b)J_n(\gamma_n a)}{J_n(\gamma_n b)}J_n(\gamma_n\rho)\right.$$
$$\left.+\frac{n\omega^2\gamma_n^2\mu_0 ea}{8\pi}\cdot\frac{H_n^{(2)\prime}(\gamma_n a)J_n'(\gamma_n b)-H_n^{(2)\prime}(\gamma_n b)J_n'(\gamma_n a)}{J_n'(\gamma_n b)}J_n'(\gamma_n\rho)\right\} \quad \text{(III18)}$$

$$H_{\rho n}=-\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega eh}{8\pi\rho}\cdot\frac{H_n^{(2)}(\gamma_n a)J_n(\gamma_n b)-H_n^{(2)}(\gamma_n b)J_n(\gamma_n a)}{J_n(\gamma_n b)}J_n(\gamma_n\rho)\right.$$
$$\left.+\frac{h\gamma_n^2 ae\omega}{8\pi}\cdot\frac{H_n^{(2)\prime}(\gamma_n a)J_n'(\gamma_n b)-H_n^{(2)\prime}(\gamma_n b)J_n'(\gamma_n a)}{J_n'(\gamma_n b)}J_n'(\gamma_n\rho)\right\} \quad \text{(III19)}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{jn\omega\gamma_n he}{8\pi}\cdot\frac{H_n^{(2)}(\gamma_n a)J_n(\gamma_n b)-H_n^{(2)}(\gamma_n b)J_n(\gamma_n a)}{J_n(\gamma_n b)}J_n'(\gamma_n\rho)\right.$$
$$\left.+\frac{jhn\gamma_n ae\omega}{8\pi\rho}\cdot\frac{H_n^{(2)\prime}(\gamma_n a)J_n'(\gamma_n b)-H_n^{(2)\prime}(\gamma_n b)J_n'(\gamma_n a)}{J_n'(\gamma_n b)}J_n(\gamma_n\rho)\right\}. \quad \text{(III20)}$$

Similarly, substituting Eqs. (III12) to (III16) into Eqs. (III8) to (III12), we find the following for the nth Fourier component of the electromagnetic field in the region $\rho>a$:

$$E_{\rho n}=-\frac{1}{\gamma_n^2}\left\{\frac{jh^2e\gamma_n}{8\pi\epsilon_0}J_n(\gamma_n a)\left[H_n^{(2)\prime}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_n b)}{J_n(\gamma_n b)}J_n'(\gamma_n\rho)\right]\right.$$
$$\left.+\frac{je\omega^2n^2\gamma_n\mu_0 a}{8\pi\rho}J_n'(\gamma_n a)\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)\prime}(\gamma_n b)}{J_n'(\gamma_n b)}J_n(\gamma_n\rho)\right]\right\} \quad \text{(III21)}$$

$$E_{\varphi n}=-\frac{1}{\gamma_n^2}\left\{\frac{nh^2e}{8\pi\epsilon_0\rho}J_n(\gamma_n a)\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_n b)}{J_n(\gamma_n b)}J_n(\gamma_n\rho)\right]\right.$$
$$\left.+\frac{n\omega^2\gamma_n^2\mu_0 ea}{8\pi}J_n'(\gamma_n a)\left[H_n^{(2)\prime}(\gamma_n\rho)-\frac{H_n^{(2)\prime}(\gamma_n b)}{J_n'(\gamma_n b)}J_n'(\gamma_n\rho)\right]\right\} \quad \text{(III22)}$$

$$H_{\rho n}=-\frac{1}{\gamma_n^2}\left\{\frac{n^2\omega eh}{8\pi\rho}J_n(\gamma_n a)\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_n b)}{J_n(\gamma_n b)}J_n(\gamma_n\rho)\right]\right.$$
$$\left.+\frac{h\gamma_n^2 ae\omega}{8\pi}J_n'(\gamma_n a)\left[H_n^{(2)\prime}(\gamma_n\rho)-\frac{H_n^{(2)\prime}(\gamma_n b)}{J_n'(\gamma_n b)}J_n'(\gamma_n\rho)\right]\right\} \quad \text{(III23)}$$

$$H_{\varphi n}=\frac{1}{\gamma_n^2}\left\{\frac{jn\omega\gamma_n he}{8\pi}J_n(\gamma_n a)\left[H_n^{(2)\prime}(\gamma_n\rho)-\frac{H_n^{(2)}(\gamma_n b)}{J_n(\gamma_n b)}J_n'(\gamma_n\rho)\right]\right.$$
$$\left.+\frac{jhn\gamma_n ae\omega}{8\pi\rho}J_n'(\gamma_n a)\left[H_n^{(2)}(\gamma_n\rho)-\frac{H_n^{(2)\prime}(\gamma_n b)}{J_n'(\gamma_n b)}J_n(\gamma_n\rho)\right]\right\}. \quad \text{(III24)}$$

The total electric and magnetic field in the ρ and φ direction is expressed by Eq. (II27).

The intensity of radiation through the cross section of the circular cylinder averaged over one revolution of the electron is given by

$$I=\frac{1}{T}\int_{-T/2}^{T/2}\int_0^b\int_0^{2\pi}(E_{\rho T}H_{\varphi T}-H_{\rho T}E_{\varphi T})\rho d\rho d\varphi, \tag{III25}$$

where $T=2\pi/\omega$. Substituting from Eq. (II27), interchanging the order of summation and integration, and making use of the orthogonality of the Fourier series, the above equation becomes

$$I=\int_0^b\int_0^{2\pi}\sum_{n=-\infty}^{\infty}\left\{\int_{-\infty}^{\infty}E_{\rho n}e^{jhz}dh\int_{-\infty}^{\infty}H_{\varphi(-n)}e^{jh'z}dh'-\int_{-\infty}^{\infty}H_{\rho n}e^{jhz}dh\int_{-\infty}^{\infty}E_{\varphi(-n)}e^{jh'z}dh'\right\}\rho d\rho d\varphi. \tag{III26}$$

For the evaluation of the integrals of the electromagnetic fields over the wave number h, we have

$$\psi_n=\int_{-\infty}^{\infty}\mathfrak{A}_n e^{jhz}dh \tag{III27}$$

where $\mathfrak{A}_n$ stands for $E_{\rho n}$, $E_{\varphi n}$, $H_{\rho n}$, and $H_{\varphi n}$. Integrating in the complex h plane, it follows from Eqs. (III17) to (III24) that the contribution to the integral with respect to h comes only from the single poles of the Bessel functions, $J_n(\gamma_n b)=0$, and $J_n'(\gamma_n b)=0$, i.e.,

$$\gamma_n b=b\mu_{nm},\ \gamma_n b=b\mu_{nm'} \tag{III28}$$

with the subscript $m=1, 2, 3, \cdots$ and $m'=1, 2, 3, \cdots$ arranged so that the zeros of the Bessel functions are in increasing order of magnitude. Thus, the poles are given by

$$h_{nm}{}^2=k_n{}^2-\mu_{nm}{}^2,\ h_{nm'}{}^2=k_n{}^2-\mu_{nm'}{}^2. \tag{III29}$$

When $|\mu_{nm}|<|k_n|$ then h_{nm} is real and when $|\mu_{nm}|>|k_n|$ then h_{nm} is imaginary. Similarly when $|\mu_{nm'}|<|k_n|$ then $h_{nm'}$ is real and when $|\mu_{nm'}|>|k_n|$ then $h_{nm'}$ is imaginary.

For $z\to\infty$, the expression $\exp(jh_{nm}z)$ becomes zero when h_{nm} is a positive imaginary number. Consequently, there is no contribution to the integral ψ_n of Eq. (III27) from the positive imaginary poles of h_{nm}. Since for a given finite value of n there is only a finite number of real poles h_{nm} on the real axis, i.e., $\pm|h_{nm}|$ $(m=1,2,\cdots,M_n)$, we have to sum the residues of the function $\mathfrak{A}_n$ over a finite number of poles h_{nm} and $h_{nm'}$.

For $z>0$ we perform the integration in the upper-half h plane (see Fig. 1) and we have that the integral on the upper semicircle is equal to zero. Thus the integral ψ_n becomes

$$\psi_n=\int_{-\infty}^{\infty}\mathfrak{A}_n e^{jhz}dh=\sum_{m=1}^{M_n}\mathfrak{A}_{nm}e^{jh_{nm}z}+\sum_{m'=1}^{M_n'}\mathfrak{A}_{nm'}e^{jh_{nm'}z}. \tag{III30}$$

From the Sommerfield condition for outgoing waves, we have that for $z>0$ and $n>0$, only the negative poles h_{nm} and $h_{nm'}$ should be included in the integration path. On the other hand, for $z>0$ and $n<0$, only the positive poles h_{nm} and $h_{nm'}$ should be included in the integration path.

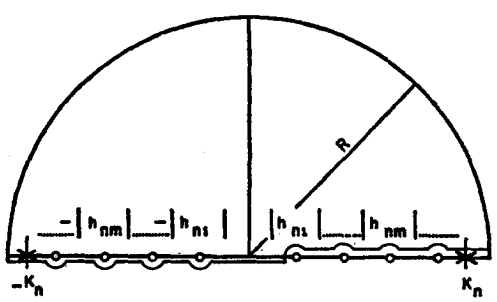


FIG. 1. Complex integration.

Similarly, for $z<0$ we perform the integration in the lower h half-plane and we evaluate the integral ψ_n by summing over all the residues of the positive poles h_{nm} and $h_{nm'}$ for $n>0$. For $n<0$ we sum over all the residues of the negative poles h_{nm} and $h_{nm'}$.

Substituting Eq. (III30) into Eq. (III26), interchanging summation and integration, and making use of the orthogonality property of the modes h_{nm} and $h_{nm'}$ and of the conjugate relation $\mathfrak{A}_{-nm}=\mathfrak{A}_{nm'}{}^*$ we find,

$$I=\sum_{n=1}^{\infty}\int_0^b\int_0^{2\pi}\sum_{m=1}^{N_n}\{[E_{\rho nm}H_{\varphi nm}{}^*+E_{\rho nm}{}^*H_{\varphi nm}]-[H_{\rho nm}E_{\varphi nm}{}^*+H_{\rho nm}{}^*E_{\varphi nm}]\}\rho d\rho d\varphi, \tag{III31}$$

where $\mathfrak{A}_{\rho nm}$ is the residue of $\mathfrak{A}_{\rho n}$ for the m pole h_{nm}, multiplied by $2\pi j$, and $N_n=M_n+M_n'$. Consequently, the nth Fourier component of the radiation intensity becomes

$$I_n=2\int_0^b\int_0^{2\pi}\sum_{m=1}^{N_n}\mathrm{Re}[E_{\rho nm}H_{\varphi nm}{}^*-H_{\rho nm}{}^*E_{\varphi nm}]\rho d\rho d\varphi, \tag{III32}$$

where Re stands for the real part of the expression within the square brackets and

$$E_{\rho nm}=2\pi\frac{eh_{nm}}{8\pi\epsilon_0}J_n(\mu_{nm}a)\frac{H_n^{(2)}(b\mu_{nm})}{bJ_n'(b\mu_{nm})}J_n'(\mu_{nm\rho}) \tag{III33}$$

$$E_{\varphi nm}=-2\pi j\frac{1}{\mu_{nm}}\cdot\frac{neh_{nm}}{8\pi\epsilon_0\rho}J_n(\mu_{nm}a)\frac{H_n^{(2)}(b\mu_{nm})}{bJ_n'(b\mu_{nm})}J_n(\mu_{nm\rho}) \tag{III34}$$

$$H_{\rho nm}=-2\pi j\frac{1}{\mu_{nm}}\cdot\frac{en^2\omega}{8\pi\rho}J_n(\mu_{nm}a)\frac{H_n^{(2)}(b\mu_{nm})}{bJ_n'(b\mu_{nm})}J_n(\mu_{nm\rho}) \tag{III35}$$

$$H_{\varphi nm}=-2\pi\frac{n\omega e}{8\pi}J_n(\mu_{nm}a)\frac{H_n^{(2)}(b\mu_{nm})}{bJ_n'(b\mu_{nm})}J_n'(\mu_{nm\rho}) \tag{III36}$$

$$E_{\rho nm'}=2\pi\frac{n^2\omega^2ea\mu_0}{8\pi\rho}J_n'(\mu_{nm'}a)\frac{H_n^{(2)}(b\mu_{nm'})}{bh_{nm'}J''(b\mu_{nm'})}J_n(\mu_{nm'\rho}) \tag{III37}$$

$$E_{\varphi nm'}=-2\pi j\frac{n\omega^2ea\mu_0\mu_{nm'}}{8\pi}J_n'(\mu_{nm'}a)\frac{H_n^{(2)}(b\mu_{nm'})}{bh_{nm'}J''(b\mu_{nm'})}J_n'(\mu_{nm'\rho}) \tag{III38}$$

$$H_{\rho nm'}=-2\pi j\frac{e\omega a\mu_{nm'}}{8\pi}J'(\mu_{nm'}a)\frac{H_n^{(2)\prime}(b\mu_{nm'})}{bJ''(b\mu_{nm'})}J_n'(\mu_{nm'\rho}) \tag{III39}$$

$$H_{\varphi nm'}=-2\pi\frac{ne\omega a}{8\pi\rho}J_n'(\mu_{nm'}a)\frac{H_n^{(2)\prime}(b\mu_{nm'})}{bJ''(b\mu_{nm'})}J_n(\mu_{nm'\rho}). \tag{III40}$$

Substituting Eqs. (III33) to (III40) into Eq. (III32), we get

$$I_n=\frac{n\pi\omega e^2}{4\epsilon_0b^2}\sum_{n=1}^{M_n}[-h_{nm}]J_n^2(a\mu_{nm})\frac{[H_n^{(2)}(b\mu_{nm})]^2}{J_n'^2(b\mu_{nm})}\int_0^b\left[J_n'^2(\mu_{nm\rho})+\frac{n^2}{\rho^2\mu_{nm}^2}J_n^2(\mu_{nm\rho})\right]\rho d\rho$$
$$+\frac{n\pi\omega^3e^2a^2\mu_0}{4b^2}\sum_{n=1}^{M_{n'}}\frac{\mu_{nm'}^2}{[-h_{nm'}]}J_n'^2(a\mu_{nm})\frac{[H_n^{(2)}(b\mu_{nm'})]^2}{J''^2(b\mu_{nm'})}\int_0^b\left[J_n'^2(\mu_{nm'\rho})+\frac{n^2}{\rho^2\mu_{nm'}^2}J_n^2(\mu_{nm'\rho})\right]\rho d\rho. \tag{III41}$$

Performing the integration with respect to ρ in the above Eq. (III41), we find the following for the nth Fourier component of the radiation intensity:

$$I_n=\frac{n\pi\omega e^2}{2\epsilon_0}\sum_{m=1}^{M_n}[-h_{nm}]J_n^2(a\mu_{nm})J_n'^2(b\mu_{mn})\left[\frac{H_n^{(2)}(b\mu_{nm})}{J_n'(b\mu_{nm})}\right]^2$$
$$+\frac{n\pi\omega^3e^2a^2\mu_0}{2b^2}\sum_{m=1'}^{M}\frac{b^2\mu_{nm'}^2-n^2}{[-h_{nm'}]}J_n'^2(a\mu_{nm'})J_n^2(b\mu_{nm'})\left[\frac{H_n^{(2)}(b\mu_{nm'})}{J_n''^2(b\mu_{nm'})}\right]^2. \tag{III42}$$

This expression is not valid for a practical wave guide when $h_{nm'}$ is very small. In this case the effects of attenuation will alter the answer materially.

IV. RADIATION FROM AN ACCELERATED ELECTRON

We consider the radiation from an electron in the presence of electric and magnetic fields parallel to each other. The motion consists of a circular motion with angular velocity ω_0 superimposed on a rectilinear motion with initial velocity v_0 and acceleration α. We assume that v_0 and α have the same sign and that the accelerating orbit persists only for time T such that $\omega_0T\gg1$. The velocity of the electron is given by

$$\mathbf{v}=a\omega_0\mathbf{i}_\varphi+(v_0+\alpha t)\mathbf{i}_z, \tag{IV1}$$

where $\mathbf{i}_\varphi$ and $\mathbf{i}_z$ are the unit vectors in the φ and z direction, respectively.

The current density of the electron $\mathbf{J}$ is a product of delta functions, i.e.,

$$\mathbf{J}=e\mathbf{v}[\delta(\rho-a)/a]\delta(z-v_0t-\tfrac{1}{2}\alpha t^2)\delta(\varphi-\omega_0t). \tag{IV2}$$

For $t<0$ and $t>T$ we have $J=0$. Substituting Eq. (IV1) into Eq. (IV2), we find the following for the

components of the current in the time interval $0 \leqslant t \leqslant T$.

$$J_\rho = 0$$
$$J_\varphi = e\omega_0 \delta(\rho - a)\delta(z - v_0 t - \tfrac{1}{2}\alpha t^2)\delta(\varphi - \omega_0 t) \qquad \text{(IV3)}$$
$$J_z = e[(v_0 + \alpha t)/a]\delta(\rho - a)\delta(z - v_0 t - \tfrac{1}{2}\alpha t^2)\delta(\varphi - \omega_0 t).$$

Fourier analyzing the delta functions in Eq. (IV3), first with respect to time t and then with respect to z, we find the following for the current density $\mathbf{J}$:

$$\mathbf{J} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\mathbf{J}_n dh d\omega, \qquad \text{(IV4)}$$

where the components of the traveling wave $\mathbf{J}_n$ are given by

$$J_{n\rho} = 0$$
$$J_{n\varphi} = \frac{e\omega_0}{(2\pi)^3} Z_1(\omega,h,T)\delta(\rho - a)e^{j\omega v_0/\alpha} e^{j[(\omega+n\omega_0)t - n\varphi + hz]} \qquad \text{(IV5)}$$
$$J_{nz} = \frac{e}{(2\pi)^3}\frac{1}{a} Z_0(\omega,h,T)\delta(\rho - a)e^{j\omega v_0/\alpha} e^{j[(\omega+n\omega_0)t - n\varphi + hz]}.$$

The functions Z_1 and Z_0 are given by

$$Z_1(\omega,h,T) = \int_0^Z \frac{\exp[-j(\omega/\alpha)(v_0^2+2\alpha z)^{\frac{1}{2}}]}{(v_0^2+2\alpha z)^{\frac{1}{2}}} e^{-jhz} dz = \left(\frac{2}{\alpha h}\right)^{\frac{1}{2}} \exp\left[j\frac{h}{2\alpha}\left(v_0^2 + \frac{\omega^2}{h^2}\right)\right] \int_{(h/2\alpha)^{\frac{1}{2}}(v_0+\omega/h)}^{(h/2\alpha)^{\frac{1}{2}}(v_T+\omega/h)} e^{-jy^2} dy$$
$$Z_0(\omega,h,T) = \int_0^Z \exp\left[-j\frac{\omega}{\alpha}(v_0^2+2\alpha z)^{\frac{1}{2}}\right] e^{-jhz} dz = \frac{j}{h}\left[\exp(-jhZ)\exp\left(-j\frac{\omega}{\alpha}v_T\right) - \exp\left(-j\frac{\omega}{\alpha}v_0\right)\right] - \frac{\omega}{h}Z_1, \qquad \text{(IV6)}$$

where

$$Z = v_0 T + \tfrac{1}{2}\alpha T^2, \quad v_T = v_0 + \alpha T. \qquad \text{(IV7)}$$

The electromagnetic fields are obtained from Maxwell's equations due to a traveling-wave current of the form of Eqs. (IV5) and are given by Eqs. (II6) to (II11) with the substitution of $(\omega+n\omega_0)$ for $(n\omega)$.

Since J_z and J_φ are delta functions, the solution of Maxwell's equations are very simple for the regions $\rho < a$ and $\rho > a$.

1. Region $\rho < a$

For this region the solution of Eqs. (II6) to (II11) is given by Eqs. (II13) to (II17) with the substitution of $(\omega+n\omega_0)$ for $(n\omega)$ and with the subscript (2) at the coefficients A_2 and B_2.

2. Region $\rho > a$

Similarly, for the region $\rho > a$, the solution of Eqs. (II6) to (II11) is given by Eqs. (II19) to (II23) with the substitution of $(\omega+n\omega_0)$ for $(n\omega)$ and with the subscript (2) at the coefficients C_2 and D_2.

The constants A_2, B_2, C_2, and D_2 are computed from the matching of the tangential fields E_z, E_φ, H_z and H_φ at the radius $\rho = a$ with the proper jump conditions which are obtained by integrating across the point $\rho = a$ with the aid of the delta functions.

$$A_2 = \frac{-e}{(4\pi)^2}\frac{e^{j\omega v_0/\alpha}}{\epsilon_0(\omega+n\omega_0)}[\omega_0 hnZ_1(\omega,h,T) + \gamma_n^2 Z_0(\omega,h,T)]H_n^{(2)}(\gamma_n a) \qquad \text{(IV8)}$$

$$B_2 = \frac{je\omega_0\gamma_n a}{(4\pi)^2} e^{j\omega v_0/\alpha} Z_1(\omega,h,T) H_n^{(2)\prime}(\gamma_n a) \qquad \text{(IV9)}$$

$$C_2 = \frac{-e}{(4\pi)^2}\frac{e^{j\omega v_0/\alpha}}{\epsilon_0(\omega+n\omega_0)}[\omega_0 hnZ_1(\omega,h,T) + \gamma_n^2 Z_0(\omega,h,T)]J_n(\gamma_n a) \qquad \text{(IV10)}$$

$$D_2 = \frac{je\omega_0\gamma_n a}{(4\pi)^2} e^{j\omega v_0/\alpha} Z_1(\omega,h,T) J_n'(\gamma_n a), \qquad \text{(IV11)}$$

where

$$\gamma_n^2 = k_n^2 - h^2 = (\omega+n\omega_0)^2\mu_0\epsilon_0 - h^2. \qquad \text{(IV12)}$$

3. Region $\rho \gg a$

Substituting Eqs. (IV8) to (IV10) into Eqs. (II21) and (II23), we find that the nth Fourier component of the electric and magnetic field in the φ direction becomes

$$E_{\varphi n}=-\frac{e}{(4\pi)^2}e^{j\omega v_0/\alpha}\left\{\frac{hn(\omega_0 hnZ_1+\gamma_n Z_0}{\rho\gamma_n{}^2\epsilon_0(\omega+n\omega_0)}J_n(\gamma_n a)H_n{}^{(2)}(\gamma_n\rho)+(\omega+n\omega_0)\mu_0\omega_0 aZ_1J_n{}'(\gamma_n a)H_n{}^{(2)\prime}(\gamma_n\rho)\right\} \quad \text{(IV13)}$$

$$H_{\varphi n}=\frac{je}{(4\pi)^2\gamma_n}e^{j\omega v_0/\alpha}\left\{\frac{\omega_0 hna Z_1}{\rho}J_n{}'(\gamma_n a)H_n{}^{(2)}(\gamma_n\rho)+(\omega hnZ_1+\gamma_n{}^2Z_0)J_n(\gamma_n a)H_n{}^{(2)\prime}(\gamma_n\rho)\right\}. \quad \text{(IV14)}$$

The total electric and magnetic field in the φ direction is expressed as

$$\mathcal{C}_T=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{n=-\infty}^{\infty}\mathcal{C}_n e^{j[(\omega+n\omega_0)t+hz-n\varphi]}dhd\omega. \quad \text{(IV15)}$$

Following the same procedure for the total electric and magnetic field in the φ direction as in Sec. II, we find

$$E_{\varphi T}=\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}-\frac{e\mu_0 a}{8\pi^2}e^{jn\pi/2}e^{j\omega v_0/\alpha}e^{j[(\omega+n\omega_0)t-n\varphi]}\cdot(\omega+n\omega_0)\omega_0 J_n{}'(\gamma_{n0}a)Z_1(\omega,h_0,T)\frac{e^{-jk_n\zeta}}{\zeta}d\omega \quad \text{(IV16)}$$

$$H_{\varphi T}=\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{je}{8\pi^2}e^{jn\pi/2}e^{j\omega v_0/\alpha}e^{j[(\omega+n\omega_0)t-n\varphi]}\cdot J_n(\gamma_{n0}a)\cdot\frac{1}{\gamma_{n0}}[\gamma_{n0}{}^2Z_0(\omega,h_0,T)-\omega_0 nk_n\cos\theta\cdot Z_1(\omega,h_0,T)]\frac{e^{-jk_n\zeta}}{\zeta}d\omega. \quad \text{(IV17)}$$

The radiated energy per unit solid angle during the time T averaged over the angle φ is,

$$[\text{R.E.}]=\frac{1}{2\pi}\int_0^T\int_0^{2\pi}\left[\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}H_{\varphi T}{}^2+\left(\frac{\epsilon_0}{\mu_0}\right)^{\frac{1}{2}}E_{\varphi T}{}^2\right]\zeta^2 d\varphi dt. \quad \text{(IV18)}$$

In order to express the radiation field per frequency $d\Omega$ where $\Omega=\omega+n\omega_0$, we write Eqs. (IV16) and (IV17) for $E_{\varphi T}$ and $H_{\varphi T}$, respectively, as follows.

$$E_{\varphi T}=\int_{-\infty}^{\infty}-\frac{e\mu_0 a\omega_0}{8\pi^2\zeta}\sum_{n=-\infty}^{\infty}\exp\left(jn\frac{\pi}{2}\right)\exp\left[-jn\left(\omega_0\frac{v_0}{\alpha}+\varphi\right)\right]\exp(-jk_\Omega\zeta)J_n{}'(\gamma_{\Omega 0}a)\,\Omega e^{j\Omega v_0/\alpha}Z_1(\Omega-n\omega_0,h_0,T)e^{j\Omega t}d\Omega \quad \text{(IV19)}$$

$$H_{\varphi T}=\int_{-\infty}^{\infty}\frac{je}{8\pi^2\zeta}\sum_{n=-\infty}^{\infty}e^{jn\pi/2}\exp\left[-jn\left(\omega_0\frac{v_0}{\alpha}+\varphi\right)\right]e^{-jk_\Omega\zeta}J_n(\gamma_{\Omega 0}a)$$
$$\times\exp\left(j\Omega\frac{v_0}{\alpha}\right)\frac{\gamma_{\Omega 0}{}^2Z_0(\Omega-n\omega_0,h_0,T)-n\omega_0 k_\Omega\cos\theta\cdot Z_1(\Omega-n\omega_0,h_0,T)}{k_\Omega\sin\theta}e^{j\Omega t}d\Omega. \quad \text{(IV20)}$$

Performing the integration with respect to time t in Eq. (IV18), we have an integral of the form

$$\int_0^T e^{j(\Omega+\Omega')t}dt=2e^{j(\Omega+\Omega')T/2}\frac{\sin(\Omega+\Omega')T/2}{\Omega+\Omega'}. \quad \text{(IV21)}$$

Using the fact that, for moderately high values of T, the expression

$$\frac{\sin(\Omega+\Omega')T/2}{\Omega+\Omega'} \quad \text{is essentially } \pi\delta(\Omega+\Omega'), \quad \text{(IV22)}$$

and making use of Eq. (IV21), Eq. (IV18) becomes

$$[\text{R.E.}]=4\pi\left(\frac{e}{8\pi^2\zeta}\right)^2\int_0^\infty\sum_{n=-\infty}^{\infty}J_n^2(\gamma_{\Omega 0}a)\,|k_\Omega\sin\theta Z_0-n\omega_0\cot\theta\cdot Z_1|^2d\Omega$$
$$+4\pi\left(\frac{-e\mu_0 a\omega_0}{8\pi^2\zeta}\right)^2\int_0^\infty\sum_{n=-\infty}^{\infty}J_n'^2(\gamma_{\Omega 0}a)\Omega^2|Z_1|^2d\Omega. \quad \text{(IV23)}$$

Thus, the Ω-frequency component of the radiated energy S_Ω per unit solid angle and unit frequency, averaged over the azimuthal angle φ is

$$S_\Omega=\frac{e^2}{2(2\pi)^3}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\sum_{n=-\infty}^{\infty}\left\{\left(\frac{a\omega_0}{c}\right)^2\Omega^2J_n'^2\left(\frac{a\Omega}{c}\sin\theta\right)|Z_1|^2+J_n^2\left(\frac{a\Omega}{c}\sin\theta\right)|k_\Omega\sin\theta\cdot Z_0-n\omega_0\cot\theta\cdot Z_1|^2\right\}. \quad \text{(IV24)}$$

The radiated power for small values of T is small since Z_0 and Z_1 are small.

V. CONCLUSIONS

The spherical-wave technique is suitable for free space and inadequate for bounded cylindrical regions.

The traveling-wave technique is effective in handling radiation problems in free space and in bounded cylindrical regions. This has been illustrated by radiation problems both in free space and bounded regions, and both steady and pulsed orbits.

Long-Time Behavior of a Random Lattice

JUN-ICHI HORI
Department of Physics, Hokkaido University, Sapporo, Japan
(Received August 11, 1961)

The time-dependent behavior of a linear isotopic random lattice is studied. The asymptotic average value of the solution of the equation of motion for the amplitude z_k of a normal mode of the virtual regular lattice, composed of atoms with mass equal to the harmonic average of those of the isotopes, is calculated in the limit $N\to\infty$, $t\to\infty$, $\gamma_n\to 0$, and $\gamma_n t$=a finite constant ($n=1, 2, \cdots$). Here N is the number of normal modes, and γ_n the nth cumulant of the distribution of inverse mass. The initial conditions are $z_j(0)=\delta_{kj}$ and $(dz_j/dt)_{t=0}=0 (j=-N/2+1, -N/2+2, \cdots, N/2-1, N/2)$. The result comes out to be

$$\langle z_k\rangle=\exp(-i\alpha\omega_k+\sum_{n=1}^{\infty}\gamma_{2n}Q_k{}^n q^{n-1}/n!)t,$$

where $\alpha\omega_k\equiv 2f^{\frac{1}{2}}\mathfrak{IC}\omega_k$ (f: force constant, $\mathfrak{IC}^2$: average inverse mass) is the eigenfrequency of the kth normal mode, $Q_k\equiv 8i\omega_k f^2/\alpha^3$, and $q=t/N$, assumed to be a constant. If $q\to 0$, there remains only the first term, and the result coincides with that for the case of a Gaussian distribution.

1. INTRODUCTION

WHILE there have appeared a large number of papers dealing with the frequency spectrum of a disordered lattice, there are only few investigations on its time-dependent behavior.[1-5] These are, moreover, restricted to the case of a lattice containing few impurities at some given positions. No attempt seems to have been made to study the motion of a lattice which contains many impurities at random. In this paper we treat the simplest case of a one-dimensional lattice composed of several isotopes which occupy each site independently of one another.

Our aim is to calculate the average propagator matrix of the normal modes. Let $z_i (i=-N/2+1, -N/2+2, \cdots, N/2-1, N/2)$ be the amplitudes of the normal modes of the "virtual regular lattice" composed of atoms whose masses equal the harmonic average of those of the isotopes. The elements of the propagator matrix $z_{kj}(t)$ are defined to be the solutions of the equations of motion for z_k corresponding to the initial conditions $z_k(0)=\delta_{kj}$ and $(dz_k/dt)_{t=0}=0$ $(j, k=-N/2+1, -N/2+2, \cdots, N/2-1, N/2)$. If the average propagator matrix could be calculated accurately, we should be able to obtain the exact frequency spectrum of a random lattice by computing the Fourier transform of its trace.[1] Such a calculation proves to be difficult (Sec. 2). We can, however, compute the asymptotic behavior of the average propagator matrix in the limit

$$N\to\infty,\quad \gamma_n\to 0,\quad t\to\infty, \tag{1.1}$$

and

$$\gamma_n t=\text{finite const},\quad n=1, 2, \cdots,$$

where γ_n is the nth cumulant of the distribution function of inverse mass of the isotopes. This limiting procedure is analogous to that used by Prigogine and Henin[6] to derive the master or Boltzmann equation from the Liouville equation, and the results are of theoretical interest in themselves.

The average propagator matrix comes out to be a diagonal one whose elements are

$$\langle z_{kk}\rangle=\exp(-i\alpha\omega_k+\sum_{n=1}^{\infty}\gamma_{2n}Q_k{}^n q^{n-1}/n!)t, \tag{1.2}$$

where $\alpha\omega_k\equiv 2f^{\frac{1}{2}}\mathfrak{IC}\omega_k$ (f: force constant, assumed to be independent of the kind of participating atoms, $\mathfrak{IC}^2$: average inverse mass) is the eigenfrequency of the kth normal mode of the virtual regular lattice, and $Q_k\equiv 8i\omega_k f^2/\alpha^3$. Further, $q=t/N$ is assumed to be a finite constant. Odd-order terms of the series give the shift of the eigenfrequency, while even-order terms give the damping of the normal mode. If $q\to 0$, i.e., if we let N increase faster than t, there remains only the first term, and the result becomes the same as that obtained when the distribution of inverse mass is assumed to be Gaussian. In this case there appear no damping terms.

The method of calculation is explained in Sec. 2. In Sec. 3 some discussions about the results obtained will be given.

2. METHOD OF CALCULATION

The equation of motion of a one-dimensional lattice composed of N isotopes is

$$\ddot{u}_m=-f(2u_m-u_{m+1}-u_{m-1})/M_m, \tag{2.1}$$

where u_m is the displacement of the mth atom from its equilibrium position and M_m is its mass. We adopt the cyclic boundary condition, and introduce new coordinates $\eta_n (n=-N/2+1, -N/2+2, \cdots, N/2-1, N/2)$ through

$$u_m=\left(\frac{1}{N}\right)^{\frac{1}{2}}\sum_{n=-\frac{1}{2}N+1}^{\frac{1}{2}N}\eta_n\exp\left(\frac{2\pi inm}{N}\right). \tag{2.2}$$

Substituting (2.2) into (2.1) and using the orthogonality

[1] E. Teramoto and S. Takeno, Prog. Theoret. Phys. (Kyoto) 24, 1349 (1960).
[2] E. Teramoto, Busseiron Kenkyu 8, 385, 459 (1960).
[3] R. J. Rubin, J. Math. Phys. 1, 309 (1960).
[4] R. J. Rubin, J. Math. Phys. 2, 373 (1961).
[5] R. E. Turner, Physica 26, 269, 274 (1960).
[6] I. Prigogine and F. Henin, J. Math. Phys. 1, 349 (1960).

relation

$$\sum_m \exp\{2\pi m(n-k)/N\} = N\delta_{nk}, \tag{2.3}$$

we get

$$\ddot{\eta}_k = -\left(\frac{4f}{N}\right)\sum_n \eta_n\omega_n{}^2 \sum_m \mu_m \exp\left\{\frac{2\pi i(n-k)m}{N}\right\}, \tag{2.4}$$

in which $\mu_m = 1/M_m$ and $\omega_n{}^2 \equiv \sin^2(\pi n/N)$. Now put

$$B_s \equiv (1/N)\sum_m \mu_m \exp(2\pi i s m/N) (= B_{-s}{}^*). \tag{2.5}$$

We assume that the μ_m's form a stationary stochastic process, and are statistically independent of one another. Then the B_s's are random variables with order of magnitude $1/N^{\frac{1}{2}}$, except that B_0 is equal to $\mathfrak{M}^2 \equiv \langle \mu_m \rangle$. To investigate the cumulants of the distribution of B_s's (excluding B_0), introduce the characteristic function $\phi(\mathbf{t})$ of the distribution of μ_m's where $\mathbf{t} \equiv (t_1, t_2, \cdots, t_N)^T$ is an N-dimensional vector. Since the B_s's are linear transformations of the original variables μ_m with the transformation matrix

$$\mathbf{C} \equiv (c_{sm}) = [(1/N)\exp(2\pi i s m/N)],$$

their characteristic function is given by $\phi(\mathbf{C}^T\mathbf{u}) \equiv \psi(\mathbf{u})$, $\mathbf{u}$ being the vector given by $\mathbf{t} = \mathbf{C}^T\mathbf{u}$. By virtue of the independence of μ_m's, we have

$$\phi(\mathbf{t}) = \prod_{i=1}^{N} \phi(t_i) = \prod_{i=1}^{N} \phi\Big(\sum_j c_{ji}u_j\Big) = \psi(\mathbf{u}), \tag{2.6}$$

where $\phi(t_m)$ is the characteristic function of a single variable μ_m, and consequently

$$\ln\phi(\mathbf{t}) = \sum_{i=1}^{N} \ln\phi(t_i) = \sum_{i=1}^{N} \ln\phi\Big(\sum_j c_{ji}u_j\Big) = \ln\psi(\mathbf{u}). \tag{2.7}$$

This shows that all the terms which give the mth cumulants of the B_s's as their coefficients are collected in the expression

$$\sum_i t_i{}^m = \sum_{i,j=1}^{N} (c_{ji}u_j)^m.$$

For example, the terms giving the variances are found in

$$\sum_{i=1}^{N} t_i{}^2 = \sum_{j,k=1} u_j u_k \sum_{i=1}^{N} \left(\frac{1}{N^2}\right) \exp\left(\frac{2\pi i i j}{N}\right) \exp\left(\frac{2\pi i i k}{N}\right)$$

$$= \sum_{j,k} u_j u_k \left(\frac{1}{N}\right)\delta_{k,-j} = \left(\frac{1}{N}\right)\sum_k u_{-k}u_k. \tag{2.8}$$

This means that only variances of the form[7] $\langle B_{-k}B_k \rangle_C$ are nonzero. Moreover, they are all equal to γ_2/N, in harmony with the result of the theory of stationary stochastic processes. It is easily seen, in the same way, that only the cumulants which are composed of pairs of the type $\{B_{-k}, B_k\}$, for example, $\langle B_{-k}B_{-m}B_{-l}B_lB_mB_k \rangle_C$, $\langle B_{-k}B_kB_{-l}B_{-m}B_lB_mB_{-n}B_n \rangle_C$, and so on, are nonzero, and all the nonzero $2m$th cumulants are equal to γ_{2m}/N^{2m-1}.

Putting $\alpha^2 \equiv 4f\mathfrak{M}^2$, $\alpha_k{}^2 \equiv \alpha^2\omega_k{}^2$, and considering that $B_0 = \mathfrak{M}^2$, (2.4) can be written as

$$\ddot{\eta}_k = -\alpha_k{}^2\eta_k - 4f\sum_{n\neq k} \eta_n\omega_n{}^2 B_{n-k}. \tag{2.9}$$

The second term of the right-hand side vanishes for the virtual regular lattice composed of atoms with mass $1/\mathfrak{M}^2$. Hence, the η_k's and α_k's are the amplitudes and eigenfrequencies, respectively, of its normal modes. In the following we omit the restriction $n \neq k$ in the summation, regarding B_0 as zero.

Now put

$$d\eta_k/dt \equiv \alpha_k\zeta_k, \tag{2.10}$$

and

$$\eta_k + i\zeta_k \equiv z_k, \quad \eta_k - i\zeta_k \equiv z_k{}^*, \tag{2.11}$$

then (2.9) becomes

$$\begin{aligned} dz_k/dt &= \dot{\eta}_k + i\dot{\zeta}_k = -i\alpha_k z_k \\ &\quad -2if\textstyle\sum_n B_{n-k}\omega_n{}^2(z_n + z_n{}^*)/\alpha_n, \\ dz_k{}^*/dt &= \dot{\eta}_k - i\dot{\zeta}_k = i\alpha_k z_k{}^* \\ &\quad +2if\textstyle\sum_n B_{n-k}\omega_n{}^2(z_n + z_n{}^*)/\alpha_n. \end{aligned} \tag{2.12}$$

The z_k's play the role of normal mode amplitudes equally as do the η_k's.

Turning to the "interaction representation" by putting

$$z_k = y_k(t)\exp\{-i\alpha_k(t)\}, \tag{2.13}$$

we have

$$\begin{aligned} \dot{y}_k &= -2if\textstyle\sum_n B_{n-k}(\omega_n{}^2/\alpha_k)[\exp\{i(\alpha_k - \alpha_n)t\} \\ &\quad \times y_n + \exp\{i(\alpha_k + \alpha_n)t\}y_n{}^*], \\ \dot{y}_k{}^* &= 2if\textstyle\sum_n B_{n-k}(\omega_n{}^2/\alpha_k)[\exp\{-i(\alpha_k + \alpha_n)t\} \\ &\quad \times y_n + \exp\{-i(\alpha_k - \alpha_n)t\}y_n{}^*]. \end{aligned} \tag{2.14}$$

If we introduce the matrix

$$\mathbf{H} \equiv \begin{bmatrix} 0 & 0 & -2ifB_1(\omega_2{}^2/\alpha_1)e^{i(\alpha_1-\alpha_2)t} & -2ifB_1(\omega_2{}^2/\alpha_1)e^{i(\alpha_1+\alpha_2)t} & \cdots \\ 0 & 0 & 2ifB_1(\omega_2{}^2/\alpha_1)e^{-i(\alpha_1+\alpha_2)t} & 2ifB_1(\omega_2{}^2/\alpha_1)e^{-i(\alpha_1-\alpha_2)t} & \cdots \\ -2ifB_{-1}(\omega_1{}^2/\alpha_2)e^{i(\alpha_2-\alpha_1)t} & -2ifB_{-1}(\omega_1{}^2/\alpha_2)e^{i(\alpha_2+\alpha_1)t} & 0 & 0 & \cdots \\ 2ifB_{-1}(\omega_1{}^2/\alpha_2)e^{-i(\alpha_2+\alpha_1)t} & 2ifB_{-1}(\omega_1{}^2/\alpha_2)e^{-i(\alpha_2-\alpha_1)t} & 0 & 0 & \cdots \\ -2ifB_{-2}(\omega_1{}^2/\alpha_3)e^{i(\alpha_3-\alpha_1)t} & -2ifB_{-2}(\omega_1{}^2/\alpha_3)e^{i(\alpha_3+\alpha_1)t} & \cdots & \cdots & \cdots \\ 2ifB_{-2}(\omega_1{}^2/\alpha_3)e^{-i(\alpha_3+\alpha_1)t} & 2ifB_{-2}(\omega_1{}^2/\alpha_3)e^{-i(\alpha_3-\alpha_1)t} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{2.15}$$

[7] We denote in general the mth cumulants by $\langle B_{i(1)}B_{i(2)}\cdots B_{i(m)} \rangle_C$.

(2.14) can be written in a vector-matrix form $\dot{\mathbf{y}}=\mathbf{H}\mathbf{y}$, where $\mathbf{y}\equiv(y_1,y_1^*,\cdots,y_N,y_N^*)^T$. (Here we momentarily relabel, for simplicity of notation, the normal modes so that the index runs from 1 to N.) The matrix solution $\mathbf{Y}(t)$ of this equation with initial condition $\mathbf{Y}(0)=\mathbf{I}$ ($\mathbf{I}$: unit matrix) is, formally,

$$\mathbf{Y}(t)=\exp\left\{\int_0^t \mathbf{H}(t)dt\right\}, \tag{2.16}$$

which gives the required propagator matrix. According to Kubo's cumulant expansion theorem,[8] the average value of its diagonal element is given by

$$\langle k|\mathbf{Y}(t)|k\rangle=\exp\left\{\sum_{n=1}^{\infty}\int_0^t dt_1\int_0^{t_1}dt_2\cdots\int_0^{t_{n-1}}dt_n \times\langle k|\mathbf{H}(t_1)\mathbf{H}(t_2)\cdots\mathbf{H}(t_n)|k\rangle_C\right\}. \tag{2.17}$$

As an example, let us calculate the fourth-order term in the exponent. The time integral in this term is of the form

$$I(t)=\int_0^t dt_1\int_0^{t_1}dt_2\int_0^{t_2}dt_3\int_0^{t_3}dt_4\exp\{i(\alpha_k-\alpha_l)t_1\} \times\exp\{i(\alpha_l-\alpha_m)t_2\}\exp\{i(\alpha_m-\alpha_n)t_3\} \exp\{i(\alpha_n-\alpha_k)t_4\}, \tag{2.18}$$

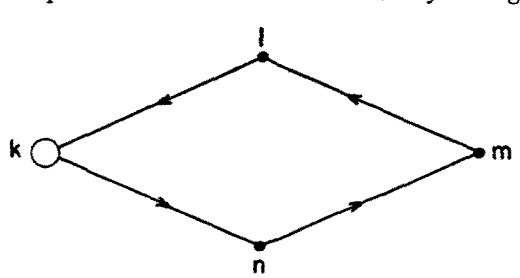
which is obtained by putting $\beta_1=\alpha_k-\alpha_l$, $\beta_2=\alpha_k-\alpha_m$, $\beta_3=\alpha_k-\alpha_n$, and $\beta_4=0$ in the integral

$$\int_0^t dt_1\int_0^{t_1}dt_2\int_0^{t_2}dt_3\int_0^{t_3}dt_4\exp\{i\beta_1(t_1-t_2)\} \times\exp\{i\beta_2(t_2-t_3)\}\exp\{i\beta_3(t_3-t_4)\}\exp\{i\beta_4 t_4\}, \tag{2.19}$$

which was treated by Prigogine and Henin.[6] According to their theorem, (2.18) becomes

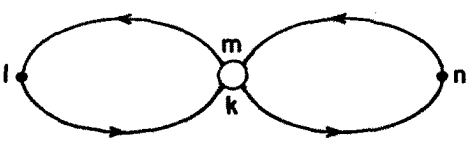
$$I(t)=\frac{1}{2\pi i}\int_{-\infty-i\epsilon}^{\infty-i\epsilon}dz \times\frac{e^{izt}}{z^2(z-\alpha_k+\alpha_l)(z-\alpha_k+\alpha_m)(z-\alpha_k+\alpha_n)}. \tag{2.20}$$

The factor $\langle B_{l-k}B_{m-l}B_{n-m}B_{k-n}\rangle_C$ accompanying this integral is nonzero only if $m=k$ or $n=l$. In other words, if we represent the above series of B_k's by a diagram



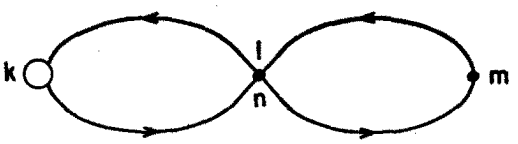
[8] R. Kubo (unpublished).

only those of the forms



(a)

and



(b)

contribute to the term considered. According to the results obtained above, their contributions are both equal to γ_4/N, independent of l and n. On the other hand, the integral $I(t)$ corresponding to the diagram (a) gives a term quadratic in t

$$(it)^2/2!(\alpha_l-\alpha_k)(\alpha_n-\alpha_k), \tag{2.21}$$

while that corresponding to (b) gives, at most, linear terms in t. Contributions of terms of lower order in t will become negligible, compared to that of (2.21), in the limit (1.1). Only the diagrams of the type (a) contribute, therefore, in this limit. Picking up only (a)-type diagrams, and in these only the terms proportional to t^2, we get, if $k\neq N/2$, as the fourth-order term of the exponent,

$$(it)^2(2^4if)^4\gamma_4\omega_k^2/2!\alpha^6N=\gamma_4t^2Q_k^2/2!N=\gamma_4tQ_k^2q/2!, \tag{2.22}$$

which is finite so far as q is finite. Here we used the formula[9]

$$\sum_l\frac{\omega_l^2}{\omega_l^2-\omega_k^2}=\sum_{l=-(N/2)+1}^{N/2}\frac{\sin^2(\pi l/N)}{\sin^2(\pi l/N)-\sin^2(\pi k/N)} =N+1-1/[2\cos^2(\pi k/N)],\ k\neq N/2, \tag{2.23}$$
$$=N-\tfrac{1}{3}N^2,\quad k=N/2,$$

neglecting the terms other than N.

The argument for the other terms in the exponent runs completely parallel. All the odd-order terms vanish, and in the $2m$th-order term, only the diagrams for the type of an m-petal flower (see Fig. 1) give a contribution in our limit, its value being

$$\gamma_{2m}tQ_k^mq^{m-1}/m!. \tag{2.24}$$

Thus, if $k\neq N/2$, we get finally

$$\langle k|\mathbf{Y}(t)|k\rangle=\exp\left(t\sum_{n=1}^{\infty}\gamma_{2n}Q_k^nq^{n-1}/n!\right), \tag{2.25}$$

[9] A. A. Maradudin, G. H. Weiss, and D. W. Jepsen, J. Math. Phys. 2, 349 (1961).

and, for the diagonal element of the propagator matrix of z_m's,

$$\langle k|\mathbf{Z}(t)|k\rangle=\langle z_{kk}\rangle=\exp(-i\alpha_k t)$$
$$\times\exp(t\sum_{n=1}^{\infty}\gamma_{2n}Q_k{}^n q^{n-1}/n!). \quad (2.26)$$

If we let $q\to 0$, the series in the exponent terminates after the first term. The same result is obtained if we leave q finite but instead assume the distribution of μ_m's to be Gaussian. It will be easily seen, from the properties of the cumulants of the B_k's and the time integral $I(t)$, that the average nondiagonal elements of the propagator matrix vanish or become negligible compared with the diagonal ones.

When $k=N/2$, i.e., at the band edge of the virtual lattice, the term $-(1/3)N^2$ in (2.23) can no longer be neglected. Here all the terms diverge in the exponent of (2.25).

If we wish to avoid any limiting procedure like (1.1), we are obliged to calculate contributions from general diagrams and the terms of lower order in t. In general, however, these involve improper integrals which are more singular than the principal value integral (2.23). It is not easy to assess the values of such integrals and as a result to calculate the average propagator matrix accurately for all times.

Odd-order terms of the series in the exponent give the shifts of frequencies from the eigenfrequencies of the virtual regular lattice, while the even-order terms give damping of the amplitudes of normal modes. In the second limit $q\to 0$, or in the Gaussian case, there appear no damping terms.

3. DISCUSSIONS

In the above calculation we altogether neglected the lower-order terms in t. Since such terms become rapidly numerous as n increases, however, the sum of all such terms may possibly give a finite contribution. This ambiguity probably cannot be avoided so far as q remains finite. If, on the other hand, we let $q\to 0$, or let N become infinite faster than t, such ambiguity does not arise, for this limiting procedure is equivalent to making all the terms for $n>1$ zero before carrying out the summation. That the result becomes the same as in the Gaussian case means that, when N grows fast indefinitely, the distribution of B_k's becomes Gaussian according to the central limit theorem.

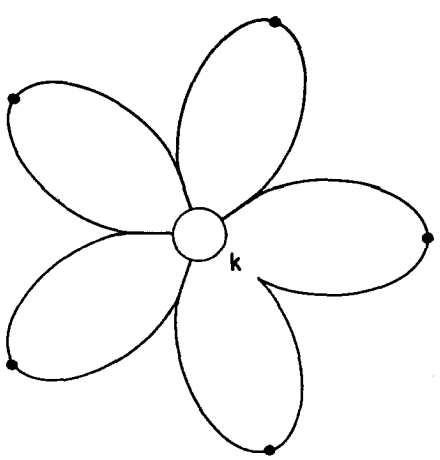


FIG. 1. m-petal flower.

If, however, the distribution of B_k's is such that only a finite number of cumulants are not zero or is nearly Gaussian so that it becomes rapidly small as n increases, (2.26) will be valid even if q remains finite.

It should be remarked also that we neglected the terms other than N in (2.23). It was already mentioned that such a neglect cannot be allowed at the band edge, and our theory breaks down here. Even if $k\neq N/2$, however, the neglected terms may become very large near the band edge. Although in the limit $N\to\infty$, especially if we let $q\to 0$, this does not formally give rise to any difficulty, and the above conclusions remain the same, some ambiguities will be brought about if we want to interpret the above results in more physical terms, keeping N large but finite. The width of the region near the band edge within which such a question arises will, however, in general, be very small.

At any rate, the above results mean physically that if the fluctuation from the virtual regular lattice is small, the motion of the random lattice will, at first, be nearly the same as that of the virtual lattice, but there will appear a shift of frequency and, occasionally, a damping of the normal mode after a sufficiently long time.

The result for $q\to 0$ can also be deduced by another method: Putting in (2.9) first

$$\eta_k=\zeta_k(t)\exp(i\alpha_k t), \quad (3.1)$$

and then

$$\dot{\zeta}_k=\xi_k, \quad \xi_k=\gamma_k(t)\exp(-2i\alpha_k t), \quad (3.2)$$

we get

$$\begin{aligned}\dot{\zeta}_k&=\gamma_k\exp(-2i\alpha_k t),\\ \dot{\gamma}_k&=-4f\textstyle\sum_n\omega_n{}^2B_{n-k}\zeta_n\exp\{i(\alpha_n+\alpha_k)t\}.\end{aligned} \quad (3.3)$$

This is another form of the equation of motion in the interaction representation. Solving (3.3) by iteration with the initial conditions

$$\zeta_n(0)=\delta_{ns}, \quad \gamma_n(0)=0 \quad \text{for all } n, \quad (3.4)$$

we have

$$\zeta_k=\sum_{n=0}^{\infty}\Delta\zeta_k{}^{(n)},$$

$$\Delta\zeta_k^{(n)} = (-4f)^n \sum_{j(n-1)} \sum_{j(n-2)} \cdots \sum_{j(1)} \omega_{j(n-1)}^2 \cdots \omega_{j(1)}^2 \omega_s^2$$

$$\times B_{j(n-1)-k} B_{j(n-2)-j(n-1)} B_{j(n-3)-j(n-2)} \cdots$$

$$\times \int_0^t \int_0^{t_{2n}} \int_0^{t_{2n-1}} \cdots \int_0^{t_2} dt_1 dt_2 \cdots dt_{2n}$$

$$\times [\exp i\alpha \{ -2\omega_k t_{2n} + (\omega_{j(n-1)} + \omega_k) t_{2n-1}$$

$$- 2\omega_{j(n-1)} t_{2n-2} + (\omega_{j(n-2)} + \omega_{j(n-1)}) t_{2n-3}$$

$$- 2\omega_{j(1)} t_2 + (\omega_s + \omega_{j(1)}) t_1 \}]. \quad (3.5)$$

The time integral appearing here can again be reduced to Prigogine and Henin's integral. If a diagram is associated with each series of B_k's in the same manner as in Sec. 2, it turns out that here again only diagrams of the form of an m-petal flower (Fig. 1) contribute to the value of $\Delta\zeta_k^{(n)}$, in the limit $\gamma_2 t = \text{finite}$, $\gamma_2 \to 0$, and $t \to \infty$. Consequently, all odd-order terms vanish, and there remains only the "diagonal response" $\langle \eta_s \rangle$, which corresponds to a diagonal element of the propagator matrix. Statistical parameters other than γ_2 do not appear in the present calculation. The result is

$$\langle \eta_s \rangle = \exp\{ i\omega_s f^{\frac{1}{2}} t (2\mathfrak{K} - \gamma_2/3\mathfrak{C}^3) \}, \quad (3.6)$$

which is coincident with (2.36) for $q=0$.

Also, in this calculation the lower-order terms in t are neglected entirely, although they may possibly give, when summed up, a finite contribution. The result (3.6) indicates that this neglect amounts to ignoring the deviation from the Gaussian distribution.

ACKNOWLEDGMENTS

The author wishes to express his grateful thanks to Professor T. Matsubara for taking interest in this work and giving helpful suggestions as to the method of calculation. He also sincerely thanks the members of the Symposium on Vibration of Disordered Lattices held at the Research Institute for Fundamental Physics of Kyoto University in June 1961 for keen discussions on this subject.

Finally, it should be mentioned that after having worked out the major part of the present paper we became aware of the work of Maradudin, Weiss, and Jepsen[9] on a similar problem, and we owe the formula (2.23) and a part of the discussions given in Sec. 3 to their paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH-APRIL, 1962

Addendum: Almost Periodicity and the Quantal H Theorem

[J. Math. Phys. 2, 235 (1961)]

IAN C. PERCIVAL
Physics Department, University College, London, England

NEARLY all the results of the above paper were proved in an article by Ono,[1] where the significance of almost periodicity was fully recognized. In our proof of the H theorem slightly more general sets of states were used to define the entropy, which led to the subsidiary conditions on the spectrum which Ono was able to avoid.

Bocchieri and Loinger[2] have also proved some of Ono's results. I should like to thank these authors for drawing attention to their article and for helping me to find the article by Ono.

[1] S. Ono, Mem. Fac. Eng., Kyushu Univ. **11**, 125 (1949).

[2] P. Bocchieri and A. Loinger, Phys. Rev. **107**, 337 (1957).

First Reprinting, November 1962

$$\Delta\zeta_k^{(n)} = (-4f)^n \sum_{j(n-1)} \sum_{j(n-2)} \cdots \sum_{j(1)} \omega_{j(n-1)}^2 \cdots \omega_{j(1)}^2 \omega_s^2$$

$$\times B_{j(n-1)-k} B_{j(n-2)-j(n-1)} B_{j(n-3)-j(n-2)} \cdots$$

$$\times \int_0^t \int_0^{t_{2n}} \int_0^{t_{2n-1}} \cdots \int_0^{t_2} dt_1 dt_2 \cdots dt_{2n}$$

$$\times [\exp i\alpha \{ -2\omega_k t_{2n} + (\omega_{j(n-1)} + \omega_k) t_{2n-1}$$

$$- 2\omega_{j(n-1)} t_{2n-2} + (\omega_{j(n-2)} + \omega_{j(n-1)}) t_{2n-3}$$

$$- 2\omega_{j(1)} t_2 + (\omega_s + \omega_{j(1)}) t_1 \}]. \quad (3.5)$$

The time integral appearing here can again be reduced to Prigogine and Henin's integral. If a diagram is associated with each series of B_k's in the same manner as in Sec. 2, it turns out that here again only diagrams of the form of an m-petal flower (Fig. 1) contribute to the value of $\Delta\zeta_k^{(n)}$, in the limit $\gamma_2 t = \text{finite}$, $\gamma_2 \to 0$, and $t \to \infty$. Consequently, all odd-order terms vanish, and there remains only the "diagonal response" $\langle \eta_s \rangle$, which corresponds to a diagonal element of the propagator matrix. Statistical parameters other than γ_2 do not appear in the present calculation. The result is

$$\langle \eta_s \rangle = \exp\{ i\omega_s f^{\frac{1}{2}} t (2\mathfrak{K} - \gamma_2/3\mathfrak{C}^3) \}, \quad (3.6)$$

which is coincident with (2.36) for $q=0$.

Also, in this calculation the lower-order terms in t are neglected entirely, although they may possibly give, when summed up, a finite contribution. The result (3.6) indicates that this neglect amounts to ignoring the deviation from the Gaussian distribution.

ACKNOWLEDGMENTS

The author wishes to express his grateful thanks to Professor T. Matsubara for taking interest in this work and giving helpful suggestions as to the method of calculation. He also sincerely thanks the members of the Symposium on Vibration of Disordered Lattices held at the Research Institute for Fundamental Physics of Kyoto University in June 1961 for keen discussions on this subject.

Finally, it should be mentioned that after having worked out the major part of the present paper we became aware of the work of Maradudin, Weiss, and Jepsen[9] on a similar problem, and we owe the formula (2.23) and a part of the discussions given in Sec. 3 to their paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 2 MARCH-APRIL, 1962

Addendum: Almost Periodicity and the Quantal H Theorem

[J. Math. Phys. 2, 235 (1961)]

IAN C. PERCIVAL
Physics Department, University College, London, England

NEARLY all the results of the above paper were proved in an article by Ono,[1] where the significance of almost periodicity was fully recognized. In our proof of the H theorem slightly more general sets of states were used to define the entropy, which led to the subsidiary conditions on the spectrum which Ono was able to avoid.

Bocchieri and Loinger[2] have also proved some of Ono's results. I should like to thank these authors for drawing attention to their article and for helping me to find the article by Ono.

[1] S. Ono, Mem. Fac. Eng., Kyushu Univ. **11**, 125 (1949).

[2] P. Bocchieri and A. Loinger, Phys. Rev. **107**, 337 (1957).

First Reprinting, November 1962