JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 3, NUMBER 4 JULY–AUGUST 1962

Reduction of Operator Rings and the Irreducibility Axiom in Quantum Field Theory*

THOMAS F. JORDAN AND E. C. G. SUDARSHAN
University of Rochester, Rochester, New York
(Received January 4, 1962)

The mathematical theory of the reduction of operator rings is used to investigate some structures which can occur in quantum field theory when the postulate that the field operators generate an irreducible ring is relaxed. In particular, it is shown that if a quantum field theory has a commutator which commutes with all field operators it is a direct integral of theories in each of which the commutator is a scalar. If in addition it satisfies the postulates of Lorentz covariance, existence of an invariant vacuum, and mass and energy spectra, then it is a direct integral of generalized free field theories whenever the unitary representation of the Lorentz group can be constructed in terms of functions of the field operators and every state can be constructed by applying field operators on the vacuum. It is also shown that the latter two assumptions together with the requirement of a unique invariant vacuum state are sufficient to prove that the ring generated by the field operators is irreducible. In other words, under these conditions the irreducibility postulate is redundant.

I. INTRODUCTION

ONE of the postulates which is often included in studies of quantum field theory is that the ring generated by "smeared polynomials in the field operators" be irreducible.[1] But for many purposes it is neither necessary nor desirable to limit the scope of investigations by this assumption. For example, if a field theory is defined in terms of its Wightman functions[2] it is possible for the resulting set of field operators to be reducible. In fact a weighted mean of two sequences of Wightman functions is again a sequence of Wightman functions so that the set of all sequences of Wightman functions forms a convex set.[3] But only the extremal points of this convex set yield theories for which the field operators generate an irreducible ring.[4] It is also interesting to note that the theories which do not correspond to extremal points will not necessarily have a vacuum state which is the unique invariant state.

In general, the operators representing the observables for a quantum mechanical system generate an irreducible ring if and only if the system has no superselection rules (or equivalently admits no supersymmetry transformation).[5] However we do not wish to assume that the ring of observables and the ring generated by the field operators are identical; the latter may contain operators, e.g., a baryon creation operator, connecting different superselection subspaces.

In this paper we explore some of the situations which can occur when the irreducibility postulate

* Supported in part by the Atomic Energy Commission.

[1] This is listed as postulate 6 by R. Haag and B. Schroer, J. Math Phys. **3**, 248 (1962), which contains a rigorous survey of axiomatic quantum field theory.

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

[3] E. C. G. Sudarshan and K. Bardakci, J. Math Phys. **2**, 767 (1961).

[4] As was discussed in detail in reference 1 by Haag and Schroer, this is a general property of the construction of a representation of a ring from a positive linear functional on the ring.

[5] J. M. Jauch, Helv. Phys. Acta **33**, 711 (1960).

of quantum field theory is relaxed. It is to be expected that the presently available models of field theories can be extended to yield examples of structure richer than have so far been evident. We consider in particular the extension of the generalized free field theory of Greenberg.[6] Under the assumption of irreducibility, it has been shown that if a field theory has a commutator which either commutes with all field operators or is translation invariant and vanishes for space-like separation of its arguments, then the commutator is a scalar. It has also been shown that if, in addition to the latter property, the field theory satisfies the customary postulates of Lorentz covariance, the existence of the vacuum state, and mass and energy spectra, then the theory is a generalized free field theory.[7] We will see that these results can be generalized in a straightforward way when the irreducibility postulate is relaxed.

Our method of investigation is based on the mathematical theory of rings of operators. In the next section we define some of the main concepts of this theory and outline the results which we will use. In Sec. III we review the defining properties of field operators in terms of this mathematical language. The structure of commutators which commute with all field operators is treated in the following section. The result is that the field theory is a direct integral of field theories in each of which the commutator is a scalar. In the final section, in addition to this property of the commutator, we postulate Lorentz invariance, an invariant vacuum state, and positive energy and square of the mass. If we also assume that the unitary representation of the Lorentz group can be constructed as functions of the field operators, and that all of the states of the theory can be obtained by operating with field operators on the invariant vacuum state, then we can deduce that the theory is a direct integral of generalized free field theories. If with these latter two assumptions we require the vacuum state to be the unique invariant state we can prove that the field operators are irreducible. In other words, if in a field theory the unitary representation of the Lorentz group is constructed in terms of functions of the field operators, if there exists a vacuum state which is the unique invariant state, and if all states can be obtained by applying field operators to the vacuum state, then the ring generated by the field operators is irreducible.

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

[7] A. L. Licht and J. S. Toll, Nuovo cimento **21**, 346 (1961); G. F. Dell'Antonio, J. Math. Phys. **2**, 759 (1961); see also R. Acharya, Nuovo cimento **23**, 580 (1962).

II. MATHEMATICAL BACKGROUND

In this section we give a brief review of the theory of operator rings and their reduction. This survey is not intended to be complete in any way. Its purpose is only to provide enough information to define the main concepts used in this paper. No proofs are given. For further details we refer the reader to the original papers by Murray and von Neumann,[8] Naimark and Fomin,[9] and to the lucid book by Naimark.[10]

We call either a finite dimensional linear inner-product space or an infinite dimensional separable Hilbert space simply a Hilbert space.[11]

Definition. A set R of bounded linear operators on a Hilbert space $\mathfrak{H}$ is a *ring* if it is closed under multiplication by scalars, addition, multiplication, and taking the adjoint. That is, if A and B belong to R and a and b are scalars, then $aA + bB$, AB, and A^+ belong to R. (This is also referred to as a *algebra, symmetric ring, ring with involution, etc.)

We will be mainly interested in rings which contain the identity. An example is the ring of all bounded operators on the space.

Definitions. A ring R is *weakly closed* if it is closed in the weak operator topology, that is if A_n are a sequence of operators in R such that for any vectors f, g in $\mathfrak{H}$, $(f, A_n g) \to (f, Ag)$ for some bounded operator A, then A belongs to R. A vector f in $\mathfrak{H}$ is a *cyclic vector* for a ring R if the set of vectors Af for all A in R is dense in $\mathfrak{H}$, that is given any vector g in $\mathfrak{H}$ there exists a sequence of operators A_n in R such that $||A_n f - g|| \to 0$. A ring R is *irreducible* if there is no proper subspace of $\mathfrak{H}$ which is invariant under R, that is if there is no subspace $\mathfrak{M}$ of $\mathfrak{H}$, different from $\mathfrak{H}$ and from the zero element, such that Af belongs to $\mathfrak{M}$ for every A in R and f in $\mathfrak{M}$.

A ring R is irreducible if and only if every bounded operator which commutes with every operator in R is a scalar multiple of the identity operator (scalar operator). An irreducible ring is identical with the ring of all bounded operators. Every non-zero vector is a cyclic vector for a ring if and only if the ring is irreducible.

[8] F. V. Murray and J. von Neumann, Ann. Math. **37**, 116 (1936); J. von Neumann, ibid. **50**, 401 (1949).

[9] M. A. Naimark and S. V. Fomin, Trans. Am. Math. Soc. **5**, 35 (1957).

[10] M. A. Naimark, *Normed Rings*, translated by L. F. Boron (P. Noordhoff, Groningen, 1959).

[11] The assumption of separability of the Hilbert space is probably not needed to obtain most of the results of this paper, but we need it in order to be able to directly apply the considerable work of von Neumann.

Definitions. If F is a set of bounded linear operators on $\mathfrak{H}$, the *commutant* F' of F is the set of bounded linear operators on $\mathfrak{H}$ which commute with all of the operators in F and with the adjoints of all of the operators in F. The *center* Z_R of a ring R is the intersection $Z_R = R \cap R'$ of the ring and its commutant ring, that is Z_R is the set of all elements of R which commute with all of the elements of R.

F' is a weakly closed ring containing the identity operator. By applying this operation a second time we can form F''. Clearly F is contained in F''. In fact F'' is the smallest weakly closed ring which contains F and the identity operator. If F itself is a weakly closed ring containing the identity operator then $F'' = F$. Clearly the center Z_R of a ring R is an Abelian ring; it is weakly closed if R is. If a ring R is Abelian then R is contained in R' or $Z_R = R$. A ring R is irreducible if and only if R' is the ring containing only scalar operators. This is the most simple structure possible for a ring. The next most simple structure is that the center of the ring contain only scalar operators.

Definition. A weakly closed ring R is a *factor* if the center Z_R of R contains only scalar operators, that is if every operator in the ring which commutes with every operator in the ring is a scalar.

Clearly an irreducible ring is a factor, as is the ring of scalar operators. If R is a factor then R' is also a factor.

It is far from true that all rings have the simple structure of irreducible rings and factors. However it turns out that any ring can be built up as a generalized direct sum of either of these kinds of building blocks, just as all representations of many groups can be formed as generalized direct sums of irreducible representations. In order to characterize the general structure of operator rings we need then the concept of a direct integral of Hilbert spaces.

Definition. Let $\sigma(t)$ be the weight function for a Lebesgue-Stieltjes measure on the real line [$\sigma(t)$ is a real, nondecreasing, right continuous bounded function of t for all real t]. For each t let $\mathfrak{H}(t)$ be a Hilbert space, and let $\mathfrak{H}$ be the set of all vector-valued functions f of t, with $f(t)$ a vector in $\mathfrak{H}(t)$, which satisfy the conditions:

(i) For any two functions f and g in $\mathfrak{H}$, $(f(t), g(t))$ is a σ measurable function of t;

(ii) For any f in $\mathfrak{H}$, $||f(t)||^2$ is a σ measurable and also a σ summable function of t, that is

$$\int ||f(t)||^2 \, d\sigma(t) < \infty .$$

An inner product is defined in $\mathfrak{H}$ by

$$(f, g) = \int (f(t), g(t)) \, d\sigma(t).$$

The space $\mathfrak{H}$ with this inner product is called the *direct integral* of the Hilbert spaces $\mathfrak{H}(t)$ with respect to the measure σ.

It can be shown that $\mathfrak{H}$ is a linear space if addition and multiplication by scalars are defined as for functions of t,

$$(af + bg)(t) = af(t) = bg(t)$$

for f, g in $\mathfrak{H}$ and a, b scalars. Two vectors f and g are considered to be identical if $f(t) = g(t)$ for σ almost all t (that is except for a set of σ measure zero). With this identification it can be shown that $\mathfrak{H}$ is a Hilbert space.

Two Hilbert spaces are equivalent if there exists an isometric linear mapping of one onto the other. We will freely identify and interchange equivalent spaces. If in forming the Hilbert space $\mathfrak{H}$ as the direct integral of the Hilbert spaces $\mathfrak{H}(t)$, as in the above definition, we were to change a set of the spaces $\mathfrak{H}(t)$ corresponding to a set of t of σ measure zero, or if we exchanged the measure σ for another measure which is completely continuous with σ (has the same sets of zero measure), then we would get a Hilbert space equivalent to $\mathfrak{H}$. For all f in $\mathfrak{H}$ the vectors $f(t)$ form a linear manifold in $\mathfrak{H}(t)$. We can assume that the closure of this manifold is $\mathfrak{H}(t)$. We will use the notation $\mathfrak{H} = \int \mathfrak{H}(t) \, [d\sigma(t)]^{1/2}$ to denote that $\mathfrak{H}$ is the direct integral of the $\mathfrak{H}(t)$, and we will call $\mathfrak{H}(t)$ the *component spaces* of the direct integral *decomposition* of $\mathfrak{H}$. For an element f in $\mathfrak{H}$ we will write $f = \int f(t) \, [d\sigma(t)]^{1/2}$.

Definitions. Let A be a bounded linear operator on $\mathfrak{H}$. A is *reduced* by the direct integral decomposition $\mathfrak{H} = \int \mathfrak{H}(t) \, [d\sigma(t)]^{1/2}$ if for every f in $\mathfrak{H}$, $(Af)(t) = A(t)f(t)$ where $A(t)$ is a bounded linear operator on $\mathfrak{H}(t)$ for σ almost all t. A set R of bounded linear operators on $\mathfrak{H}$ is reduced if every operator in R is reduced.

We will write $A = \int A(t) \, d\sigma(t)$ and call $A(t)$ the *part* of A in the component space $\mathfrak{H}(t)$. A particular class of operators which are reduced are those which are scalar operators in each component space. These have the form $(Af)(t) = a(t)f(t)$ for any f in $\mathfrak{H}$, where $a(t)$ is a complex valued, σ measurable, essentially bounded function of t. These operators form a weakly closed Abelian ring containing the

identity operator which we will call the *kernel ring* P associated with the given direct integral decomposition. A necessary and sufficient condition for a bounded linear operator A to be reduced is that A be in P'. A necessary and sufficient condition for a ring R to be reduced is then that R be contained in P'.

Conversely, if we are given a weakly closed Abelian ring P containing the identity operator, there exists a direct integral decomposition of the space $\mathfrak{H}$ for which P is the kernel ring. Now our main question is this: Suppose we are given a weakly closed ring R containing the identity operator. Can we find a direct integral decomposition $\mathfrak{H} = \int \mathfrak{H}(t)\ [d\sigma(t)]^{1/2}$ which reduces R such that the part of R in each component space $\mathfrak{H}(t)$ is a factor or is an irreducible ring? The answer found by von Neumann[12] is that any decomposition which has a kernel ring P equal to the center $R \cap R'$ of the ring R will reduce R such that the part of R in the component space $\mathfrak{H}(t)$ is a factor for σ almost all t. A necessary and sufficient condition for the part of R in almost every component space to be an irreducible ring as well as a factor is that the center $R \cap R'$ of R be a maximal Abelian subring of R'.[9] The latter condition is true whenever R contains a subring Q which has the property that $Q = Q'$ (Q is called a maximal Abelian ring).[5] It is also sufficient that R' be equal to the center Z of R, or equivalently that R' be Abelian. In any case we can always find a decomposition of $\mathfrak{H}$ which reduces R into factors, and every operator in the center $R \cap R'$ of R will be reduced such that its part in each component space is a scalar operator.

If, instead of the Lebesgue-Stieltjes measure σ on the real line, we use any Borel measure ρ on a locally compact Hausdorff space, we can define the direct integral of Hilbert spaces and the reduction of operators and of a ring in a similar manner. Within this more general framework there always exists a direct integral decomposition of the Hilbert space which reduces a given weakly closed ring R containing the identity operator for which there is a cyclic vector, such that the part of R in ρ almost every component space is an irreducible ring.[13] ρ can be taken to be a measure on a compact Hausdorff space X which has the property that open sets have positive measure. This has the consequence that either X consists of a single point or else can be divided into two disjoint measurable sets each having positive measure. (Assume that X has at least two distinct points, x and y. Then there exist disjoint open sets V and W in X with x in V and y in W. Since W has positive measure and is contained in the complement of V, both V and the complement of V have positive measure.)

So far we have considered only bounded operators. But the unbounded self-adjoint operators, which are often of interest in physics, can be handled very easily.

Definition. An unbounded self-adjoint operator A is *associated* with the ring R of bounded operators, if every projection operator E_x in the spectral decomposition $A = \int x\, dE_x$ of A belongs to R.

If A is associated with R then we can say that R contains all bounded functions of A. If R is reduced by a direct integral decomposition of the Hilbert space then each projection operator E_x will be reduced and A will act as a reduced operator. In such a case we will say that A is reduced.

III. FIELD OPERATOR RING AND ITS REDUCTION

For every point x of space-time let $\phi(x)$ be a neutral scalar field operator on the separable Hilbert space $\mathfrak{H}$. The rigorous version of this statement is that ϕ is an operator valued distribution over space-time which is defined as follows. Let S be some suitable class of complex testing functions of one or several space-time variables, for example those which are infinitely many times differentiable and vanish at infinity faster than any power of a space-time variable. Then ϕ is a linear mapping of S into linear operators on $\mathfrak{H}$ which we denote symbolically by

$$\int f(x)\phi(x)\, d^4x, \tag{3.1}$$

if the element f of S is a function of a single space-time variable, and by

$$\int f^{(n)}(x_1 \cdots x_n)\phi(x_1) \cdots \phi(x_n)\, d^4x_1 \cdots d^4x_n, \tag{3.2}$$

if $f^{(n)} \in S$ is a function of n variables. It is postulated that all of these operators have a common dense domain so that they can be added to form "smeared polynomials in the field operators." It is also postulated that an operator of the form (3.1) is self-adjoint whenever f is real, and in general that an operator of the form (3.2) is self-adjoint whenever

$$f^{(n)*}(x_1, \cdots x_n) = f^{(n)}(x_n, \cdots x_1). \tag{3.3}$$

We denote the set of all such self-adjoint operators by F, Since any function $f^{(n)}$ belonging to S can be written as

[12] J. von Neumann, reference 8, Theorem VII, p. 460.
[13] M. A. Naimark, reference 10, p. 515.

$$f^{(n)} = g^{(n)} + ih^{(n)},$$

where $g^{(n)}$ and $h^{(n)}$ belong to S and satisfy the reality condition (3.3) [set $g^{(n)}(x_1 \cdots x_n) = \frac{1}{2}\{f^{(n)}(x_1, \cdots x_n) + f^{(n)*}(x_n, \cdots x_1)\}$ and

$$h^{(n)}(x_1, \cdots x_n) = -i/2\ \{f^{(n)}(x_1 \cdots x_n) - f^{(n)*}(x_n, \cdots x_1)\}],$$

we see that every operator of the form (3.2), or any smeared polynomial in the field operators, has form $A + iB$ where A and B are self-adjoint operators which are members of the set F.

Now all of the operators in F will not necessarily be bounded so we can not form a ring containing F. But if we let F' be the set of all bounded linear operators which commute with every operator in F, that is commute with every projection operator which occurs in the spectral decomposition of an operator in F, and if we let $R = F''$, then R and F' $(= R')$ are weakly closed rings containing the identity operator which have the following properties.[14] Every bounded operator in F is in R, as is every projection operator which occurs in the spectral decomposition of an operator in F. In fact, R is the minimal weakly closed ring containing these projection operators. We may say that R is the smallest weakly closed ring containing bounded functions of the operators in F, or containing bounded functions of "smeared polynomials in the field operators."

Every unbounded operator in F is associated with R, so that if R is reduced by a direct integral decomposition of the Hilbert space $\mathfrak{H}$ then F is also reduced according to the terminology introduced in the preceding section. In such a situation we will say that the field operator ϕ is reduced since it gives "smeared polynomials" all of which are reducible.

IV. FIELDS WITH COMMUTATORS WHICH REDUCE TO SCALARS

We will take the statement that the commutator $[\phi(x), \phi(y)]_-$ is a scalar (c number) to mean that for every testing function $f^{(2)}(x, y)$ belonging to S the operator

$$C_{f^{(2)}} = \int f^{(2)}(x, y)[\phi(x)\phi(y) - \phi(y)\phi(x)]\, d^4x\, d^4y$$
$$= \int [f^{(2)}(x, y) - f^{(2)}(y, x)]\phi(x)\phi(y)\, d^4x\, d^4y,$$

is a scalar operator on $\mathfrak{H}$. Similarly, the statement that the field has a vanishing double commutator,

$$[[\phi(x), \phi(y)]_-, \phi(z)]_- = 0, \tag{4.1}$$

implies that when integrated with any testing function $f^{(3)}(x, y, z)$ belonging to S the left-hand side of the above equation gives the zero operator on $\mathfrak{H}$. But, Eq. (4.1) is also taken to imply that

$$[[\phi(x), \phi(y)]_-, \phi(z_1)\phi(z_2)]_- = \phi(z_1)[[\phi(x), \phi(y)]_-, \phi(z_2)]_- + [[\phi(x), \phi(y)]_-, \phi(z_1)]_-\ \phi(z_2) = 0,$$

and by induction that

$$[[\phi(x), \phi(y)]_-, \phi(z_1)\phi(z_2) \cdots \phi(z_n)]_- = 0. \tag{4.2}$$

By integrating the latter equation with testing functions of the form $f^{(n+2)}(x, y, z_1, \cdots z_n) = f^{(2)}(x, y)f^{(n)}(z_1, \cdots z_n)$ where $f^{(2)}$ and $f^{(n)}$ belong to S we can deduce that for any $f^{(2)}$ in S the operator $C_{f^{(2)}}$ commutes with every operator in F, or commutes with every operator in R. This implies that $C_{f^{(2)}}$ is associated with the center $Z_R = R \cap R'$ of R.

Now if R is irreducible we have that $C_{f^{(2)}}$ is a scalar operator. But this conclusion can be drawn from the weaker assumption that R is a factor. In general it is expected that the assumption that R is a factor will be sufficient to prove most of the statements of this type which are of interest in field theory. In any case if R is not a factor we can find a direct integral decomposition of the Hilbert space $\mathfrak{H}$ which reduces R into component rings which are factors.[15] Since for each $f^{(2)}$ in S the commutator operator $C_{f^{(2)}}$ is in the center of R, it will be reduced to component operators each of which is a scalar operator on the component space. Hence, the field theory has the structure of a direct integral of field theories, in each of which the commutator is a scalar, whenever Eq. (4.1) is true.

By the proof of Licht and Toll,[7] using the Jacobi identity, it can be shown that Eq. (4.1) is valid if the commutator $[\phi(x), \phi(y)]_-$ is translation invariant and vanishes for space-like $x - y$. We can summarize the results of this section then in the following.

Theorem If for a set of field operators the double commutator vanishes, that is if equation (4.1) is true, then there exists a direct integral decomposition of the Hilbert space $\mathfrak{H}$ which reduces R into factors, such that for every $f^{(2)}$ belonging to S the commutator operator $C_{f^{(2)}}$ is reduced such that its part in each component space is a scalar operator. A particular case in which this is true is when the commutator $[\phi(x), \phi(y)]_-$ is translation invariant and vanishes for space-like $(x - y)$.

[14] M. A. Naimark, reference 10, pp. 444–450.

[15] These factors will be irreducible if we assume that R contains a complete set of commuting observables, that is a maximal Abelian ring. See Sec. II and J. M. Jauch, reference 5.

V. DIRECT INTEGRALS OF GENERALIZED FREE FIELDS

In this section we are interested in some particular cases where the reduction of the field operator ring R into factors yields a generalized free field theory in each component space of the associated direct integral decomposition of the Hilbert space $\mathfrak{H}$. Hence, we postulate that the following conditions, having to do with Lorentz invariance, the vacuum state, and the mass and energy spectrum, are satisfied:

(a) There exists a set of unitary operators $U(a, \Lambda)$ on $\mathfrak{H}$ which form a true representation of the proper inhomogeneous Lorentz group, that is

$$U(a_1, \Lambda_1)U(a_2, \Lambda_2) = U(a_1 + \Lambda_1 a_2, \Lambda_1\Lambda_2).$$

(b) The field operators transform according to

$$U(a, \Lambda)\phi(x)U^+(a, \Lambda) = \phi(\Lambda x + a).$$

Let U be the ring generated by the $U(a, \Lambda)$. That is $U = \{U(a, \Lambda)\}''$ is the smallest weakly closed ring containing all of the $U(a, \Lambda)$, Clearly U contains the identity operator. We assume that the operators $U(a, \Lambda)$ can be formed as functions of the field operators, or more precisely we postulate that

(c) U is contained in R.[16]

We also assume that there exists at least one invariant state which we call the vacuum state, that is we postulate:

(d) There exists a vector w in $\mathfrak{H}$, with $(w, w) = ||w||^2 = 1$, which is invariant under the inhomogeneous Lorentz transformation operators $U(a, \Lambda)$, that is $U(a, \Lambda)w = w$ for every $U(a, \Lambda)$.[17]

While we have chosen not to postulate that the field operators form an irreducible ring, we need to limit the size of the Hilbert space $\mathfrak{H}$ with respect to the ring R by assuming that all of the state vectors can be obtained by operating with "smeared polynomials in the field operators" on the vacuum state. Thus we postulate the "completeness" property of the field operators that

(e) The vector w of postulate (d) is cyclic for R, that is the vectors of the form Aw, with A in R, are dense in $\mathfrak{H}$.

The energy-momentum operators P_μ are defined as the generators of the translation operators $U(a, 1) = e^{iP\cdot a}$, where $P\cdot a = P_0a_0 - \mathbf{P}\cdot\mathbf{a}$ and the mass operator M is defined by $M^2 = P^2 = P_0^2 - \mathbf{P}^2$. We postulate that the energy-momentum four-vector lies in the forward light cone.

(f) The self-adjoint operators P_0 and M^2 are positive.

Now suppose we decompose the Hilbert space $\mathfrak{H}$ into a direct integral of Hilbert spaces such that the ring R is reduced into factors. We would like to know that the conditions (a) through (f) are satisfied by the parts of the field operators in each component space, so that the theory can be thought of as a direct integral of field theories in each of which these conditions, plus the condition that the ring generated by the field operator is a factor, are satisfied. Since U is contained in R each operator $U(a, \Lambda)$ will be reduced, and the parts of these operators in any one component space will form a representation of the proper inhomogeneous Lorentz group and will transform the parts of the field operators covariantly. It is clear that the part of the ring U in any component space will be contained in the part of the ring R in that component space. Also the properties of the generators of the translation operators, in particular the positiveness of the energy and square of the mass, will hold for the parts of these operators in each component space. Hence we have that conditions (a), (b), (c), and (f) are satisfied for the part of the field theory in each component space. The question which remains then is whether each component space has a vacuum state. If we can find a nonzero, cyclic, invariant state we can always normalize it; so a positive answer to this question is given by the following.

Theorem. Let U be contained in R and let w be a vector which is cyclic for R and invariant under each $U(a, \Lambda)$. Then if R is reduced by a direct integral decomposition of the Hilbert space $\mathfrak{H} = \int \mathfrak{H}(t)\,[d\sigma(t)]^{1/2}$, the component $w(t)$ of the vector $w = \int w(t)\,[d\sigma(t)]^{1/2}$ is a nonzero (normalizable) vector in $\mathfrak{H}(t)$ which is invariant under the part in $\mathfrak{H}(t)$ of each $U(a, \Lambda)$ and is cyclic for the part of R in $\mathfrak{H}(t)$ for σ almost all t.

[16] If we assume that R can be reduced into irreducible rings by a decomposition of the Hilbert space into a *discrete* direct sum then we can prove that U is contained in R. For then, each unitary operator $U(a, \Lambda)$ must map each component space onto a single component space, that is, it must at worst introduce a permutation of the component spaces. But, since each of these operators is continuously connected to the identity operator, it must leave every component space invariant, which implies that U must be reduced and hence be contained in R. See R. Hagedorn, Nuovo cimento Suppl. 12, 73 (1959). This assumption has also been proved, for the case of a Wightman theory with local commutativity, by H. J. Borchers, "On Structure of the Algebra of Field Operators," Institute for Advanced Study, Princeton, New Jersey, (preprint). Borchers has also proved, for a Wightman field, that the Hilbert space is separable. In this regard see also D. Ruelle, "On the Asymptotic Condition in Quantum Field Theory," (to be published, Helv. Phys. Acta), Appendix.

[17] Note that we do not assume that w is the *unique* vector invariant under every $U(a, \Lambda)$.

Proof. Let K be the set of all t for which $||w(t)||^2 > 0$. Since $||w(t)||^2$ is a measurable function, K is a measurable set, and w belongs to $\int_K \mathfrak{H}(t)\ [d\sigma(t)]^{1/2}$ which is a subspace invariant under R. But then we have that

$$Rw \subset R \int_K \mathfrak{H}(t)[d\sigma(t)]^{1/2} = \int_K \mathfrak{H}(t)[d\sigma(t)]^{1/2}.$$

Since w is cyclic this implies that

$$\mathfrak{H} = \int_K \mathfrak{H}(t)[d\sigma(t)]^{1/2},$$

which means that K differs from the space of all t only by a set of σ measure zero. Hence we have proved that $w(t)$ is nonzero for σ almost all t. Let $V = \int V(t)\, d\sigma(t)$ be any operator for which $Vw = w$, which we can write as $||Vw - w||^2 = 0$, or $\int ||V(t)w(t) - w(t)||^2\, d\sigma(t) = 0$, which implies that $V(t)w(t) = w(t)$ for all t except possibly a set of σ measure zero. By letting V be in turn each of a finite number of generating elements of the representation of the proper inhomogeneous Lorentz group by the unitary operators $U(a, \Lambda)$, we can deduce that $w(t)$ is invariant under the part in $\mathfrak{H}(t)$ of each operator $U(a, \Lambda)$ for all t not in the union of the finite number of corresponding sets of zero σ measure, that is for σ almost all t. Finally, for all operators A belonging to R, the vectors

$$\psi_A = Aw = \int A(t)w(t)[d\sigma(t)]^{1/2} = \int \psi_A(t)[d\sigma(t)]^{1/2}$$

form a dense set in $\mathfrak{H}$. For any t the vectors $\psi_A(t) = A(t)w(t)$ form a linear manifold $\mathfrak{M}(t)$ in $\mathfrak{H}(t)$. The closure Cl $(\mathfrak{M}(t))$ of this linear manifold is a subspace of $\mathfrak{H}(t)$ and $\int$ Cl $(\mathfrak{M}(t))\ [d\sigma(t)]^{1/2}$ is a subspace of $\mathfrak{H}$ which contains the dense set of vectors ψ_A. Hence this subspace is equal to the whole of $\mathfrak{H}$ and Cl $(\mathfrak{M}(t)) = \mathfrak{H}(t)$ for σ almost all t, or the vectors $A(t)w(t)$ are dense in $\mathfrak{H}(t)$ for σ almost all t. Hence we have shown that $w(t)$ is cyclic for the part of R in $\mathfrak{H}(t)$ for σ almost all t, which completes the proof of the theorem.

We next note that, if we make the additional assumption that the vacuum vector w represents the unique state which is invariant under the representation of the proper inhomogeneous Lorentz group, we can prove that the ring R generated by the field operators is a factor or even that it is irreducible. In other words, under the assumptions we have made, the postulate of the irreducibility of the field operators is implied by the postulate of the uniqueness of the vacuum.

Theorem. Let the ring U generated by the representation of the proper inhomogeneous Lorentz group belong to the ring R generated by the field operators. If there exists a vector w which is cyclic for R and which is the unique (up to a scalar factor) vector invariant under each $U(a, \Lambda)$, then R is a factor which is in fact irreducible.[18]

Proof. Let $\mathfrak{H} = \int \mathfrak{H}(t)\ [d\sigma(t)]^{1/2}$ be a decomposition of the Hilbert space which reduces R into factors. Then in the space of all t there is either only one set, consisting of a single point, which has positive measure, in which case the reduction of R is trivial and R is itself a factor, or there is a set K of t such that both K and its complement K^c have positive measure. In the latter case let E_K be the projection operator defined by

$$(E_K f)(t) = I_K(t)f(t)$$

for any $f = \int f(t)\ [d\sigma(t)]^{1/2}$ belonging to $\mathfrak{H}$, where $I_K(t)$ is the real function of t which is equal to one when t is in K and is equal to zero otherwise. Then E_K belongs to R'. For any inhomogeneous Lorentz transformation operator $U(a, \Lambda)$ we then have that

$$U(a, \Lambda)E_K w = E_K U(a, \Lambda)w = E_K w,$$

or $E_K w$ is invariant under each $U(a, \Lambda)$. Now by the preceding theorem we have that $||E_K w||^2 = \int_K ||w(t)||^2\, d\sigma(t) \neq 0$, for otherwise $||w(t)||^2$ would vanish on the set K of positive measure. Similarly, since K^c has positive measure, we can deduce that $w - E_K w = (1 - E_K)w = E_{K^c} w \neq 0$. But this contradicts the uniqueness of w. Hence we must conclude that R is a factor.

By using a direct integral decomposition of the Hilbert space that reduces R into irreducible rings instead of just into factors we can construct a completely analogous argument to show that R must be irreducible. This completes the proof of the theorem. [Note that the above proofs do not depend on the specific transformation properties of the field operators or the mass and energy conditions of our postulates (b) and (f). Neither do they depend on any properties of the Lorentz group. The proof remains valid for any symmetry group, e.g., the Galilei group.]

We have chosen not to postulate that the ring R generated by the field operators is a factor (or an irreducible ring), but to investigate the consequences of the weaker postulates that the ring U generated

[18] This result is also contained in the work of H. J. Borchers, and of D. Ruelle,[16] within the framework of Wightman field theory. In addition, the converse of this theorem, that irreducibility implies uniqueness of the vacuum, has been proved for Wightman theories by Borchers. The work reported in this paper was done independently of that of Borchers and that of Ruelle.

by the unitary representation of the proper inhomogeneous Lorentz group is contained in R and that there exists a vector w which is invariant under each $U(a, \Lambda)$ and cyclic for R. We have seen that we can have two kinds of situations. If we postulate that the vacuum vector w is the unique vector which is invariant under each $U(a, \Lambda)$, we can prove that R is irreducible. If we do not postulate the uniqueness of the vacuum, R may not be a factor, but we can always reduce it into factors and the properties associated with Lorentz covariance and the vacuum will hold in the part of the field theory in each component space of the direct integral decomposition which effects this reduction.

In particular, suppose that a field theory satisfies postulates (a) through (f) and in addition has the property that the commutator commutes with all field operators, or more precisely $C_{f^{(2)}}$ belongs to $R \cap R'$ for all $f^{(2)}$ in S. Then, if we decompose the Hilbert space so that R is reduced into factors, the commutator operator $C_{f^{(2)}}$ will be a scalar in each component space. But the postulates (a) through (f) will also be valid for the part of the theory in each component space. By the argument of Licht and Toll[7] the commutator then has the generalized free field form in each component space and the part of the theory in each component space is equivalent to a generalized free field theory. Thus we summarize and conclude with the following statement: If a field theory satisfies postulates (a) through (f) it is equivalent to a direct integral of generalized free field theories if it has the property that $[[\phi(x), \phi(y)]_-, \phi(z)]_- = 0$. In particular, it will have the latter property whenever the commutator is translation invariant and vanishes for space-like separations of the arguments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Singularities of the Second Type

D. B. FAIRLIE*
Department of Applied Mathematics and Theoretical Physics, University of Cambridge
P. V. LANDSHOFF†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
J. NUTTALL*
St. John's College, Cambridge
J. C. POLKINGHORNE‡
Institute for Advanced Study, Princeton, New Jersey
(Received February 27, 1962)

A new class of singularities associated with Landau-Cutkosky diagrams is investigated. They correspond to solutions of the Landau equations with infinite internal momenta. Some of their properties depend on the spins of the participating particles. It is shown that a subset of these singularities, called pure second-type singularities, does not appear on the physical sheet.

1. INTRODUCTION

BECAUSE of the great difficulty of the rigorous derivation of the analytic properties of matrix elements from the general axioms of quantum field theory, much attention has been given to the corresponding problem in perturbation theory.[1] Here the situation is somewhat simpler because the contribution from a given Feynman graph is known, in the form of a multiple integral. With the use of a generalization of the now well-known lemma of Hadamard,[2] the positions of all possible singularities of such an integral may be obtained, in principle, by

* N.A.T.O. Fellow.

† On leave of absence from St. John's College, Cambridge. Supported in part by the U. S. Air Force Office of Scientific Research, Air Research and Development Command.

‡ On leave of absence from Trinity College and Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Supported in part by the U. S. Air Force under grant No. AF-AFOSR-61-19 monitored by the Air Force Office of Scientific Research, Air Research and Development Command.

[1] See, for example, the lectures of J. C. Polkinghorne and of R. J. Eden in *Brandeis University Summer Institute, Lecture Notes I, 1961* (New York, 1962).

[2] This lemma was first used in quantum field theory by R. J. Eden, Proc. Roy. Soc. (London) **A210**, 388 (1952). For derivation of the Landau equations see J. C. Polkinghorne and G. R. Screaton, Nuovo cimento **15**, 289 (1960).

by the unitary representation of the proper inhomogeneous Lorentz group is contained in R and that there exists a vector w which is invariant under each $U(a, \Lambda)$ and cyclic for R. We have seen that we can have two kinds of situations. If we postulate that the vacuum vector w is the unique vector which is invariant under each $U(a, \Lambda)$, we can prove that R is irreducible. If we do not postulate the uniqueness of the vacuum, R may not be a factor, but we can always reduce it into factors and the properties associated with Lorentz covariance and the vacuum will hold in the part of the field theory in each component space of the direct integral decomposition which effects this reduction.

In particular, suppose that a field theory satisfies postulates (a) through (f) and in addition has the property that the commutator commutes with all field operators, or more precisely $C_{f^{(2)}}$ belongs to $R \cap R'$ for all $f^{(2)}$ in S. Then, if we decompose the Hilbert space so that R is reduced into factors, the commutator operator $C_{f^{(2)}}$ will be a scalar in each component space. But the postulates (a) through (f) will also be valid for the part of the theory in each component space. By the argument of Licht and Toll[7] the commutator then has the generalized free field form in each component space and the part of the theory in each component space is equivalent to a generalized free field theory. Thus we summarize and conclude with the following statement: If a field theory satisfies postulates (a) through (f) it is equivalent to a direct integral of generalized free field theories if it has the property that $[[\phi(x), \phi(y)]_-, \phi(z)]_- = 0$. In particular, it will have the latter property whenever the commutator is translation invariant and vanishes for space-like separations of the arguments.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Singularities of the Second Type

D. B. FAIRLIE*
Department of Applied Mathematics and Theoretical Physics, University of Cambridge
P. V. LANDSHOFF†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
J. NUTTALL*
St. John's College, Cambridge
J. C. POLKINGHORNE‡
Institute for Advanced Study, Princeton, New Jersey
(Received February 27, 1962)

A new class of singularities associated with Landau-Cutkosky diagrams is investigated. They correspond to solutions of the Landau equations with infinite internal momenta. Some of their properties depend on the spins of the participating particles. It is shown that a subset of these singularities, called pure second-type singularities, does not appear on the physical sheet.

1. INTRODUCTION

BECAUSE of the great difficulty of the rigorous derivation of the analytic properties of matrix elements from the general axioms of quantum field theory, much attention has been given to the corresponding problem in perturbation theory.[1] Here the situation is somewhat simpler because the contribution from a given Feynman graph is known, in the form of a multiple integral. With the use of a generalization of the now well-known lemma of Hadamard,[2] the positions of all possible singularities of such an integral may be obtained, in principle, by

* N.A.T.O. Fellow.

† On leave of absence from St. John's College, Cambridge. Supported in part by the U. S. Air Force Office of Scientific Research, Air Research and Development Command.

‡ On leave of absence from Trinity College and Department of Applied Mathematics and Theoretical Physics, University of Cambridge. Supported in part by the U. S. Air Force under grant No. AF-AFOSR-61-19 monitored by the Air Force Office of Scientific Research, Air Research and Development Command.

[1] See, for example, the lectures of J. C. Polkinghorne and of R. J. Eden in *Brandeis University Summer Institute, Lecture Notes I, 1961* (New York, 1962).

[2] This lemma was first used in quantum field theory by R. J. Eden, Proc. Roy. Soc. (London) **A210**, 388 (1952). For derivation of the Landau equations see J. C. Polkinghorne and G. R. Screaton, Nuovo cimento **15**, 289 (1960).

solving a set of simultaneous equations. These equations are commonly known as the Landau equations.[3]

The Landau equations are in most cases too complicated to solve algebraically and a prescription has been given for obtaining solutions by means of a geometrical construction.[3,4] This is done by drawing a so-called dual diagram, which is a vector diagram in which each four-momentum of the Feynman graph is on the mass shell.

One is tempted to suppose that these dual diagrams have a general applicability, independent of perturbation theory.[3] This hope is strengthened by the methods found for evaluating the discontinuity across the cut attached to a given Landau singularity: in the prescription for this evaluation no reference need be made to perturbation theory.[5] In fact it has recently been shown that the presence of all the singularities of perturbation theory is necessary in any unitary theory.[6]

Our main purpose here is to show that the conventional dual diagrams do not represent all possible solutions of the Landau equations. The extra solutions will be called *second-type* solutions and correspond to infinite values for some of the components of the internal momenta in the Feynman graph. This does not imply that the resulting singularities of the Feynman integral depend on the range of integration over internal momenta being infinite. They would be present also if the range of integration were finite and arise from distortions of the integration contours, consequent on analytic continuation, extending to infinity. In the case of the finite integral there are additional singularities associated with the end points of the integration. The latter are absent in the infinite integral because the infinite contours of integration are topologically closed.

After a review of conventional dual diagram analysis in Sec. 2, we give in Sec. 3 some simple examples of second-type singularities. These are all of pure second type, for which the positions of the singularities are independent of the masses of the internal particles. They are given by the Gram determinant equation[7]

$$\det \vec{p}_i \cdot \vec{p}_j = 0. \tag{1}$$

Here the $\vec{p}_i$ represent any $(E-1)$ of the E external momenta of the graph.

A general analysis of pure second-type singularities is described in Secs. 4 and 5. This analysis requires some modification when applied to single-loop graphs and this is discussed in Sec. 6. Here there arises the property that the singularity may be absent from all Riemann sheets of the function. The condition for the presence of the singularity involves the dimensionality of space, the spins of the participating particles, and the details of their interactions. It has previously been observed that the dimensionality of space has an important influence on analyticity properties,[8,9] but the question of spins has been supposed to be irrelevant.

In Sec. 7 mixed second-type singularities are analyzed in which some internal momenta are infinite and others not. The positions of the singularities depend on the internal masses associated with the noninfinite momenta. In Sec. 8 we discuss the application to second-type singularities of the methods that have been developed[10,11] for investigating the physical sheet properties of Landau singularities. It is concluded that for the three- and four-point function pure second-type singularities do not appear on the physical sheet.

2. CONVENTIONAL DUAL DIAGRAMS

We first briefly review the conventional dual diagram analysis and establish our notation. The contribution from a general Feynman graph is the multiple integral[7]

$$I(z) = c \int \frac{\nu(\vec{q})\ \delta(\sum_i \alpha_i - 1) \prod_i d\alpha_i \prod_j d^n\vec{k}_j}{\psi^N}, \tag{2}$$

with

$$\psi(\vec{p}, \vec{k}, \alpha) = \sum_i \alpha_i(\vec{q}_i - m_i^2). \tag{3}$$

Here N is the number of internal lines of the graph and α_i, $\vec{q}_i$, m_i are, respectively, the Feynman integration parameter, the momentum, and the mass associated with the ith line. $\nu(\vec{q})$ is a polynomial whose structure depends on the spins of the internal particles; to begin with we shall suppose all particles to be spinless, so that $\nu(\vec{q}) = 1$. The integral I is a function

[3] L. D. Landau, Nuclear Phys. **13**, 181 (1959).
[4] J. C. Taylor, Phys. Rev. **117**, 261 (1960).
[5] R. Cutkosky, J. Math. Phys. **1**, 429 (1960).
[6] J. C. Polkinghorne, Nuovo cimento **23**, 360 (1962); H. P. Stapp, Phys. Rev. **125**, 2139 (1962).
[7] We shall use arrows throughout to denote vectors in Lorentz space. Bold type is used to denote vectors and matrices in the space introduced in Sec. 2.
[8] P. V. Landshoff, Nuclear Phys. **20**, 129 (1960).
[9] G. Källèn and H. Wilhelmsson, Kgl. Danske Videnskb. Selskab, Mat-fys. Skrifter **1**, No. 9 (1959).
[10] R. J. Eden, Phys. Rev. **121**, 1566 (1960); P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Nuovo cimento **19**, 939 (1961); R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, Phys. Rev. **122**, 307 (1961).
[11] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. **2**, 656 (1961).

of various scalar products z of the external vectors $\vec{p}$ for the graph. n is the dimensionality of Lorentz space and so of course is usually taken to be equal to 4.

The integration variables $\vec{k}_i$ in (2) are a set of independent momenta, one running around each independent closed loop of the Feynman graph. Each momentum $\vec{q}_i$ is composed of sums and differences of the $\vec{k}$ and $\vec{p}$. Hence ψ in (3) has the structure[7]

$$\psi(\vec{p}, \vec{k}, \alpha) = \vec{\mathbf{k}}^T \cdot \mathbf{A}\vec{\mathbf{k}} - 2\vec{\mathbf{k}}^T \cdot \mathbf{B}\vec{\mathbf{p}} + (\vec{\mathbf{p}}^T \cdot \mathbf{\Gamma}\vec{\mathbf{p}} - \sigma), \quad (4)$$

where $\sigma = \sum_i \alpha_i m_i^2$.

Here $\mathbf{A}$, $\mathbf{B}$, $\mathbf{\Gamma}$ are, respectively, $l \times l$, $l \times (E-1)$, $(E-1) \times (E-1)$ matrices whose elements are linear in α, and l is the number of independent closed loops in the Feynman graph. We recall that E is the number of external lines of the graph, so that there are $(E-1)$ independent vectors $\vec{p}$ occurring in (4). $\vec{\mathbf{k}}$ and $\vec{\mathbf{p}}$ are vectors in the space of the matrices; in this space $\vec{\mathbf{k}}$ has l components which are the $\vec{k}_i$, and $\vec{\mathbf{p}}$ has $(E-1)$ components which are the independent $\vec{p}$. Each such component is itself a Lorentz vector. Thus, for example, $\vec{\mathbf{k}}^T \cdot \mathbf{A}\vec{\mathbf{k}}$ represents a double sum

$$k_i^\mu A_{ij} k_j^\mu$$

both over the matrix indices i, j, and over the Lorentz index μ.

It is a simple matter to perform the integration over the $\vec{k}$ in (2) explicitly. The result, apart from a numerical factor, is

$$I(z) = \int \frac{C^{N-(1/2)n(l+1)}\, \delta(\sum_i \alpha_i - 1) \prod_i d\alpha_i}{D^{N-(1/2)nl}}. \quad (5)$$

Here

$$C = \det \mathbf{A} \quad (6)$$

and $D = CD'$, where D' is the result of eliminating $\vec{k}$ from ψ according to the equations

$$\partial\psi/\partial\vec{k}_j = 0 \quad \text{for each} \quad j. \quad (7)$$

In the notation of (4) these equations are

$$\mathbf{A}\vec{\mathbf{k}} = \mathbf{B}\vec{\mathbf{p}} \quad (8)$$

and lead, writing $\mathbf{X} = \text{adj } \mathbf{A}$, to the equation

$$D = -(\mathbf{B}\vec{\mathbf{p}})^T \cdot \mathbf{X}(\mathbf{B}\vec{\mathbf{p}}) + (\vec{\mathbf{p}}^T \cdot \mathbf{\Gamma}\vec{\mathbf{p}} - \sigma)C. \quad (9)$$

This last result is valid even when $C = 0$, though D' is not then defined.

According to the generalized Hadamard lemma[2] the necessary conditions for a singularity of I are, using the representation (5),

$$\alpha_i\, \partial D/\partial\alpha_i = 0 \quad \text{for each} \quad i. \quad (10)$$

Equations (10), together with (7), are equivalent to those obtained on application of the lemma directly to the representation (2):

$$\partial\psi/\partial\vec{k}_j = 0 \quad \text{for each} \quad j, \quad (7)$$

and

$$\alpha_i\, \partial\psi/\partial\alpha_i = 0 \quad \text{for each} \quad i. \quad (11)$$

Using (7) and (11) we obtain from (3) the Landau equations

$$\sum_i \alpha_i \vec{q}_i = 0 \quad \text{for each} \quad j, \quad (12)$$

and

$$\alpha_i(\vec{q}_i^2 - m_i^2) = 0 \quad \text{for each} \quad i, \quad (13)$$

where $\sum_i$ in (12) denotes summation round the jth closed loop of the graph.

Disregarding for the moment the possibility of any $\alpha_i = 0$, we see that (13), together with momentum conservation, leads to a vector diagram in which each line is on the mass shell. The conditions (12) determine the α_i and in addition impose certain geometrical constraints on the diagram. The vector diagram with these constraints is commonly known as the dual diagram for the graph. For example, in the case of the single-loop triangle graph, (12) implies a linear relation among the three internal vectors, so that these vectors and therefore the whole dual diagram are in a plane. (Unfortunately the geometrical constraints implied by (12) are not always as simple as this). Solutions of (12) and (13), for which some $\alpha_i = 0$, lead to dual diagrams for graphs obtained from the given one by "short circuiting" the corresponding internal lines.

The general dual diagram contains a closed polygon formed by the external vectors, thus ensuring the momentum is conserved overall. The squares of the lengths of the various diagonals of the polygon are just the scalar products z. The remaining lines of the dual diagram, the internal vectors $\vec{q}$, impose a single constraint on the shape of the polygon and thus one obtains a single equation, relating the z and involving the internal masses m_i, that is the manifold of possible singularities in multidimensional complex z space. This manifold is known as the Landau curve.

It will be found that for a second-type singularity the equation of the manifold in z space is just the determinantal relation (1) that arises when the dimensions of the space spanned by the E external vectors is less than $(E-1)$, or more generally when

there exists a subspace of dimension less than $(E-1)$ such that the component of each external vector $\vec{p}$ perpendicular to this subspace has zero length. The equation does not involve the internal mass m_i and so does not correspond to a dual diagram of the conventional type.

3. EXAMPLES OF SECOND TYPE SINGULARITIES

The simplest graph that one can consider is that of Fig. 1, the single-loop self-energy graph. Unfortunately this is divergent in four dimensions but we can consider it when $n = 3$ and it converges. The conventional dual diagram analysis reviewed in Sec. 2 gives as the equation of the Landau curve

$$\lambda(s, m_1^2, m_2^2) \equiv [s-(m_1+m_2)^2][s-(m_1-m_2)^2] = 0, \quad (14)$$

where $\vec{p}^2 = s$ is the square of the external momentum and m_1, m_2 are the internal masses. This equation combines the normal threshold $s = (m_1+m_2)^2$ with the pseudothreshold $s = (m_1-m_2)^2$, the former representing a singularity on the physical sheet and the latter only on unphysical sheets.

According to the prescription of Cutkosky[5] the discontinuity of the Feynman amplitude across the cut attached to either of these singularities is

$$\int \delta(\vec{k}^2 - m_1^2)\ \delta[(\vec{p}+\vec{k})^2 - m_2]^2\, d^n\vec{k}.$$

For $n = 3$ the result of this integration is proportional to

$$1/\sqrt{s}. \quad (15)$$

This expression has a singularity at $s = 0$ that is not given by the conventional dual diagram analysis. Since (15) is the difference between the values of the original Feynman function on two of its Riemann sheets the original function possesses this singularity also, though not on the physical sheet. This provides an example of a second-type singularity.

A simple explanation for its occurrence can be given in this particular case. The Feynman function is

$$I = \int d^3\vec{k}\, \frac{1}{(\vec{k}^2 - m_1^2)[(\vec{k}+\vec{p})^2 - m_2^2]}. \quad (16)$$

The content of the Landau equations (12) and (13) is that the spheres $\vec{k}^2 = m_1^2$ and $(\vec{k}+\vec{p})^2 = m_2^2$ in $\vec{k}$ space, that are the surfaces of singularity of the integrand, should touch. The condition for them to touch at a finite point is

$$\vec{p}^2 = (m_1 \pm m_2)^2.$$

However, if their centers coincide or are joined by a zero-length vector, that is if $\vec{p}^2 = 0$, they touch at infinity, giving the second-type singularity. We shall find that all the new singularities are associated with pinches occurring at infinity in the $\vec{k}$ integrations.

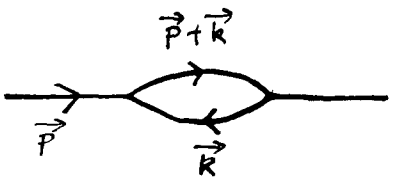

FIG. 1. The single-loop self-energy part.

Another instructive example is provided by the single-loop triangle graph. For this graph the integration for a derivative of the Feynman function for $n = 4$ was performed by Källèn and Wightman[12] and only the singularities of the conventional Landau analysis appear.[13] The function itself for $n = 4$ has been computed by Wu[14] and contains an extra singularity which is a second-type singularity. Its presence was pointed out by Cutkosky,[5] who finds that the discontinuity across the cut attached to the leading singularity corresponding to all the lines being on the mass shell is

$$1/[4\sqrt{\lambda(z_1, z_2, z_3)}],$$

where the z are the squares of the three external momenta. Thus the Feynman integral has a singularity given by

$$\lambda(z_1, z_2, z_3) = 0 \quad (17)$$

on the unphysical sheet reached by passing from the physical sheet through the cut attached to the leading singularity.[15] Equation (17) is just the Gram determinantal condition (1).

The integral computed by Källèn and Wightman differs from the Feynman integral in that for it the power of D in the representation (5) is two instead of one. It is of interest that for $n = 2$ the KW integral is just the correct Feynman integral, so that the presence of the second-type singularity for the triangle graph depends on the dimensionality of Lorentz space. The four-point single-loop graph has the same property: for $n = 4$ the second-type singularity does not appear, but it does for $n = 5$.

4. SECOND TYPE SOLUTIONS OF THE LANDAU EQUATIONS

We consider now graphs with more than one loop and show that when C, defined by (6), is put equal

[12] G. Källèn and A. S. Wightman, Kgl. Danske. Videnskb. Selskab Mat.-fys. Skrifter 1, No. 6 (1958).
[13] M. Fowler, P. V. Landshoff, and R. W. Lardner, Nuovo cimento 17, 956 (1960).
[14] A. C. T. Wu (private communication).
[15] In the integral of Källèn and Wightman this singularity is a pole and so has no cut attached to it.

to zero there are many solutions of Eqs. (10). Since the latter equations are the Hadamard equations for the representation (5), corresponding singularities for $I(z)$ are expected to occur in general. Our new solutions of Eqs. (10) also satisfy Eqs. (12) but are not at first sight generally compatible with (13). This is paradoxical since we have said that (12) and (13) are equivalent to (10); the resolution of this difficulty will be given in Sec. 5.

Let $C = 0$ and suppose that $l \times l$ matrix A is of rank $(l - 1)$. Then the adjoint of A can be written

$$\mathbf{X} = \mathbf{K}\mathbf{K}^T, \tag{18}$$

where the column matrix $\mathbf{K}$ satisfies

$$\mathbf{A}\mathbf{K} = \mathbf{0}. \tag{19}$$

Let $\mathbf{\Lambda}$ be any column matrix and $\vec{\lambda}$ a zero-length vector in Lorentz space:

$$\vec{\lambda}^2 = 0. \tag{20}$$

Suppose also

$$\vec{\lambda}\cdot\vec{p} = 0 \tag{21}$$

for each external vector $\vec{p}$ of the graph under consideration.

Let $\vec{k}$ be vectors that are linear combinations of the $\vec{p}$ and satisfy the equation

$$\mathbf{A}\vec{\mathbf{k}} = \mathbf{B}\vec{\mathbf{p}} - \mathbf{\Lambda}\vec{\lambda}. \tag{22}$$

A necessary and sufficient condition for such $\vec{k}$ to exist is that

$$\mathbf{K}^T\mathbf{B}\,\vec{\mathbf{p}} = \vec{\lambda}\mathbf{K}^T\mathbf{\Lambda}. \tag{23}$$

Equations (21) and (23) together require that there be a linear combination of the vectors $\vec{p}$ that is equal to zero-length vector whose scalar product with each $\vec{p}$ is zero. Thus they lead to the Gram condition (1) characteristic of all pure second-type singularity curves.

If we denote differentiation with respect to α_i by the subscript i we obtain from (9)

$$D_i = -(\mathbf{B}\vec{\mathbf{p}})^T\cdot\mathbf{X}_i\mathbf{A}\vec{\mathbf{k}} + (\vec{\mathbf{p}}^T\cdot\mathbf{\Gamma}\vec{\mathbf{p}} - \sigma)C_i. \tag{24}$$

Here we have used the condition $C = 0$ and Eqs. (18) and (19), together with the fact that the matrix $\mathbf{A}$ is symmetric, to eliminate several terms. Now

$$\mathbf{X}\mathbf{A} = C,$$

so that

$$\mathbf{X}_i\mathbf{A} + \mathbf{X}\mathbf{A}_i = C_i,$$

and hence

$$D_i = C_i\psi(\vec{p}, \vec{k}, \alpha). \tag{25}$$

Thus the Hadamard equations $D_i = 0$ are satisfied for $\vec{k}$ chosen as above provided that $\psi = 0$ also. Thus there are many solutions to these equations.

If in the above we choose $\vec{\lambda} = 0$ we may dispense with the condition that the solution $\vec{\mathbf{k}}$ to (22) be a linear combination of the $\vec{p}$ and we obtain similar results. Thus the Hadamard equations are solved by applying (12) to any vector diagram such that the $(E - 1)$ vectors $\vec{p}$ that are generally independent fulfil the Gram condition (1) and the single constraint

$$\psi(\vec{p}, \vec{k}, \alpha) = 0. \tag{26}$$

It is not necessary that the internal vectors be on the mass shell, so that (13) need not hold.

5. PINCHES AT INFINITY

We now show how these results obtained for the representation (5) can be reconciled with representation (2). We shall find that the second-type solutions arise out of solutions of the Landau equations for which the $\vec{k}$ are infinite. We find that the vector diagram of the end of Sec. 4 has a different significance from a conventional dual diagram.

In order to make infinity accessible to our analysis, it is convenient to introduce homogeneous coordinates[16]:

$$\vec{\mathbf{k}} = \vec{\mathbf{K}}/\zeta, \tag{27}$$

so that the hyperplane at infinity is

$$\zeta = 0.$$

Define

$$\psi(\vec{p}, \vec{k}, \zeta, \alpha) = \vec{\mathbf{K}}^T\cdot\mathbf{A}\vec{\mathbf{K}} - 2\vec{\mathbf{K}}^T\cdot\mathbf{B}\vec{\mathbf{p}}\zeta + (\vec{\mathbf{p}}^T\cdot\mathbf{\Gamma}\vec{p} - \sigma)\zeta^2. \tag{28}$$

The conditions for an extremum of ψ are

$$\partial\psi/\partial\vec{\mathbf{K}} = 0, \qquad \partial\psi/\partial\zeta = 0, \qquad \partial\psi/\partial\alpha = 0, \tag{29}$$

which on $\zeta = 0$ give

$$\begin{aligned} \mathbf{A}\vec{\mathbf{K}} &= 0, \\ \vec{\mathbf{K}}^T\cdot\mathbf{B}\vec{\mathbf{p}} &= 0, \\ \vec{\mathbf{K}}^T\cdot\mathbf{A}_i\vec{\mathbf{K}} &= 0. \end{aligned} \tag{30}$$

Equations (30) are given a solution in the following way. Choose the α so that $C = 0$. There then exists a column matrix $\mathbf{K}$ satisfying (19). Take

$$\vec{\mathbf{K}} = \mathbf{K}\vec{\lambda}', \tag{31}$$

where $\vec{\lambda}'$ is a Lorentz vector, chosen such that

$$\vec{\lambda}'\cdot(\mathbf{K}^T\mathbf{B}\vec{\mathbf{p}}) = 0 \tag{32}$$

[16] This may be regarded as a transformation of the integration variables. Then there is one variable too many, so that one of the components of one of the $\vec{\mathbf{K}}$ must be taken constant. The Jacobian is simple to evaluate.

to satisfy the second equation in (30) and such that

$$\vec{\lambda}'^2 = 0 \tag{33}$$

to satisfy the third equation in (30).

The above solution exists for all $\vec{p}$ and we can therefore deduce that it does not correspond to a singularity of the Feynman integral.

The reason for this is as follows. The general idea behind the Hadamard lemma[2] is that a singularity of the multiple integral arises for z such that the hypercontour over which the integration is performed cannot be continuously distorted so as to avoid intersecting the manifolds on which the integrand is singular, that is, the hypercontour is "pinched" by the singularities. The Landau equations are necessary conditions for such a pinch to occur. But if these equations are satisfied for all z and if the integral is known to exist for some value z_0 of z, they are not sufficient to produce a singularity. This is because, since the integral is defined at z_0, the hypercontour is not trapped at z_0 and so as one continues analytically from z_0 it does not become trapped unless there is a charge in the topological character of the intersection of the hypercontour with the manifold of singularity of the integrand.

We therefore seek the condition for Eqs. (29) to be satisfied also at a point in the integration space adjacent to one of the solutions already found. Let this point be

$$\vec{\mathbf{K}} + \delta\vec{\mathbf{K}}, \qquad \delta\zeta, \qquad \alpha + \delta\alpha.$$

Then (29) requires

$$\begin{aligned} \mathbf{A}\,\delta\vec{\mathbf{K}} &= \mathbf{B}\vec{\mathbf{p}}\,\delta\zeta - \delta\mathbf{A}\mathbf{K}\vec{\lambda}' - \delta\vec{\mathbf{K}}^T\cdot\mathbf{B}\vec{\mathbf{p}} \\ &\quad + (\vec{\mathbf{p}}^T\cdot\boldsymbol{\Gamma}\vec{\mathbf{p}} - \sigma)\,\delta\zeta - \vec{\lambda}'\mathbf{K}^T\cdot\delta\mathbf{B}\vec{\mathbf{p}} = 0, \qquad (34) \\ \vec{\lambda}'\cdot\mathbf{K}^T\mathbf{A}_i\,\delta\mathbf{K} &- \vec{\lambda}'\cdot\mathbf{K}^T\mathbf{B}_i\vec{\mathbf{p}}\,\delta\zeta = 0. \end{aligned}$$

The first of these equations is equivalent to (22) with $\vec{\mathbf{k}}$ set equal to $\delta\vec{\mathbf{K}}/\delta\zeta$. If $\vec{\lambda}'$ is made to satisy (21) rather than the weaker condition (32), the second equation in (34) is equivalent to (26). The third equation of (34) is most simply satisfied by taking $\vec{\lambda}' = \vec{0}$, or otherwise by making $\delta\vec{\mathbf{K}}$ a linear combination of the $\vec{p}$.

We have thus reproduced the conditions for the construction of the vector diagram discussed at the end of Sec. 4. The internal lines of this vector diagram are not on the mass shell, but since they do not in fact represent the internal momenta for the Feynman graph, only increments thereof, this does not imply that for second-type singularities the internal vectors are not on the mass shell.

Finally, we note that for graphs with six or more external lines all points in a four-dimensional space lie on the Gram determinant curve. This does not mean, of course, that the whole of space is singular on some sheet of the function. As before, it is necessary to seek further conditions expressing an increase in the degeneracy of the singularity manifold of the integrand at infinity.

6. SINGLE-LOOP GRAPHS

The general analysis of the preceding two sections does not immediately apply to single-loop graphs for two reasons. In the first place, for a single loop

$$C = \sum_i \alpha_i,$$

and so its vanishing is inconsistent with the δ function in the integral representation. This means that our solutions will correspond to infinite α. The second point of difference is that, for a single-loop graph, the condition (26) cannot in general be met on the Gram determinant curve, despite the considerable freedom in the choice of the α. This is shown in the Appendix.

We briefly summarize the analysis applicable to single loops. The squares of the partial sums of the external momenta $\vec{p}_1 \cdots \vec{p}_E$ are denoted by

$$\begin{aligned} \nu_{ij} &= (\vec{p}_i + \vec{p}_{i+1} + \cdots + \vec{p}_{j-1})^2, \qquad i < j, \\ &= \nu_{ji}, \qquad i > j, \qquad (35) \\ &= 0, \qquad i = j. \end{aligned}$$

Then

$$D = \tfrac{1}{2}\sum_{i,j=1}^{E} \nu_{ij}\alpha_i\alpha_j - \sigma C, \tag{36}$$

where $\sigma = \sum_i \alpha_i m_i^2$ as before. We eliminate the δ function in the integral (5) by replacing

$$\alpha_n = 1 - \sum_{i=1}^{E-1} \alpha_i,$$

transform to homogeneous coordinates[17]

$$\alpha_i = \alpha_i'/\zeta, \qquad i = 1, \cdots E-1,$$

and consider

$$D'(\alpha', \zeta) = \zeta^2\, D(\alpha).$$

The conditions

$$\partial D'/\partial\alpha_i' = 0, \qquad i = 1, \cdots E-1,$$

yield after some manipulation the Gram determinant equation (1). However there is $\partial D'/\partial\zeta$ which must also vanish. In fact it is found to be nonzero, being proportional to ψ. Thus there will only be a singularity if there is a net positive power of ζ in the

[17] Compare with footnote 16.

denominator of the integrand outside D', since then a pinch can occur between $D' = 0$ and $\zeta = 0$ without requiring $\partial D'/\partial\zeta = 0$. The condition for this is obtained from the power of the denominator in (4) and from the Jacobian.[17] It is

$$E < n, \tag{37}$$

where n is the dimensionality of Lorentz space, as before. This result is in accord with the examples discussed in Sec. 3.

The condition (37) is derived on the basis of ν being unity in (2). This is not the case when fermions are involved nor when derivative interactions are introduced. In this case, since for example

$$\frac{\vec{k}^2}{\psi^\rho} = \frac{1}{(\sum\alpha)\psi^{\rho-1}} + \frac{\text{form linear in } \vec{k}}{\psi^\rho},$$

the effective power of D may be reduced. Thus the presence of spins may introduce second-type singularities which otherwise would be absent.[18]

7. MIXED SINGULARITIES

For graphs that contain more than one loop, singularities may occur which correspond to some of the loop momenta being pinched at infinity and the remainder at finite points. We call such singularities *mixed* second-type singularities.

Their occurrence may be understood in the following way. A general graph G may be divided (in a number of ways) into two subgraphs G_1 and G_2. Let the Feynman function for G_1 be f. Then the Feynman function for G is equal to a multiple integral whose integrand is f multiplied by a propagator for each line that is internal to G but not to G_1. A mixed singularity of F occurs when a pure second-type singularity of f pinches with the poles of the propagators.

One may set up a graphical scheme, similar to dual diagram analysis, to obtain equations of the mixed singularity curves. The lines corresponding to the propagators above are put on the mass shell and the external vectors of the subgraph G_1 are drawn in the configuration appropriate to a pure second-type singularity for G_1. There are also geometrical constraints similar to those associated with Eq. (12).

These ideas may be illustrated by the graph of

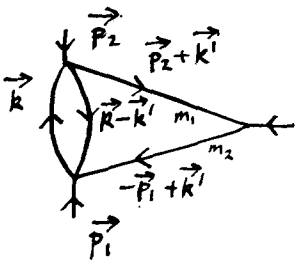

FIG. 2. A simple two-loop diagram.

Fig. 2.[18a] The result of performing the $\vec{k}$ integration in the Feynman integral is

$$F = \int f(\vec{k}'^2)\,\frac{1}{(\vec{p}_2+\vec{k}')^2 - m_1^2}\cdot\frac{1}{(\vec{p}_3-\vec{k}')^2 - m_2^2}\,d\vec{k}',$$

where f is the two-point function for the $\vec{k}$ loop. f has a pure second-type singularity when $\vec{k}^2$ vanishes and the mixed singularity of F arises from the additional conditions that the two internal momenta at the $\vec{p}_1$ vertex be on the mass shell and coplanar with $\vec{k}$. The resulting vector diagram is drawn in Fig. 3 and leads to the equation for the mixed-type solution.

$$2m_2^2(m_1^2 - z_2)/m_2^2 - z_1 = m_2^2 + m_1^2 - z_3 \pm [(m_2^2 + m_1^2 - z_3)^2 - 4m_1^2m_2^2]^{1/2}.$$

It is not known what is the condition analogous to (37) for the actual presence of this singularity.

The analysis of the mixed singularities in terms of the representations (2) or (5) presents some subtleties. This is because it turns out that all the α associated with the subgraph that is in the pure second-type configuration vanish. The general method is conveniently illustrated by a discussion of the specific example we have already considered.

We transform some of the variables:

$$\vec{k} = \vec{K}/\zeta, \qquad \alpha_3 = \zeta\beta_3, \qquad \alpha_4 = \zeta\beta_4, \tag{38}$$

and consider

$$\psi' = \zeta\psi = \zeta\alpha_1[(\vec{p}_2+\vec{k}')^2 - m_1^2] + \zeta\alpha_2[(\vec{p}_1-\vec{k}')^2 - m_2^2] + \beta_3[(\vec{K}-\vec{k}'\zeta)^2 - m_3^2\zeta^2] + \beta_4[\vec{K}^2 - m_4^2\zeta^2]. \tag{39}$$

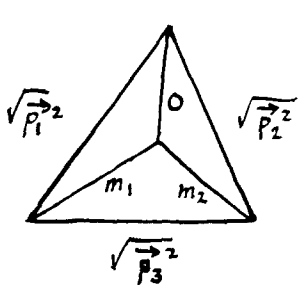

FIG. 3. A mixed singularity dual diagram.

[18] That introduction of a numerator function can have an important effect on analyticity properties is well illustrated by the case of a gauge-invariant theory. Then the generalized Ward identity [V. Takahashi, Nuovo cimento **6**, 371 (1957); K. Nishijima, Phys. Rev. **119**, 485 (1960)] relates the vertex function to the propagator. The latter cannot have complex singularities and therefore also not the former, which is different from the usual situation [reference 13 and P. V. Landshoff and S. B. Treiman, Nuovo cimento **19**, 1249 (1961)].

[18a] The existence of this mixed singularity has been explicitly verified by I. T. Drummond. It has also been discussed by M. Fowler (private communications).

The Hadamard equations for a pinch when $\zeta = 0$ yield

$$\begin{aligned} \beta_3 + \beta_4 &= 0, \\ \vec{K}^2 &= 0, \\ \alpha_1[(\vec{p}_2 + \vec{k}')^2 - m_1^2] + \alpha_2[(\vec{p}_1 - \vec{k}')^2 - m_2^2] & \\ - 2\beta_3 \vec{K}\cdot\vec{k}' &= 0. \end{aligned} \tag{40}$$

These equations do not suffice to determine a curve in z space. It is necessary, therefore, to seek the conditions that they be satisfied also at a neighboring point of the integration space. Thus we change incrementally the variables associated with the pure second-type singular subgraph and obtain:

$$\begin{aligned} (\delta\beta_3 + \delta\beta_4)\vec{K} &= \beta_3\, \delta\zeta\vec{k}', \\ \vec{K}\cdot(\delta\vec{K} - \delta\zeta\vec{k}') &= 0, \\ \vec{K}\cdot\delta\vec{K} &= 0, \\ \vec{k}'\cdot[\beta_3(\delta\vec{K} - \vec{k}'\,\delta\zeta) + \delta\beta_3\vec{K}] &= 0; \end{aligned} \tag{41}$$

$$\begin{aligned} (\vec{p}_2 + \vec{k}')^2 - m_1^2 &= 0, \\ (\vec{p}_1 - \vec{k}')^2 - m_2^2 &= 0; \end{aligned} \tag{42}$$

$$\alpha_1(\vec{p}_2 + \vec{k}') + \alpha_2(\vec{p}_3 - \vec{k}') - \beta_3\vec{K} = 0. \tag{43}$$

Equations (42) are just the conditions stated above that the two internal vectors at the $\vec{p}_1$ vertex be on the mass shell and (43) makes them coplanar with $\vec{K}$. Equations (40) and (41) together are those appropriate to a pure second-type singularity of the $\vec{K}$ loop.

8. PHYSICAL SHEET PROPERTIES

We now consider whether second-type singularities are to be found on the physical sheet. In particular we consider first the pure second-type singularity associated with three- and four-point functions.

For definiteness, consider the four-point function. The existence of single variable dispersion relations for this case shows that not all the second-type curve is singular on the physical sheet. According to previous analysis,[10,11] this is sufficient to ensure that none of the curve is singular on the physical sheet unless either

(i) it has effective intersection with some other curve that is itself singular on the physical sheet or

(ii) the curve contains either acnodes or cusps.

The second possibility can be rejected for the pure second-type curve [Eq. (1)]. Also Γ, the pure second-type curve for scattering, cannot have effective intersection with normal or anomalous threshold curves since these are straight lines parallel to the s, t, or u axes whose position depends on the internal masses and so cannot in general touch Γ whose equation is independent of these masses. There remains the possibility that Γ has effective intersection with leading curves for four-point graphs. However, we can assert that these curves are never singular at these intersections when approached along Γ. This is because Γ is the curve to which the invariants z would be confined if Lorentz space were two dimensional, and if space were two dimensional it would not be possible to draw dual diagrams for the leading curve of a scattering graph.[8,19] Hence one may conclude that the pure second-type curve for the four-point function is not singular on the physical sheet. A similar argument applies to the three-point function.

A corollary of this argument is that starting from a nonsingular point of Γ and *moving on* Γ one cannot enter a sheet in which Γ is singular. This is of significance in connection with analytic properties of partial wave amplitudes since Γ is the boundary of the region of integration when the partial wave projection is made.

The physical sheet properties of mixed singularities can be discussed by the hierarchical type of argument[10] employed for conventional Landau singularities. Acnodes and cusps[11] may be expected to occur in this case.

9. CONCLUSION

We have shown that perturbation theory, or any unitary theory,[6] possesses a new and numerous class of singularities corresponding to solutions of the Landau equations with infinite momenta. A question that awaits analysis is the evaluation of the discontinuity across the cuts attached to these singularities: the Cutkosky analysis[5] does not seem directly applicable to this case.[20]

ACKNOWLEDGMENTS

We wish to thank I. T. Drummond, M. Fowler, and Dr. M. Froissart for useful discussions. This work was begun while the first three of us were at the Summer Institute of Theoretical Physics in the

[19] The analogous result for the contact of the second-type curve $\lambda(z_1, z_2, z_3) = 0$ with the leading curve for the single-loop triangle graph is simple to verify explicitly.

Some apparent exceptions to this argument may seem to occur for diagrams like the square with diagonal (see reference 11) for which the whole dual diagram can be plane at some intersections with Γ. However, these are not effective intersections.

[20] See, however, J. C. Polkinghorne, Nuovo cimento (to be published).

University of Wisconsin and we wish to thank Professor R. G. Sachs for the hospitality we received there. One of us (J. C. P.) is grateful to Professor J. R. Oppenheimer for the hospitality of the Institute for Advanced Study.

APPENDIX

In this Appendix we wish to show that ψ does not vanish for a single loop at a general point on the Gram determinant curve. The reason will be sufficiently illustrated by the particular example of Fig. 4 for the case of a triangular loop.

The α's are determined by

$$\begin{aligned}(\alpha_1 + \alpha_2 + \alpha_3)y &= 0,\\ \alpha_1(x + M_1) + \alpha_2 x + \alpha_3(x - M_3) &= 0,\end{aligned} \tag{A1}$$

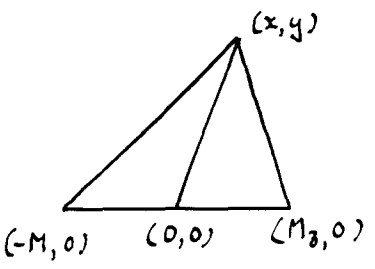

FIG. 4. A diagram for constructing second-type solutions for the triangle loop.

and are $1/M_1$, $-(1/M_1 + 1/M_3)$, $1/M_3$. Then

$$\psi = M_1 + M_3 - \frac{m_1^2}{M_1} + m_2^2\left(\frac{1}{M_1} + \frac{1}{M_3}\right) - \frac{m_3^2}{M_3},$$

and is independent of x and y.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

On Forces and Interactions between Fields

DAVE PANDRES, JR.
The Martin Company,
Denver, Colorado
(Received December 9, 1961)

After a discussion of certain properties of multivalued functions is given, these functions are used to give a scalar representation of electromagnetic theory. In this representation, the four-vector potential **A** appears as the four-gradient of a multivalued function ϕ. The problem of solving the time-dependent Schrödinger equation for a particle of charge e in an electromagnetic field is equivalent to the problem of solving the corresponding field-free equation in a space of multivalued functions $\bar{\psi}$, which is connected to the space of single-valued functions ψ by the relation $\bar{\psi} = U\psi$ where $U = \exp(-ie\phi)$. Further, if $L_0(\bar{\psi}, \mathbf{A})$ is the Lagrangian for a noninteracting particle field $\bar{\psi}$ and photon field **A**, then $L_0(U\psi, \mathbf{A})$, regarded as a functional of ψ rather than $\bar{\psi}$, is the Lagrangian for *interacting* fields ψ and **A**. This suggests that other interactions (e.g., strong and weak interactions) should be introduced by using a transformation T which is more general than U.

More fundamental reasons for believing that this is the case arise when one considers multivalued coordinate transformations. It is shown that any curved Riemannian space may be connected with a flat space by a multivalued coordinate transformation. This is possible because the metric transforms like a tensor under these transformations, but the Riemann curvature symbol does not. If one writes the equations of motion of a free, classical, relativistic particle in flat-space coordinates and assumes that the equations transform covariantly under multivalued coordinate transformations (this is a natural generalization of the equivalence principle), one obtains equations of motion in the curved-space coordinates which describe a particle in a gravitational and electromagnetic field. Unfortunately, however, the electromagnetic field at any point of space-time depends on the four-velocity of the test particle at that point. This defect, together with the known existence of essentially quantum-mechanical forces and the fact that the uncertainty principle prevents one from measuring velocity when position is measured exactly, suggests that one should write the Klein-Gordon (or Dirac) equation for a free particle in flat-space coordinates, and then make a change of independent variables into curved-space coordinates. When this is done, it is seen that a corresponding change of dependent variable $\bar{\psi} = T\psi$ is required in order that the curved-space wave function shall be a single-valued function of the curved-space coordinates. The fully transformed equation describes a particle in a gravitational and electromagnetic field as well as certain other fields whose properties are not fully understood.

I. INTRODUCTION

THE success of Einstein in geometrizing the gravitational field led him and others to search for a way to geometrize electromagnetism. However, no generally accepted theory has appeared, and it now seems that the task of constructing a true unified field theory is more formidable than ever, since the strong and weak interactions would have to be included.

There is a problem in quantum field theory which

University of Wisconsin and we wish to thank Professor R. G. Sachs for the hospitality we received there. One of us (J. C. P.) is grateful to Professor J. R. Oppenheimer for the hospitality of the Institute for Advanced Study.

APPENDIX

In this Appendix we wish to show that ψ does not vanish for a single loop at a general point on the Gram determinant curve. The reason will be sufficiently illustrated by the particular example of Fig. 4 for the case of a triangular loop.

The α's are determined by

$$\begin{aligned}(\alpha_1 + \alpha_2 + \alpha_3)y &= 0,\\ \alpha_1(x + M_1) + \alpha_2 x + \alpha_3(x - M_3) &= 0,\end{aligned} \tag{A1}$$

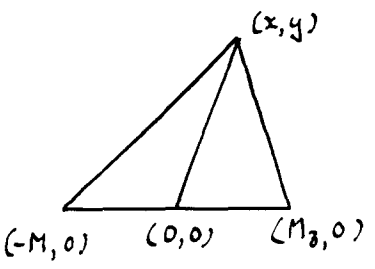

FIG. 4. A diagram for constructing second-type solutions for the triangle loop.

and are $1/M_1$, $-(1/M_1 + 1/M_3)$, $1/M_3$. Then

$$\psi = M_1 + M_3 - \frac{m_1^2}{M_1} + m_2^2\left(\frac{1}{M_1} + \frac{1}{M_3}\right) - \frac{m_3^2}{M_3},$$

and is independent of x and y.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

On Forces and Interactions between Fields

DAVE PANDRES, JR.
The Martin Company,
Denver, Colorado
(Received December 9, 1961)

After a discussion of certain properties of multivalued functions is given, these functions are used to give a scalar representation of electromagnetic theory. In this representation, the four-vector potential **A** appears as the four-gradient of a multivalued function ϕ. The problem of solving the time-dependent Schrödinger equation for a particle of charge e in an electromagnetic field is equivalent to the problem of solving the corresponding field-free equation in a space of multivalued functions $\bar{\psi}$, which is connected to the space of single-valued functions ψ by the relation $\bar{\psi} = U\psi$ where $U = \exp(-ie\phi)$. Further, if $L_0(\bar{\psi}, \mathbf{A})$ is the Lagrangian for a noninteracting particle field $\bar{\psi}$ and photon field **A**, then $L_0(U\psi, \mathbf{A})$, regarded as a functional of ψ rather than $\bar{\psi}$, is the Lagrangian for *interacting* fields ψ and **A**. This suggests that other interactions (e.g., strong and weak interactions) should be introduced by using a transformation T which is more general than U.

More fundamental reasons for believing that this is the case arise when one considers multivalued coordinate transformations. It is shown that any curved Riemannian space may be connected with a flat space by a multivalued coordinate transformation. This is possible because the metric transforms like a tensor under these transformations, but the Riemann curvature symbol does not. If one writes the equations of motion of a free, classical, relativistic particle in flat-space coordinates and assumes that the equations transform covariantly under multivalued coordinate transformations (this is a natural generalization of the equivalence principle), one obtains equations of motion in the curved-space coordinates which describe a particle in a gravitational and electromagnetic field. Unfortunately, however, the electromagnetic field at any point of space-time depends on the four-velocity of the test particle at that point. This defect, together with the known existence of essentially quantum-mechanical forces and the fact that the uncertainty principle prevents one from measuring velocity when position is measured exactly, suggests that one should write the Klein-Gordon (or Dirac) equation for a free particle in flat-space coordinates, and then make a change of independent variables into curved-space coordinates. When this is done, it is seen that a corresponding change of dependent variable $\bar{\psi} = T\psi$ is required in order that the curved-space wave function shall be a single-valued function of the curved-space coordinates. The fully transformed equation describes a particle in a gravitational and electromagnetic field as well as certain other fields whose properties are not fully understood.

I. INTRODUCTION

THE success of Einstein in geometrizing the gravitational field led him and others to search for a way to geometrize electromagnetism. However, no generally accepted theory has appeared, and it now seems that the task of constructing a true unified field theory is more formidable than ever, since the strong and weak interactions would have to be included.

There is a problem in quantum field theory which

is analogous to the classical relativistic problem of constructing a unified field theory. This is the problem of finding general principles which determine the interactions between fields. Interactions are conventionally introduced by adding to the Lagrangian L_0 of the free uncoupled fields an interaction term L_I which must satisfy the requirements of hermiticity, relativistic invariance, and the limitation that it contain no space-time derivatives of field quantities higher than the first. However, it is well known that these requirements are not generally sufficient to determine L_I. In this paper we give an argument which suggests another requirement that must be satisfied, and which displays the connection between the classical and the quantum mechanical problems in a rather direct way. The theory to be developed here depends upon the assumption that the equivalence principle should be generalized to include covariance under multivalued coordinate transformations. Since very little has appeared in the literature concerning multivalued functions, we begin by discussing some of their properties.

II. PROPERTIES OF MULTIVALUED FUNCTIONS

Let $\mathbf{x}$ be a vector whose Cartesian components are $x^1, \cdots, x^4$ and let $\mathbf{A}$ be a vector whose Cartesian components are the single-valued functions $A_1(\mathbf{x}), \cdots, A_4(\mathbf{x})$. We may define a function ϕ by the relation

$$d\phi = A_i\, dx^i. \tag{1}$$

In Eq. (1) and henceforth, we use the summation convention that a repeated lower case index is summed over the range 1 to 4. The usual point of view is that Eq. (1) is integrable [defines $\phi(\mathbf{x})$ up to an arbitrary constant] only if $\mathbf{A}$ is conservative i.e., only if $\partial A_i/\partial x^i = \partial A_i/\partial x^i$. However, we adopt a more general point of view according to which we write from Eq. (1),

$$\phi(\mathbf{x}) = \int_{\mathbf{x}_0}^{\mathbf{x}} \mathbf{A}(\mathbf{x}')\cdot d\mathbf{x}' + \text{const}, \tag{2}$$

where $\mathbf{x}_0$ is some fixed point in $\mathbf{x}$ space. If $\mathbf{A}$ is not conservative, $\phi(\mathbf{x})$ may be regarded as a multivalued function whose value depends on the path taken from $\mathbf{x}_0$ to $\mathbf{x}$. Consider the change in ϕ when $\mathbf{x}$ moves once around an arbitrary closed curve C. Clearly, we have

$$\Delta\phi = \int_C \mathbf{A}\cdot d\mathbf{x}. \tag{3}$$

If we define an antisymmetric tensor F_{ij} (called the curl of $\mathbf{A}$) by

$$F_{ij} = \partial A_j/\partial x^i - \partial A_i/\partial x^j \tag{4}$$

then we have only to apply a 4-dimensional Stokes theorem[1] to Eq. (3) in order to obtain

$$\Delta\phi = \sum_{i<j} \int_S F_{ij}\, dx^i\, dx^j, \tag{5}$$

where S is a two-parameter hypersurface bounded by C. The nature of the multivaluedness of ϕ is exhibited quite clearly by Eq. (5), which states that the change in ϕ that is produced by moving $\mathbf{x}$ around a closed curve equals the total flux of F_{ij} encircled by the curve. ϕ is single-valued only if F_{ij} vanishes. It follows from Eq. (1) that

$$\partial\phi/\partial x^i = A_i, \tag{6}$$

i.e., that the gradient of ϕ is $\mathbf{A}$. Thus, *any* vector field may be represented as the gradient of a scalar function, the function being multivalued if the vector field is nonconservative.

It is clear from Eqs. (4) and (6) that

$$[\partial/\partial x^i,\ \partial/\partial x^j]\phi = F_{ij}. \tag{7}$$

We see from Eq. (7) that if ϕ is a multivalued function of $\mathbf{x}$ then partial differentiation with respect to the components of $\mathbf{x}$ is not a commutative operation, although the commutators are well defined. This is one of the most important properties of multivalued functions.

III. MULTIVALUED FUNCTIONS AND ELECTROMAGNETIC INTERACTIONS

It is well known that an electromagnetic field is represented by an antisymmetric tensor F_{ij} which satisfies the Maxwell equations

$$\begin{aligned} &\partial F_{nk}/\partial x^k = 4\pi j_n \\ &\partial F_{ij}/\partial x^k + \partial F_{ki}/\partial x^j + \partial F_{jk}/\partial x^i = 0. \end{aligned} \tag{8}$$

Here j_n denotes the nth component of the four-current density $\mathbf{j}$, and the velocity of light is taken as unity. We may use Eqs. (7) and (8) to obtain a scalar representation of electromagnetic theory which is quite different from that due to Wolf,[2] and highly suggestive for further theoretical developments. We obtain

$$\frac{\partial}{\partial x^k}\left[\frac{\partial}{\partial x^n}\,,\ \frac{\partial}{\partial x^k}\right]\phi = 4\pi j_n. \tag{9}$$

The second set of Eqs. (8) and the continuity equation $\partial j_n/\partial x^n = 0$ are automatically satisfied

[1] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, Canada, 1956), p. 274.
[2] E. Wolf, Proc. Phys. Soc. (London) **74**, 269 (1959).

because the four-gradient of ϕ is single-valued. Thus, we see that to every multivalued function with a single-valued gradient there corresponds a unique electromagnetic field. Conversely, to every electromagnetic field there corresponds a multivalued function ϕ which possesses a single-valued gradient and which is unique up to an arbitrary additive single-valued function. The gradient of ϕ is the usual vector potential $\mathbf{A}$ which is fixed by F_{ij} to within a gauge transformation (i.e., up to the gradient of a single-valued function). Thus, it is clear that we can introduce an electromagnetic field into a field-free equation by making a gauge transformation with a multivalued gauge function. This is quite straightforward classically, but deserves special attention in quantum mechanics. The point here is that if ψ is a single-valued wave function and ϕ represents an (nontrivial) electromagnetic field, then the unitary (gauge) transformation $U = \epsilon^{-ie\phi}$ is multivalued. Indeed, U is not an operator on a single Hilbert space. It maps the space of single-valued functions ψ onto a space of multivalued functions $\bar{\psi} = U\psi$. Nevertheless, we may express the Schrödinger equation for a particle in an external electromagnetic field in the form

$$(H_0 - i\,\partial/\partial x^4)\bar{\psi} = 0 \tag{10}$$

where x^4 denotes the time t and H_0 is the *field-free* Hamiltonian for the system. If we transform Eq. (10) by $U^{-1} = U^* = \epsilon^{ie\phi}$ we obtain

$$(H - i\,\partial/\partial x^4)\psi = 0, \tag{11}$$

where

$$H - i\,\partial/\partial x^4 = U^{-1}(H_0 - i\,\partial/\partial x^4)U.$$

Since $U^{-1}\,\nabla\,U = \nabla - ie\mathbf{A}$, it follows that H is the Hamiltonian for a particle of charge e in the electromagnetic field F_{ij}. From a slightly different point of view, we see that the standard problem of solving the time-dependent Schrödinger equation for a system in an external electromagnetic field is equivalent to that of solving the corresponding field-free equation, with the provision that the solutions shall be sought in a space of multivalued functions, the detailed character of which is fixed by the external field. In other words, Eq. (10) is equivalent to Eq. (11), but in an *admissibility condition representation*. Bohm and Aharonov[3] have noted the connection between Eqs. (10) and (11) for the special case where the fields vanish in some multiply connected region.

[3] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).

In quantum electrodynamics, the situation is similar to that descibed above. If $L_0(\bar{\psi}, \mathbf{A})$ is the Lagrangian for a noninteracting particle field $\bar{\psi}$ and photon field $\mathbf{A}$, and we define

$$L(\psi, \mathbf{A}) = L_0(U\psi, \mathbf{A}), \tag{12}$$

then L is the Lagrangian for the *coupled* fields ψ and $\mathbf{A}$. This suggests strongly that other interactions (e.g., strong and weak interactions) should be introduced by using a transformation T which is more general than the multivalued gauge transformation U. We shall see in Sec. V that there are further reasons for believing that this is the case.

IV. MULTIVALUED COORDINATE TRANSFORMATIONS

Let $\bar{g}_{ij}(\bar{\mathbf{x}})$ be a metric for some general Riemannian space-time. The element of length ds is then defined by

$$ds^2 = \bar{g}_{ij}\,d\bar{x}^i\,d\bar{x}^j. \tag{13}$$

Now, let $S_{ij}(\bar{\mathbf{x}})$ be a real and unitary (orthogonal) matrix which diagonalizes $\bar{g}$. We have

$$\tilde{g}_{kn} = S_{ki}\bar{g}_{ij}S_{nj} \tag{14}$$

with $\tilde{g}$ diagonal. The element $\tilde{g}_{NN}$ is an eigenvalue of $\bar{g}$ corresponding to the eigenvector S_{Nj}. It follows from Eq. (14) that

$$\bar{g}_{ij} = S_{ki}\tilde{g}_{kn}S_{nj}. \tag{15}$$

We assume that $\bar{x}^1$, $\bar{x}^2$, $\bar{x}^3$ are space-like, while $\bar{x}^4$ is time-like. It then follows from Eq. (15) that

$$ds^2 = g_{ij}\,dx^i\,dx^j, \tag{16}$$

where

$$g_{ij} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}$$

and

$$dx^N = |\tilde{g}_{NN}|^{1/2}\,S_{Ni}\,d\bar{x}^i. \tag{17}$$

Upon solving Eq. (17) for $d\bar{x}^i$ we obtain

$$d\bar{x}^i = |\tilde{g}_{nn}|^{-1/2}\,S_{ni}\,dx^n. \tag{18}$$

Now, if we adopt the same point of view toward Eqs. (17) and (18) that we adopted toward Eq. (1), we see that these equations define a multivalued coordinate transformation. Consistent with this point of view we have

$$\partial x^N/\partial \bar{x}^i = |\tilde{g}_{NN}|^{1/2} S_{Ni} \qquad (19)$$
$$\partial \bar{x}^j/\partial x^N = |\tilde{g}_{NN}|^{-1/2} S_{Nj}$$

and, we see that

$$\frac{\partial x^n}{\partial \bar{x}^i}\frac{\partial \bar{x}^i}{\partial x^k} = \frac{\partial \bar{x}^n}{\partial x^i}\frac{\partial x^i}{\partial \bar{x}^k} = \delta^n_k, \qquad (20)$$

where δ^n_k denotes the usual Kronecker symbol. This means that we may use tensor methods to investigate the relation between the x and the $\bar{x}$ systems even though the transformation may be multivalued. If we define g^{ij} and $\bar{g}^{ij}$ by

$$g^{ij}g_{jk} = \bar{g}^{ij}\bar{g}_{jk} = \delta^i_k, \qquad (21)$$

we see that g^{ij} and $\bar{g}^{ij}$ transform like the contravariant components of a tensor while g_{ij} and $\bar{g}_{ij}$ transform like the covariant components of a tensor. For example, we have

$$\bar{g}_{ij} = \frac{\partial x^k}{\partial \bar{x}^i}\frac{\partial x^n}{\partial \bar{x}^j} g_{kn}. \qquad (22)$$

Equation (22) shows that a curved-space metric and a flat-space metric are connected by a tensor transformation corresponding to a multivalued coordinate transformation. Now a well-known theorem due to Riemann states that it is not possible to connect a curved space and a flat space with a (single-valued) coordinate transformation. Since we have seen that this *is* possible with a multivalued coordinate transformation, we enquire just how the Riemann theorem breaks down. The answer is that when one takes the covariant derivative of a contravariant vector in the usual way, the object one obtains does *not* transform like a tensor under multivalued coordinate transformation. Hence, the Riemann curvature symbol is not a tensor under these transformations. In particular, it may vanish in the x coordinate system but not in the $\bar{x}$ coordinate system if the two systems are connected by a multivalued transformation.

V. TRANSFORMATION OF THE EQUATIONS OF MOTION OF A PARTICLE

Consider the motion of a classical relativistic particle. The ideas of the general theory of relativity suggest that force terms appear in the equations of motion because of the curvature of space. This implies that the equations should be those of a free particle when expressed in terms of flat-space coordinates. We therefore write

$$d^2x^i/ds^2 = 0. \qquad (23)$$

We now seek the corresponding equations of motion in curved-space coordinates. These are easily found to be

$$\frac{d\bar{x}^k}{dS^2} + \left\{ \begin{matrix} k \\ i \quad j \end{matrix} \right\} \frac{d\bar{x}^i}{ds}\frac{d\bar{x}^j}{ds} = \bar{g}^{kn}F_{nj}\frac{d\bar{x}^j}{ds}\frac{e}{m} \qquad (24)$$

where $\{ {}_i{}^k{}_j \}$ is the usual Christoffel symbol of the second kind, and

$$F_{nj} = \frac{m}{e} g_{ia}\frac{dx^i}{ds}\left(\frac{\partial^2 x^a}{\partial \bar{x}^n\, \partial \bar{x}^j} - \frac{\partial^2 x^a}{\partial \bar{x}^j\, \partial \bar{x}^n}\right). \qquad (25)$$

In passing from Eq. (23) to Eqs. (24) and (25) we have assumed that the equations transform covariantly under multivalued coordinate transformations. This is a natural generalization of the principle of equivalence, and we notice that it leads to results which are at variance with Einstein's assumption that particles move along geodesics. Eq. (24) defines a geodesic only if F_{nj} vanishes, i.e., only if x^a is a single-valued function of the $\bar{x}$ coordinates, so that the $\bar{x}$ space is flat. For nonvanishing F_{nj}, Eq. (24) is the standard form for the equations of motion of a charged particle in a gravitational and electromagnetic field. Since Eq. (23) is satisfied, it is clear that dx^i/ds is constant along the trajectory, and we see that F_{nj} satisfies Maxwell's equations. Unfortunately, however, the dependence of F_{nj} on dx^i/ds is not in agreement with experiment. This is clear since the quantities dx^i/ds determine, and are determined by, the quantities $d\bar{x}^i/ds$. Hence, Eq. (25) asserts that the electromagnetic field seen by a particle at any point of space-time is dependent upon the four-velocity of the particle. This result is unsatisfactory, but it does suggest that we should write down the *quantum-mechanical* equations of motion for a free relativistic particle in flat-space coordinates and then transform this equation into curved-space coordinates. The reason this is suggested is that, in quantum mechanics, the uncertainty principle prevents us from measuring the four-velocity of a particle if we measure its position in space-time exactly. Hence we have some reason to hope that the curved-space equation will describe the motion of a particle in a gravitational and electromagnetic field, and that velocity dependent fields will not appear in quantum theory. There is a second reason to turn to quantum mechanics in the search for a unified field theory. A true unified field theory should account for strong and weak interactions as well as electromagnetic and gravitational interactions, and the former appear to be essentially quantum mechanical in nature.

The field-free Klein-Gordon equation in flat-space

coordinates is

$$-g^{ij}\,\partial^2\psi/\partial x^i\,\partial x^j = m^2\psi, \tag{26}$$

while the field-free Dirac equation is

$$(-i\gamma^\mu\,\partial/\partial x^\mu + m)\psi = 0$$
$$\gamma^\mu\gamma^\nu + \gamma^\nu\gamma^\mu = 2g^{\mu\nu}. \tag{27}$$

When Eqs. (26) and (27) are expressed in terms of curved-space coordinates they become

$$-\bar{g}^{kn}\frac{\partial^2\psi}{\partial\bar{x}^k\,\partial\bar{x}^n} + \bar{g}^{rs}\frac{\partial^2 x^j}{\partial\bar{x}^r\,\partial\bar{x}^s}\frac{\partial\bar{x}^n}{\partial x^j}\frac{\partial\psi}{\partial\bar{x}^n} = m^2\psi \tag{28}$$

and

$$(-i\bar{\gamma}^k\,\partial/\partial\bar{x}^k + m)\psi = 0$$
$$\bar{\gamma}^k\bar{\gamma}^n + \bar{\gamma}^n\bar{\gamma}^k = 2\bar{g}^{kn}, \tag{29}$$

respectively. Now, there are terms in Eqs. (28) and (29) which may easily be associated with a gravitational field, but there is no term which may be associated with an electromagnetic field. The reason for this (apparent) discrepancy with the classical case is that the transformed Dirac and Klein-Gordon equations exhibit a flaw which we must eliminate before proceeding further with the analysis. The point here is that if ψ is a single-valued function of the x coordinates, then it is a multivalued function of the $\bar{x}$ coordinates. This may be seen from the identity

$$\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right]\psi = \frac{\partial\bar{x}^k}{\partial x^i}\frac{\partial\bar{x}^n}{\partial x^j} \times \left[\frac{\partial}{\partial\bar{x}^k}, \frac{\partial}{\partial\bar{x}^n}\right]\psi + \frac{\partial\psi}{\partial\bar{x}^n}\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right]\bar{x}^n \tag{30}$$

It is clear from Eq. (30) that $[\partial/\partial x^i, \partial/\partial x^j]\psi$ and $[\partial/\partial\bar{x}^k, \partial/\partial\bar{x}^n]\psi$ cannot both vanish unless $[\partial/\partial x^i, \partial/\partial x^j]\bar{x}^n$ vanishes also. In other words ψ cannot be a single-valued function of the x and $\bar{x}$ coordinates unless the coordinate transformation is single-valued. If the transformation is multivalued, the change of independent variables from x to $\bar{x}$ coordinates must be accompanied by a corresponding change of dependent variable from ψ to $\bar{\psi}$, where $\bar{\psi}$ is a single-valued function of the $\bar{x}$ coordinates. Thus, we write

$$[\partial/\partial x^i, \partial/\partial x^j]\psi = 0$$
$$[\partial/\partial\bar{x}^i, \partial/\partial\bar{x}^j]\bar{\psi} = 0 \tag{31}$$
$$\psi = T\bar{\psi}.$$

Equations (31) simply state that ψ is a single-valued function of the x coordinate, that $\bar{\psi}$ is a single-valued function of the $\bar{x}$ coordinates, and that T is a (generally multivalued) transformation which connects the barred and the unbarred function spaces. It is clear that the transformation $\psi = T\bar{\psi}$ introduces fields in Eqs. (28) and (29) in much the same way as the transformation $\bar{\psi} = U\psi$ introduced fields into Eq. (10). On the other hand, it is easily verified that Eqs. (31) cannot be satisfied by taking T to be simply multiplication by some multivalued scalar function. This means that although T may introduce electromagnetic fields into Eqs. (28) and (29) it must introduce other fields as well. The properties of these fields are defined by Eqs. (31), but the question is completely open whether they correspond to any forces which are observed in nature. Indeed, a rigorous analysis of Eqs. (31) probably would require the development of some new mathematical tools, capable of treating the calculus of multivalued operators.

APPENDIX

The following argument shows that although the function ϕ defined in Sec. II is "path dependent" rather than "multivalued" in the conventional sense, it can be expressed as a linear combination of conventional multivalued functions. In the interest of clarity the argument is given for a two-dimensional case. Let a vector $\mathbf{A}$ be defined by $\mathbf{A} = A_x t_{\hat{x}} + A_y t_{\hat{y}}$ where $A_x(x, y)$ and $A_y(x, y)$ are single-valued and $t_{\hat{x}}$, $t_{\hat{y}}$ are unit vectors in the x and y directions. Now, let $B(x, y) = (\partial Ay/\partial x - \partial Ax/\partial y)$. It is well known that if B does not vanish, there is no (single-valued) function whose gradient is $\mathbf{A}$. However, consider the function

$$\theta = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x', y')\arctan\frac{y - y'}{x - x'}\,dx'\,dy'.$$

It seems appropriate to call θ a multivalued function since it is a linear combination of the arctan functions. Upon computing $\partial\theta/\partial x$ and $\partial\theta/\partial y$ by differentiation under the integral sign, we have

$$a_x = \frac{\partial\theta}{\partial x} = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{y - y'}{(x - x')^2 + (y - y')^2}\,dx'\,dy'$$

$$a_y = \frac{\partial\theta}{\partial y} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{x - x'}{(x - x')^2 + (y - y')^2}\,dx'\,dy'$$

Clearly, a_x and a_y are single-valued, since they are linear combinations of single-valued functions. Now, consider the quantity

$$\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right]\theta.$$

We have

$$\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x', y')\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right] f \, dx' \, dy'$$

where $f = \arctan (y - y')/(x - x')$. The change in $f(x, y)$ which results from moving in a counter clockwise direction around a closed curve C in the x, y plane is

$$\Delta f = \int_C \left(\frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy\right),$$

and by Stoke's theorem

$$\Delta f = \int_S \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right] f \, dx \, dy,$$

where S is a surface bounded by C. Now, Δf equals 2π or zero according as C does or does not encircle the point x', y', i.e., according as that point does or does not lie on S. Hence, we may write

$$\int_S \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right] f \, dx \, dy = 2\pi \int_S \delta(x - x', y - y') \, dx \, dy,$$

where $\delta(x - x', y - y')$ is the usual two-dimensional Dirac delta function. Since the above relation must hold for an arbitrary surface S, it follows that

$$\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right] f = 2\pi \, \delta(x - x', y - y')$$

and that

$$\frac{\partial a_y}{\partial x} - \frac{\partial a_x}{\partial y} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} B(x', y') \, \delta(x - x', y - y') \, dx' \, dy' = B(x, y).$$

It is now clear that if ϕ is defined as in Eq. (2), then $\mathbf{A} = \nabla\phi$ and $\mathbf{a} = \nabla\theta$ are vector potentials corresponding to the same fields. Hence, they can differ only by a gauge transformation, i.e., by the gradient of a single-valued scalar. This means that θ and ϕ differ only by a function which is single-valued. Since θ is a superposition of (conventional) multivalued functions, it seems appropriate to call ϕ a multivalued function also. In conclusion, it should be emphasized that derivatives of ϕ are defined by giving the path of integration a slight increment at its end point while holding the rest of it completely fixed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Renormalization of a Classical Gravitational Field Interacting with Quantized Matter Fields*

RYOYU UTIYAMA† AND BRYCE S. DEWITT
Institute of Field Physics,
Department of Physics, University of North Carolina, Chapel Hill, North Carolina
(Received November 10, 1961)

The behavior of the mean value of the energy-momentum tensor of a set of quantized matter fields interacting with a classical gravitational field which is, in turn, produced by this mean value, is investigated. Singularities appear in the energy-momentum tensor corresponding to divergences of three different orders: ∞^4, ∞^2, and $\log \infty$. These can be removed by the introduction of counter terms into Einstein's equation. The ∞^4 singularity is removed by a "cosmological term," the ∞^2 singularity by a renormalization of the gravitation constant, and the $\log \infty$ singularity by a counter term derivable from a Lagrangian which is quadratic in the Riemann tensor.

The gravitational Green's function corresponding to this semiclassical approximation to the fully quantized theory is found to have the asymptotic behavior $1/p^4$ instead of $1/p^2$, and thus to have a much weaker singularity in the coordinate representation than the Green's function of the "bare" linearized theory.

INTRODUCTION

IN a previous paper[1] it was shown that the conventional renormalization procedures of quantum electrodynamics suffice also to remove the singularities in the photon and electron Green's functions in a given classical gravitational field. Furthermore the renormalization constants were shown to be identical with those of the Lorentz invariant theory.

As the next step towards the quantum theory of gravitation, we shall consider in the present paper the case in which the gravitational field is not an arbitrarily given c-number field but is regarded as being produced by an appropriate matrix element of the energy-momentum density of quantized matter fields.[2] Now, a classical gravitational field can induce a nonvanishing energy-momentum density (which actually diverges) in the vacuum state.[3,4] In the present approximation this phenomenon gives rise to closed-loop Feynman diagrams which lead to a new type of divergence, characteristic of gravitational interactions.

Strictly speaking, this divergence also occurs in I, if one actually calculates the vacuum expectation value of the S matrix. However, in the definition of the Green's functions, which are discussed in I, such a divergence appears as a factor in both numerator and denominator and consequently cancels out.

In the present paper, quantities depending on the gravitational potential (especially the expectation value of the energy-momentum density) are expanded in power series in the gravitation constant κ. Each term of such a series corresponds to a Feynman diagram which, in the case of the energy-momentum density, consists of a loop of internal matter lines yielding divergences of three different types: ∞^4, ∞^2, and $\log \infty$. The most singular part of the series, each term of which is proportional to ∞^4, can be reduced to a simple expression removable by a counter term in the Lagrangian analogous to the "cosmological-term" of general relativity, with an infinite coefficient. Furthermore, by employing the methods of I, it will be shown that the singular part proportional to ∞^2 can be removed by renormalizing the gravitation constant. The singular part proportional to $\log \infty$, on the other hand, can be handled only by introducing counter terms of a completely new type.

The infinite constants appearing in the counter terms corresponding to the ∞^2 and $\log \infty$ divergences are new ones. They are characteristic of the gravitational interaction and are not encountered in Lorentz invariant field theories.

1. PRELIMINARY DISCUSSION

In order to develop a quantum theory of matter fields interacting with a c-number gravitational field,

* This research was supported in part by the Department of the Navy, Office of Naval Research, under contract Nonr-855(07) and in part by the Air Force Office of Scientific Research under contract AFOSR 61-72.

† On leave of absence from Osaka University, Osaka, Japan.

[1] Ryoyu Utiyama, Phys. Rev. **125**, 1727 (1962). This work will be referred to as I.

[2] The term "matter field" here means not only the conventional matter fields, such as those describing electrons and mesons, but also includes the radiation field.

[3] Bryce S. DeWitt, Ph.D. Thesis, Harvard (1950).

[4] Here the "vacuum" means a state where neither material particles nor photons are present.

it will be sufficient to postulate

(1) A system of coordinates $x^\mu(\mu = 1, 2, 3, 4)$ can be established for which every hypersurface $x^4 =$ constant is everywhere space-like.

This postulate enables one to set up canonical commutation relations for the operators of the matter fields. In addition, for local field theories, one can define a state vector as a functional of field variables at any time x^4, and the total energy-momentum operators will act as generators of infinitesimal displacements even though these operators are not constants of the motion.

If in addition to (1) we also postulate

(2) The interaction constants between matter fields adiabatically vanish as $x^4 \to \pm\infty$ in the system of coordinates introduced in (1),

then we may introduce the so-called interaction picture.[5]

It should be noted that in dealing with questions of general covariance it suffices to restrict our attention to coordinate transformations which maintain the conditions of postulate (1).

The state vector in the interaction picture is a functional on any space-like surface σ, and the functional dependence of the state vector upon σ is governed by the Tomonaga-Schwinger equation. The formal integration of this equation gives us the S matrix, the derivation of which needs postulate (1).

In order that it be possible to introduce a complete set of orthonormal vectors in the Hilbert space at $x^4 = -\infty$, it is convenient (although perhaps unnecessary) to require, in addition to postulates (1) and (2), condition:

$$\lim_{x^4 \to -\infty} g_{\mu\nu} = \eta_{\mu\nu}. \tag{3}$$

Postulates (2) and (3) together allow us to employ the conventional incoming-field-representation in terms of which every field operator in the Heisenberg picture can be represented by means of the so-called $U(x^4, -\infty)$ matrix.

All the above requirements can be satisfied if the gravitational field is sufficiently well behaved. In the present case, however, since the gravitational field is regarded as being produced by the matter fields according to Einstein's equation, one must investigate its behavior *a posteriori* in order to check consistency.

[5] In this case the free Hamiltonian does not involve any interaction between matter fields. The interaction of each field with the gravitational field, however, remains.

Of course, it is very hard to give more than a qualitative criterion for the validity of the postulates. Roughly speaking, however, if the deviation $|g_{\mu\nu}(x) - \eta_{\mu\nu}|$ is smaller than unity everywhere in the "world," it is reasonable to regard all the postulates as being satisfied. Accordingly, one must first of all remove the singularities of the energy-momentum tensor. Otherwise this tensor will give rise to a very large deviation of $g_{\mu\nu}$ from the Minkowskian metric $\eta_{\mu\nu}$.

When the removal of divergences of the energy-momentum tensor has been performed, the gravitational effect of one electron may be regarded as that due to a mass m spread over region of dimension $\lambda = \hbar/mc$. This produces a deviation of the metric from flatness of order

$$\Delta g_{\mu\nu} = Gm^2/\hbar c = 0.3 \times 10^{-45} \ll 1,$$

where G is the gravitation constant.

Consequently, we may reasonably assume that all the postulates are satisfied and proceed to apply conventional methods of quantum field theory to the present case. The self consistency of this procedure will then be verified at the end.

2. FIELD EQUATIONS

Our first problem is to replace the classical energy-momentum tensor of the matter field, which appears in Einstein's equation, by a suitably defined mean value of the corresponding operator.

The most straightforward way is to adopt the simple expectation value. However, if one applies Schwinger's method to the quantized Einstein equation in order to obtain the fundamental equation for the Green's functions of the q-number gravitational field, one is led to adopt the following mean value:

$$(\Psi_a^* S \cdot T_H^{\mu\nu}(x)\Psi_b)/(\Psi_a^* S \Psi_b)$$

where Ψ_a and Ψ_b are two state vectors belonging to the Hilbert space established at $x^4 = -\infty$, $T_H^{\mu\nu}(x)$ is the q-number energy-momentum tensor of the matter fields in the Heisenberg picture, and S is the S matrix which is assumed to exist in virtue of the postulates mentioned in Sec. 1.

In the present section we shall be exclusively concerned with the special, but important, case $\Psi_a = \Psi_b = \Psi_0$ (the vacuum state at $x^4 = -\infty$)

$$\langle T^{\mu\nu}(x)\rangle \equiv \frac{(\Psi_0^* S \cdot T_H^{\mu\nu}(x)\Psi_0)}{(\Psi_0^* S \Psi_0)} \tag{2.1}$$

which can be alternatively written as

$$= \frac{\{\Psi_0^*, T[S \cdot T^{\mu\nu}(x)]\Psi_0\}}{(\Psi_0^* S \Psi_0)},$$

the symbol T denoting the T product which can be well defined in virtue of postulate (1). The $T^{\mu\nu}(x)$ in the alternative expression is taken in the interaction picture.

The more general case, where Ψ_b and Ψ_a are arbitrarily chosen initial and final states, will be discussed briefly in the final section. It will be seen, however, that the analysis of the special value $\langle T^{\mu\nu}\rangle$ taken here is sufficient for the discussion of the removal of divergences appearing in $T^{\mu\nu}_H$, which is the main purpose of the present paper.

Our starting equation is

$$(-g)^{1/2}(R^{\mu\nu} - (1/2)g^{\mu\nu}R) = -\kappa[-g(x)]^{1/2}\langle T^{\mu\nu}(x)\rangle, \qquad (2.2)$$

where

$$R_{\mu\nu} = \partial_\mu\Gamma_{\nu\rho}{}^{\rho} - \partial_\rho\Gamma_{\mu\nu}{}^{\rho} + \Gamma_{\mu\rho}{}^{\lambda}\Gamma_{\nu\lambda}{}^{\rho} - \Gamma_{\mu\nu}{}^{\rho}\Gamma_{\rho\lambda}{}^{\lambda},$$

$$R = g^{\mu\nu}R_{\mu\nu},$$

κ = Einstein's gravitational constant.

Equation (2.2) can be derived by the variation principle from the action integral

$$I = \int (-g)^{1/2}L\, d^4x$$

$$L = (1/2\kappa)R + \langle L_M\rangle,$$

where $\langle L_M(x)\rangle$ stands for the mean value of the Lagrange function of the matter fields, which is assumed to be a scalar under the general coordinate transformations. The symmetric energy-momentum tensor $T^{\mu\nu}_H(x)$ is obtained from L_M in the following manner:

$$(-g)^{1/2}T^{\mu\nu}_H = -2\frac{\delta[(-g)^{1/2}L_M(x)]}{\delta g_{\mu\nu}} \equiv -2\left[\frac{\partial[(-g)^{1/2}L_M]}{\partial g_{\mu\nu}} - \frac{\partial}{\partial x^\rho}\left\{\frac{\partial[(-g)^{1/2}L_M]}{\partial(\partial g_{\mu\nu}/\partial x^\rho)}\right\}\right].$$

Let us now put

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \kappa\phi_{\mu\nu}(x)$$

$$g^{\mu\nu}(x) = \eta^{\mu\nu} - \kappa\phi^{\mu\nu}(x) + \cdots.$$

Following the argument given in Sec. 1, we assume the deviation $|\kappa\phi_{\mu\nu}(x)|$ to be smaller than unity (at least when the singularity of $\langle T^{\mu\nu}\rangle$ has been removed) even in the vicinity of matter particles. We also assume the convergence of the expansion of any quantity depending on $g_{\mu\nu}$ in powers of κ.

Equation (2.2) becomes in such an expansion

$$\partial_\mu\partial_\nu\phi + \Box\phi_{\mu\nu} - \partial_\mu\partial^\rho\phi_{\rho\nu} - \partial_\nu\partial^\rho\phi_{\rho\mu} - \eta_{\mu\nu}\Box\phi + \eta_{\mu\nu}\partial^\lambda\partial^\sigma\phi_{\lambda\sigma} = -S_{\mu\nu}(x) - t_{\mu\nu}(x), \qquad (2.3)$$

where the following abbreviations have been employed:

$$\partial^\mu = \eta^{\mu\nu}\partial_\nu, \qquad \phi = \eta^{\mu\nu}\phi_{\mu\nu}$$

$$S_{\mu\nu} = 2(-g)^{1/2}\langle T_{\mu\nu}(x)\rangle + (2/\kappa)0_{\mu\nu}(\kappa^2)$$

$$0_{\mu\nu}(\kappa^2) = (-g)^{1/2}(R_{\mu\nu} - (1/2)g_{\mu\nu}R) - \kappa/2 \text{ [left-hand side of (2.3)]} = \text{order of } \kappa^2$$

On the right-hand side of (2.3) an additional term $t_{\mu\nu}$ has been introduced for the sake of convenience. It is assumed to be a well-behaved tensor density satisfying the conservation law

$$\partial_\mu t^{\mu\nu} + \Gamma_{\rho\lambda}{}^{\nu}t^{\rho\lambda} = 0, \qquad t^{\mu\nu} = g^{\mu\rho}g^{\nu\sigma}t_{\rho\sigma}. \qquad (2.4)$$

Because of (2.4), $t_{\mu\nu}(x)$ cannot be independent of $g_{\mu\nu}$, and must be represented by a power series in κ.

The divergence of the left-hand side of (2.3) vanishes identically. Consequently we have

$$\partial^\mu(S_{\mu\nu} + t_{\mu\nu}) = 0, \qquad (2.5)$$

when (2.3) itself is satisfied.

The left-hand side of (2.3) can be rewritten in the simpler form

$$\Box h_{\mu\nu} = -(S_{\mu\nu} + t_{\mu\nu}), \qquad (2.6)$$

if we introduce a new field quantity defined by

$$h_{\mu\nu} = \phi_{\mu\nu} - (1/2)\eta_{\mu\nu}\phi,$$

satisfying the subsidiary condition

$$\partial^\mu h_{\mu\nu} = 0. \qquad (2.7)$$

It is easily seen that condition (2.7) is compatible with the field equation (2.6) when (2.5) is satisfied.

The formal solution of (2.6) is

$$h_{\mu\nu}(x) = h^0_{\mu\nu}(x) + \int \Delta(x - x')\{S_{\mu\nu}(x') + t_{\mu\nu}(x')\}\, dx',$$

where

$$\Box h^0_{\mu\nu} = 0,$$

and $\Delta(x - x')$ is defined by

$$\Box\Delta(x - x') = -\delta^4(x - x')$$

together with appropriate boundary conditions. The above solution for h leads to

$$\phi_{\mu\nu} = \phi^0_{\mu\nu}(x) + \partial_\mu\lambda_\nu(x) + \partial_\nu\lambda_\mu(x) + \int \Delta(x - x')\{S_{\mu\nu}(x') + t_{\mu\nu}(x') - (1/2)\eta_{\mu\nu}S(x') - (1/2)\eta_{\mu\nu}t(x')\}\, d^4x', \qquad (2.8)$$

where $\lambda_\nu(x)$ is an arbitrary covariant four-vector (under Lorentz transformations) and S and t are defined as

$$S = \eta^{\mu\nu}S_{\mu\nu}, \qquad t = \eta^{\mu\nu}t_{\mu\nu}.$$

The terms involving $\lambda_\mu(x)$ in (2.8) give zero when inserted into the left-hand side of (2.3). The ambiguity in $\phi_{\mu\nu}$ due to the arbitrary functions λ_μ reflects the invariance of the theory under general coordinate transformations.

Since the gravitational field is expected to vanish in the absence of the source $t^{\mu\nu}$ (when the infinities appearing in the vacuum expectation value $\langle T^{\mu\nu}\rangle$ have been subtracted), $\phi^0_{\mu\nu}(x)$ in (2.8) should be put equal to zero. In addition, the terms involving λ_μ can be removed by a suitable transformation of the coordinate system.

Therefore, our basic equation is now

$$\phi_{\mu\nu}(x) = \int \Delta(x - x')\{M_{\mu\nu}(x') + N_{\mu\nu}(x')\}\, d^4x,$$

$$M_{\mu\nu} = S_{\mu\nu} - (1/2)\eta_{\mu\nu}S,$$

$$N_{\mu\nu} = t_{\mu\nu} - (1/2)\eta_{\mu\nu}t. \tag{2.9}$$

Here the quantities M and N are functionals of ϕ and can be expanded in power series in κ in the following way:

$$M = M_0 + \kappa\int M_1\phi + \frac{\kappa^2}{2!}\iint M_2\phi\phi + \cdots,$$

$$N = N_0 + \kappa\int N_1\phi + \frac{\kappa^2}{2!}\iint N_2\phi\phi + \cdots,$$

where tensor indices and integration variables have been suppressed for compactness.

Inserting these expansions into (2.9) and iterating, we obtain the solution of (2.9) as a power series in κ, the first few terms of which are pictorially represented in Fig. 1.

In contrast to the case of full quantization of the gravitational field there occur in Fig. 1 no closed loops of dotted lines (which will be called internal gravitational lines or *g lines*). Consequently $\phi_{\mu\nu}$ becomes finite if all the M_n's ($n \geq 1$) are finite and M_0 is equal to zero. (Since M_0 is independent of x, as will be seen later, the integration with respect to x diverges unless $M_0 = 0$.) Furthermore, if the external source $t^{\mu\nu}$ vanishes $\phi_{\mu\nu}$ becomes zero as was expected.

3. GENERAL PROPERTIES OF $\langle T^{\mu\nu}\rangle$

Before beginning the detailed investigation of $\langle T^{\mu\nu}\rangle$, it is useful to give a concrete expression of $\langle T^{\mu\nu}\rangle$ in a special case.[3]

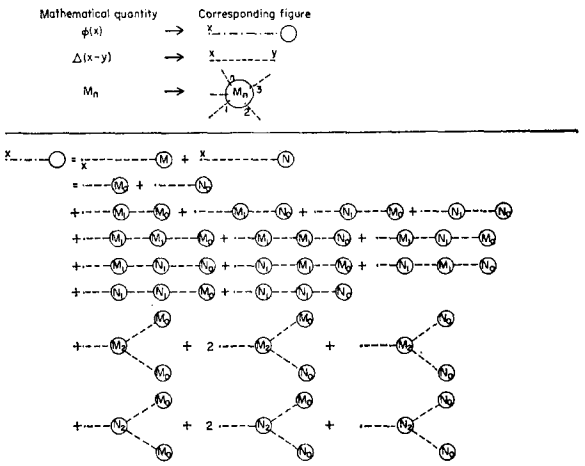

FIG. 1. Feynman diagram of $\phi_{\mu\nu}(X)$.

As an example, let us consider the pure radiation field in the absence of electric charges; i.e.,

$$L_M = -(1/4)g^{\mu\rho}g^{\nu\sigma}f_{\mu\nu}f_{\rho\sigma}$$

$$f_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

which gives

$$\langle T^{\mu\nu}(x)\rangle = -\langle f^{\mu\alpha}f^{\nu}_{\alpha}\rangle - g^{\mu\nu}\langle L_M\rangle$$

$$= \lim_{x'\to x}[-g^{\mu\rho}(x)g^{\alpha\sigma}g^{\nu\beta}(\partial_\rho\,\delta^\lambda_\sigma - \partial_\sigma\,\delta^\lambda_\rho)(\partial'_\beta\,\delta^\tau_\alpha - \partial'_\alpha\,\delta^\tau_\beta)$$

$$+ (1/4)g^{\mu\nu}g^{\alpha\rho}g^{\beta\sigma}(\partial_\alpha\,\delta^\lambda_\beta - \partial_\beta\,\delta^\lambda_\alpha)(\partial'_\rho\,\delta^\tau_\sigma - \partial'_\sigma\,\delta^\tau_\rho)]$$

$$\times\, \mathcal{G}_{\lambda\tau}(x, x') \tag{3.1}$$

$$\mathcal{G}_{\lambda\tau}(x, x') = \langle A_\lambda(x),\, A_\tau(x')\rangle.$$

The most interesting property of (3.1) is the fact that its trace vanishes;

$$g_{\mu\nu}\langle T^{\mu\nu}(x)\rangle \equiv 0. \tag{3.2}$$

This identity can be used to reduce the number of counter terms needed for the removal of the singularities appearing in (3.1).

More generally $\langle T^{\mu\nu}\rangle$ has the typical structure

$$\langle T^{\mu\nu}(x)\rangle = \lim_{x'\to x}\sum_n D^{\mu\nu}_{(n)}(g(x),\, \partial,\, \partial')G_{(n)}(x, x'), \tag{3.3}$$

where $D_{(n)}$ involves $g_{\mu\nu}(x)$, $\partial g/\partial x$, $\partial/\partial x$, and $\partial/\partial x'$ in a certain order and $G_{(n)}$ is the Green's function for the nth matter field in the presence of the gravitational field. This general form holds for fermion as well as boson fields and is maintained also in the presence of interactions between these fields provided the $G_{(n)}$ are then understood as generalized propagators.

Let us suppose that the Green's function for the

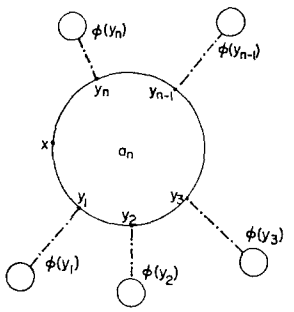

FIG. 2. Diagram of a_n.

nth field satisfies the equation

$$F_{(n)}G_{(n)}(x, x') = \delta(x - x') \qquad n = 1, 2, \cdots N,$$

where $F_{(n)}$ is a certain differential operator of, say the αth order. In addition, let us assume that all the singularities due to mutual interactions have been removed from the $G_{(n)}$, by a suitable renormalization technique. Then, in momentum space the Fourier transform of $G_{(n)}$ has the following asymptotic behavior:

$$G_{(n)}(p, p') \sim O(p^{-\alpha}) \quad \text{or} \quad O(p'^{-\alpha})$$

$$\text{for} \quad p; p' \gg \text{mass of particles.}$$

The degree of the highest derivative appearing in the differential operator $D_{(n)}$ in (3.3) is closely related to that involved in $F_{(n)}$, owing to the special character of the gravitational interaction. Namely, $D_{(n)}$ behaves asymptotically in the momentum representation like

$$D_{(n)} \sim O(p^{\alpha}).$$

Therefore a very rough estimate leads to the following result in momentum space

$$\langle T^{\mu\nu}(x)\rangle \cong \sum \int D_{(n)}(p)G_{(n)}(p)\, d^4p \cong \sum \int O(p^{\alpha})\cdot O(p^{-\alpha})\, d^4p = \infty^4, \tag{3.3a}$$

regardless of the properties of individual matter fields.

Let us expand $(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle$ in a power series in κ:

$$(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle = \sum_{n=0}^{\infty} \frac{\kappa^n}{n!} \int \cdots \int a_n \times [\mu\nu, \rho_1\sigma_1, \cdots \rho_n\sigma_n; x - y, x - y_2, \cdots x - y_n] \prod_{j=1}^{n} \phi_{\rho_j\sigma_j}(y_j) \prod_{j=1}^{n} d^4y_j,$$

with $a_0[\mu\nu] = \lim_{\kappa\to 0} (-g)^{1/2}\langle T^{\mu\nu}(x)\rangle = A\eta^{\mu\nu}$ where the quantity $a_n[\cdots]$ is a contravariant tensor of the $2n + 2$th rank under Lorentz transformations whose tensor indices are indicated in the square brackets.

As is easily seen (cf. I), a_n can be represented by a diagram of the form shown in Fig. 2. The lines (xy_1), (y_1y_2), $\cdots$ (y_nx) correspond to Green's functions of the matter fields, including the effect of their mutual interactions, but excluding the effect of the gravitational field itself.

The functions a_n are to some extent restricted by relations stemming from the transformation character of $\langle T^{\mu\nu}\rangle$ under general coordinate transformations and from the law of conservation of energy-momentum. Thus, under

$$x^{\mu} \to x'^{\mu} = x^{\mu} + \epsilon\kappa\xi^{\mu}(x),$$

we have

$$\begin{aligned}\kappa(-g)^{1/2}&\{\langle T^{\rho\nu}\rangle\, \partial_\rho\xi^{\mu}(x) + \langle T^{\mu\rho}\rangle\, \partial_\rho\xi^{\nu} - \langle T^{\mu\nu}\rangle\, \partial_\rho\xi^{\rho}\} \\ &= -\int \frac{\delta[(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle]}{\delta\phi_{\rho\sigma}(y)} \{\partial_\rho\xi_\sigma(y) \\ &\quad + \partial_\sigma\xi_\rho(y) + \kappa\, \partial_\rho\xi^{\lambda}(y)\cdot\phi_{\lambda\sigma}(y) \\ &\quad + \kappa\, \partial_\sigma\xi^{\lambda}(y)\cdot\phi_{\lambda\rho}(y) + \kappa\xi^{\lambda}(y)\cdot\partial_\lambda\phi_{\rho\sigma}(y)\}\, d^4y \\ &\quad + \kappa\xi^{\rho}(x)\, \partial_\rho\{(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle\}\end{aligned} \tag{3.4}$$

and

$$\partial_\mu\{(-g)^{1/2}\langle T^{\mu\nu}\rangle\} + \Gamma^{\nu}_{\rho\sigma}(-g)^{1/2}\langle T^{\rho\sigma}\rangle = 0. \tag{3.5}$$

It is to be noted that in general $\langle T^{\mu\nu}\rangle$ is a functional of ϕ and $t^{\mu\nu}$. It will be seen, however, that the Green's functions of the matter fields depend only upon ϕ, and since the effect of the external source t is mediated by ϕ, it is unnecessary to take account of the variation of $t^{\mu\nu}$, the latter being already contained in the variation of ϕ.

The expansion of (3.4) and (3.5) in power series in κ yields the recurrence formulas

$$\begin{aligned}a_0[\rho\nu]\, \partial_\rho\xi^{\mu}(x) &+ a_0[\mu\rho]\, \partial_\rho\xi^{\nu}(x) \\ &- a_0[\mu\nu]\, \partial_\rho\xi^{\rho}(x) = \xi^{\rho}(x)(\partial/\partial x^{\rho})a_0[\mu\nu] \\ &- \int a_1[\mu\nu, \rho\sigma; x - y]\{\partial_\rho\xi_\sigma(y) + \partial_\sigma\xi_\rho(y)\}\, d^4y,\end{aligned} \tag{3.6}$$

etc.,

and

$$\begin{aligned}\frac{\partial}{\partial x^{\mu}} &\int a_1[\mu\nu, \rho\sigma, x - y]\phi_{\rho\sigma}(y)\, d^4y \\ &+ (1/2)a_0[\rho\sigma]\eta^{\nu\lambda}(\partial_\sigma\phi_{\lambda\rho} + \partial_\rho\phi_{\sigma\lambda} - \partial_\lambda\phi_{\rho\sigma}) = 0, \quad \text{etc.}\end{aligned} \tag{3.7}$$

Since $a_0[\mu\nu]$ is independent of the coordinate x [as is easily seen from graphical considerations],

(3.6) can be rewritten in the simpler form

$$A[\eta^{\mu\sigma}\partial^{\nu}+\eta^{\nu\sigma}\partial^{\mu}-\eta^{\mu\nu}\partial^{\sigma}]\,\delta(x-y) = -2(\partial/\partial x^{\rho})a_1[\mu\nu,\ \rho\sigma,\ x-y], \qquad (3.6')$$

where $a_0[\mu\nu]$ has been put as

$$a_0[\mu\nu] = A\eta^{\mu\nu} = \lim_{\kappa\to 0}(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle, \qquad (3.8)$$

$$(A = \text{const}).$$

In a similar manner (3.7) becomes

$$2(\partial/\partial x^{\mu})a_1[\mu\nu,\ \rho\sigma;\ x-y] + A[\eta^{\nu\rho}\partial^{\sigma}+\eta^{\nu\sigma}\partial^{\rho}-\eta^{\rho\sigma}\partial^{\nu}]\,\delta(x-y)=0. \qquad (3.7')$$

Equations (3.6') and (3.7') will later play an important role in the segregation of the singular parts of $(-g)^{1/2}\langle T^{\mu\nu}\rangle$.

4. SEPARATION OF DIVERGENCES FROM $\langle T^{\mu\nu}\rangle$

As has already been pointed out, it is necessary to remove singularities from the M_n's in order to obtain a finite solution of $\phi_{\mu\nu}$. The definition (2.9) and the definition of $S_{\mu\nu}$ show that the singularity of M_n arises solely from $\langle T^{\mu\nu}\rangle$ because the quantity $0_{\mu\nu}(\kappa^2)$ involved in $S_{\mu\nu}$ is a sum of products of ϕ and its derivatives and gives no trouble if ϕ is finite.

Following the conventional approach, let us consider the momentum representation of $(-g)^{1/2}\langle T^{\mu\nu}\rangle$:

$$(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle = \frac{1}{(2\pi)^2}\int \mathfrak{T}^{\mu\nu}(p)e^{ipx}\,d^4p$$

$$\phi_{\mu\nu}(x) = \frac{1}{(2\pi)^2}\int \phi_{\mu\nu}(k)e^{ikx}\,d^4k$$

$$a_n[\mu\nu,\ \rho_1\sigma_1,\ \cdots,\ \rho_n\sigma_n;\ z_1,\ \cdots,\ z_n] = \frac{1}{(2\pi)^{2n}}\int b_n[\mu\nu,\ \rho_1\sigma_1,\ \cdots\ \rho_n\sigma_n;\ q_1 q\ \cdots\ q_n] \times \exp\left[i\sum_{j=1}^{n}q_j z_j\right]\prod_{j=1}^{n}d^4q_j \qquad (n\geq 1),$$

and

$$\mathfrak{T}^{\mu\nu}(p) = (2\pi)^2\sum_{n=0}^{\infty}\frac{\kappa^n}{n!}\int\delta\left[p-\sum_{j=1}^{n}p_j\right] \times b_n[\mu\nu,\ \rho_1\sigma_1,\ \cdots\ \rho_n\sigma_n;\ p_1, p_2,\ \cdots\ p_n] \times \prod_{j=1}^{n}\phi_{\rho_j\sigma_j}(p_j)\,dp_j. \qquad (4.1)$$

The quantity b_n in (4.1) is represented by the diagram shown in Fig. 3 and diverges like ∞^4 regardless of the number of g lines.

The singular part of b_n can be defined as follows:

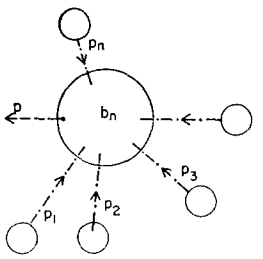

FIG. 3. Diagram of b_n.

$$^{\text{sing}}b_n[\mu\nu,\ \rho_1\sigma_1\ \cdots\ \rho_n\sigma_n;\ p_1, p_2\ \cdots\ p_n] = A_n[\mu\nu,\ \rho_1\sigma_1,\ \cdots,\ \rho_n\sigma_n] + \frac{1}{2!}\sum_{i,k=1}^{n}p_{(i),\alpha}\cdot p_{(k),\beta}\cdot B_{n,(ik)} \times[\mu\nu,\ \rho_1\sigma_1,\ \cdots,\ \rho_n\sigma_n,\ \alpha\beta] + \frac{1}{4!}\sum_{i,j,h,k=1}^{n}p_{(i),\alpha}\cdot p_{(j),\beta}\cdot p_{(h),\gamma}\cdot p_{(k),\delta}\cdot C_{n(ijhk)} \times[\mu\nu,\ \rho_1\sigma_1,\ \cdots,\ \rho_n\sigma_n,\ \alpha\beta\gamma\delta] \qquad (n\geq 1),$$

and

$$A_0[\mu,\nu]\equiv A\eta^{\mu\nu} = b_0[\mu\nu],$$

where

$$A_n[\mu\nu,\ \cdots,\ \sigma_n] \equiv b_n[\mu\nu,\ \cdots\sigma_n;\ p_1,\ \cdots,\ p_n]_{\text{all }p=0} = O(\infty^4),$$

$$B_{n,(jk)}\equiv\left[\frac{\partial^2}{\partial p_{(j)\alpha}\,\partial p_{(k)\beta}}b_n[\mu\nu\ \cdots\ p_n]\right]_{\text{all }p=0} = O(\infty^2)$$

$$C_{n,(ijhk)} \equiv \partial^4 b_n/\partial p_{(i)\alpha}\cdot\partial p_{(j)\beta}\cdot\partial p_{(h)\gamma}\cdot\partial p_{(k)\delta}\,\text{all }p=0 = O(\log\infty).$$

The quantity

$$b_n - {}^{\text{sing}}b_n = {}^{\text{finite}}b_n$$

is finite provided all the singularities involved in Green's functions of the matter fields, due to the mutual interactions, have been removed beforehand by a suitable subtraction technique.

The singular coefficients A, B, and C are sums of products of $\eta^{\mu\nu}$. The terms which are cubic and linear in p do not appear in $^{\text{sing}}b_n$ because there exist no constant tensors of odd rank.

The singular parts of $\langle T^{\mu\nu}\rangle$ must by themselves satisfy the relations (3.4) and (3.5). This then allows us to obtain concrete expressions for all the $^{\text{sing}}b_n$'s $(n\geq 2)$ in terms of $^{\text{sing}}b_0$ and $^{\text{sing}}b_1$.

We introduce the following notation:

$$^{\text{sing}}\{(-g)^{\frac{1}{2}}\langle T^{\mu\nu}\rangle\} = {}^{\text{sing}}\mathfrak{T}^{\mu\nu}_A(x) + {}^{\text{sing}}\mathfrak{T}^{\mu\nu}_B(x) + {}^{\text{sing}}\mathfrak{T}^{\mu\nu}_C(x),$$

where

$$
\begin{aligned}
{}^{\text{sing}}\mathfrak{J}_A^{\mu\nu}(p) &= \frac{1}{(2\pi)^2}\int e^{-ipx}\,{}^{\text{sing}}\mathfrak{J}_A^{\mu\nu}(x)\,dx \\
&= (2\pi)^2\sum_{n=0}^{\infty}\frac{\kappa^n}{n!}\int \delta\Big[p-\sum_{j=1}^{n}p_j\Big]\cdot A_n[\mu\nu,\ \rho_1\cdots\sigma_n] \\
&\quad\times\prod_{j=1}^{n}\phi_{\rho_j\sigma_j}(p_j)\,d^4p_j, \qquad (4.2)
\end{aligned}
$$

etc.

${}^{\text{sing}}\mathfrak{J}_A^{\mu\nu}(x)$ is a tensor density under general coordinate transformations and satisfies the conservation law (3.5). Since

$$
\lim_{\kappa\to 0}{}^{\text{sing}}\mathfrak{J}_A^{\mu\nu}(x) = A\eta^{\mu\nu} = \lim_{\kappa\to 0}[(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle], \qquad (4.3)
$$

it therefore follows that in general

$$
{}^{\text{sing}}\mathfrak{J}_A^{\mu\nu}(x) = A(-g)^{1/2}g^{\mu\nu}(x), \qquad (4.4)
$$

where the infinite constant can be calculated from the expression

$$
\lim_{\kappa\to 0}(-g)^{1/2}\langle T^{\mu\nu}(x)\rangle = A\eta^{\mu\nu}.
$$

Similar considerations may be applied to the determination of ${}^{\text{sing}}\mathfrak{J}_B$ and ${}^{\text{sing}}\mathfrak{J}_C$. We have

$$
\begin{aligned}
\lim_{\kappa\to 0}\frac{1}{\kappa}\,{}^{\text{sing}}\mathfrak{J}_B^{\mu\nu}(p) &= (2\pi)^2\int \delta[p-p']\,\frac{1}{2!}\,p'_\alpha p'_\beta B_1 \\
&\quad\times[\mu\nu,\ \rho\sigma,\ \alpha\beta]\phi_{\rho\sigma}(p')\,d^4p' \\
&= ((2\pi)^2/2!)p_\alpha p_\beta B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta]\phi_{\rho\sigma}(p), \qquad (4.5)
\end{aligned}
$$

and

$$
\begin{aligned}
&\lim_{\kappa\to 0}\frac{1}{\kappa}\,{}^{\text{sing}}\mathfrak{J}_C^{\mu\nu}(p) \\
&= ((2\pi)^2/4!)p_\alpha p_\beta p_\gamma p_\delta C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta,\ \gamma\ \delta]\phi_{\rho\sigma}(p). \qquad (4.6)
\end{aligned}
$$

We now note that the general structure of the coefficients B_1 and C_1 is completely determined by relations (3.6′) and (3.7′), which, in the momentum representation, take the forms

$$
\begin{aligned}
&A[\eta^{\mu\sigma}p^\nu + \eta^{\nu\sigma}p^\mu - \eta^{\mu\nu}p^\sigma] = -2(2\pi)^2p_\rho\,\{A_1[\mu\nu,\ \rho\sigma] \\
&\quad + (1/2!)p_\alpha p_\beta B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta] \\
&\quad + (1/4!)p_\alpha p_\beta p_\gamma p_\delta C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta\gamma\delta]\}, \qquad (4.7)
\end{aligned}
$$

and

$$
\begin{aligned}
&2(2\pi)^2p_\mu\{A_1[\mu\nu,\ \rho\sigma] + (1/2!)p_\alpha p_\beta B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta] \\
&\quad + (1/4!)p_\alpha p_\beta p_\gamma p_\delta C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta\gamma\delta]\} \\
&\quad + A\{\eta^{\nu\rho}p^\sigma + \eta^{\nu\sigma}p^\rho - \eta^{\rho\sigma}p^\nu\} = 0. \qquad (4.8)
\end{aligned}
$$

Comparing terms of equal order in p on both sides, we find

$$
\begin{aligned}
&A_1[\mu\nu,\ \rho\sigma] \\
&\quad = (A/2(2\pi)^2)[\eta^{\mu\nu}\eta^{\rho\sigma} - \eta^{\mu\rho}\eta^{\nu\sigma} - \eta^{\nu\rho}\eta^{\mu\sigma}] \qquad (4.9) \\
&\left.\begin{aligned} p_\rho B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta]p_\alpha p_\beta &= 0 \\ p_\mu B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta]p_\alpha p_\beta &= 0 \end{aligned}\right\}, \qquad (4.10) \\
&\left.\begin{aligned} p_\rho C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta\gamma\ \delta]p_\alpha p_\beta p_\gamma p_\delta &= 0 \\ p_\mu C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta\gamma\ \delta]p_\alpha p_\beta p_\gamma p_\delta &= 0, \end{aligned}\right\}. \qquad (4.11)
\end{aligned}
$$

The expression of A_1 given by (4.9) agrees with that derived from (4.4), whereas Eqs. (4.10) and (4.11) have the general solutions

$$
\begin{aligned}
&(1/2!)B_1[\mu\nu,\ \rho\sigma,\ \alpha\beta]p_\alpha p_\beta = (B/2(2\pi)^2)[p^2\{\eta^{\mu\nu}\eta^{\rho\sigma} \\
&\quad - (1/2)(\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho})\} \\
&\quad - (p^\mu p^\nu\eta^{\rho\sigma} + p^\rho p^\sigma\eta^{\mu\nu}) + (1/2)(p^\mu p^\rho\eta^{\nu\sigma} \\
&\quad + p^\mu p^\sigma\eta^{\nu\rho} + p^\nu p^\rho\eta^{\mu\sigma} + p^\nu p^\sigma\eta^{\mu\rho})], \qquad (4.12)
\end{aligned}
$$

and

$$
\begin{aligned}
&(1/4!)C_1[\mu\nu,\ \rho\sigma,\ \alpha\beta\gamma\ \delta]p_\alpha p_\beta p_\gamma p_\delta \\
&\quad = (2C/(2\pi)^2)[p^2\{\eta^{\mu\nu}\eta^{\rho\sigma}p^2 - p^\mu p^\nu\eta^{\rho\sigma} \\
&\quad - p^\rho p^\sigma\eta^{\mu\nu}\} + p^\mu p^\nu p^\rho p^\sigma] \\
&\quad + (D/(2\pi)^2)[(p^2)^2(\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho}) \\
&\quad - p^2(p^\mu p^\rho\eta^{\nu\sigma} + p^\mu p^\sigma\eta^{\nu\rho} + p^\nu p^\rho\eta^{\mu\sigma} \\
&\quad + p^\nu p^\sigma\eta^{\mu\rho}) + 2p^\mu p^\nu p^\rho p^\sigma]. \qquad (4.13)
\end{aligned}
$$

Explicit expressions for the infinite constants B, C, and D may be obtained directly from $b_1[\mu\nu,\ \rho\sigma,\ p]$ by the method of regulators.

Substituting (4.12) and (4.13) into (4.5) and (4.6), we get

$$
\begin{aligned}
&\lim_{\kappa\to 0}[1/\kappa\ {}^{\text{sing}}\mathfrak{J}_B^{\mu\nu}(x)] = -(B/2) \\
&\quad\times[\eta^{\mu\nu}\,\Box\phi - \Box\phi^{\mu\nu} - \partial^\mu\partial^\nu\phi - \eta^{\mu\nu}\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma} \\
&\quad + \partial^\mu\,\partial_\rho\phi^{\rho\nu} + \partial^\nu\,\partial_\rho\phi^{\rho\mu}], \qquad (4.14)
\end{aligned}
$$

$$
\begin{aligned}
&\lim_{\kappa\to 0}[1/\kappa\ {}^{\text{sing}}\mathfrak{J}_C^{\mu\nu}(x)] = 2C[\eta^{\mu\nu}\,\Box^2\phi - \Box\partial^\mu\,\partial^\nu\phi \\
&\quad - \eta^{\mu\nu}\,\Box\partial^\rho\,\partial^\sigma\phi_{\rho\sigma} + \partial^\mu\,\partial^\nu\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma}] \\
&\quad + 2D[\Box^2\phi^{\mu\nu} - \Box\partial^\mu\,\partial_\rho\phi^{\rho\nu} \\
&\quad - \Box\partial^\nu\,\partial_\rho\phi^{\rho\mu} + \partial^\mu\,\partial^\nu\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma}]. \qquad (4.15)
\end{aligned}
$$

Consider now the following invariants:

$$
I_0 = \int (-g)^{1/2}R\,d^4x,
$$

$$
I_1 = \int (-g)^{1/2}R^2\,d^4x,
$$

and

$$I_2 = \int (-g)^{1/2} R^{\alpha\beta\gamma\delta} R_{\alpha\beta\gamma\delta}\, d^4x.$$

The variational derivatives of these invariants with respect to $g_{\mu\nu}$ are readily found to be

$$\begin{aligned}\delta I_0/\delta g_{\mu\nu}(x) &= -(-g)^{1/2}(R^{\mu\nu} - (1/2)g^{\mu\nu}R) \equiv -(-g)^{1/2}G_{(0)}{}^{\mu\nu} \\ &= (\kappa/2)\,[\eta^{\mu\nu}\Box\phi - \Box\phi^{\mu\nu} - \partial^\mu\,\partial^\nu\phi - \eta^{\mu\nu}\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma} \\ &\quad + \partial^\nu\,\partial_\rho\phi^{\rho\mu} + \partial^\mu\,\partial_\rho\phi^{\rho\nu}] + 0(\kappa^2), \qquad (4.16)\end{aligned}$$

$$\begin{aligned}\delta I_1/\delta g_{\mu\nu}(x) &\overset{d}{\equiv} (-g(x))^{1/2}G^{\mu\nu}_{(1)}(x) \\ &= 2\kappa[\eta^{\mu\nu}\Box^2\phi - \eta^{\mu\nu}\Box\partial^\rho\,\partial^\sigma\phi_{\rho\sigma} \\ &\quad - \partial^\mu\,\partial^\nu\Box\phi + \partial^\mu\,\partial^\nu\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma}] + 0(\kappa^2), \qquad (4.17)\end{aligned}$$

$$\begin{aligned}\delta I_2/\delta g_{\mu\nu}(x) &\overset{d}{\equiv} (-g)^{1/2}G^{\mu\nu}_{(2)}(x) \\ &= 2\kappa[\Box^2\phi^{\mu\nu} + \partial^\mu\,\partial^\nu\,\partial^\rho\,\partial^\sigma\phi_{\rho\sigma} \\ &\quad - \Box\partial^\mu\,\partial_\rho\phi^{\rho\nu} - \Box\partial^\nu\,\partial_\rho\phi^{\rho\mu}] + 0(\kappa^2). \qquad (4.18)\end{aligned}$$

Since the I_α ($\alpha = 0, 1, 2$) are invariant under general coordinate transformations, we have the identities

$$\partial_\mu[(-g)^{1/2}G^{\mu\nu}_{(\alpha)}] + \Gamma^\nu_{\rho\sigma}(-g)^{1/2}G^{\rho\sigma}_{(\alpha)} \equiv 0 \qquad (\alpha = 0, 1, 2). \qquad (4.19)$$

By comparing (4.14) and (4.15) with (4.16), (4.17), and (4.18) we are therefore enabled to infer, to all orders of κ,

$$\begin{aligned}{}^{\text{sing}}\mathfrak{T}^{\mu\nu}_B(x) &= B(-g)^{1/2}G_0^{\mu\nu}(x) \\ &= B(-g)^{1/2}(R^{\mu\nu} - 1/2g^{\mu\nu}R), \qquad (4.20)\end{aligned}$$

$${}^{\text{sing}}\mathfrak{T}^{\mu\nu}_C(x) = \{CG^{\mu\nu}_{(1)}(x) + DG^{\mu\nu}_{(2)}(x)\}(-g)^{1/2}. \qquad (4.21)$$

Thus we arrive at the final expression

$$\begin{aligned}\langle T^{\mu\nu}(x)\rangle &= Ag^{\mu\nu}(x) + BG^{\mu\nu}_{(0)} \\ &\quad + CG^{\mu\nu}_{(1)}(x) + DG^{\mu\nu}_{(2)} + {}^{\text{finite}}\langle T^{\mu\nu}\rangle, \qquad (4.22)\end{aligned}$$

where $A(\approx \infty^4)$, $B(\approx \infty^2)$, $C(\approx \log \infty)$, and $D(\approx \log \infty)$ are constants the precise "values" of which depend on the number and character of the matter fields which have been included.

5. REMOVAL OF THE SINGULARITIES OF $\langle T^{\mu\nu}\rangle$

Consider the following renormalization of κ and $\phi_{\mu\nu}$:

$$\begin{aligned}\kappa &\to \kappa' = Z_\kappa, \\ \phi_{\mu\nu} &\to \phi'_{\mu\nu} = Z^{-1}\phi_{\mu\nu}.\end{aligned} \qquad (5.1)$$

This renormalization leaves $g_{\mu\nu}$ unchanged:

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa\phi_{\mu\nu} = \eta_{\mu\nu} + \kappa'\phi'_{\mu\nu} = g'_{\mu\nu}.$$

In place of the original action integral in Sec. 2, let us now introduce[6,7]

$$I = (1/2\kappa)I_0 + \alpha I_1 + \beta I_2 + \gamma I_3 + I_M, \qquad (5.2)$$

where α, β, and γ are indeterminate constants and

$$I_3 = \int (-g)^{1/2}\, d^4x,$$

$$I_M = \int (-g)^{1/2}\langle L_M\rangle\, d^4x.$$

Applying next the transformation (5.1) and keeping in mind that all the I_α's ($\alpha = 0, 1, 2, 3$ and M) are left unchanged thereby, we have

$$I = (Z/2\kappa')I'_0 + \alpha I'_1 + \beta I'_2 + \gamma I'_3 + I'_M, \qquad (5.2')$$

where

$$I'_\alpha \equiv I_\alpha[g_{\mu\nu} = \eta + \kappa'\phi'_{\mu\nu}].$$

The action (5.2′) leads to the new Einstein equation

$$\begin{aligned}&-2\kappa'(\delta I/\delta g'_{\mu\nu})/(-g')^{1/2} \\ &= ZG'^{\mu\nu}_0 - 2\kappa'\alpha G'^{\mu\nu}_1 - 2\kappa'\beta G'^{\mu\nu}_2 - \kappa'\gamma g'^{\mu\nu} \\ &\quad + \kappa'\langle T^{\mu\nu}(x, g')\rangle = 0. \qquad (5.3)\end{aligned}$$

Since the arguments of Sec. 4 which led to the conclusion (4.22), remain unchanged in the primed system, the quantity $\langle T^{\mu\nu}(x, g')\rangle \equiv \langle T'^{\mu\nu}\rangle$ retains the features of (4.22).

The new field equation may therefore be written in the form

$$\begin{aligned}&G'^{\mu\nu}_0 + \kappa'\,{}^{\text{finite}}\langle T'^{\mu\nu}\rangle \\ &\quad + \kappa'(A - \gamma)g'^{\mu\nu} + (\kappa'B + Z - 1)G'^{\mu\nu}_{(0)} \\ &\quad + \kappa'(C - 2\alpha)G'^{\mu\nu}_{(1)} + \kappa'(D - 2\beta)G'^{\mu\nu}_{(2)} = 0,\end{aligned}$$

[6] At the meeting of the American Physical Society in New York in January of 1961, Professor Feynman gave a talk in which he introduced similar counter terms into Einstein's equation.

[7] There is a linear relation

$$G^{\mu\nu}_{(1)} + G^{\mu\nu}_{(2)} - 4G^{\mu\nu}_{(4)} = 0,$$

where

$$G^{\mu\nu}_{(4)} = \frac{\delta}{\delta g_{\mu\nu}}\int (-g)^{1/2}R^{\rho\sigma}R_{\rho\sigma}\, d^4x,$$

$$G^{\mu\nu}_{(2)} = \frac{\delta}{\delta g_{\mu\nu}}\int (-g)^{1/2}R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}\, d^4x,$$

$$G^{\mu\nu}_{(1)} = \frac{\delta}{\delta g_{\mu\nu}}\int (-g)^{1/2}R^2\, d^4x.$$

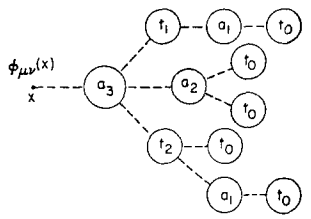

FIG. 4. Diagram corresponding to a typical term of the functional series for $\phi_{\mu\nu}(x)$. The dotted lines correspond to the Δ function.

where A, B, C, and D are the infinite constants previously introduced. If we now choose Z, α, β, and γ as follows:

$$\begin{aligned} Z &= 1 - \kappa'\beta, \\ \alpha &= C/2, \\ \beta &= D/2, \\ \gamma &= A, \end{aligned} \tag{5.4}$$

then we obtain the equation

$$G_0'^{\mu\nu} \equiv R'^{\mu\nu} - \tfrac{1}{2}g'^{\mu\nu}R' = -\kappa'\,{}^{\text{finite}}\langle T'^{\mu\nu}(x)\rangle, \tag{5.5}$$

or, more generally,

$$R'^{\mu\nu} - \tfrac{1}{2}g'^{\mu\nu}R' = -\kappa'\,{}^{\text{finite}}\langle T'^{\mu\nu}(x)\rangle - t^{\mu\nu}, \tag{5.5'}$$

which is free of singularities.

Recalling the definition of M_0 introduced in Sec. 2, and taking account of (4.3), we see now that

$$M_0 = 0.$$

Consequently, $\phi'_{\mu\nu}$ can be written as a functional series in the external source $t^{\mu\nu}$ starting with the linear term.

It is obvious that any term in this series can be represented by a tree-like diagram similar to the one shown in Fig. 4.

It will be noted that at the end of every branch of the tree $t_0^{\mu\nu}$ appears, where

$$t^{\mu\nu} = \sum_{n=0}^{\infty} \frac{\kappa^n}{n!} \int t_n^{\mu\nu}(\phi)^n,$$

$$\partial t_0^{\mu\nu}(x)/\partial x^\mu = 0.$$

Accordingly the gravitational potentials vanish in the special coordinate system chosen in this paper if the external source $t^{\mu\nu}$ vanishes.

6. DISCUSSION

(A) Simple example of ${}^{\text{sing}}(-g)^{1/2}\langle T^{\mu\nu}\rangle$

We have seen that in general four independent counter terms are needed for the removal of singularities appearing in $(-g)^{1/2}\langle T^{\mu\nu}\rangle$. Sometimes, however, this number is reduced if the original energy-momentum tensor satisfies certain algebraic identities.

As an example,[3] consider the pure electromagnetic field. In this case the relation

$$g_{\mu\nu}\langle T^{\mu\nu}\rangle \equiv 0$$

is well known. In order that ${}^{\text{finite}}\langle T^{\mu\nu}\rangle$ separately satisfy this relation, it is necessary to postulate

$$g_{\mu\nu}\,{}^{\text{sing}}\langle T^{\mu\nu}\rangle = 0,$$

which leads to the following relations in the momentum representation:

$$\eta_{\mu\nu}\,{}^{\text{sing}}b_0[\mu\nu] = 4A = 0$$

$$\eta_{\mu\nu}\,{}^{\text{sing}}b_1[\mu\nu,\ \rho\sigma,\ p] + \phi_{\mu\nu}(p)\,{}^{\text{sing}}b_0[\mu\nu] = 0.$$

These relations in turn yield

$$A = 0 \qquad B = 0 \qquad D = -3C, \tag{6.1}$$

whence

$$Z = 1 \quad \alpha = C/2 \quad \beta = -3C/2 = -3\alpha \quad \gamma = 0.$$

Therefore in this case only one renormalization constant is necessary.

(B) Specification of Δ

The Δ function introduced in Sec. 2 becomes well defined when a definite boundary condition is selected. In Sec. 1 we postulated, for the sake of convenience, that

$$\lim_{x^4 \to -\infty} g_{\mu\nu} \to \eta_{\mu\nu},$$

which can be satisfied if we choose the retarded function $\Delta_r(x - x')$. In the quantum theory, on the other hand, the Feynman function Δ^F is more appropriate. If Δ^F is employed, it may be necessary to apply an "adiabetic switching" process to the gravitation constant κ in order to satisfy the initial condition for $g_{\mu\nu}$.

(C) The Use of More General Matrix Elements for $\langle T_H{}^{\mu\nu}\rangle$

We have thus far restricted our attention to a particular type of expectation value for $T_H^{\mu\nu}(x)$. A more general one is given by

$$\langle T^{\mu\nu}\rangle_{ab} \equiv \frac{(\Psi_a^* S \cdot T_H^{\mu\nu}(x)\Psi_b)}{(\Psi_a^* \cdot S\Psi_b)},$$

where Ψ_a and Ψ_b are any pair of states belonging to the Hilbert space established at $x^4 = -\infty$. Graphical considerations allow us to split $\langle T^{\mu\nu}(x)\rangle_{ab}$ into two parts. The first part corresponds to the

set of diagrams in which the point x is not connected by any internal line with the incoming or outgoing material particles; the second part corresponds to the diagrams in which the point x is connected with the initial of final states and the operators of the matter fields involved in $T^{\mu\nu}_H$ create or destroy the incoming or outgoing particles.

Accordingly we can write

$$\langle T^{\mu\nu}(x)\rangle_{ab} = \frac{(\Psi_0^* S T^{\mu\nu}_H \Psi_0)}{(\Psi_0^* S \Psi_0)} \frac{(\Psi_a^* S \Psi_b)}{(\Psi_a^* S \Gamma_b)} + \frac{[\Psi_a^* S T^{\mu\nu}_H(x) \Psi_b]_{\text{connected}}}{(\Psi_a^* S \Psi_b)} = \langle T^{\mu\nu}(x)\rangle + T^{\mu\nu}(x, ab), \tag{6.2}$$

where $(\Psi_n^* S T^{\mu\nu} \Psi_b)_{\text{connected}}$ means the contributions from the "second part" mentioned above and $T(x, ab)$ is

$$T^{\mu\nu}(x, ab) = \frac{[\Psi_a^* S T^{\mu\nu}_H(x) \Psi_b]_{\text{connected}}}{(\Psi_a^* S \Psi_b)}.$$

It is easily seen that the two terms in (6.2) separately satisfy conservation law (2.4). Furthermore, $T^{\mu\nu}(ab)$ has no singularities provided the singularities due to the coupling of the matter fields to one another have been removed beforehand. This is because the insertion of the operator $T^{\mu\nu}_H$ into the S-matrix element has an effect similar to the simple insertion of an external g line into the corresponding diagram.

If the external source $t^{\mu\nu}(x)$ is replaced by $T^{\mu\nu}(x, ab)$, then the solution $\phi'_{\mu\nu}$ of (5.5′) gives the gravitational field produced by the material system during the transition from Ψ_b to Ψ_a. If this solution is inserted in the equation for the Green's functions of the matter fields, one then obtains a semiclassical description of the scattering of particles due to their gravitational interaction.

(D) Green's Function of the Gravitational Field

The external source $t^{\mu\nu}$ must satisfy

$$\partial_\mu t^{\mu\nu} + \Gamma^\nu_{\rho\sigma} t^{\rho\sigma} = 0, \tag{6.3}$$

which shows that $t^{\mu\nu}$ cannot be independent of $g_{\mu\nu}$. If we expand $t^{\mu\nu}$ in a functional power series in ϕ,

$$t^{\mu\nu}(x) = \sum_{n=0}^{\infty} (\kappa^n/n!) \times \int t_n[\mu\nu, \rho_1\sigma_1, \cdots \rho_n\sigma_n, x, y_1, \cdots y_n] \times \prod_{j=1}^{n} \{\phi_{\rho_j\sigma_j}(y_j)\, dy_j\},$$

and insert the expansion into (6.3) we can find, by a sequential process, expressions for the t_n's in terms of t_0 and a set of arbitrary functions which enter because (6.3) is not sufficient by itself to determine each term t_n completely. If we postulate, however, that each t_n should tend to zero when t_0 becomes zero, then the arbitrary functions become unique, and each t_n becomes proportional to t_0.

Consider a very small $t_0(x)$ (or $t^{\mu\nu}$). Since we know that $\phi_{\mu\nu}$ can be represented by a series in t beginning with the linear term, we can write the field equation (5.5′), correct to lowest order, in the form

$$\{\eta^{\rho\sigma}\partial^\mu\partial^\nu + \delta^{\mu\nu,\rho\sigma}\Box - \tfrac{1}{2}(\eta^{\nu\sigma}\partial^\mu\partial^\rho + \eta^{\mu\sigma}\partial^\nu\partial^\rho + \eta^{\nu\rho}\partial^\mu\partial^\sigma + \eta^{\mu\rho}\partial^\nu\partial^\sigma) - \eta^{\rho\sigma}\eta^{\mu\nu}\Box + \eta^{\mu\nu}\partial^\rho\partial^\sigma\} \int G_{\rho\sigma,\alpha\beta}(x, y) t_0^{\alpha\beta}(y)\, dy = -2\int \mathfrak{I}^{\mu\nu,\rho\sigma}_{(2)}(x - z) G_{\rho\sigma,\alpha\beta}(z, y) t_0^{\alpha\beta}(y)\, dy\, dz - 2t_0^{\mu\nu}(x). \tag{6.4}$$

The following abbreviations have been employed here:

$$\delta^{\mu\nu,\rho\sigma} \equiv \tfrac{1}{2}(\eta^{\mu\rho}\eta^{\nu\sigma} + \eta^{\mu\sigma}\eta^{\nu\rho}),$$
$$t_0^{\mu\nu}(x) \equiv t_0[\mu\nu, x],$$
$$G_{\rho\sigma,\alpha\beta}(x, y) \equiv \left[\frac{\delta\phi_{\rho\sigma}(x, t)}{\delta t_0^{\alpha\beta}(y)}\right]_{t_0=0}, \tag{6.5}$$
$$\mathfrak{I}^{\mu\nu,\rho\sigma}_{(2)}(x - z) \equiv \left[\frac{\delta}{\delta\phi_{\rho\sigma}(z)}\{(-g(x))^{1/2\ \text{finite}}\langle T^{\mu\nu}(x)\rangle\}\right]_{t_0=0} = \kappa^{\prime\ \text{finite}} a_1[\mu\nu, \rho\sigma, x - z]. \tag{6.6}$$

In virtue of the conditions

$$(\partial/\partial x^\mu)\mathfrak{I}^{\mu\nu,\rho\sigma}_{(2)}(x - z) = 0,$$

and

$$(\partial/\partial x^\mu) t_0^{\mu\nu}(x) = 0. \tag{6.7}$$

we can introduce the supplementary condition

$$\partial^\mu H_{\mu\nu,\rho\sigma}(x, y) = 0,$$
$$H_{\mu\nu,\rho\sigma} \equiv G_{\mu\nu,\rho\sigma} - \tfrac{1}{2}\eta_{\mu\nu}\eta^{\tau\lambda}G_{\tau\lambda,\rho\sigma},$$

which transforms (6.4) to

$$\int \Box H^{\mu\nu}_{,\rho\sigma}(x, y) t_0^{\rho\sigma}(y)\, dy = -2\int \mathfrak{I}^{\mu\nu,\rho\sigma}_{(2)}(x - z) G_{\rho\sigma,\alpha\beta}(z, y) \times t_0^{\alpha\beta}(y)\, dy\, dz - 2t_0^{\mu\nu}(x), \tag{6.4′}$$

whence

$$\Box H^{\mu\nu}_{\cdot\,\alpha\beta}(x, x') = -2\int \mathfrak{I}^{\mu\nu,\rho\sigma}_{(2)}(x - z)G_{\rho\sigma,\alpha\beta}(z, x')\,dz - 2\,\delta^{\mu\nu}_{\alpha\beta}\,\delta(x - x'), \tag{6.8}$$

or, alternatively,

$$\Box G_{\mu\nu,\alpha\beta}(x, x') = -2\int [\mathfrak{I}^{\rho\sigma}_{(2)\mu\nu,}(x - z) - \tfrac{1}{2}\eta_{\mu\nu}\eta_{\lambda\tau}\mathfrak{I}^{\lambda\tau,\rho\sigma}_{(2)}(x - z)]G_{\rho\sigma,\alpha\beta}(z, x')\,dz - 2[\eta_{\mu\nu,\alpha\beta} - \tfrac{1}{2}\eta_{\mu\nu}\eta_{\alpha\beta}]\,\delta(x - x') \tag{6.8′}$$

$$\eta_{\mu\nu,\alpha\beta} = \tfrac{1}{2}(\eta_{\mu\alpha}\eta_{\nu\beta} + \eta_{\mu\beta}\eta_{\nu\alpha}).$$

In the momentum representation $G(p)$ evidently has, roughly speaking, the form

$$G(p) \cong 1/\{p^2 + \mathfrak{I}_{(2)}(p)\}.$$

If the matter field is that of the neutral scalar meson, for example, one finds

$$\mathfrak{I}^{\mu\nu\rho\sigma}_{(2)}(p) \propto \int \frac{(p - k)^\mu k^\nu k^\rho (p - k)^\sigma}{[(p - k)^2 + \mu^2](k^2 + \mu^2)}\,d^4k.$$

In the high-energy limit ($p^2 \gg \mu^2$) the scaling relation

$$\mathfrak{I}_{(2)}(\lambda p) \approx \lambda^4 \mathfrak{I}_{(2)}(p) \tag{6.9}$$

is easily verified. The same relation also holds for *any* matter field. Now, as has already been noted, $\mathfrak{I}_{(2)}(p)$ always diverges like ∞^4. Let us therefore regularize it according to the Pauli-Villars prescription. As long as the regulating masses remain finite the scaling relation (6.9) is still satisfied. In order to segregate the eventually divergent parts we may expand ${}^{\text{reg}}\mathfrak{I}_{(2)}(p)$ in a power series in p, obtaining an expression of the form

$${}^{\text{reg}}\mathfrak{I}_{(2)}(p) = A + (1/2!)\sum Bpp + (1/4!)\sum Cpppp + {}^{\text{finite}}\mathfrak{I}_{(2)}(p).$$

Of the eventually diverging terms the third dominates in the high energy limit (although it has the weakest divergence). Since this term satisfies the same scaling relation as ${}^{\text{reg}}\mathfrak{I}_{(2)}(p)$ so also does ${}^{\text{finite}}\mathfrak{I}_{(2)}(p)$. Therefore[8,9]

$${}^{\text{finite}}\mathfrak{I}_{(2)}(p) \approx O(p^4), \qquad p^2 \gg \mu^2.$$

This is in contrast to the renormalized electron-mass operator in electrodynamics, for example, which has the asmptotic behavior

$$\Sigma(p) \approx O(p),$$

which leaves the nature of the singularity of the Green's function

$$G(p) \approx 1[\not{p} + m + \Sigma(p)]$$

essentially unchanged in the coordinate representation from that of the bare electron. In the present case the singularity of the Green's function becomes much weaker. This interesting feature suggests that the fully quantized theory, of which the "bubble diagram" approximation corresponds to the semiclassical theory presented here, may contain within itself the automatic "cutoff" which has been speculated upon many times. That this is, in fact, the case will be the subject of a subsequent paper. Once it is recognized that the asymptotic behavior of $G(p)$ is $1/p^4$, it may then be shown that this behavior is maintained to all orders of perturbation theory, no matter how complicated the corresponding diagrams become, because of the very special property of the gravitational interaction that the insertion of any number of external g lines leaves the degree of divergence of a given diagram unaffected. The "cutoff" is *not* sharp enough to eliminate all divergences, but no others beyond those already enumerated in the present paper appear.

[8] Let us consider, in general, some quantity $F(p)$, a function of the external momenta p's, which corresponds to some type of Feynman graph. Suppose that $F(p)$ is divergent of the order ∞^n. Then one can show that the renormalized $F'(p)$ which is defined by

$$F'(p) = F(p) - \sum_{k=0}^{n} (p^k/k!)\cdot[\partial^k F(q)/\partial q^k]_{q=0},$$

has the asymptotic behavior

$$F'(p) \approx O(p^n) \quad \text{for} \quad p \to \infty,$$

where ∞^0, $O(p^0)$ mean $\log \infty$ and $O(\log p)$, respectively.

[9] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*. (Interscience Publishers, Inc., New York, 1959), see Sec. 26.3.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Definition of Commutators via the Uncertainty Principle

BRYCE S. DEWITT
Institute of Field Physics, Department of Physics, University of North Carolina, Chapel Hill, North Carolina

(Received October 20, 1961)

A general demonstration is given of the fact that the quantization of any system, as expressed by the uncertainty principle, implies the quantization of all other systems to which it can be coupled. It is also shown that the precise form which the uncertainty principle takes is uniquely specified. That is to say, the uncertainty principle effectively defines the commutator of any pair of observables. The argument serves to detach completely from its canonical origins a previously given definition of the commutator (which is a generalization of one given by Peierls) applicable to systems with infinite-dimensional invariance groups, for which the identification of canonical variables is an inconvenience. The measurements of a single observable and of a pair of observables are analyzed according to the Bohr-Rosenfeld scheme.

INTRODUCTION

IN a previous paper[1] a means of constructing the commutator of any two observables was given which is independent of the explicit discovery of pairs of canonically conjugated variables. The method is essentially a generalization of Peierls' technique[2] to the case of dynamical systems possessing infinite dimensional invariance groups. The category of "observables" is restricted to the set of group invariants, and the main analytical tools are the Green's functions which describe the propagation of small disturbance in the system in question.

Although the connection with canonical theory was established in (I), it was pointed out there that the Green's function approach is a logically independent structure which can be based directly on the quantum theory of measurement. The uncertainty principle serves to *define* the commutator in terms of the interference between two measurements. It is the purpose of the present paper to demonstrate this fact in detail.

The Peierls definition of the Poisson bracket is briefly reviewed in Sec. 1, and a careful statement of the uncertainty principle is given. In Sec. 2 the theory is applied to the measurement of a single observable, and the necessity for the introduction of compensation mechanisms supplementing the measuring apparatus is shown. The measurement of two observables is studied in Sec. 3, where it is shown that even with the introduction of correlation as well as compensation mechanisms, the mutual interference of the two measurements prevents the achievement of a simultaneous accuracy better than that allowed by the uncertainty principle. Limitations imposed by the strictly classical nature of the analysis are briefly considered at the end of Sec. 3.

The demonstration is essentially a paraphrase, applicable to completely arbitrary systems, of the Bohr-Rosenfeld paper[3] on the electromagnetic commutators. It is shown in a quite general manner that the quantization of a given system implies also the quantization of any other system to which it can be coupled. By a principle of induction, therefore, the quantum theory must be extended to all physical systems. Moreover, the precise form of the commutator between any two observables is uniquely specified, with, however, one qualification: The uncertainty principle, like the correspondence principle, is only a guide to the exact theory and cannot, for example, resolve the factor ordering ambiguity which occurs in the transition from a classical expression to a quantum operator. Except where otherwise indicated the notation is that of (I).

1. THE POISSON BRACKET

The theory of the Poisson bracket given in (I) was based on a transformation of the action functional of the system in question, of the form

$$S \rightarrow S + \epsilon B, \tag{1.1}$$

where B is any invariant of the system and ϵ is an infinitesimal constant. Such a transformation may be regarded as yielding the most rudimentary description of the coupling of the original system weakly to an external system. The smallness of ϵ corresponds to the weakness of the coupling and

[1] B. S. DeWitt, J. Math. Phys. **2**, 151 (1961). This paper will be referred to as (I).
[2] R. E. Peierls, Proc. Roy. Soc. (London) **A214**, 143 (1952).
[3] N. Bohr and L. Rosenfeld, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **12**, 8 (1933).

its constancy corresponds to the neglect of the reaction of the original system back on the external system, a neglect which is justified if the external system is sufficiently macroscopic and the primary object of interest is the disturbance which the external system itself produces. The choice of B as an invariant of the original system implies that the coupling maintains all invariance properties intact.

The change (1.1) induces advanced and retarded changes $\delta_B^{\pm}\psi^i$ in the dynamical variables ψ^i, which satisfy the equation

$$\int S_{,ij'}\, \delta_B^{\pm}\psi^{j'}\, d^4x' = -\epsilon B_{,i} \tag{1.2}$$

[cf. (I, 3.9)]. Although the $\delta_B^{\pm}\psi^i$ are not uniquely determined by this equation when an infinite-dimensional invariance group is present, the corresponding variations

$$\delta_B^{\pm}A = \int A_{,i}\, \delta_B^{\pm}\psi^i\, d^4x \tag{1.3}$$

in any group invariant A are well defined. Following Peierls[2] it is convenient to introduce the quantity

$$D_BA \equiv \lim_{\epsilon\to 0} (1/\epsilon)\, \delta_B^{-}A. \tag{1.4}$$

The choice of the retarded boundary condition here, rather than the advanced, anticipates the "one way" character of the measurement process in the description of which this quantity will be used in the following sections.

From the linearity of Eq. (1.2), which permits the application of the superposition principle, the following identities, involving group invariants A, B, C, may be readily inferred:

$$D_A(B + C) = D_AB + D_AC, \tag{1.5}$$

$$D_A(BC) = (D_AB)C + BD_AC, \tag{1.6}$$

$$D_{A+B}C = D_AC + D_BC, \tag{1.7}$$

$$D_{AB}C = (D_AC)B + AD_BC. \tag{1.8}$$

The Poisson bracket of two invariants A and B will, for *all* physical systems, be defined by

$$(A, B) \equiv D_AB - D_BA. \tag{1.9}$$

In view of the reciprocity theorem

$$\delta_A^{\pm}B = \delta_B^{\mp}A \tag{1.10}$$

[cf. (I, 1.14), (I, 3.20), and footnote 11 of (I)] this definition is identical to that given in (I). In the case of systems possessing no canonical constraints it has been shown by Peierls[2] to reduce to the conventional one. It will be noted immediately that the usual identities

$$(A, B) = -(B, A), \tag{1.11}$$

$$(A, B + C) = (A, B) + (A, C), \tag{1.12}$$

$$(A, BC) = (A, B)C + B(A, C), \tag{1.13}$$

are satisfied. The proof of the Poisson-Jacobi identity is given in (I).

The system S is formally quantized by relating the commutator to the Poisson bracket in the familiar manner

$$[A, B] = i(A, B) = i(D_AB - D_BA), \qquad (\hbar = 1), \tag{1.14}$$

which leads immediately to the uncertainty principle

$$\Delta A\ \Delta B \geq \tfrac{1}{2}\, |\langle (A, B)\rangle|, \tag{1.15}$$

where ΔA and ΔB are the root-mean-square deviations of A and B, respectively, from their average values $\langle A\rangle$ and $\langle B\rangle$ in the quantum state in question. It is important in what follows to have a clear understanding of exactly what the uncertainty principle says in physical terms. The uncertainty principle is a statement about the fundamental limitations imposed by the quantum theory on the relation between measurements and prediction—more precisely, between measurements and the possibilities of making predictions expressed in classical language. Suppose the observable[4] A has been measured with an accuracy ΔA; what does this imply in the way of restrictions on the accuracy of predictions concerning the ourcome of subsequent measurements? Taking note of the fact that the measurement of a given observable will, in general, occupy a finite amount of time, which may itself be involved in the definition of the observable, let us first consider the case in which the interval associated with the observable B is subsequent to that associated with A. The uncertainty principle then states that as a result of the uncontrollable disturbance in the system produced by the measurement of A, the use of a classical value for B in making predictions about the outcome of subsequent measurements of quantities which depend on B is limited to the extent of an uncertainty ΔB which is given by relation (1.15). The "classical value" to be used for B in this case is its average value $\langle B\rangle$ in the quantum state resulting from (or "prepared by") the measurement of A. For simplicity in the subsequent discussion the brackets $\langle\ \rangle$ will be omitted

[4] Any invariant is here regarded as an observable.

whenever it is clear from the context that the "classical value" is meant. In particular, we shall write the uncertainty relation in the loose form

$$\Delta A\ \Delta B \gtrsim \tfrac{1}{2}\,|(A, B)| = \tfrac{1}{2}\,|D_A B - D_B A|. \qquad (1.16)$$

The relation between A and B is a completely reciprocal one, and because of the time reversibility of quantum mechanics the above situation may be equally well described in terms of a limitation ΔA on *retrodictions* conditioned by a measurement of B with accuracy ΔB in the future. The necessity for this reciprocity is revealed with particular keenness in the case in which the time intervals associated with A and B overlap. The simplicity of the previous description, in which the measurements of A and B could be ordered in temporal sequence, is missing in this case, and the state of the system must here be regarded as conditioned simultaneously by the results of both measurements, together with their mutual interference.

2. MEASUREMENT OF A SINGLE OBSERVABLE

Measurements are performed on a system S through coupling with a second system S called the *apparatus*. In principle, any invariant can be measured through suitable choice of apparatus and coupling. We shall assume that the uncertainty principle holds in the form (1.16) for the apparatus. It will then be shown (in the next section) to hold in the same form for the system S. By induction, therefore, it may be extended in this form to all physical systems. Consequently, if a description of nature is demanded which avoids the use of "hidden variables," the commutator must in all cases be given by (1.14).

The coupling of system to apparatus which is mathematically the most convenient for analyzing the measurement of a single observable A is that which is expressed (at least approximately) by a total action functional of the form

$$S + \mathbf{S} + gxA.$$

Here g is an adjustable "coupling constant" and x is some convenient apparatus variable. For example, in the Stern-Gerlach experiment, where A is an atomic spin, x is effectively a finite time integral of the z component of the position of the atom in a magnetic field which is inhomogeneous in the z direction, the strength of the field and the magnitude of the atomic magnetic moment being described by g. Similarly, in an electromagnetic field measurement, where A is an average of a field component over some space-time domain, x is a space-time average over the positions of the constituent particles of an appropriate test body, the electric or magnetic charge on the test body being contained in g. It is true, of course, that the classical description of the measurement process which is given in what follows is strictly applicable only to the observables A possessing continuous or at least quasi-continuous spectra. The uncertainty principle itself, however, can be extended in well-known ways to discrete-spectrum situations.

The measurement of A is carried out by determining the deviation in the value of some other suitable apparatus variable π, as a result of the coupling, from the value it would have had in the absence of coupling ($g = 0$). The variable π is said to be "suitable" if it satisfies the conditions

$$D_\pi x = 0, \qquad (x, \pi) = D_x \pi \neq 0. \qquad (2.1)$$

That is, π describes a dynamical state of affairs subsequent to the coupling process so that although x has a retarded effect on π, π has no retarded effect on x. For example, in the Stern-Gerlach experiment π might be the position of the point at which the atom strikes a photographic plate after having passed through the magnetic field, while in an electromagnetic field measurement π may be the total momentum of the test body at the end of the time interval involved in the coupling term gxA, as observed via the Doppler shift of photons, for example.[5] Of course, it is in the analysis of the final observation, performed upon the apparatus variable π, that the source of many of the polemics concerning the conceptual foundations of the quantum theory lies. But the resolution of the difficulties inherent in this analysis, whether in terms of a discontinuous "collapsible" behavior of wave functions, as demanded by the Copenhagen school,[6] or by insistence on an isomorphism between the real world and an infinitely "branching" universal wave function,[7,8] or with the aid of some other viewpoint, is largely a metaphysical problem, irrelevant to the present discussion.

---

[5] In reference 3, Bohr and Rosenfeld actually found it more convenient to work with a π which satisfies the opposite conditions

$$D_x \pi = 0, \qquad (x, \pi) = -D_\pi x \neq 0.$$

Their π is the difference between the momenta at the beginning and end of the measurement interval. Since the final momentum is measured by them with infinite accuracy the deviation of their π from the value it would have in the absence of coupling satisfies advanced boundary conditions.

[6] W. Heisenberg, in *Niels Bohr and the Development of Physics*, edited by W. Pauli (Pergamon Press, New York, 1955).

[7] H. Everett, III, Revs. Modern Phys. **29**, 454 (1957).

[8] J. A. Wheeler, Revs. Modern Phys. **29**, 463 (1957).

Let us denote the complete set of apparatus variables by Φ^α. Without essential loss of generality they may, like the ψ^i, be regarded as functions of space as well as time. In the absence of coupling the dynamical variables ψ^i, Φ^α satisfy the equations

$$S_{,i}[\psi] = 0, \tag{2.2}$$

$$\mathbf{S}_{,\alpha}[\Phi] = 0. \tag{2.3}$$

The presence of coupling induces deviations $\delta^-\psi^i$, $\delta^-\Phi^\alpha$ which satisfy

$$\begin{aligned} 0 &= S_{,i}[\psi + \delta^-\psi] + gx[\Phi + \delta^-\Phi]A_{,i}[\psi + \delta^-\psi] \\ &= \int S_{,ij'}\,\delta^-\psi^{j'}\,d^4x' + \tfrac{1}{2}\int d^4x' \\ &\quad \times \int d^4x''S_{,ij'k''}\,\delta^-\psi^{j'}\,\delta^-\psi^{k''} + \cdots \\ &\quad + gxA_{,i} + gx\int A_{,ij'}\,\delta^-\psi^{j'}\,d^4x' \\ &\quad + gA_{,i}\int x_{,\alpha'}\,\delta^-\Phi^{\alpha'}\,d^4x' + \cdots, \end{aligned} \tag{2.4}$$

$$\begin{aligned} 0 &= \mathbf{S}_{,\alpha}[\Phi + \delta^-\Phi] + gx_{,\alpha}[\Phi + \delta^-\Phi]A[\psi + \delta^-\psi] \\ &= \int \mathbf{S}_{,\alpha\beta'}\,\delta^-\Phi^{\beta'}\,d^4x' + \tfrac{1}{2}\int d^4x' \\ &\quad \times \int d^4x''S_{,\alpha\beta'\gamma''}\,\delta^-\Phi^{\beta'}\,\delta^-\Phi^{\gamma''} + \cdots \\ &\quad + gx_{,\alpha}A + gA\int x_{,\alpha\beta'}\,\delta^-\Phi^{\beta'}\,d^4x' \\ &\quad + gx_{,\alpha}\int A_{,i'}\,\delta^-\psi^{i'}\,d^4x' + \cdots \end{aligned} \tag{2.5}$$

correct to second order in g. We now assume g to be small. In many analyses of the measurement process precisely the opposite is assumed, the coupling between system and apparatus being taken as intense, although of brief duration. The weak coupling limit, however, corresponds to the classical idea that a careful measurement should be able to impart a negligible disturbance to the object system. Our purpose here is to determine the limitations on this idea and to find the uncontrollable uncertainties which remain in spite of all precautions.

In the weak coupling limit we have

$$\int S_{,ij'}\,\delta^-\psi^{j'}\,d^4x' = -gxA_{,i}, \tag{2.6}$$

and hence,

$$\delta^-A = gx\,D_AA. \tag{2.7}$$

In the corresponding apparatus equation, however' it would not be correct to retain only first-order terms. In order to allow for the very disturbance (2.7) which the measurement produces in A, and hence to observe A in the truly quantum domain, it is necessary to go to the second order. On the other hand, by choosing the apparatus sufficiently macroscopic ($|\mathbf{S}_{,\alpha\beta'}| \gg |S_{,ij'}|$), we may ignore those second-order terms which involve the $\delta^-\Phi^\alpha$. We have, therefore,

$$\begin{aligned} 0 &= \int \mathbf{S}_{,\alpha\beta'}\,\delta^-\Phi^{\beta'}\,d^4x' + gx_{,\alpha}A \\ &\quad + gx_{,\alpha}\int A_{,i'}\,\delta^-\psi^{i'}\,d^4x' \\ &= \int \mathbf{S}_{,\alpha\beta'}\,\delta^-\Phi^{\beta'}\,d^4x' + g(A + \delta^-A)x_{,\alpha}, \end{aligned} \tag{2.8}$$

and hence,[9]

$$\delta^-\pi = g(A + \delta^-A)\,D_x\pi. \tag{2.9}$$

Solving Eq. (2.9) we get

$$A = \frac{\delta^-\pi}{g\,D_x\pi} - gx\,D_AA, \tag{2.10}$$

which expresses A in terms of the "experimental data," and from which it follows that the mean square uncertainty in the measurement of A satisfies

$$\begin{aligned} (\Delta A)^2 \gtrsim\ & \frac{1}{2}\left(\frac{\Delta\pi}{g\,|D_x\pi|} - g\,|D_AA|\,\Delta x\right)^2 \\ & + \frac{1}{2}\left(\frac{\Delta\pi}{g\,|D_x\pi|} + g\,|D_AA|\,\Delta x\right)^2, \end{aligned} \tag{2.11}$$

where $\Delta\pi$ and Δx are the root-mean-square uncertainties in the values of π and x in the original apparatus state. Since the uncertainty principle is assumed to hold for the apparatus, we have

$$\Delta x\,\Delta\pi \gtrsim \tfrac{1}{2}\,|(x, \pi)| = \tfrac{1}{2}\,|D_x\pi|, \tag{2.12}$$

whence

$$\begin{aligned} (\Delta A)^2 \gtrsim\ & \frac{1}{2}\left(\frac{1}{2g\,\Delta x} - g\,|D_AA|\,\Delta x\right)^2 \\ & + \frac{1}{2}\left(\frac{1}{2g\,\Delta x} + g\,|D_AA|\,\Delta x\right)^2, \end{aligned} \tag{2.13}$$

[9] When π satisfies the conditions of footnote 5, Eq. (2.9) is replaced by

$$\delta^+\pi = g(A + \delta^-A)D_\pi x,$$

in which use of the reciprocity theorem (1.10) is made. This leads to the same uncertainties (2.13) and (2.14) as those which follow from the conditions adopted in the text. We note that in view of the reciprocity theorem we have

$$\delta^-A = \delta^+A,$$

and hence the specification of boundary conditions on the measurement-induced disturbance in A is really unnecessary here. Boundary conditions are important, however, when two observables are involved as in Sec. 3.

which, upon minimization with respect to the product $g\Delta x$, yields

$$\Delta A \gtrsim |D_A A|^{1/2}. \tag{2.14}$$

Bohr and Rosenfeld[3] have shown that (2.14) does not really represent the maximum accuracy attainable in the measurement of a single observable. By the introduction of a suitable "compensation mechanism" the effect of the disturbance which the measurement itself induces in A can be canceled. Mathematically, the introduction of the compensation mechanism corresponds (at least approximately) to the addition of a second-order term, $-\frac{1}{2}g^2x^2D_AA$, to the action functional, so that the total action now takes the form[10]

$$S + \mathbf{S} + gxA - \tfrac{1}{2}g^2x^2\, D_AA.$$

With this form the first-order disturbance in A is the same as before [Eq. (2.7)]. The effect of the coupling on π, however, is now given, correct to second order, by

$$\begin{aligned}\delta^-\pi &= g(A + \delta^-A)\, D_x\pi \\ &\quad - g^2x\, D_AA\, D_x\pi = gA\, D_x\pi,\end{aligned} \tag{2.15}$$

i.e., by the classical weak-coupling limit. Therefore,

$$\Delta A \gtrsim \frac{\Delta\pi}{g\,|D_x\pi|} \gtrsim \frac{1}{2g\,\Delta x}, \tag{2.16}$$

and we see that the accuracy with which A may be measured is limited only by how big the product $g\Delta x$ can be made without destroying either the observable A itself or the validity of the weak coupling approximation. The only fundamental limitation of the latter kind which is presently known to exist is that associated with the existence of a "smallest" length in the quantum theory of gravitation, a limit, however, which lies within the ordinary quantum domain by many orders of magnitude.[11]

3. MEASUREMENT OF TWO OBSERVABLES

For the measurement of two observables, A and B, it is necessary to introduce variables x_1, π_1 and x_2, π_2 from each of two independent apparatuses $\mathbf{S}_1$ and $\mathbf{S}_2$, satisfying the conditions (2.1). The systems $\mathbf{S}_1$ and $\mathbf{S}_2$ may be regarded as forming together a single apparatus

$$\mathbf{S} = \mathbf{S}_1 + \mathbf{S}_2, \tag{3.1}$$

for which the uncertainty principle is again assumed to hold. Introducing compensation mechanisms as before, we are led to consider a total action functional of the form

$$\begin{aligned}S + \mathbf{S} + g_1x_1A + g_2x_2B \\ - \tfrac{1}{2}g_1^2x_1^2\, D_AA - \tfrac{1}{2}g_2^2x_2^2\, D_BB.\end{aligned}$$

The deviations in the quantities A, B, π_1, π_2 from the values they would have had in the absence of coupling ($g_1 = g_2 = 0$) are now given, correct to their appropriate orders, by

$$\delta^-A = g_1x_1\, D_AA + g_2x_2\, D_BA, \tag{3.2}$$

$$\delta^-B = g_1x_2\, D_AB + g_2x_2\, D_BB, \tag{3.3}$$

$$\begin{aligned}\delta^-\pi_1 &= g_1(A + \delta^-A)\, D_{x_1}\pi_1 - g_1^2x_1\, D_AAD_{x_1}\pi_1 \\ &= g_1AD_{x_1}\pi_1 + g_1g_2x_2\, D_BAD_{x_1}\pi_1,\end{aligned} \tag{3.4}$$

$$\begin{aligned}\delta^-\pi_2 &= g_2(B + \delta^-B)\, D_{x_2}\pi_2 - g_2^2x_2\, D_BBD_{x_2}\pi_2 \\ &= g_2BD_{x_2}\pi_2 + g_1g_2x_1\, D_ABD_{x_2}\pi_2.\end{aligned} \tag{3.5}$$

A and B are therefore given, in terms of the experimental data, by

$$A = \frac{\delta^-\pi_1}{g_1\, D_{x_1}\pi_1} - g_2x_2\, D_BA, \tag{3.6}$$

$$B = \frac{\delta^-\pi_2}{g_2\, D_{x_2}\pi_2} - g_1x_1\, D_AB. \tag{3.7}$$

Applying the uncertainty principle to the apparatus as before, we get

$$\begin{aligned}\Delta A^2 \gtrsim \frac{1}{2}\left(\frac{1}{2g_1\,\Delta x_1} - g_2\,|D_BA|\,\Delta x_2\right)^2 \\ + \frac{1}{2}\left(\frac{1}{2g_1\,\Delta x_1} + g_2\,|D_BA|\,\Delta x_2\right)^2,\end{aligned} \tag{3.8}$$

$$\begin{aligned}\Delta B^2 \gtrsim \frac{1}{2}\left(\frac{1}{2g_2\,\Delta x_2} - g_1\,|D_AB|\,\Delta x_1\right)^2 \\ + \frac{1}{2}\left(\frac{1}{2g_2\,\Delta x_2} + g_1\,|D_AB|\,\Delta x_1\right)^2.\end{aligned} \tag{3.9}$$

It is readily verified that the product $\Delta A^2\Delta B^2$ is minimized by setting[12]

$$g_1g_2\,\Delta x_1\,\Delta x_2 = \tfrac{1}{2}\,|D_AB|^{-1/2}\,|D_BA|^{-1/2}. \tag{3.10}$$

[10] In many important cases D_AA either vanishes or is equal to a constant determined by the geometry and parameters of the measuring arrangement, and it is therefore usually not difficult to invent a conceptual device to serve as a satisfactory compensation mechanism. The quadratic dependence of the compensation term on x suggests the frequent suitability of mechanical springs. (See references 3 and 11.)

[11] B. S. DeWitt, in *Gravitation, An Introduction to Current Research*, edited by L. Witten [John Wiley & Sons, Inc., New York, (to be published)].

[12] If $D_AB = D_BA = 0$, it is obvious that A and B can be measured to a simultaneous accuracy limited only by the weak coupling conditions on g_1Ax_1 and g_2Ax_2.

Therefore

$$\begin{aligned}\Delta A\ \Delta B \gtrsim & [\tfrac{1}{4}(-\tfrac{1}{2}\,|D_A B| \\ & - \tfrac{1}{2}\,|D_B A| + |D_A B|^{1/2}\,|D_B A|^{1/2})^2 \\ & + \tfrac{1}{4}(\tfrac{1}{2}\,|D_A B| - \tfrac{1}{2}\,|D_B A|)^2 \\ & + \tfrac{1}{4}(-\tfrac{1}{2}\,|D_A B| + \tfrac{1}{2}\,|D_B A|)^2 \\ & + \tfrac{1}{4}(\tfrac{1}{2}\,|D_A B| + \tfrac{1}{2}\,|D_B A| \\ & + |D_A B|^{1/2}\,|D_B A|^{1/2})^2]^{1/2} \\ = & \tfrac{1}{2}(|D_A B| + |D_B A|). \end{aligned} \tag{3.11}$$

If either $D_A B$ or $D_B A$ vanishes (e.g., if the time intervals associated with A and B do not overlap) then this uncertainty relation reduces to (1.16). In the general overlapping case, however, (3.11) contradicts the uncertainty principle. In particular, if $A = B$ we are led to the inadmissable implication that the results of simultaneous measurements performed on the same quantity are not necessarily identical.

Bohr and Rosenfeld[3] showed that the plus sign appearing on the right-hand side of (3.11) can be changed to the minus sign of relation (1.16) by introducing, in addition to the individual compensation mechanisms already present, devices which also correlate the disturbances produced by the two measurements. Such "correlation mechanisms" are described by an additional term in the action of the form $-\frac{1}{2}g_1 g_2 x_1 x_2 (D_A B + D_B A)$. The resulting total action,

$$\begin{aligned} S + \mathbf{S} + g_1 x_1 A + g_2 x_2 B - \tfrac{1}{2} g_1^2 x_1^2\, D_A A \\ - \tfrac{1}{2} g_2^2 x_2^2\, D_B B - \tfrac{1}{2} g_1 g_2 x_1 x_2 (D_A B + D_B A), \end{aligned}$$

yields the maximum simultaneous accuracy for the two measurements. With this action functional we have

$$\begin{aligned}\delta^- \pi_1 &= g_1[A + \delta^- A - g_1 x_1\, D_A A \\ &\quad - \tfrac{1}{2} g_2 x_2 (D_A B + D_B A)]\, D_{x_1} \pi_1 \\ &= g_1[A - \tfrac{1}{2} g_2 x_2 (D_A B - D_B A)]\, D_{x_1} \pi_1. \end{aligned} \tag{3.12}$$

$$\begin{aligned}\delta^- \pi_2 &= g_2[B + \delta^- B - g_2 x_2\, D_B B \\ &\quad - \tfrac{1}{2} g_1 x_1 (D_A B + D_B A)]\, D_{x_2} \pi_2 \\ &= g_2[B + \tfrac{1}{2} g_1 x_1 (D_A B - D_B A)]\, D_{x_1} \pi_2, \end{aligned} \tag{3.13}$$

Eqs. (3.2) and (3.3) remaining unchanged. Therefore

$$A = \frac{\delta^- \pi_1}{g_1\, D_{x_1} \pi_1} + \tfrac{1}{2} g_2 x_2 (A, B), \tag{3.14}$$

$$B = \frac{\delta^- \pi_2}{g_2\, D_{x_2} \pi_2} - \tfrac{1}{2} g_1 x_1 (A, B), \tag{3.15}$$

$$\Delta A^2 \gtrsim \frac{1}{2}\left(\frac{1}{2 g_1\, \Delta x_1} + \tfrac{1}{2} g_2\, |(A, B)|\, \Delta x_2\right)^2 + \frac{1}{2}\left(\frac{1}{2 g_1\, \Delta x_1} - \tfrac{1}{2} g_2\, |(A, B)|\, \Delta x_2\right)^2, \tag{3.16}$$

$$\Delta B^2 \gtrsim \frac{1}{2}\left(\frac{1}{2 g_2\, \Delta x_2} - \tfrac{1}{2} g_1\, |(A, B)|\, \Delta x_1\right)^2 + \frac{1}{2}\left(\frac{1}{2 g_2\, \Delta x_2} + \tfrac{1}{2} g_1\, |(A, B)|\, \Delta x_1\right)^2, \tag{3.17}$$

$$\Delta A\ \Delta B \gtrsim \tfrac{1}{2}\,|(A, B)|, \tag{3.18}$$

the minimum value of the product $\Delta A \Delta B$ being attained when[13]

$$g_1 g_2\ \Delta x_1\ \Delta x_2 = |(A, B)|^{-1}. \tag{3.19}$$

The assertion made in the Introduction, that the uncertainty principle effectively determines the commutator (or Poisson bracket) of two observables, is thus confirmed.

It remains necessary only to call attention to the fact that the use of "classical" or "average" values of observables such as A and B in the above description has important limitations. Some of these observables may occur in products or may themselves be expressible as products of other variables. Now, the average value of a product may be equated to the product of the average values only in the limit of high quantum numbers, and then only in the case of systems possessing a finite number of degrees of freedom. A direct classical description of the quantities in question will therefore not be strictly valid, particularly in the case of quantized fields. Such a description neglects a number of important purely quantum effects, namely, those which give rise to the phenomena of vacuum polarization and level shifts as well as to mathematical infinities in the formalism. However, the technical procedure of "renormalization" can be expected to reinstate the approximate validity (i.e., to lowest order) of the classical description *provided* the coupling of the field to its sources is sufficiently weak and/or there exists a fundamental invariance group which sufficiently dominates the formalism (e.g., electrodynamics,[14] gravitation).

[13] If $(A, B) = 0$ only the weak coupling condition limits the simultaneous accuracy of the two measurements.

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

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Quantization of Fields with Infinite-Dimensional Invariance Groups. II. Anticommuting Fields*

BRYCE S. DEWITT
Institute of Field Physics,
Department of Physics,
University of North Carolina, Chapel Hill, North Carolina
(Received December 9, 1961)

The Green's function approach to the definition of commutators for fields possessing infinite-dimensional invariance groups is extended to the case of anticommuting fields. The discussion is restricted to fields which provide linear homogeneous or inhomogeneous representations of the group, a restriction which excludes no case of practical interest and facilitates setting up the formalism in a manifestly covariant way. Self-consistency of supplementary conditions, Huygens' principle, and reciprocity relations are established just as for commuting fields. Careful attention must be paid to the ordering of anticommuting factors, particularly in the demonstration of the Poisson-Jacobi identity. The invariance properties of the Poisson bracket are investigated in detail and the notion of *conditional invariant* is introduced. A special class of conditional invariants called *asymptotic invariants*, which give a complete physical characterization of initial and final states of the dynamical system, is studied in the final section.

INTRODUCTION

In a previous paper[1] [to be referred to as (I)] a general theory of the Poisson bracket (commutator) was developed which dispenses entirely with Hamiltonian ideas and which treats space-time in a completely homogeneous fashion. The formalism is especially suited to the study of fields possessing infinite-dimensional invariance groups, for which the identification of canonical variables is an inconvenience. By introducing the Green's functions which describe the propagation of small disturbances about a given solution of the dynamical equations, and by restricting the definition of the commutator to group invariants (i.e., actual physical observables), the quantization of such fields is accomplished without the use of subsidiary conditions on the physical state vectors. Although the connection between the adopted definition of the Poisson bracket and that of the conventional canonical theory was indicated in (I), it was pointed out that the Green's-function definition is capable of standing independently on its own as a corollary of the uncertainty principle and the theory of measurement. This assertion has been given detailed justification in another paper[2] in which it is shown explicitly how the uncertainty principle serves uniquely to define the commutator via analysis of coupling with a measuring apparatus.

In the present paper we wish to extend the previous work in a different direction. The discussion of (I) was limited to the case of commuting fields, for which direct classical forms exist. It was remarked, however, that the methods introduced were capable of straightforward generalization to the case of anticommuting fields. We shall here carry out this generalization. At the same time we shall revise and specialize some of our earlier procedures with an eye on future applications.

The notation of (I) will be somewhat altered and condensed. The use of primes on indices to distinguish space-time points will be reserved for special instances in which the display of details is desirable (Sec. 3). More commonly we shall simply omit both primes and integration signs, allowing the indices to do double duty as (1) discrete labels for field components and, (2) continuous labels over space-time points, and extending the summation convention for repeated indices so as to imply integration as well. Thus, for example, Eq. (I, 3.5) will suffer the condensation

$$\int A_{,i} R^i_A \, d^4x = 0 \rightarrow A_{,i} R^i_\alpha = 0.$$

Functional differentiation will be denoted by the comma as before, but Greek letters from the first part of the alphabet will be used in place of capital Latin letters to denote indices associated with the invariance group. For the achievement of maximum generality both commuting and anticommuting fields will be considered at once, and the symbol ψ will now be reserved for the latter. Commuting

* This research was supported in part by the Department of the Navy, Office of Naval Research, under contract Nonr-855(07) and in part by the Air Force Office of Scientific Research under contract AFOSR 61-72.
[1] B. S. DeWitt, J. Math. Phys. **2**, 151 (1961).
[2] B. S. DeWitt, J. Math. Phys. **3**, 619 (1962).

fields will be identified by the symbol ϕ. Furthermore, indices associated with the two types of fields will be distinguished by writing those associated with the anticommuting fields in boldface and leaving the others in lightface type. Thus, for example,

$$\delta A/\delta\phi^i \equiv A_{,i}, \qquad \delta A/\delta\psi^i \equiv A_{,i}.$$

In performing functional differentiations with respect to anticommuting field components a distinction must be made between *right* and *left* derivatives. The comma will here always be understood as signifying the *right* derivative, and indices following it will be written in the order in which the differentiations are performed:

$$A_{,ij\ldots} \equiv \cdots \frac{\delta}{\delta\psi^j}\frac{\delta}{\delta\psi^i} A.$$

The variation in A due to anticommuting variations $\delta\psi^i$ in the ψ^i is then given by

$$\delta A = A_{,i}\,\delta\psi^i,$$

with the factors taken in the order indicated.

The measurement theoretical analysis of the uncertainty principle will again be taken as the basis for the definition of commutators of observables. It is to be noted that it is the commutator rather than the anticommutator which is involved in this analysis, for only commutators map into the Poisson brackets of the uncertainty relation. This implies that an observable (and also the action) must not only be a group invariant, but must involve combinations of the ψ^i of even degree only. The "quasi-classical approximation" of (I), in which all quantities (except those appearing in the primary commutator under discussion) are treated as rigorously commuting, will be extended by treating the ψ^i as rigorously anticommuting among themselves while commuting with the ϕ^i. The ψ^i must therefore be contained either linearly, not at all, or in completely antisymmetric combinations in all dynamical quantities. From this it follows that boldface indices induced by repeated functional differentiation with respect to the ψ^i will anticommute among themselves while commuting with any lightface indices induced by functional differentiation with respect to the ϕ^i. The ψ^i, like the ϕ^i, will be assumed to be real (Hermitian). The functional derivatives of order 1, 2, 5, 6, 9, 10, etc. with respect to the ψ^i of any real (Hermitian) observable will, therefore, be imaginary (anti-Hermitian) while the functional derivatives of order 3, 4, 7, 8, 11, 12, etc., will be real (Hermitian). Furthermore the functional derivatives of even order with respect to the ψ^i will, in the quasi-classical approximation, commute with everything while those of odd order will anticommute among themselves. Reflecting this rule, the formalism of the following sections will be set up in such a way that quantities bearing an even number of boldface indices will be of the commuting type while those bearing an odd number will be of the anticommuting type. The symmetry and reality properties of the two kinds of indices will *not*, however, generally follow the pattern which holds for functional differentiation.

The representations of the invariance group provided by the ϕ^i and ψ^i will be restricted, in the present paper, to be *linear* (homogeneous or inhomogeneous). Reasons for this restriction, which excludes no case of practical interest, are given in Sec. 1, and a number of its simplifying consequences will be noted. In Sec. 2 supplementary conditions are imposed on the variables representing small disturbances in the dynamical system, so that Green's functions describing the propagation of these disturbances may be introduced. The self-consistency of these conditions is verified in detail. Huygens' principle and the reciprocity relations for the nonsymmetric wave operators which appear when anticommuting variables are present are established in Sec. 3. Section 4 is devoted to a study of the invariance properties of the Poisson bracket and the conditions under which it is unique. The notion of *conditional invariant* is introduced in this connection. The steps in the derivation of the Poisson-Jacobi identity are outlined in Sec. 5. Finally in Sec. 6 an important class of conditional invariants is constructed, namely, the *asymptotic invariants* which serve to characterize initial and final states of the dynamical system.

1. RESTRICTION TO LINEAR REPRESENTATIONS OF THE INVARIANCE GROUP

In the classical domain the theory of dynamical systems possessing infinite-dimensional invariance groups is not strongly affected by the nature of the representation of the invariance group which the dynamical variables provide. The theory of the Poisson bracket can be developed for any type of representation, and with any choice of variables (provided they are local). Even *transitive* representations are permitted, although in this case the theory becomes completely trivial, since the only invariants are then numerical constants and all Poisson

brackets vanish.[3] In the quantum theory, however, *linear* representations appear to be mandatory. It is known from the work of Pais and Uhlenbeck[4] on field theories with nonlocalized action that if the vacuum is to exist as a stable state and if negative probabilities are to be avoided, the covariant Green's functions of the theory must generally satisfy differential equations of degree no higher than the second. This means that the Lagrangian must be similarly limited in the degree of its space-time derivatives. Only for linear representations can Lagrangians of this type, which yield invariant action functionals, be easily found.[5] Such action functionals usually take the form of quadratic or bilinear invariants, the only familiar exception being that of the gravitational field.

Of course, a given representation may be rendered nonlinear by transforming to a different set of variables, and since the dynamical equations themselves are always nonlinear when the invariance group is non-Abelian it would seem to matter little whether such a transformation is performed or not, at least in the non-Abelian case. However, when a linear form for the representation exists, it is a convenience to use it—quite apart, it may be added, from any aid it may be in the resolution of the factor ordering ambiguity of the rigorous quantum theory, a problem which remains presently rather obscure. Furthermore, it may be noted that from a purely group-theoretical viewpoint the restriction to linear representations is relatively minor, since any representation may be rendered linear by adding a sufficient number of extra variables and performing a suitable transformation. It must be admitted, of course, that the extra variables, being invariants, would be in principle observable. However, they can always be rendered dynamically innocuous by suitable choice of the action.

The infinitesimal group transformation law for the ϕ^i, ψ^{i} will be written in the form

$$\delta\phi^i = R^i_\alpha\,\delta\xi^\alpha \qquad \delta\psi^{\mathrm{i}} = R^{\mathrm{i}}_\alpha\,\delta\xi^\alpha \tag{1.1}$$

where the $\delta\xi^\alpha$ are group parameters which are here assumed to be real numbers which commute with everything.[6] We shall assume that the functions R^i_α, R^{i}_α provide a faithful representation of the maximal infinite-dimensional invariance group of the system, so that

$$\begin{bmatrix} R^i_\alpha \\ R^{\mathrm{i}}_\alpha \end{bmatrix} \delta\xi^\alpha = 0 \quad \text{if and only if} \quad \delta\xi^\alpha = 0. \tag{1.2}$$

The linearity requirement takes the form

$$\left.\begin{aligned} R^i_{\alpha,\mathrm{j}} &= 0, \quad R^{\mathrm{i}}_{\alpha,j} = 0, \\ R^i_{\alpha,jk} &= 0, \quad R^{\mathrm{i}}_{\alpha,\mathrm{jk}} = 0. \end{aligned}\right\} \tag{1.3}$$

The first pair of equations arises from the fact that the representation cannot mix variables of different type. That is, the variations $\delta\phi^i$, $\delta\psi^{\mathrm{i}}$ must have the same (anti-)commutation properties as the ϕ^i, ψ^{i}. This requirement implies also that although the representation provided by the ϕ^i may be linear inhomogeneous, that provided by the ψ^{i} must be strictly homogeneous. That is,

$$R^{\mathrm{i}}_\alpha = R^{\mathrm{i}}_{\alpha,\mathrm{j}}\,\psi^j. \tag{1.4}$$

The integrability conditions for the representation take the form

$$R^i_{\alpha,i}R^i_\beta - R^i_{\beta,i}R^i_\alpha = R^i_\gamma c^\gamma_{\alpha\beta}, \tag{1.5}$$

$$R^{\mathrm{i}}_{\alpha,\mathrm{j}}R^{\mathrm{j}}_\beta - R^{\mathrm{i}}_{\beta,\mathrm{j}}R^{\mathrm{j}}_\alpha = R^{\mathrm{i}}_\gamma c^\gamma_{\alpha\beta}, \tag{1.6}$$

where the $c^\gamma_{\alpha\beta}$ are the structure constants of the group. These conditions may also be regarded as expressing linear homogeneous transformation laws for the R^i_α, R^{i}_α themselves:

$$\delta R^i_\alpha \equiv R^i_{\alpha,i}R^i_\beta\,\delta\xi^\beta = (-R^i_\gamma c^\gamma_{\beta\alpha} + R^i_{\beta,i}R^i_\alpha)\,\delta\xi^\beta, \tag{1.7}$$

$$\delta R^{\mathrm{i}}_\alpha \equiv R^{\mathrm{i}}_{\alpha,\mathrm{j}}R^{\mathrm{j}}_\beta\,\delta\xi^\beta = (-R^{\mathrm{i}}_\gamma c^\gamma_{\beta\alpha} + R^{\mathrm{i}}_{\beta,\mathrm{j}}R^{\mathrm{j}}_\alpha)\,\delta\xi^\beta. \tag{1.8}$$

The representations which R^i_α and R^{i}_α provide are evidently direct products of (1) the representation contragredient to the adjoint representation, and (2) the homogeneous parts of the representations provided by ϕ^i and ψ^{i}, respectively.

A group invariant A is characterized by the condition

$$A_{,i}R^i_\alpha + A_{,\mathrm{i}}R^{\mathrm{i}}_\alpha \equiv 0. \tag{1.9}$$

[3] For Lie groups [see L. Pontrjagin, *Topological Groups* (Princeton University Press, Princeton, New Jersey, 1946)] each transitive representation can be represented as a transformation group over a manifold of "points" identifiable with the left cosets of a certain subgroup containing no normal subgroups of the original Lie group other than the identity. The classification of transitive representations reduces to the determination of subgroups possessing this property. The same is presumably true in the case of infinite dimensional continuous groups. The only group possessing representations which are simultaneously transitive and linear is the full linear group.

[4] A. Pais and G. E. Uhlenbeck, Phys. Rev. 79, 145 (1950).

[5] The limitation to linear representations is further reinforced by the requirement that the field variables shall asymptotically describe free relativistic particles having definite mass and spin. For a discussion of the transformation properties of the asymptotic fields see Sec. 6.

[6] The possibility of anticommuting group parameters also exists, but since it has no apparent physical interest we do not consider it. It is encountered, for example, in the gauge groups of massless fields having spins 3/2, 5/2, etc. Since these groups are Abelian, the structure constants vanish. The general non-Abelian case would involve anticommuting structure constants.

Differentiating this identity and making use of the relation

$$(XY)_{,\mathrm{i}} = -X_{,\mathrm{i}}Y + XY_{,\mathrm{i}}, \tag{1.10}$$

which holds for functionals Y of odd degree in the ψ^{i}, we get

$$\begin{aligned} 0 &\equiv A_{,ji}R^i_\alpha + A_{,j}R^i_{\alpha,i} + A_{,\mathrm{j}i}R^{\mathrm{j}}_\alpha \\ &= A_{,ij}R^j_\alpha + A_{,j}R^j_{\alpha,i} + A_{,i\mathrm{j}}R^{\mathrm{j}}_\alpha, \end{aligned} \tag{1.11}$$

$$\begin{aligned} 0 &\equiv A_{,j\mathrm{i}}R^j_\alpha - A_{,\mathrm{j}i}R^{\mathrm{j}}_\alpha + A_{,\mathrm{j}}R^{\mathrm{j}}_{\alpha,\mathrm{i}} \\ &= A_{,\mathrm{i}j}R^j_\alpha + A_{,\mathrm{i}\mathrm{j}}R^{\mathrm{j}}_\alpha + A_{,\mathrm{j}}R^{\mathrm{j}}_{\alpha,\mathrm{i}}, \end{aligned} \tag{1.12}$$

which, when applied to the action functional, yield

$$\begin{pmatrix} S_{,ij} & S_{,i\mathrm{j}} \\ S_{,\mathrm{i}j} & S_{,\mathrm{i}\mathrm{j}} \end{pmatrix} \begin{pmatrix} R^j_\alpha \\ R^{\mathrm{j}}_\alpha \end{pmatrix} = 0 \tag{1.13}$$

when the dynamical equations

$$S_{,i} = 0 \qquad S_{,\mathrm{i}} = 0 \tag{1.14}$$

are satisfied. For arbitrary invariants Eqs. (1.11) and (1.12) also yield the transformation laws

$$\begin{aligned} \delta A_{,i} &\equiv (A_{,ij}R^j_\alpha + A_{,i\mathrm{j}}R^{\mathrm{j}}_\alpha)\ \delta\xi^\alpha \\ &= -A_{,j}R^j_{\alpha,i}\ \delta\xi^\alpha, \end{aligned} \tag{1.15}$$

$$\begin{aligned} \delta A_{,\mathrm{i}} &\equiv (A_{,\mathrm{i}j}R^j_\alpha + A_{,\mathrm{i}\mathrm{j}}R^{\mathrm{j}}_\alpha)\ \delta\xi^\alpha \\ &= -A_{,\mathrm{j}}R^{\mathrm{j}}_{\alpha,\mathrm{i}}\ \delta\xi^\alpha. \end{aligned} \tag{1.16}$$

By repeatedly differentiating Eqs. (1.11) and (1.12) it is easy to show that the functional derivatives, to all orders, of any invariant satisfy homogeneous linear transformation laws identifiable as direct products of representations contragredient to the homogeneous parts of the representations provided by ϕ^i, ψ^{i},

Denoting by $\delta^*_A\phi^i$, $\delta^*_A\psi^{\mathrm{i}}$ the advanced and retarded changes produced in a given solution of the dynamical equations by the infinitesimal change

$$S \to S + \epsilon A \tag{1.17}$$

in the action functional, we have [cf. (I,3.9)]

$$\begin{pmatrix} S_{,ij} & S_{,i\mathrm{j}} \\ S_{,\mathrm{i}j} & S_{,\mathrm{i}\mathrm{j}} \end{pmatrix} \begin{pmatrix} \delta^*_A\phi^j \\ \delta^*_A\psi^{\mathrm{j}} \end{pmatrix} = -\epsilon \begin{pmatrix} A_{,i} \\ A_{,\mathrm{i}} \end{pmatrix}. \tag{1.18}$$

Because of the singularity of the matrix on the left, as expressed by Eq. (1.13), the solutions of this equation are determined only *modulo* an infinitesimal group transformation (1.1). Particular solutions may be obtained by imposing supplementary conditions. For this purpose we introduce a continuous *local*[7] matrix

$$\begin{pmatrix} g_{ij} & g_{i\mathrm{j}} \\ g_{\mathrm{i}j} & g_{\mathrm{i}\mathrm{j}} \end{pmatrix} \tag{1.19}$$

which has the same symmetry and reality properties as the matrix of Eqs. (1.13) and (1.18) and which is such that the continuous matrix

$$\begin{aligned} F_{\alpha\beta} &\equiv (R^i_\alpha\, R^{\mathrm{i}}_\alpha) \begin{pmatrix} g_{ij} & g_{i\mathrm{j}} \\ -g_{\mathrm{i}j} & -g_{\mathrm{i}\mathrm{j}} \end{pmatrix} \begin{pmatrix} R^j_\beta \\ R^{\mathrm{j}}_\beta \end{pmatrix} \\ &= (R_{i\alpha}\, R_{\mathrm{i}\alpha}) \begin{pmatrix} R^i_\beta \\ R^{\mathrm{i}}_\beta \end{pmatrix}, \end{aligned} \tag{1.20}$$

with

$$\begin{pmatrix} R_{i\alpha} \\ R_{\mathrm{i}\alpha} \end{pmatrix} \equiv \begin{pmatrix} g_{ij} & g_{i\mathrm{j}} \\ g_{\mathrm{i}j} & g_{\mathrm{i}\mathrm{j}} \end{pmatrix} \begin{pmatrix} R^j_\alpha \\ R^{\mathrm{j}}_\alpha \end{pmatrix}, \tag{1.21}$$

is a proper wave operator possessing unique Green's functions $G^{*\alpha\beta}$. Here again, as in (I), such matrices are easily found in actual cases, their existence depending, just as in the theory of finite matrices, on the linear independence of the R^i_α, R^{i}_α [faithful representation; cf. Eq. (1.2)]. It is not actually necessary to assume, as was done in (I), that the matrix (1.19) is nonsingular. We shall have no need of its inverse, and only use it to lower the indices i, i on R^i_α, R^{i}_α. Its choice is to a considerable extent arbitrary, but we shall show later (Sec. 4) that this arbitrariness leaves the Poisson bracket unaffected. We note that although R^i_α, R^{i}_α, and $R_{i\alpha}$ are real (Hermitian), the anticommuting quantity $R_{\mathrm{i}\alpha}$ is imaginary (anti-Hermitian).

It is a great practical convenience, although not strictly necessary, to select the matrix (1.19) in such a way that it has the same linear group transformation properties as the matrix of Eqs. (1.13) and (1.18). This makes it possible to keep the manifest covariance of the formalism always clearly in the foreground, for the transformation character of all expressions appearing in the formalism is then rendered evident at a glance. This is one of the major virtues of the restriction to linear representations. Under these conditions the Green's functions $G^{*\alpha\beta}$ transform according to the adjoint representation (taken twice) and the wave operator $F_{\alpha\beta}$ transforms contragrediently.

2. THE SUPPLEMENTARY CONDITIONS

As supplementary conditions on the $\delta^*_A\phi^i$, $\delta^*_A\psi^{\mathrm{i}}$ we choose [cf. (I, 3.10)]

$$R_{i\alpha}\ \delta^*_A\phi^i + R_{\mathrm{i}\alpha}\ \delta^*_A\psi^{\mathrm{i}} = 0. \tag{2.1}$$

[7] By "local" we mean having delta-function type behavior. The need for this condition, which insures the maintenance of the sharp distinction between advanced and retarded Green's functions, was not sufficiently emphasized in (I).

If these conditions are not already satisfied it is easily seen that they may nevertheless be imposed by carrying out an infinitesimal group transformation having the parameters

$$\delta^* \xi^\alpha = G^{*\alpha\beta}(R_{i\beta}\,\delta_A^* \phi^i + R_{\mathrm{i}\beta}\,\delta_A^* \psi^{\mathrm{i}}). \tag{2.2}$$

Under these conditions Eq. (1.18) may be replaced by

$$\begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} \begin{pmatrix} \delta_A^* \phi^j \\ \delta_A^* \psi^{\mathrm{j}} \end{pmatrix} = -\epsilon \begin{pmatrix} A_{,i} \\ A_{,\mathrm{i}} \end{pmatrix}, \tag{2.3}$$

where

$$\begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} \equiv \begin{pmatrix} S_{,ij} & S_{,i\mathrm{j}} \\ S_{,\mathrm{i}j} & S_{,\mathrm{ij}} \end{pmatrix} + g^{\alpha\beta} \begin{pmatrix} R_{i\alpha}R_{j\beta} & R_{i\alpha}R_{\mathrm{j}\beta} \\ R_{\mathrm{i}\alpha}R_{j\beta} & R_{\mathrm{i}\alpha}R_{\mathrm{j}\beta} \end{pmatrix}, \tag{2.4}$$

$g^{\alpha\beta}$ being an arbitrary local symmetric real nonsingular matrix having elements of the commuting type and transforming like the Green's functions $G^{*\alpha\beta}$. The matrix $g^{\alpha\beta}$ will be used together with its inverse $g_{\alpha\beta}$ to raise and lower group indices.

The matrix (2.4) has the same symmetry, reality and transformation properties as the matrix formed from $S_{,ij}$, $S_{,i\mathrm{j}}$, etc. Unlike the latter, however, it is nonsingular. The proof of this statement is based on the assumption that the $(R_\alpha^i\ R_\alpha^{\mathrm{i}})$ constitute a complete linearly independent set of zero-eigenvalue eigenvectors of the matrix of Eq. (1.13). We may therefore adjoin to them another linearly independent set of vectors $(R_A^i\ R_A^{\mathrm{i}})$, no linear combination of which are zero-eigenvalue eigenvectors, such that the matrix

$$\begin{pmatrix} R_\alpha^i & R_A^i \\ R_\alpha^{\mathrm{i}} & R_A^{\mathrm{i}} \end{pmatrix} \tag{2.5}$$

is nonsingular. Moreover, under the transformation

$$\begin{pmatrix} R'^i_A \\ R'^{\mathrm{i}}_A \end{pmatrix} = \begin{pmatrix} R_A^i \\ R_A^{\mathrm{i}} \end{pmatrix} - G^{*\alpha\beta} \begin{pmatrix} R_\alpha^i R_{j\beta} & R_\alpha^i R_{\mathrm{j}\beta} \\ R_\alpha^{\mathrm{i}} R_{j\beta} & R_\alpha^{\mathrm{i}} R_{\mathrm{j}\beta} \end{pmatrix} \begin{pmatrix} R_A^j \\ R_A^{\mathrm{j}} \end{pmatrix}, \tag{2.6}$$

which does not destroy their linear independence, these vectors can be made orthogonal to the vectors $(R_{i\alpha}\ \ R_{\mathrm{i}\alpha})$.[8] We assume such a transformation already to have been carried out, so that

$$R_{i\alpha}R_A^i + R_{\mathrm{i}\alpha}R_A^{\mathrm{i}} = 0. \tag{2.7}$$

We then have

$$\begin{pmatrix} R_\alpha^i & -R_\alpha^{\mathrm{i}} \\ R_A^i & -R_A^{\mathrm{i}} \end{pmatrix} \begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} \begin{pmatrix} R_\beta^j & R_B^j \\ R_\beta^{\mathrm{j}} & R_B^{\mathrm{j}} \end{pmatrix} = \mathrm{diag}\left[F_{\alpha\gamma} g^{\gamma\delta} F_{\delta\beta},\ (R_A^i\ -R_A^{\mathrm{i}}) \begin{pmatrix} S_{,ij} & S_{,i\mathrm{j}} \\ S_{,\mathrm{i}j} & S_{,\mathrm{ij}} \end{pmatrix} \begin{pmatrix} R_B^j \\ R_B^{\mathrm{j}} \end{pmatrix} \right], \tag{2.8}$$

which is nonsingular by construction. Since the matrix (2.5) is nonsingular, therefore, so is (2.4). It will be observed that the conventional notions of linear independence and nonsingularity have here been extended to anticommuting as well as commuting quantities. This extension is perfectly justified if proper attention is paid, as we have done here, to the ordering of factors, which accounts for the occurrence of a number of minus signs.

The matrix (2.4) is not only nonsingular but is a wave operator possessing unique Green's functions satisfying

$$\begin{pmatrix} F_{ik} & F_{i\mathrm{k}} \\ F_{\mathrm{i}k} & F_{\mathrm{ik}} \end{pmatrix} \begin{pmatrix} G^{*kj} & G^{*k\mathrm{j}} \\ G^{*\mathrm{k}j} & G^{*\mathrm{kj}} \end{pmatrix} = \begin{pmatrix} -\delta_i^j & 0 \\ 0 & -\delta_{\mathrm{i}}^{\mathrm{j}} \end{pmatrix}, \tag{2.9}$$

$$\begin{pmatrix} G^{*ik} & G^{*i\mathrm{k}} \\ G^{*\mathrm{i}k} & G^{*\mathrm{ik}} \end{pmatrix} \begin{pmatrix} F_{kj} & F_{k\mathrm{j}} \\ F_{\mathrm{k}j} & F_{\mathrm{kj}} \end{pmatrix} = \begin{pmatrix} -\delta_j^i & 0 \\ 0 & -\delta_{\mathrm{j}}^{\mathrm{i}} \end{pmatrix}. \tag{2.10}$$

That it should possess well-defined left inverses and right inverses of the advanced and retarded variety follows from its nonsingularity together with its locality. That the left and right inverses of a given variety should be identical, with the matrix and operator ordering taken as in Eqs. (2.9) and (2.10), cannot, however, be inferred from the theory of (I) which applies only to symmetric (self-adjoint) wave operators having elements of the commuting type. The generalization of the discussion of (I) which is needed in order to cover the case of nonsymmetric wave operators having elements of the anticommuting as well as commuting type is given in the next section. The reality and symmetry properties (reciprocity relations) of the Green's functions will also be noted there. We merely observe here that the G^{*ij}, $G^{*i\mathrm{j}}$, etc., must transform contragrediently to the F_{ij}, $F_{i\mathrm{j}}$, etc.

The theory of (I) *does* apply to the wave operator $F_{\alpha\beta}$, which is symmetric by construction and possesses elements of the commuting type only. Its Green's functions therefore satisfy the reciprocity relations

$$G^{*\alpha\beta} = G^{*\beta\alpha}. \tag{2.11}$$

Using the Green's functions G^{*ij}, $G^{*i\mathrm{j}}$, etc., we may now express the variations $\delta_A^* \phi^i$, $\delta_A^* \psi^{\mathrm{i}}$ in the

[8] The locality of the $R_\alpha{}^i$, $R_\beta{}^{\mathrm{i}}$, and hence of $F_{\alpha\beta}$, ensures that no problem arises concerning integrations by parts for either choice of the Green's functions $G^{\pm\alpha\beta}$ in (2.6).

form

$$\begin{pmatrix} \delta_A^{\pm}\phi^i \\ \delta_A^{\pm}\psi^{\mathrm{i}} \end{pmatrix} = \epsilon \begin{pmatrix} G^{\pm ij} & G^{\pm i\mathrm{j}} \\ G^{\pm \mathrm{i}j} & G^{\pm \mathrm{ij}} \end{pmatrix} \begin{pmatrix} A_{,i} \\ A_{,\mathrm{i}} \end{pmatrix} \tag{2.12}$$

whenever they satisfy the supplementary conditions (2.1). As in (I) we must, for the sake of consistency, check that these solutions of Eq. (2.3) in fact satisfy the supplementary conditions. Again we derive an auxiliary lemma. Writing

$$\begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} \begin{pmatrix} R_\beta^j \\ R_\beta^{\mathrm{j}} \end{pmatrix} = \begin{pmatrix} R_i^\alpha R_{j\alpha} & R_i^\alpha R_{\mathrm{j}\alpha} \\ R_{\mathrm{i}}^\alpha R_{j\alpha} & R_{\mathrm{i}}^\alpha R_{\mathrm{j}\alpha} \end{pmatrix} \begin{pmatrix} R_\beta^i \\ R_\beta^{\mathrm{j}} \end{pmatrix} = \begin{pmatrix} R_i^\alpha \\ R_{\mathrm{i}}^\alpha \end{pmatrix} F_{\alpha\beta}, \tag{2.13}$$

and multiplying this equation on the left by the matrix formed from the $G^{\pm ij}$, $G^{\pm i\mathrm{j}}$, etc., and on the right by the Green's functions $G^{\pm\beta\alpha}$, we get

$$\begin{pmatrix} R_\beta^i \\ R_\beta^{\mathrm{i}} \end{pmatrix} G^{\pm\beta\alpha} = \begin{pmatrix} G^{\pm ij} & G^{\pm i\mathrm{j}} \\ G^{\pm \mathrm{i}j} & G^{\pm \mathrm{ij}} \end{pmatrix} \begin{pmatrix} R_i^\alpha \\ R_{\mathrm{j}}^\alpha \end{pmatrix}. \tag{2.14}$$

The $\pm$ signs must go together in this lemma for obvious kinematic reasons.[9] The lemma may also be written in a transposed form which, when account is taken of the reciprocity relations (2.11) and (3.12) and the (anti-)commutativity of the various factors, is given by

$$G^{\pm\alpha\beta}(R_\beta^i \;\; -R_\beta^{\mathrm{j}}) = (R_i^\alpha \;\; R_{\mathrm{i}}^\alpha) \begin{pmatrix} G^{\pm ij} & G^{\pm i\mathrm{j}} \\ G^{\pm \mathrm{i}j} & G^{\pm \mathrm{ij}} \end{pmatrix}. \tag{2.15}$$

Lowering the index α and applying the result to Eq. (2.12), we get

$$\begin{aligned} (R_{i\alpha} \;\; R_{\mathrm{i}\alpha}) \begin{pmatrix} \delta_A^{\pm}\phi^i \\ \delta_A^{\pm}\psi^{\mathrm{i}} \end{pmatrix} &= \epsilon G_\alpha^{\pm\beta}(R_\beta^i \;\; -R_\beta^{\mathrm{j}}) \begin{pmatrix} A_{,i} \\ A_{,\mathrm{j}} \end{pmatrix} \\ &= \epsilon G_\alpha^{\pm\beta}(A_{,i}R_\beta^i + A_{,\mathrm{i}}R_\beta^{\mathrm{i}}), \end{aligned} \tag{2.16}$$

which vanishes in virtue of the group invariance of A [Eq. (1.9)], thus establishing the self-consistency of the supplementary conditions.

3. HUYGENS' PRINCIPLE AND RECIPROCITY RELATIONS

Consider two arbitrary vectors $(\Phi_1^i \;\; \Phi_1^{\mathrm{i}})$, $(\Phi_2^i \;\; \Phi_2^{\mathrm{i}})$ which appear together with the wave operator (2.4) in the following combination:

$$\int \left[(\Phi_1^i \;\; \Phi_1^{\mathrm{i}}) \begin{pmatrix} F_{ij'} & F_{i\mathrm{j}'} \\ F_{\mathrm{i}j'} & F_{\mathrm{ij}'} \end{pmatrix} \begin{pmatrix} \Phi_2^{j'} \\ \Phi_2^{\mathrm{j}'} \end{pmatrix} - (\Phi_1^{j'} \;\; \Phi_1^{\mathrm{j}'}) \begin{pmatrix} F_{j'i} & F_{j'\mathrm{i}} \\ F_{\mathrm{j}'i} & F_{\mathrm{j}'\mathrm{i}} \end{pmatrix} \begin{pmatrix} \Phi_2^i \\ \Phi_2^{\mathrm{i}} \end{pmatrix} \right] d^4x'. \tag{3.1}$$

[9] The integrations by parts needed in the derivation of (2.14) cannot be performed if the signs do not match.

If the functions Φ_1^i, Φ_1^{i}, Φ_2^i, Φ_2^{i} vanish sufficiently rapidly in remote regions of space-time, the integral of this expression over all x will vanish by symmetry. Since the wave operator (2.4) is a local matrix (i.e., differential operator), this implies that (3.1) must be re-expressible in the form

$$\int d^4x' \int d^4x'' \frac{\partial}{\partial x^\mu} \left[(\Phi_1^{i'} \;\; \Phi_1^{\mathrm{i}'}) \begin{pmatrix} f^\mu_{i'i''} & f^\mu_{i'\mathrm{j}''} \\ f^\mu_{\mathrm{i}'i''} & f^\mu_{\mathrm{i}'\mathrm{j}''} \end{pmatrix} \begin{pmatrix} \Phi_2^{i''} \\ \Phi_2^{\mathrm{j}''} \end{pmatrix} \right], \tag{3.2}$$

where the $f^\mu_{i'i''}$, $f^\mu_{i'\mathrm{j}''}$, etc. are certain homogeneous *quadratic* combinations of the delta function and its derivatives, with coefficients involving the ϕ^i, ψ^{i} and their derivatives, having the same symmetry, reality, and (anti-)commutation properties as the $F_{i'i''}$, $F_{i'\mathrm{j}''}$, etc. Since equality of (3.1) and (3.2) involves the properties of the functions Φ_1^i, Φ_1^{i}, etc., only locally, it must evidently hold for arbitrary functions. Moreover, as long as the ordering of factors is left as indicated, the functions Φ_1^i, Φ_1^{i}, etc., may each be of either the commuting or anti-commuting type.

We now introduce the propagation functions

$$\begin{pmatrix} G^{ij} & G^{i\mathrm{j}} \\ G^{\mathrm{i}j} & G^{\mathrm{ij}} \end{pmatrix} \equiv \begin{pmatrix} G^{+ij} & G^{+i\mathrm{j}} \\ G^{+\mathrm{i}j} & G^{+\mathrm{ij}} \end{pmatrix} - \begin{pmatrix} G^{-ij} & G^{-i\mathrm{j}} \\ G^{-\mathrm{i}j} & G^{-\mathrm{ij}} \end{pmatrix}, \tag{3.3}$$

and write Huygens' principle in the form

$$\begin{pmatrix} \delta\phi^i \\ \delta\psi^{\mathrm{i}} \end{pmatrix} = \int_\Sigma d\Sigma_{\mu'} \int d^4x'' \int d^4x''' \begin{pmatrix} G^{ij''} & G^{i\mathrm{j}''} \\ G^{\mathrm{i}j''} & G^{\mathrm{ij}''} \end{pmatrix} \times \begin{pmatrix} f^{\mu'}_{i''k'''} & f^{\mu'}_{i''\mathrm{k}'''} \\ f^{\mu'}_{\mathrm{j}''k'''} & f^{\mu'}_{\mathrm{j}''\mathrm{k}'''} \end{pmatrix} \begin{pmatrix} \delta\phi^{k'''} \\ \delta\psi^{\mathrm{k}'''} \end{pmatrix}. \tag{3.4}$$

Here the value at an arbitrary point x of a solution of the equation of small disturbances,

$$\begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} \begin{pmatrix} \delta\phi^j \\ \delta\psi^{\mathrm{j}} \end{pmatrix} = 0, \tag{3.5}$$

is expressed in terms of initial values (Cauchy data)

$$\int \begin{pmatrix} f^{\mu'}_{i''k'''} & f^{\mu'}_{i''\mathrm{k}'''} \\ f^{\mu'}_{\mathrm{j}''k'''} & f^{\mu'}_{\mathrm{j}''\mathrm{k}'''} \end{pmatrix} \begin{pmatrix} \delta\phi^{k'''} \\ \delta\psi^{\mathrm{k}'''} \end{pmatrix} d^4x''' \tag{3.6}$$

on an arbitrary space-like hypersurface Σ having directed surface element $d\Sigma_{\mu'}$. The proof of (3.4) is carried out by changing the surface integral into a volume integral (Gauss' theorem) and using (2.10), (3.5), and the equality of (3.1) and (3.2). For $x > \Sigma$ the right-hand side of Eq. (3.4) becomes

$$\int_{\Sigma}^{\text{future}} d^4x' \int d^4x'' \Bigg[\begin{pmatrix} G^{-ij'} & G^{-ij'} \\ G^{-ij'} & G^{-ij'} \end{pmatrix} \times \begin{pmatrix} F_{i'k''} & F_{i'\mathbf{k}''} \\ F_{\mathbf{j}'k''} & F_{\mathbf{j}'\mathbf{k}''} \end{pmatrix} \begin{pmatrix} \delta\phi^{k''} \\ \delta\psi^{\mathbf{k}''} \end{pmatrix} - \begin{pmatrix} G^{-ik''} & G^{-i\mathbf{k}''} \\ G^{-i\mathbf{k}''} & G^{-i\mathbf{k}''} \end{pmatrix} \times \begin{pmatrix} F_{k''j'} & F_{k''\mathbf{j}'} \\ F_{\mathbf{k}''j'} & F_{\mathbf{k}''\mathbf{j}'} \end{pmatrix} \begin{pmatrix} \delta\phi^{j'} \\ \delta\psi^{\mathbf{j}'} \end{pmatrix} \Bigg] = \begin{pmatrix} \delta\phi^{i} \\ \delta\psi^{\mathbf{i}} \end{pmatrix}, \tag{3.7}$$

while for $x < \Sigma$ it becomes

$$\int_{\text{past}}^{\Sigma} d^4x' \int d^4x'' \Bigg[\begin{pmatrix} G^{+ij'} & G^{+ij'} \\ G^{+ij'} & G^{+ij'} \end{pmatrix} \times \begin{pmatrix} F_{i'k''} & F_{i'\mathbf{k}''} \\ F_{\mathbf{j}'k''} & F_{\mathbf{j}'\mathbf{k}''} \end{pmatrix} \begin{pmatrix} \delta\phi^{k''} \\ \delta\psi^{\mathbf{k}''} \end{pmatrix} - \begin{pmatrix} G^{+ik''} & G^{+i\mathbf{k}''} \\ G^{+i\mathbf{k}''} & G^{+i\mathbf{k}''} \end{pmatrix} \times \begin{pmatrix} F_{k''j'} & F_{k''\mathbf{j}'} \\ F_{\mathbf{k}''j'} & F_{\mathbf{k}''\mathbf{j}'} \end{pmatrix} \begin{pmatrix} \delta\phi^{j'} \\ \delta\psi^{\mathbf{j}'} \end{pmatrix} \Bigg] = \begin{pmatrix} \delta\phi^{i} \\ \delta\psi^{\mathbf{i}} \end{pmatrix}, \tag{3.8}$$

Q.E.D. The extension of the domains of integration arbitrarily far into the future and past, respectively, is permitted because the Green's functions themselves vanish beyond the point x in each case.

It will be noted that Eq. (2.10) was used in the above proof but not Eq. (2.9). If we therefore take (2.10) as the defining equation for the Green's functions we may infer the validity of (2.9) through the following considerations: Since the $\delta\phi^i$, $\delta\psi^{\mathbf{i}}$ satisfy Eq. (3.5), and since the Cauchy data (3.6) may be chosen completely arbitrarily on Σ, it follows from the form of Eq. (3.4) that the propagation functions also satisfy (3.5), i.e.,[10]

$$\begin{pmatrix} F_{ik} & F_{i\mathbf{k}} \\ F_{\mathbf{i}j} & F_{\mathbf{ik}} \end{pmatrix} \begin{pmatrix} G^{kj} & G^{kj} \\ G^{\mathbf{k}j} & G^{\mathbf{k}j} \end{pmatrix} = 0. \tag{3.9}$$

Equations (2.9) are then obtained by splitting this equation into its advanced and retarded parts. The kinematics of these parts ensure that it is only the delta functions δ^j_i, $\delta^{\mathbf{j}}_{\mathbf{i}}$ or their derivatives which can make an appearance on the right-hand side, while dimensional considerations eliminate the latter. The coefficient of the delta functions is determined as -1 from the identity

$$\begin{pmatrix} G^{*ik} & G^{*i\mathbf{k}} \\ G^{*i\mathbf{k}} & G^{*i\mathbf{k}} \end{pmatrix} \begin{pmatrix} F_{kl} & F_{k\mathbf{l}} \\ F_{\mathbf{k}l} & F_{\mathbf{kl}} \end{pmatrix} \begin{pmatrix} G^{*lj} & G^{*lj} \\ G^{*\mathbf{l}j} & G^{*\mathbf{l}j} \end{pmatrix} = -\begin{pmatrix} G^{*ij} & G^{*ij} \\ G^{*ij} & G^{*ij} \end{pmatrix}, \tag{3.10}$$

in which integrations by parts and interchanges of orders of integration are permitted by the kinematics of the Green's functions only when the $\pm$ signs go together.

Equations (2.9) and (2.10) together with the symmetry properties of the F_{ij}, F_{ij}, etc., lead to the reciprocity relations. We have

$$F_{ij} = F_{ji}, \quad F_{ij} = F_{\mathbf{j}i}, \quad F_{\mathbf{ij}} = -F_{\mathbf{ij}}, \tag{3.11}$$

and hence

$$\begin{pmatrix} G^{*ij} & G^{*ij} \\ G^{*ij} & G^{*ij} \end{pmatrix} - \begin{pmatrix} G^{*ji} & -G^{*ji} \\ -G^{*ji} & -G^{*ji} \end{pmatrix} = \begin{pmatrix} G^{*ik} & G^{*i\mathbf{k}} \\ G^{*i\mathbf{k}} & G^{*i\mathbf{k}} \end{pmatrix} \Bigg[\begin{pmatrix} F_{kl} & F_{k\mathbf{l}} \\ F_{\mathbf{k}l} & F_{\mathbf{kl}} \end{pmatrix} - \begin{pmatrix} F_{lk} & F_{l\mathbf{k}} \\ F_{l\mathbf{k}} & -F_{l\mathbf{k}} \end{pmatrix} \Bigg] \begin{pmatrix} G^{*jl} & -G^{*jl} \\ -G^{*jl} & -G^{*jl} \end{pmatrix} = 0,$$

or

$$G^{*ij} = G^{*ji}, \quad G^{*ij} = -G^{*ji}, \quad G^{*ij} = -G^{*ji}, \tag{3.12}$$

$$G^{ij} = -G^{ji}, \quad G^{ij} = G^{ji}, \quad G^{ij} = G^{ji}. \tag{3.13}$$

Careful attention to the ordering of factors is important in obtaining these results. We note that since F_{ij} is real (Hermitian) while F_{ij} and $F_{\mathbf{ij}}$ are imaginary (anti-Hermitian) it follows that G^{*ij} and G^{*ij} must be real (Hermitian) while G^{*ij} is imaginary (anti-Hermitian).

If B is any observable we now have

$$\begin{aligned} \delta^*_A B &\equiv B_{,i}\, \delta^*_A\phi^i + B_{,\mathbf{i}}\, \delta^*_A\psi^{\mathbf{i}} \\ &= \epsilon(B_{,i}\ B_{,\mathbf{i}}) \begin{pmatrix} G^{*ij} & G^{*ij} \\ G^{*ij} & G^{*ij} \end{pmatrix} \begin{pmatrix} A_{,j} \\ A_{,\mathbf{j}} \end{pmatrix} \\ &= \epsilon(A_{,j}\ A_{,\mathbf{j}}) \begin{pmatrix} G^{*ij} & -G^{*ij} \\ -G^{*ij} & -G^{*ij} \end{pmatrix} \begin{pmatrix} B_{,i} \\ B_{,\mathbf{i}} \end{pmatrix} \\ &= \epsilon(A_{,j}\ A_{,\mathbf{j}}) \begin{pmatrix} G^{*ji} & G^{*ji} \\ G^{*ji} & G^{*ji} \end{pmatrix} \begin{pmatrix} B_{,i} \\ B_{,\mathbf{i}} \end{pmatrix} \\ &= \delta^*_B A, \end{aligned} \tag{3.14}$$

which is the invariant form of the reciprocity relations.

4. UNIQUENESS AND INVARIANCE OF THE POISSON BRACKET

Following the measurement theoretical arguments of reference 2 we define the Poisson bracket of two

[10] That the propagation functions satisfy the corresponding equation with the wave operator standing on the right follows already from Eqs. (2.10).

observables A and B as follows:

$$(A, B) \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} (\delta_A^- B - \delta_B^- A)$$

$$= \lim_{\epsilon \to 0} \frac{1}{\epsilon} (\delta_B^+ A - \delta_B^- A)$$

$$= (A_{,i} \quad A_{,\mathrm{i}}) \begin{pmatrix} G^{ij} & G^{i\mathrm{j}} \\ G^{\mathrm{i}j} & G^{\mathrm{ij}} \end{pmatrix} \begin{pmatrix} B_{,j} \\ B_{,\mathrm{j}} \end{pmatrix}. \tag{4.1}$$

If this Poisson bracket is now regarded as defining an actual commutator, we may write

$$i(A, B) \equiv [A, B] = (A_{,i} \quad A_{,\mathrm{i}}) \begin{pmatrix} [\phi^i, B] \\ [\psi^{\mathrm{i}}, B] \end{pmatrix}$$

$$= (A_{,i} \quad A_{,\mathrm{i}}) \begin{pmatrix} [\phi^i, \phi^j] & -[\phi^i, \psi^{\mathrm{j}}] \\ [\psi^{\mathrm{i}}, \phi^j] & -\{\psi^{\mathrm{i}}, \psi^{\mathrm{j}}\} \end{pmatrix} \begin{pmatrix} B_{,j} \\ B_{,\mathrm{j}} \end{pmatrix}, \tag{4.2}$$

and it is seen that the commutator may be computed *as if* the dynamical variables ϕ^i, ψ^{i} satisfied the (anti-)commutation relations[11]

$$[\phi^i, \phi^j] = iG^{ij}, \tag{4.3}$$

$$[\phi^i, \psi^{\mathrm{j}}] = -iG^{i\mathrm{j}}, \tag{4.4}$$

$$[\psi^{\mathrm{i}}, \phi^j] = iG^{\mathrm{i}j}, \tag{4.5}$$

$$\{\psi^{\mathrm{i}}, \psi^{\mathrm{j}}\} = -iG^{\mathrm{ij}}. \tag{4.6}$$

When infinite-dimensional invariance groups are absent these are the actual (anti-)commutation relations for the dynamical variables, and we note that the functions on the right of Eqs. (4.3)–(4.6) possess the symmetry and reality properties demanded by the (anti-)commutators on the left. There remain, of course, questions of factor ordering and the proper definition of the propagation functions as quantum operators, but these difficult questions will not be investigated here.

It is important to check the uniqueness of the quantities $\delta_A^{\pm} B$ and hence of the Poisson bracket (A, B). We note to begin with that these quantities are group invariant by construction. This follows immediately from our restriction to linear representations and our careful choice of the transformation properties of all quantities appearing in the formalism, although it is also true without these restrictions. We may in particular show that the $\delta_A^{\pm} B$ are independent of the choice of the matrices $g^{\alpha\beta}$, g_{ij}, $g_{i\mathrm{j}}$, etc. We do this by considering infinitesimal variations in the latter. We have

[11] The curly bracket denotes the anticommutator.

$$\begin{pmatrix} F_{ij} & F_{i\mathrm{j}} \\ F_{\mathrm{i}j} & F_{\mathrm{ij}} \end{pmatrix} = \begin{pmatrix} S_{,ij} & S_{,i\mathrm{j}} \\ S_{,\mathrm{i}j} & S_{,\mathrm{ij}} \end{pmatrix}$$

$$+ \begin{pmatrix} g_{ik} & g_{i\mathrm{k}} \\ g_{\mathrm{i}k} & g_{\mathrm{ik}} \end{pmatrix} \begin{pmatrix} R_\alpha^k \\ R_\alpha^{\mathrm{k}} \end{pmatrix} g^{\alpha\beta} (R_\beta^l \quad R_\beta^{\mathrm{l}}) \begin{pmatrix} g_{lj} & g_{l\mathrm{j}} \\ -g_{\mathrm{l}j} & -g_{\mathrm{lj}} \end{pmatrix}, \tag{4.7}$$

and hence, making use of (2.14) and (2.15),

$$\delta \begin{pmatrix} G^{*ij} & G^{*i\mathrm{j}} \\ G^{*\mathrm{i}j} & G^{*\mathrm{ij}} \end{pmatrix}$$

$$= \begin{pmatrix} G^{*ik} & G^{*i\mathrm{k}} \\ G^{*\mathrm{i}k} & G^{*\mathrm{ik}} \end{pmatrix} \delta \begin{pmatrix} F_{kl} & F_{k\mathrm{l}} \\ F_{\mathrm{k}l} & F_{\mathrm{kl}} \end{pmatrix} \begin{pmatrix} G^{*lj} & G^{*l\mathrm{j}} \\ G^{*\mathrm{l}j} & G^{*\mathrm{lj}} \end{pmatrix}$$

$$= \begin{pmatrix} G^{*ik} & G^{*i\mathrm{k}} \\ G^{*\mathrm{i}k} & G^{*\mathrm{ik}} \end{pmatrix} \delta \begin{pmatrix} g_{kl} & g_{k\mathrm{l}} \\ g_{\mathrm{k}l} & g_{\mathrm{kl}} \end{pmatrix} \begin{pmatrix} R_\alpha^l \\ R_\alpha^{\mathrm{l}} \end{pmatrix} G^{*\alpha\beta} (R_\beta^j \quad -R_\beta^{\mathrm{j}})$$

$$+ \begin{pmatrix} R_\gamma^i \\ R_\gamma^{\mathrm{i}} \end{pmatrix} G^{*\gamma}{}_\alpha \, \delta g^{\alpha\beta} G_\beta^{*\delta} (R_\delta^i \quad -R_\delta^{\mathrm{j}})$$

$$+ \begin{pmatrix} R_\alpha^i \\ R_\alpha^{\mathrm{i}} \end{pmatrix} G^{*\alpha\beta} (R_\beta^k \quad R_\beta^{\mathrm{k}})$$

$$\times \delta \begin{pmatrix} g_{kl} & -g_{k\mathrm{l}} \\ -g_{\mathrm{k}l} & -g_{\mathrm{kl}} \end{pmatrix} \begin{pmatrix} G^{*lj} & G^{*l\mathrm{j}} \\ G^{*\mathrm{l}j} & G^{*\mathrm{lj}} \end{pmatrix}. \tag{4.8}$$

Inserting this into the equation

$$\delta(\delta_A^{\pm} B) = \epsilon (B_{,i} \quad B_{,\mathrm{i}}) \, \delta \begin{pmatrix} G^{*ij} & G^{*i\mathrm{j}} \\ G^{*\mathrm{i}j} & G^{*\mathrm{ij}} \end{pmatrix} \begin{pmatrix} A_{,j} \\ A_{,\mathrm{j}} \end{pmatrix}, \tag{4.9}$$

we see that $\delta(\delta_A^{\pm} B)$ vanishes in virtue of the group invariance of A and B [Eq. (1.9)]. The Poisson bracket (A, B) is therefore also independent of the choice of the $g^{\alpha\beta}$, g_{ij}, $g_{i\mathrm{j}}$, etc.

In the above derivations it has been assumed that the group invariance of the observables A and B is a property independent of the dynamical equations. Actually, this need not be the case. The Poisson bracket of A and B is still well defined and unique even if A and B satisfy identities of the form

$$A_{,i} R_\alpha^i + A_{,\mathrm{i}} R_\alpha^{\mathrm{i}} \equiv S_{,i} a_\alpha^i + S_{,\mathrm{i}} a_\alpha^{\mathrm{i}}, \tag{4.10}$$

$$B_{,i} R_\alpha^i + B_{,\mathrm{i}} R_\alpha^{\mathrm{i}} \equiv S_{,i} b_\alpha^i + S_{,\mathrm{i}} b_\alpha^{\mathrm{i}}, \tag{4.11}$$

so that they become group invariant only when the dynamical equations are satisfied. This fact is of importance in the theory of asymptotic invariants (see Sec. 6) which are observables of just this type. Generally, we shall refer to such observables as *conditional invariants*. It is not difficult to see that with the exception of the argument proving the group invariance of the Poisson bracket, all previous

derivations, including those of all earlier sections, go through equally well for conditional invariants.

The group invariance of the Poisson bracket can be proved for the case of conditional invariants with the aid of the following lemmas:

$$\begin{pmatrix} S_{,ik} & S_{,i\mathbf{k}} \\ S_{,\mathbf{i}k} & S_{,\mathbf{ik}} \end{pmatrix} \begin{pmatrix} G^{kj} & G^{k\mathbf{j}} \\ G^{\mathbf{k}j} & G^{\mathbf{kj}} \end{pmatrix}$$
$$= -\begin{pmatrix} R_{i\alpha}R^{\alpha}_{k} & R_{i\alpha}R^{\alpha}_{\mathbf{k}} \\ R_{\mathbf{i}\alpha}R^{\alpha}_{k} & R_{\mathbf{i}\alpha}R^{\alpha}_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} G^{kj} & G^{k\mathbf{j}} \\ G^{\mathbf{k}j} & G^{\mathbf{kj}} \end{pmatrix}$$
$$= -\begin{pmatrix} R_{i\alpha} \\ R_{\mathbf{i}\alpha} \end{pmatrix} G^{\alpha\beta}(R^{j}_{\beta} \; -R^{\mathbf{j}}_{\beta}), \tag{4.12}$$

$$\begin{pmatrix} G^{ik} & G^{i\mathbf{k}} \\ G^{\mathbf{i}k} & G^{\mathbf{ik}} \end{pmatrix} \begin{pmatrix} S_{,kj} & S_{,k\mathbf{j}} \\ S_{,\mathbf{k}j} & S_{\mathbf{k},\mathbf{j}} \end{pmatrix}$$
$$= -\begin{pmatrix} G^{ik} & G^{i\mathbf{k}} \\ G^{\mathbf{i}k} & G^{\mathbf{ik}} \end{pmatrix} \begin{pmatrix} R^{\alpha}_{k}R_{j\alpha} & R^{\alpha}_{k}R_{\mathbf{j}\alpha} \\ R^{\alpha}_{\mathbf{k}}R_{j\alpha} & R^{\alpha}_{\mathbf{k}}R_{\mathbf{j}\alpha} \end{pmatrix}$$
$$= -\begin{pmatrix} R^{i}_{\alpha} \\ R^{\mathbf{i}}_{\alpha} \end{pmatrix} G^{\alpha\beta}(R_{j\beta} \; R_{\mathbf{j}\beta}), \tag{4.13}$$

which are corollaries of Eqs. (2.14), (2.15), and (3.9). We first note that the group transformation laws for $A_{,i}$ and $A_{,\mathbf{i}}$ (and similarly for $B_{,i}$ and $B_{,\mathbf{i}}$) are given by

$$\delta A_{,i} \equiv (A_{,ij}R^{j}_{\alpha} + A_{,i\mathbf{j}}R^{\mathbf{j}}_{\alpha})\;\delta\xi^{\alpha}$$
$$= (-A_{,j}R^{j}_{\alpha,i} + S_{,ij}a^{j}_{\alpha} + S_{,i\mathbf{j}}a^{\mathbf{j}}_{\alpha})\;\delta\xi^{\alpha}, \tag{4.14}$$

$$\delta A_{,\mathbf{i}} \equiv (A_{,\mathbf{i}j}R^{j}_{\alpha} + A_{,\mathbf{ij}}R^{\mathbf{j}}_{\alpha})\;\delta\xi^{\alpha}$$
$$= (-A_{,\mathbf{j}}R^{\mathbf{j}}_{\alpha,\mathbf{i}} + S_{,\mathbf{i}j}a^{j}_{\alpha} + S_{,\mathbf{ij}}a^{\mathbf{j}}_{\alpha})\;\delta\xi^{\alpha}, \tag{4.15}$$

as may be inferred by differentiating Eq. (4.10) and making use of the dynamical equations. Only the second and third terms on the right of Eqs. (4.14) and (4.15), which do not appear in the group transformation laws for the derivatives of a rigorous invariant, survive in the group transformation law for the Poisson bracket. We have, in fact,

$$\delta(A, B) = (a^{i}_{\alpha} \; -a^{\mathbf{i}}_{\alpha}) \begin{pmatrix} S_{,ik} & S_{,i\mathbf{k}} \\ S_{,\mathbf{i}k} & S_{,\mathbf{ik}} \end{pmatrix} \begin{pmatrix} G^{kj} & G^{k\mathbf{j}} \\ G^{\mathbf{k}j} & G^{\mathbf{kj}} \end{pmatrix} \begin{pmatrix} B_{,j} \\ B_{,\mathbf{j}} \end{pmatrix} \delta\xi^{\alpha}$$
$$+ (A_{,i} \;\; A_{,\mathbf{i}}) \begin{pmatrix} G^{ik} & G^{i\mathbf{k}} \\ G^{\mathbf{i}k} & G^{\mathbf{ik}} \end{pmatrix} \begin{pmatrix} S_{,kj} & S_{,k\mathbf{j}} \\ S_{,\mathbf{k}j} & S_{,\mathbf{kj}} \end{pmatrix} \begin{pmatrix} b^{j}_{\alpha} \\ b^{\mathbf{j}}_{\alpha} \end{pmatrix} \delta\xi^{\alpha}. \tag{4.16}$$

But this vanishes in virtue of Eqs. (4.10), (4.11), (4.12), (4.13), and the dynamical equations, Q.E.D.

The Lemmas (4.12) and (4.13) may also be used to show that observables are really defined only *modulo* the dynamical equations. Thus if we make the transformations

$$A' = A + S_{,i}a^{i} + S_{,\mathbf{i}}a^{\mathbf{i}},$$
$$B' = B + S_{,i}b^{i} + S_{,\mathbf{i}}b^{\mathbf{i}}, \tag{4.17}$$

where a^{i}, $a^{\mathbf{i}}$, b^{i}, $b^{\mathbf{i}}$ are arbitrary, we then have, after dropping terms in $S_{,i}$ and $S_{,\mathbf{i}}$,

$$(A', B') = (A, B) + (A_{,i} \;\; A_{,\mathbf{i}})$$
$$\times \begin{pmatrix} G^{ij} & G^{i\mathbf{j}} \\ G^{\mathbf{i}j} & G^{\mathbf{ij}} \end{pmatrix} \begin{pmatrix} S_{,ik} & S_{,i\mathbf{k}} \\ S_{,\mathbf{j}k} & S_{,\mathbf{jk}} \end{pmatrix} \begin{pmatrix} b^{k} \\ b^{\mathbf{k}} \end{pmatrix}$$
$$+ (a^{i} \; -a^{\mathbf{i}}) \begin{pmatrix} S_{,ij} & S_{,i\mathbf{j}} \\ S_{,\mathbf{i}j} & S_{,\mathbf{ij}} \end{pmatrix} \begin{pmatrix} G^{jk} & G^{j\mathbf{k}} \\ G^{\mathbf{j}k} & G^{\mathbf{jk}} \end{pmatrix} \begin{pmatrix} B_{,k} \\ B_{,\mathbf{k}} \end{pmatrix}$$
$$+ (a^{i} \; -a^{\mathbf{i}}) \begin{pmatrix} S_{,ij} & S_{,i\mathbf{j}} \\ S_{,\mathbf{i}j} & S_{,\mathbf{ij}} \end{pmatrix} \begin{pmatrix} G^{jk} & G^{j\mathbf{k}} \\ G^{\mathbf{j}k} & G^{\mathbf{jk}} \end{pmatrix}$$
$$\times \begin{pmatrix} S_{,kl} & S_{,k\mathbf{l}} \\ S_{,\mathbf{k}l} & S_{,\mathbf{kl}} \end{pmatrix} \begin{pmatrix} b^{l} \\ b^{\mathbf{l}} \end{pmatrix}$$
$$= (A, B) - (A_{,i} \;\; A_{,\mathbf{i}}) \begin{pmatrix} R^{i}_{\alpha} \\ R^{\mathbf{i}}_{\alpha} \end{pmatrix} G^{\alpha\beta}(R_{i\beta} \; R_{\mathbf{j}\beta}) \begin{pmatrix} b^{i} \\ b^{\mathbf{j}} \end{pmatrix}$$
$$- (a^{i} \; -a^{\mathbf{i}}) \begin{pmatrix} R_{i\alpha} \\ R_{\mathbf{i}\alpha} \end{pmatrix} G^{\alpha\beta}(R^{j}_{\beta} \; -R^{\mathbf{j}}_{\beta}) \begin{pmatrix} B_{,i} \\ B_{,j} \end{pmatrix}$$
$$- (a^{i} \; -a^{\mathbf{i}}) \begin{pmatrix} S_{,ij} & S_{,i\mathbf{j}} \\ S_{,\mathbf{i}j} & S_{,\mathbf{ij}} \end{pmatrix} \begin{pmatrix} R^{j}_{\alpha} \\ R^{\mathbf{j}}_{\alpha} \end{pmatrix} G^{\alpha\beta}(R_{k\beta} \; R_{\mathbf{k}\beta}) \begin{pmatrix} b^{k} \\ b^{\mathbf{k}} \end{pmatrix}, \tag{4.18}$$

which is seen to reduce to (A, B) in virtue of Eqs. (1.13), (4.10), (4.11), and the dynamical equations.

5. THE POISSON - JACOBI IDENTITY

In order to prove that the Poisson bracket (4.1) satisfies the Poisson-Jacobi identity one must pay careful attention to the (anti-)commutation properties, symmetries, and ordering of the factors appearing in the many terms involved. Writing

$$(A, (B, C)) = (A_{,i} \;\; A_{,\mathbf{i}}) \begin{pmatrix} G^{il} & G^{i\mathbf{l}} \\ G^{\mathbf{i}l} & G^{\mathbf{il}} \end{pmatrix}$$
$$\times \begin{pmatrix} \dfrac{\delta}{\delta\phi^{l}} \left[(B_{,i} \;\; B_{,\mathbf{j}}) \begin{pmatrix} G^{jk} & G^{j\mathbf{k}} \\ G^{\mathbf{j}k} & G^{\mathbf{jk}} \end{pmatrix} \begin{pmatrix} C_{,k} \\ C_{,\mathbf{k}} \end{pmatrix} \right] \\ \dfrac{\delta}{\delta\psi^{\mathbf{l}}} \left[(B_{,i} \;\; B_{,\mathbf{j}}) \begin{pmatrix} G^{jk} & G^{j\mathbf{k}} \\ G^{\mathbf{j}k} & G^{\mathbf{jk}} \end{pmatrix} \begin{pmatrix} C_{,k} \\ C_{,\mathbf{k}} \end{pmatrix} \right] \end{pmatrix}, \tag{5.1}$$

performing the differentiations indicated, taking note of Eq. (1.10), and expanding the matrix multiplications, one encounters 160 separate terms which

divide into two distinct groups. The first group consists of all terms in which the propagation functions remain undifferentiated. There are 32 such terms, and these are obtainable from the prototypes

$$A_{,i}G^{il}B_{,jl}G^{jk}C_{,k} \quad \text{and} \quad A_{,i}G^{il}B_{,j}G^{jk}C_{,kl} \tag{5.2}$$

by replacing the four dummy lightface indices in all possible ways with boldface indices and affixing minus signs when required by the conditions of Eq. (1.10).[12] Adding to these terms the 64 additional terms obtained by cyclic permutation of A, B, and C, reordering factors judiciously (making corresponding sign changes where necessary), and taking note of the symmetries (3.13) of the propagation functions, one can readily effect a term by term cancellation.

To carry out a similar procedure for the second group one must first obtain expressions for the functional derivatives of the propagation functions. Making use of Eqs. (2.14) and (2.15) we have

$$\begin{aligned}
&\frac{\delta}{\delta\phi^k}\begin{bmatrix} G^{*ij} & G^{*ij} \\ G^{*ij} & G^{*ij} \end{bmatrix} \\
&= \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\left[\frac{\delta}{\delta\phi^k}\begin{bmatrix} F_{lm} & F_{l\mathbf{m}} \\ F_{1m} & F_{1\mathbf{m}} \end{bmatrix}\right]\begin{bmatrix} G^{*mj} & G^{*mj} \\ G^{*mj} & G^{*mj} \end{bmatrix} \\
&= \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\left[\begin{bmatrix} S_{,lmk} & S_{,lmk} \\ S_{,1mk} & S_{,1mk} \end{bmatrix} + \begin{bmatrix} R_{l\alpha,k} \\ R_{1\alpha,k} \end{bmatrix}(R^{\alpha}_{m} \;\; R^{\alpha}_{m})\right. \\
&\quad \left. + \begin{bmatrix} R_{l\alpha} \\ R_{1\alpha} \end{bmatrix}(R^{\alpha}_{m,k} \;\; R^{\alpha}_{m,k})\right]\begin{bmatrix} G^{*mj} & G^{*mj} \\ G^{*mj} & G^{*mj} \end{bmatrix} \\
&= \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\begin{bmatrix} S_{,lmk} & S_{,lmk} \\ S_{,1mk} & S_{,1mk} \end{bmatrix}\begin{bmatrix} G^{*mj} & G^{*mj} \\ G^{*mj} & G^{*mj} \end{bmatrix} \\
&\quad + \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\begin{bmatrix} R_{l\alpha,k} \\ R_{1\alpha,k} \end{bmatrix}G^{*\alpha\beta}(R^{i}_{\beta} \;\; -R^{j}_{\beta}) \\
&\quad + \begin{bmatrix} R^{i}_{\alpha} \\ R^{i}_{\alpha} \end{bmatrix}G^{*\alpha}{}_{\beta}(R^{\beta}_{l,k} \;\; R^{\beta}_{1,k})\begin{bmatrix} G^{*lj} & G^{*lj} \\ G^{*lj} & G^{*lj} \end{bmatrix}
\end{aligned} \tag{5.3}$$

and, similarly, taking note of (1.10),

$$\begin{aligned}
&\frac{\delta}{\delta\psi^k}\begin{bmatrix} G^{*ij} & G^{*ij} \\ G^{*ij} & G^{*ij} \end{bmatrix} = \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix} \\
&\quad \times\left[\frac{\delta}{\delta\psi^k}\begin{bmatrix} F_{lm} & F_{l\mathbf{m}} \\ F_{1m} & F_{1\mathbf{m}} \end{bmatrix}\right]\begin{bmatrix} G^{*mj} & -G^{*mj} \\ -G^{*mj} & G^{*mj} \end{bmatrix} \\
&= \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\left[\begin{bmatrix} S_{,lmk} & S_{,lmk} \\ S_{,1mk} & S_{,1mk} \end{bmatrix}\right. \\
&\quad + \begin{bmatrix} R_{l\alpha,k} \\ R_{1\alpha,k} \end{bmatrix}(R^{\alpha}_{m} \;\; -R^{\alpha}_{m}) \\
&\quad \left. + \begin{bmatrix} R_{l\alpha} \\ R_{1\alpha} \end{bmatrix}(R^{\alpha}_{m,k} \;\; R^{\alpha}_{m,k})\right]\begin{bmatrix} G^{*mj} & -G^{*mj} \\ -G^{*mj} & G^{*mj} \end{bmatrix} \\
&= \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\begin{bmatrix} S_{,lmk} & S_{,lmk} \\ S_{,1mk} & S_{,1mk} \end{bmatrix}\begin{bmatrix} G^{*mj} & -G^{*mj} \\ -G^{*mj} & G^{*mj} \end{bmatrix} \\
&\quad + \begin{bmatrix} G^{*il} & G^{*il} \\ G^{*il} & G^{*il} \end{bmatrix}\begin{bmatrix} R_{l\alpha,k} \\ R_{1\alpha,k} \end{bmatrix}G^{*\alpha\beta}(R^{i}_{\beta} \;\; R^{j}_{\beta}) \\
&\quad + \begin{bmatrix} R^{i}_{\alpha} \\ R^{i}_{\alpha} \end{bmatrix}G^{*\alpha}{}_{\beta}(R^{\beta}_{l,k} \;\; R^{\beta}_{1,k})\begin{bmatrix} G^{*lj} & -G^{*lj} \\ -G^{*lj} & G^{*lj} \end{bmatrix}.
\end{aligned} \tag{5.4}$$

Inserting these results into (5.1) and making use of the identity (I, 1.20) and the group invariance of B and C, one sees that only the terms involving the third functional derivatives of the action survive. There are 128 such terms, obtainable from the prototypes

$$\left.\begin{aligned} &A_{,i}G^{ia}B_{,j}G^{jb}S_{,bca}G^{ck}C_{,k}\theta(k, c) \\ &\text{and} \\ &-A_{,i}G^{ia}B_{,j}G^{jb}S_{,bca}G^{ck}C_{,k}\theta(j, b) \end{aligned}\right\} \tag{5.5}$$

by again replacing the dummy lightface indices in all possible ways with boldface indices and changing the signs when required by the conditions of Eq. (1.10).[13] Adding to these terms the 256 additional terms obtained by cyclic permutation of A, B, and C, reordering factors (with appropriate sign changes), and taking note of the symmetries not only of the propagation functions but also of the third derivatives of the action, one again finds a term-by-term cancellation. Typical cancelling terms are, for example,

$$\left.\begin{aligned} &A_{,i}G^{ia}B_{,j}G^{jb}S_{,bca}G^{ck}C_{,k}\theta(k, c) \\ &\text{and} \\ &B_{,i}G^{ia}C_{,j}G^{jb}S_{,bca}G^{ck}A_{,k}\theta(j, b) \end{aligned}\right\}, \tag{5.6}$$

[12] Minus signs are required on the first prototype of (5.2) for the following combinations of indices:

i, j, *k*, l *i*, j, k, l i, j, *k*, l i, j, k, l.

No minus signs are required with the second prototype.

[13] Sign changes are here required on both prototypes for the following combinations of indices:

i, *j*, *k*, a, *b*, c	*i*, *j*, k, a, *b*, c
i, *j*, *k*, a, b, c	*i*, *j*, k, a, b, c
i, j, *k*, a, *b*, c	*i*, j, k, a, *b*, c
i, j, *k*, a, b, c	*i*, j, k, a, b, c
i, *j*, *k*, a, *b*, c	i, *j*, k, a, *b*, c
i, *j*, *k*, a, b, c,	i, *j*, k, a, b, c
i, j, *k*, a, *b*, c	i, j, k, a, *b*, c
i, j, *k*, a, b, c	i, j, k, a, b, c

$$\left.\begin{aligned} &A_{,i}G^{ia}B_{,j}G^{jb}S_{,bca}G^{ck}C_{,k}\theta(\mathbf{k},\mathbf{c}) \\ &\text{and} \\ &-B_{,i}G^{ia}C_{,j}G^{jb}S_{,bca}G^{ck}C_{,k}\theta(\mathbf{j},\mathbf{b}) \end{aligned}\right\}, \quad \text{etc.} \tag{5.7}$$

Therefore

$$(A,(B,C)) + (B,(C,A)) + (C,(A,B)) = 0. \tag{5.8}$$

6. ASYMPTOTIC INVARIANTS

In the practical application of quantum field theory it is important to be able to set up boundary conditions in the remote past or future. It is essential, for the physical interpretation of the theory, to be able to do this in a group invariant way, and for this purpose one needs to introduce *asymptotic invariants.*

Let us suppose that the zero points of the field variables ϕ^i, ψ^i have been chosen in such a way that the values $\phi^i = 0$, $\psi^i = 0$ are classical solutions of the dynamical equations corresponding to flat empty space-time. It is then possible to expand the action in a formal power series[14]

$$\begin{aligned} S = {} & (1/2!)\,{}^0S_{,ij}\phi^i\phi^j + (1/3!)\,{}^0S_{,ijk}\phi^i\phi^j\phi^k + \cdots \\ & + (1/2!)({}^0S_{,ij}\psi^j\psi^i + {}^0S_{,ijk}\psi^j\psi^i\phi^k \\ & + (1/2!)\,{}^0S_{,ijkl}\psi^j\psi^i\phi^k\phi^l + \cdots) \\ & + (1/4!)({}^0S_{,ijkl}\psi^l\psi^k\psi^j\psi^i \\ & + {}^0S_{,ijklm}\psi^l\psi^k\psi^j\psi^i\phi^m + \cdots) \\ & + \cdots, \end{aligned} \tag{6.1}$$

of which the quadratic terms may be taken as the action functionals for "free" asymptotic incoming and outgoing fields. These asymptotic fields, which will be denoted by ϕ^{-i}, ψ^{-i} and ϕ^{+i}, ψ^{+i}, respectively, satisfy the dynamical equations

$${}^0S_{,ij}\phi^{\pm j} = 0, \qquad {}^0S_{,ij}\psi^{\pm j} = 0. \tag{6.2}$$

Because of the group invariance of the action S, the coefficients of the power series (6.1) are not all independent of one another. In particular we have, setting $\phi^i = 0$, $\psi^i = 0$ in (1.13),

$${}^0S_{,ij}\,{}^0R^j_\alpha = 0. \tag{6.3}$$

The ${}^0R^i_\alpha$ may be regarded as the coefficients of the infinitesimal transformation law of an *Abelian* invariance group for the asymptotic fields[15]:

$$\delta\phi^{\pm i} = {}^0R^i_\alpha\,\delta\zeta^\alpha, \tag{6.4}$$

$$\delta\psi^{\pm i} = 0. \tag{6.5}$$

The group parameters $\delta\zeta^\alpha$ must differ, in general, from the $\delta\xi^\alpha$, since an Abelian group cannot be mapped isomorphically into a non-Abelian group. A relation between the two [Eq. (6.14)] can, however, be established in the following manner. We begin by imposing the supplementary condition

$${}^0R_{i\alpha}\phi^i = 0 \tag{6.6}$$

on the field variables ϕ^i and rewriting the dynamical equations in the forms

$$\left.\begin{aligned} 0 &= {}^0S_{,ij}\phi^j + S_{,i} - {}^0S_{,ij}\phi^j \\ &= {}^0F_{ij}\phi^j + (S_{,i} - {}^0S_{,ij}\phi^j), \\ 0 &= {}^0S_{,ij}\psi^j + S_{,i} - {}^0S_{,ij}\psi^j, \end{aligned}\right\} \tag{6.7}$$

the general solutions of which may be obtained by iteration of

$$\left.\begin{aligned} \phi^i &= \phi^{\pm i} + {}^0G^{\pm ij}(S_{,j} - {}^0S_{,jk}\phi^k), \\ \psi^i &= \psi^{\pm i} + {}^0G^{\pm ij}(S_{,j} - {}^0S_{,jk}\psi^k), \end{aligned}\right\} \tag{6.8}$$

with

$${}^0F_{ij}\phi^{\pm j} = 0, \qquad {}^0S_{,ij}\psi^{\pm j} = 0. \tag{6.9}$$

The $\phi^{\pm i}$, $\psi^{\pm i}$ here may be identified with the asymptotic fields of (6.2) by observing that when the dynamical equations $S_{,i} = 0$ are satisfied, the $\phi^{\pm i}$ as well as the ϕ^i satisfy the supplementary condition (6.6). Writing

$$\left.\begin{aligned} \phi^{\pm i} &= \phi^i - {}^0G^{\pm ij}(S_{,j} - {}^0S_{,jk}\phi^k), \\ \psi^{\pm i} &= \psi^i - {}^0G^{\pm ij}(S_{,j} - {}^0S_{,jk}\psi^k), \end{aligned}\right\} \tag{6.10}$$

and using (6.3) and the Lemma (2.15) with $\phi^i = 0$, $\psi^i = 0$, we have, in fact,

$$\begin{aligned} {}^0R_{i\alpha}\phi^{\pm i} &= {}^0R_{i\alpha}\phi^i - {}^0G^{\pm\beta}_\alpha\,{}^0R^j_\beta(S_{,j} - {}^0S_{,jk}\phi^k) \\ &= {}^0R_{i\alpha}\phi^i = 0. \end{aligned} \tag{6.11}$$

Because of the singularity of the matrix ${}^0S_{,ij}$, as expressed by Eq. (6.3), a distinct physical solution of Eqs. (6.2) is defined only *modulo* a group transformation (6.4). Such a solution becomes fixed as a particular solution of Eqs. (6.9) only when the supplementary conditions (6.6) and (6.11) are satisfied. Now it is to be noted that these supplementary conditions are conditions on the dynamical variables themselves and not, as in Eq. (2.1), on their variations. Conditions of this type, in the quantum theory, are usually handled by the unpleasant expedient of regarding them as conditions

[14] The superscript ^{0} standing in front of a quantity indicates that the quantity is to be evaluated with $\phi^i = 0, \psi^i = 0$.

[15] In the cases of practical interest (e.g., electrodynamics Yang-Mills field, gravitation) the ${}^0R^i_\alpha$, like the R^i_α, are linearly independent. That is, ${}^0R^i_\alpha\delta\zeta^\alpha = 0$ if and only if $\delta\zeta^\alpha = 0$.

on the state vectors of a Hilbert space which is bigger than the physical content of the theory warrants. What is important for us here is the fact that these supplementary conditions are irrelevant as far as physical observables are concerned—that we can, in particular, construct out of the quantities (6.10) a set of conditional invariants which give a complete physical description of the asymptotic fields, regardless of whether the supplementary condition (6.6) is satisfied or not. Performing a group transformation on (6.10), we obtain, with the aid of (1.15), (1.16), and (2.14), the equations

$$\begin{aligned}\delta\phi^{*i} &= R^i_\alpha\,\delta\xi^\alpha - {}^0G^{*ij}[-S_{,k}R^k_{\alpha,j} \\ &\quad - ({}^0F_{jk} - {}^0R^\beta_j\,{}^0R_{k\beta})R^k_\alpha]\,\delta\xi^\alpha \\ &= -\,{}^0R^i_\gamma\,{}^0G^{*\gamma\beta}\,{}^0R_{k\beta}R^k_\alpha\,\delta\xi^\alpha, \qquad (6.12)\end{aligned}$$

$$\begin{aligned}\delta\psi^{*i} &= R^i_\alpha\,\delta\xi^\alpha - {}^0G^{*ij} \\ &\quad \times(-S_{,k}R^k_{\alpha,j} - {}^0S_{,jk}R^k_\alpha)\,\delta\xi^\alpha = 0, \qquad (6.13)\end{aligned}$$

which have the same forms as Eqs. (6.4) and (6.5) if we make the identification

$$\delta\zeta^\alpha = -\,{}^0G^{*\alpha\beta}\,{}^0R_{i\beta}R^i_\gamma\,\delta\xi^\gamma. \qquad (6.14)$$

Therefore, as the desired set of conditional invariants we may choose the quantities

$$\phi^*_A \equiv {}^0R^i_A\,{}^0g_{ij}\phi^{*j}, \qquad (6.15)$$

where the ${}^0R^i_A$ are the R^i_A of the orthogonality relation (2.7), evaluated for $\phi^i = 0$, $\psi^i = 0$. For under a group transformation we see immediately that

$$\delta\phi^*_A = 0, \qquad \delta\psi^{*i} = 0, \qquad (6.16)$$

where the dynamical equations $S_{,i} = 0$, $S_{,i} = 0$ are satisfied.

The ϕ^*_A, ψ^{*i} constitute a complete set of *asymptotic invariants*. Although they are only conditionally invariant we know from the results of Sec. 4 that they nevertheless possess well-defined (anti-) commutators[16] and hence have unique representations in the quantum theory. Moreover, even when the ϕ^{*i} defined by (6.10) do not satisfy the supplementary condition (6.11), and hence do not satisfy Eq. (6.9), they nevertheless differ only by a group transformation from variables which do, for it is not difficult to see that they still satisfy Eq. (6.2). Thus

$$\begin{aligned}{}^0S_{,ij}\phi^{*j} &= {}^0S_{,ij}\phi^j - ({}^0F_{ij} - {}^0R_{i\alpha}\,{}^0R^\alpha_j)\,{}^0G^{*jk} \\ &\quad \times(S_{,kl} - {}^0S_{,kl}\phi^l) \\ &= {}^0S_{,ij}\phi^j + (S_{,i} - {}^0S_{,ij}\phi^j) \\ &\quad + {}^0R_{i\alpha}\,{}^0G^{*\alpha\beta}\,{}^0R^k_\beta(S_{,k} - {}^0S_{,kl}\phi^l) \\ &= 0. \qquad (6.17)\end{aligned}$$

The ϕ^*_A, ψ^{*i} are therefore invariants of the true asymptotic free fields and can be used to construct initial and final states.

---

[16] The results of Sec. 4 apply, strictly speaking, only to observables, which contain no terms of odd degree in the $\psi^{\pm i}$. Nevertheless, the commutators of observables together with Eqs. (6.2) suffice to determine uniquely the (anti-)commutators $[\phi^*_A, \phi^*_B]$, $[\phi^*_A, \psi^{*i}]$, $\{\psi^{*i}, \psi^{*j}\}$. It is evident that these (anti-)commutators, when the $\pm$ signs are taken the same throughout, are just those of the free fields:

$$\begin{aligned}[\phi^*_A, \phi^*_B] &= i\,{}^0R_{iA}\,{}^0R_{jB}\,{}^0G^{ij}, \\ [\phi^*_A, \psi^{*i}] &= 0, \\ \{\psi^{*i}, \psi^{*j}\} &= -i\,{}^0G^{ij}.\end{aligned}$$

This can, in fact, be verified by a straightforward calculation.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

The Spin-Statistics Theorem*

R. ARNOWITT
Department of Physics, Syracuse University, Syracuse, New York
AND
S. DESER
Department of Physics, Brandeis University, Waltham, Massachusetts

A derivation of the connection between spin and statistics is obtained for spin 0, ½, and 1 fields with arbitrary local interactions. The basis used is the Schwinger action principle, whose assumptions are specified; they include neither positive energy spectrum nor *TCP* invariance. The connection can be obtained without either of these two extra requirements in most cases. The remaining cases are characterized by non-*TCP* invariant *free* Lagrangians and nonpositive definite free-particle energies. Commutation relations among different fields are also briefly discussed by means of the action principle.

I. INTRODUCTION

AFTER Pauli's original derivation of the connection between spin and statistics,[1] a number of deductions of this result, using various sets of assumptions, have been presented. These derivations have removed the restriction to noninteracting fields involved in the initial proof. Thus, the work of Schwinger[2] employed *TCP* invariance, while the approach exemplified by Burgoyne and Luders and Zumino[3] assumed the existence of a vacuum state representing the lowest energy of the system. The latter postulate ensures certain analyticity properties of vacuum expectation values, which together with other quite general requirements leads to the connection. This elegant method provides a direct generalization of Pauli's proof to coupled fields. However, in the uncoupled case, the existence of a vacuum state was only invoked to forbid Bose quantization of charged half-integral spin fields. It is therefore of interest to see to what extent one may avoid such additional postulates as the energy requirement and *TCP* invariance for coupled fields.

In this note, we shall start from the Schwinger action principle,[4] which considers only systems with a local Lagrangian, but does not demand a vacuum state. An explicit statement of the principle and its assumptions is given; as in all other derivations, we require that the Hilbert space metric be positive-definite and consider only the possibilities of commutativity or anticommutativity. We shall divide the problem into four parts according to whether the field is neutral or charged and has integral or half-integral spin. Our explicit derivation will be made for 0, ½, and 1 spins only.[5] We shall show that in all but the massless neutral spin ½ case and charged spin ½ case, neither the vacuum assumption nor *TCP* invariance is needed, irrespective of interactions. In these cases, a "wrong" connection leads to purely algebraic inconsistencies, reminiscent of those originally found by Pauli for free integral spin fields.

II. THE ACTION PRINCIPLE FRAMEWORK

The action principle,[4] upon which our treatment of the spin-statistics connection is based, requires in its derivation a number of specific postulates. We therefore first list these and discuss briefly their nature. The Appendix contains a more complete treatment.

I. The conventional Hilbert space interpretation of quantum mechanics, with positive-definite metric holds.

II. The system is invariant with respect to the proper orthochronous inhomogeneous Lorentz group.

III. The characterization of a state at a given time and the equations of its time development are local in time (i.e., we are dealing with a local field theory).

In order to discuss the remaining postulates, we introduce some notation. Let $\delta\langle a_1 t_1 \mid a_2 t_2\rangle$ be the change of a transformation function[6] under infinites-

* Supported in part by the National Science Foundation and United States Air Force Office of Scientific Research.

[1] W. Pauli, Phys. Rev. **58**, 716 (1940).

[2] J. Schwinger, Phys. Rev. **82**, 914 (1951)., Proc. Natl. Acad. Sci. U. S. **44**, 223, 617 (1958).

[3] N. Burgoyne, Nuovo cimento **8**, 607 (1958); G. Luders and B. Zumino, Phys. Rev. **110**, 1450 (1958).

[4] J. Schwinger, Phys. Rev. **82**, 914 (1951); **91**, 713 (1953).

[5] From the structure of the proof, we expect that the generalization to higher spin fields should be feasible.

[6] For simplicity, we are assuming here that out operators and states are defined at a fixed time. A more general treatment, in terms of space-like surfaces can, of course, be given.

imal transformations which alter the complete set $\{A\}$ at time t (here $A \mid a_1 t\rangle = a_1 \mid a_1 t\rangle$) and move the system in time: $t \to t + \delta t$. These changes are unitary transformations on the basis vectors according to I. Defining the infinitesimal operator δW_{12} by $\delta(a_1 t_1 \mid a_2 t_2) \equiv i\langle a_1 t_1 \mid \delta W_{12} \mid a_2 t_2\rangle$, (so that δW_{12} is necessarily Hermitian) we postulate that

IV (a). There exists a finite operator W_{12} such that a unique set of variations on its operator form yield δW_{12} for the classes of transformations considered above.

One can show (see Appendix) that W_{12} has the form of a space-time integral over the region between t_1 and t_2,

$$W_{12} = \int_{t_2}^{t_1} d^4x \mathfrak{L}(x)$$

(with the scalar function $\mathfrak{L}$ Hermitian). We further postulate that

IV(b) $\mathfrak{L}(x) = \chi A^\mu \partial_\mu \chi - \partial_\mu \chi A^\mu \chi - \mathfrak{IC}(\chi) + \partial_\mu W^\mu(\chi)$,

where χ is a column symbol whose components are all the field variables and A^μ are constant numerical matrices in that space.[7]

From postulate III, $\mathfrak{IC}$ is a local function, while Hermiticity of $\mathfrak{L}$ requires the A^μ to be skew-Hermitian and $\mathfrak{IC}$ and W^μ Hermitian. The generalized Kemmer form assumed above for $\mathfrak{L}$ is no essential restriction, since any local field system obeying second order equations with at most first derivative coupling can be described by such a Lagrangian,[8] with $\mathfrak{IC}(\chi)$ independent of derivatives of χ.

In varying W_{12} to yield the required $\delta\langle a_1 t_1 \mid a_2 t_2\rangle$ it can be shown (see Appendix) that variations of the time $t \to t + \Delta t$ and of the field variables $\chi \to \chi + \bar{\delta}\chi$ must be made throughout the space-time region. The relation of these changes to specific unitary transformations carried out on the transformation function must be specified in order to give meaning to the basic postulate IV(a). We therefore assume

V(a) If δt is the time translation carried out on the transformation function, then $\Delta t = \alpha(t)\, \delta t$, α is a c number, i.e., Δt vanishes when no unitary transformations corresponding to pure time motion are made.

V(b) The variation $\bar{\delta}\chi_\alpha$ (of the field variable χ_α) in $\mathfrak{L}$ either commutes or anticommutes with a given field variable χ_β.

The most general *a priori* possibility for Δt would be the form $\Delta t = \alpha(t)\, \delta t + \delta\beta$, where $\delta\beta$ does not vanish when a change of basis at fixed time is made, i.e., when $\delta t = 0$. The requirement V(a) is thus that $\delta\beta = 0$. It can then be shown that $\alpha = 1$ and the not unexpected result $\Delta t = \delta t$ holds. Postulate V(b) is a condition only on the operator nature of $\bar{\delta}\chi$. Some rule on the operator character of $\bar{\delta}\chi$ is needed to obtain well-defined Lagrange equations.[9] As will be seen, V(b) leads to either commutation or anticommutation relations holding between field variables and hence is similar to the assumptions commonly made in other discussions of spin and statistics. At this stage, we may compare our set of postulates with other starting points of field theory. Assumptions I, II, and V(b) are postulates conventionally included in other treatments of field theory, while IV and V(a) are characteristic of the action principle. Postulates III and IV(b) (which limit us to local fields) and a final requirement VI, to be introduced below, are invoked in all discussions of local field theory.

The above postulates allow us to obtain the Lagrange equations from W_{12} (see discussion in the Appendix). One finds

$$(-\bar{\delta}\chi A^\mu \partial_\mu \chi + \partial_\mu \chi A^\mu \bar{\delta}\chi) - m(\bar{\delta}\chi B\chi + \chi B\, \bar{\delta}\chi) - \bar{\delta}\chi\, \partial\mathfrak{IC}_I/\partial\chi = 0, \quad (2.1a)$$

where we have written $\mathfrak{IC} = m\chi B\chi + \mathfrak{IC}_I(\chi)$ to exhibit a possible mass term for the field. The symbol $\bar{\delta}\chi\, \partial\mathfrak{IC}_I/\partial\chi$ stands for $\mathfrak{IC}_I(\chi + \bar{\delta}\chi) - \mathfrak{IC}_I(\chi)$ and the matrix B is necessarily Hermitian. Explicit equations of motion can now be obtained by using the commutation relations obeyed by the $\bar{\delta}\chi$ to move them all to one side in Eq. (2.1a), and equating their coefficients to zero. For each field of given spin we will see that the A^μ are all either symmetric or antisymmetric. In the symmetric (antisymmetric) case, one will obtain equations with space-time derivatives only if $\bar{\delta}\chi$ is taken to anticommute (commute). The "wrong" choice of commutation relations for $\bar{\delta}\chi$ then reduces the content of the Lagrange equations to $\bar{\delta}\chi\, \partial\mathfrak{IC}/\partial\chi = 0$ [since the first parenthesis in Eq. (2.1a) vanishes]. This is either an identity $(0 = 0)$ or an algebraic relation (constraint) among the χ's at any time (which may or

[7] We use the metric $\eta_{\mu\nu} = \text{diag}\ (1, 1, 1, -1)$, with Latin indices varying over 1, 2, 3, Greek over 1, 2, 3, 0, and natural units: $\hbar = 1 = c$.

[8] If higher derivatives than the first had been present in the interaction, one could have adjoined the derivatives as new variables to χ to reach first-order form; however, the χ space is then not irreducible, which greatly complicates the analysis. We do not consider such couplings in this work.

[9] This restriction also guarantees that the quantum Lagrange equations resemble the classical ones in form; more complicated relations between variations and fields would lose this feature.

may not be consistent). In either case, no true Lagrange equations arise from use of the "wrong" choice. If the "right" choice is taken, one gets the standard Kemmer-Dirac equations of motion with local interactions:

$$A^\mu \partial_\mu \chi + mB\chi + \tfrac{1}{2} \partial \mathfrak{K}_I / \partial \chi = 0 \qquad (2.1b)$$

From our postulates, one finds (see Appendix) that the generator of the unitary transformation for time translation ($\mathfrak{U} = 1 + iG_t$) is

$$G_t = -H\ \delta t = -\int d^3r[\tfrac{1}{2}(-\chi A^i\ \partial_i \chi + \partial_i \chi A^i \chi) + \mathfrak{K}]\ \delta t, \qquad (2.2a)$$

where H is the usual field Hamiltonian. The effect of G_t on χ is to translate it in time by an amount δt, i.e.,

$$[\chi, G_t] = i\dot{\chi}\ \delta t. \qquad (2.2b)$$

Equation (2.2b) is the Heisenberg equation of motion. The generator G' of arbitrary unitary transformations at a fixed time may be shown to be

$$G' = G_\chi + \int d^3r\ \bar{\delta}\chi\ (\partial W^0/\partial \chi) \qquad (2.3)$$

where

$$G_\chi \equiv \int d^3r \tfrac{1}{2}(\chi A^0\ \bar{\delta}\chi - \delta\chi A^0 \chi) \qquad (2.4)$$

and $\bar{\delta}\chi\ (\partial W^0/\partial\chi)$ stands for $W^0(\chi + \bar{\delta}\chi) - W^0(\chi)$. A special case of importance is the choice $W^0 = 0$, i.e., $G' = G_\chi$. The effect of G_χ on χ is to change it by an amount proportional to $\bar{\delta}\chi$. One may express this in general by writing

$$[\chi, G_\chi] = \tfrac{1}{2} i f\ \bar{\delta}\chi \qquad (2.5)$$

where f is an unknown operator. Our final postulate[9a] is

VI. f is a c number.

The "wrong" choice for $\bar{\delta}\chi$ reduces G_χ to zero identically, and as we have seen, also fails to yield valid Lagrange equations. Hence, we may drop this empty possibility and retain only the "right" choice of $\bar{\delta}\chi$ in accordance with the symmetry character of A^μ. In this case, f is necessarily unity, due to the consistency requirement between the Lagrange and Heisenberg equations. This result arises from the following considerations. The effects of the generators G_t and G_χ are given by Eqs. (2.2b) and (2.5). These relations are not independent, but are subject to the important consistency requirement that the effect of G_t on G_χ agree with that of G_χ on G_t. We have on the one hand from the action of G_χ on χ that

$$-i[G_\chi, H[\chi]] = \int d^3r[T^{00}(\chi - \tfrac{1}{2} f\ \bar{\delta}\chi) - T^{00}(\chi)], \qquad (2.6a)$$

where T^{00} is the energy density. On the other hand, we may evaluate (2.6a) through the effect of G_t on χ:

$$-i[G_\chi, H] = -i \int d^3r[\chi A^0\ \bar{\delta}\chi, H] = \int d^3r A^0 \dot{\chi}\ \bar{\delta}\chi. \qquad (2.6b)$$

Here we have used the fact (shown in the Appendix) that $\bar{\delta}\chi$ commutes with H. Equations (2.6) express $A^0\dot{\chi}$ as a function of the χ's. However, $A^0\dot{\chi}$ is also specified through the Lagrange equations (2.1b) by use of which the right member of Eq. (2.6b) may be replaced by $\int d^3r[T^{00}(\chi - \tfrac{1}{2}\bar{\delta}\chi) - T^{00}(\chi)]$, and so $f = 1$ follows from VI. In the Appendix, the more general consistency requirements between an arbitrary G' and G_t are examined and found to allow $f = 1$.

The equal-time commutation relations among χ's are established from Eq. (2.5), which reads explicitly

$$\tfrac{1}{2} \int d^3r[\chi', \chi A^0\ \bar{\delta}\chi - \bar{\delta}\chi A^0 \chi] = \tfrac{1}{2} i\ \bar{\delta}\chi', \qquad \chi' \equiv \chi(\mathbf{r}', t). \qquad (2.7)$$

If A^0 is antisymmetric, we have seen that $A^0\delta\chi$ commutes with χ, so that Eq. (2.7) becomes

$$\int d^3r[\chi', \chi] A^0\ \bar{\delta}\chi = \tfrac{1}{2} i\ \bar{\delta}\chi' \qquad (2.8)$$

For nonsingular A^0, it then follows that

$$[\chi', \chi] = \tfrac{1}{2} i A_0^{-1}\ \delta^3(\mathbf{r} - \mathbf{r}') \qquad (2.9a)$$

while if A^0 is singular, one cannot deduce a simple commutation relation between χ and χ'. A singular A^0 implies the existence of constraints in the theory.[10] Aside from fields, such as the electromagnetic one, which possess a gauge group, this situation presents no difficulty: as we shall see for the explicit cases to be treated, the $A^0\chi$ turn out to be all the independent field variables and one

[9a] *Note added in proof.* This assumption is actually derivable from the previous postulates. See, "Note on Uniqueness of Canonical Commutation Relations," J. Math. Phys. (to be published).

[10] See for example, reference 4.

gets the complete set of commutators from Eq. (2.8):

$$[A^0\chi', A^0\chi] = \tfrac{1}{2}iA^0\,\delta^3(\mathbf{r}-\mathbf{r}') \qquad (2.9b)$$

The electromagnetic field must be treated separately (see end of Sec. IV below). For symmetric nonsingular A^0, the result corresponding to Eq. (2.9a) is

$$\{\chi', \chi\} = \tfrac{1}{2}iA_0^{-1}\,\delta^3(\mathbf{r}-\mathbf{r}'). \qquad (2.10)$$

The singular case does not arise for spin $\frac{1}{2}$. We have thus found that all fields obey commutations or anticommutation relations as a result of postulate $V(a)$, and that the equal-time relations are c numbers due to postulate VI.

It should be noted that the equal-time commutation relations (2.9, 2.10) have been obtained purely from the kinetic part of the Lagrangian. The symmetry character of B and the nature of $\mathfrak{H}_I$ were not involved. We also mention that for each field, a complete set of equal time commutation relations were found; thus, in the charged scalar case for example, we shall get not only $[\varphi', \varphi^+]$ and $[\varphi', \pi]$ but also $[\varphi', \varphi]$. It is really the last commutator which is the "statistics" part of the theorem, i.e., the one which allows or forbids more than one particle per state. Without a complete particle interpretation and a derivation of $[\varphi, \varphi']$-like relations, the theorem is not fully established.[11]

III. NEUTRAL SPIN $\frac{1}{2}$

We begin with the special case of the Majorana field with nonvanishing mass. The field equations read here

$$A^\mu\,\partial_\mu\chi + mB\chi + (\tfrac{1}{2})\,\partial\mathfrak{H}_I/\partial\chi = 0. \qquad (3.1)$$

The matrix B is necessarily nonsingular in order that a Dirac equation of the form (3.1) exist. The Dirac γ^μ are then formed from A^μ, B according to

$$\gamma^\mu = iB^{-1}A^\mu, \qquad \{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu} \qquad (3.2)$$

and the A^μ must also be nonsingular, since the γ^μ are nonsingular. In general, $A^\mu = a^\mu + s^\mu$ where s^μ and a^μ are, respectively, symmetric and antisymmetric. The variation $\delta\chi_\alpha$ is assumed either to commute or anticommute[12] with a given χ_β; hence the argument in Sec. II implies that in the former case only the terms with a^μ, and in the latter case only those with s^μ, will survive in the Lagrange equations and in the generator. We may therefore examine the cases $A^\mu = s^\mu$, a^μ separately. Treating first $A^\mu = s^\mu$, (so that $s^{\mu *} = -s^\mu$), we note that all χ_α anticommute by Eq. (2.10). Thus the $\chi B\chi$ term in $\mathfrak{L}$ will vanish unless B is antisymmetric, and so without loss of generality, we require $\tilde{B} = -B$, so that $B^* = -B$ as B is Hermitian. To establish the fact that this is the Majorana field, we show that one can build up the Dirac algebra from s^μ and an antisymmetric B. This is accomplished by the choice $B = \gamma^0$, $s^\mu = -i\gamma^0\gamma^\mu$ where γ^μ are the usual Dirac matrices in the Majorana representation. With this choice, Eq. (2.10) becomes the standard anticommutation relation for the Majorana field.

We show next that the opposite symmetry assumption, i.e., $A^\mu = a^\mu$ and consequently $\tilde{B} = B$, $B^* = B$, which implies Bose relations by Eq. (2.9a), is not possible. In particular we now show that Eq. (3.2) cannot be satisfied. From the assumed symmetry properties of a^μ, B, and the Majorana γ^μ ($\tilde{\gamma}^0 = -\gamma^0$, $\tilde{\gamma}^i = \gamma^i$) one finds

$$[B, \gamma^0] = 0 = \{B, \gamma^i\}. \qquad (3.3)$$

Since the neutral spin $\frac{1}{2}$ field is a 4×4 realization of the Dirac algebra, B must be constructed from the 16 Dirac matrices; Eqs. (3.3) require $B = \eta\gamma^0$ (η a number). However, then $\tilde{B} = -B$, which contradicts the assumed symmetry of B.

The massless case needs separate treatment, since there is then no B in $\mathfrak{L}$. One is therefore free to investigate the possibility of adjoining any matrix b to the A^μ such that $\gamma^\mu = ib^{-1}A^\mu$. Again, the assumption that $A^\mu = s^\mu$ (the normal case) clearly leads to the correct connection, as in the $m \neq 0$ case, with $b = \gamma^0$ (the choice $\tilde{b} = b$ is impossible here). The other possibility is $A^\mu = a^\mu$, which leads to Bose quantization, for arbitrary $b = b^s + b^a$ ($\tilde{b}^s = b^s$, $\tilde{b}^a = -b^a$). The symmetry properties of γ^μ then imply that

$$[b^s, \gamma^0] + \{b^a, \gamma^0\} = 0 \qquad (3.6a)$$

$$\{b^s, \gamma^i\} + [b^a, \gamma^i] = 0 \qquad (3.6b)$$

from which it follows that

$$b^s = 0, \qquad b^a = i\eta\gamma^0\gamma^5 \qquad (3.7a)$$

$$a^\mu = \eta\gamma^5\gamma^0\gamma^\mu \qquad (3.7b)$$

where η is a real number and $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$. The commutation relations (2.9a) now read

$$[\chi'_\alpha, \chi_\beta] = -(i/2\eta)\gamma^5_{\alpha\beta}\,\delta^3(\mathbf{r}-\mathbf{r}') \qquad (3.8)$$

[11] G. Feinberg has pointed out that the Burgoyne derivation does not establish the connection for the $[\phi, \phi']$ relations in the charged case. These relations have recently been established within the framework of reference 3, however, by G. F. Dell' Antonio and by A. S. Wightman (private communication to G. Feinberg). Dell' Antonio's derivation appears in Ann. Phys. **16**, 153, (1961).

[12] By Lorentz covariance, all components $\delta\chi_\alpha$ of, say, a spinor will either commute with a given χ_β or all will anticommute.

while the action and Hamiltonian are given by

$$I = \int \mathfrak{L}\, d^4x = \int [\chi\eta\gamma^5\gamma^0\gamma^\mu\, \partial_\mu\chi - \mathfrak{IC}_I]\, d^4x \qquad (3.9a)$$

$$H = \int [-\eta\chi\gamma^5\gamma^0\gamma^i\, \partial_i\chi + \mathfrak{IC}_I]\, d^3r. \qquad (3.9b)$$

The field with the properties (3.8), (3.9) has no manifest inconsistencies in that the Lagrange and Heisenberg equations agree. However, it possesses a number of strange features. First, the free Hamiltonian is not a positive-definite operator, and states with opposite helicity have opposite signs of energy. Second, the theory is invariant under neither P nor TCP. The Hermitian nature of the field means that C invariance holds trivially and so the P nonconservation cannot be compensated by C. (By contrast, the usual massless Majorana field with $I = \int \chi\gamma^0\gamma^\mu\, 1/i\, \partial_\mu\chi\, d^\mu x$ conserves both P and TCP, of course.)[13]

To eliminate this case of wrong connection, we may therefore invoke the TCP requirement. Though Pauli[1] did not originally consider such a field, one would have to make the same demand within his framework to avoid it. (Alternately, for free fields, the vacuum state condition would also be sufficient.)[14]

IV. NEUTRAL SPIN 0, 1 FIELDS

We consider next neutral integral spin fields, with or without mass. The zero spin Lagrangian in first-order form is

$$\mathfrak{L} = \tfrac{1}{4}\{\phi, \partial_\mu\phi^\mu\} - \tfrac{1}{4}\{\partial_\mu\phi, \phi^\mu\} - \tfrac{1}{2}(\mu^2\phi^2 - \phi_\mu\phi^\mu) - \mathfrak{IC}_I, \qquad (4.1)$$

which clearly gives the usual field equations upon independent variation of ϕ and ϕ_μ (the anticommutators in $\mathfrak{L}$ are needed to preserve its Hermitian character).[15] The Kemmer form (2.1a) of $\mathfrak{L}$ is obtained by introducing the vector $\chi \equiv (\phi, \phi_\mu)$, so that A^μ and B are 5×5 matrices. In particular

$$A^0 = \tfrac{1}{2}\begin{pmatrix}\alpha & 0\\ 0 & 0\end{pmatrix} = -\tilde{A}^0, \qquad \alpha \equiv \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \qquad (4.2)$$

and B is diagonal, with elements $\frac{1}{2}(\mu^2, 1, -1, -1, -1)$.

[13] The theory (3.9) is invariant under chirality transformations, $\chi \to \gamma^5\chi$ just as is the normal massless Majorana case, however.

[14] The appearance of Bose commutation relations may also be understood in terms of derivations which assume invariance under TCP. In (3.9) the TCP operation reverses the sign of the free part of $\mathfrak{L}$ (instead of leaving it invariant), and the connection is reversed.

[15] The other possible Hermitian form, $i[\phi, \partial_\mu\phi^\mu]$, is easily seen to lead either to no contribution to the dynamics when $\delta\phi_\mu$ commutes or to the inconsistency $\partial_\mu\phi = 0$ when $\delta\phi_\mu$ anticommutes.

For the neutral spin one case, the Lagrangian reads

$$\mathfrak{L} = \tfrac{1}{4}\{\phi_\nu, \partial_\mu G^{\mu\nu}\} - \tfrac{1}{4}\{\partial_\mu\phi_\nu, G^{\mu\nu}\} - \tfrac{1}{2}(\mu^2\phi_\mu\phi^\mu - \tfrac{1}{2}G_{\mu\nu}G^{\mu\nu}) - \mathfrak{IC}_I \qquad (4.3)$$

where $G^{\mu\nu} = -G^{\nu\mu}$ and ϕ_μ are to be varied independently. Again, in terms of $\chi \equiv (\phi_1, G_{01}, \phi_2, G_{02}, \phi_3, G_{03}, \phi_0, G_{ij})$, Eq. (4.3) has the form (2.1a) with $\tilde{A}^\mu = -A^\mu$, $\tilde{B} = B$, the 10×10 matrix A^0 being

$$A^0 = \tfrac{1}{2}\begin{pmatrix}\alpha & & & & \\ & \alpha & & 0 & \\ & & \alpha & & \\ & 0 & & 0 & \\ & & & & 0\end{pmatrix}$$

The matrix B is diagonal, with elements $\frac{1}{2}(\mu^2, 1, \mu^2, 1, \mu^2, 1, \mu^2, -1, -1, -1)$. Unlike the Majorana case, we have here obtained a particular matrix representation for A^μ, B directly from the known form of the integral spin Lagrangians (4.1, 3). Any other representation is reached by a linear transformation on the Kemmer column symbol $\chi = S\chi'$. This replaces A^μ and B by $A^{\mu\prime} = \tilde{S}A^\mu S$, $B' = \tilde{S}BS$ thereby leaving the symmetry properties unaltered. The general results of Sec. II show directly that only *commutation* relations can occur. In particular, the singularity of A^0 only allows one to write the form (2.9b). For example, in spin 0, the last three components ϕ_i of χ do not enter in $A^0\chi$, being in fact determined from the constraint equations. The latter are defined in general to be those equations which are independent of time derivatives, and read in this case

$$\phi_i = \partial_i\phi - \partial H_I/\partial\phi_i \qquad (4.4)$$

Thus Eqs. (2.9b) represent the usual set of commutation relations between ϕ and $\phi_0 = \partial\mathfrak{L}/\partial\dot{\phi}$. For spin 1, the four quantities ϕ_0, G_{ij} are missing from $A^0\chi$, the corresponding constraints being

$$\partial_i G_{0i} + \mu^2\phi_0 + \partial H_I/\partial\phi_0 = 0 \qquad (4.5a)$$

$$G_{ij} = \partial_i\phi_j - \partial_j\phi_i + \partial H_I/\partial G_{ij}. \qquad (4.5b)$$

If $\mu \neq 0$, Eqs. (4.5) may be solved for ϕ_0 and G_{ij}, again showing that Eqs. (2.9b) are the usual commutation relations between ϕ_i and G_{0i}. For the electromagnetic case, ϕ_0 no longer appears in Eq. (4.5a).[16] Instead, Eq. (4.5a) determines

[16] This is obvious in the first-order formulation of charged fields, whose $\mathfrak{L}$, being linear in the derivatives, is therefore also linear in $(\partial_\mu - ieA_\mu)$. Hence $j^0 \equiv -\partial\mathfrak{IC}_I/\partial A_0$ is indeed independent of A_0. [Elimination of G_{ij} in j^0 by Eq. (4.5b) similarly cannot introduce any A_0 dependence since $\partial\mathfrak{IC}_I/\partial G_{ij}$ is independent of A_0.]

$\partial_i G_{0i}$ ($\equiv \nabla\cdot\mathbf{E}$) in terms of the other variables of the system. Thus, the longitudinal part of $\mathbf{E}$ is eliminated and the Bose quantization then follows in terms of the two independent transverse degrees of freedom of the photon.[17,18]

To summarize, the usual Bose quantization is valid for (massed or massless) neutral integral spin fields, while the nonoccurrence of symmetric A^μ forbids Fermi quantization.

V. CHARGED INTEGRAL SPIN

A charged field may be built up from two Hermitian fields by means of a 2×2 charge space. One simply defines $\chi \equiv (\chi_1, \chi_2)$, where $\chi_{1,2}$ are two independent Hermitian fields of the type considered in Sec. IV. Correspondingly, the dimensionality of A^μ, B is doubled in one of two possible ways. Thus, if a^μ, b are the 5×5 or 10×10 matrices of Sec. IV, then A^μ, B of the charged system are[19]

$$A^\mu = \begin{bmatrix} a^\mu & 0 \\ 0 & a^\mu \end{bmatrix}, \qquad B = \begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix} \tag{5.1}$$

or

$$A^\mu = \begin{bmatrix} 0 & ia^\mu \\ -ia^\mu & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & ib \\ -ib & 0 \end{bmatrix}. \tag{5.2}$$

The significance of the two possibilities (5.1, 2) can easily be understood in terms of the usual charged fields $\phi \equiv (\chi_1 - i\chi_1)2^{-1/2}$ and ϕ^+. Thus, for spin zero, choice (5.1) gives (to within a divergence), for $\chi A^\mu \partial_\mu \chi - \partial_\mu \chi A^\mu \chi$, the form

$$\{\phi_\mu, \partial_\mu \phi^+\} + \{\phi_\mu^+, \partial_\mu \phi\} \tag{5.3a}$$

while choice (5.2) gives[20]

$$[\partial_\mu \phi^+, \phi_\mu] + [\phi_\mu^+, \partial_\mu \phi]. \tag{5.3b}$$

Any Hermitian form in terms of the ϕ fields may always be written as a linear combination of an anticommutator (5.3a) and a commutator (5.3b), and so the derivation of the connection for such a form is automatically covered in the cases (5.1) and (5.2).

The choice (5.1) leads to the correct Bose quantization only, since it preserves the antisymmetry of A^μ, which is all that was required in Sec. IV. The generator G_χ of the system with choice (5.1) is simply the sum $G_{\chi_1} + G_{\chi_2}$. It leads necessarily, by the techniques of the previous sections, to the standard Bose relation of each of the χ's with itself. No commutation relations between χ_1, and χ_2 can be deduced, however. If one wishes to interpret the system as a single charged field, then χ_1 and χ_2 are coupled through the electromagnetic interaction term. The current is, in fact, proportional to $\chi_2 a^\mu \chi_1 - \chi_1 a^\mu \chi_2$ by the usual gauge arguments. The choice $[\chi_1, \chi_2] = 0$ (which follows from the choice $]\delta\chi_1, \chi_2] = 0$) leads to the standard charged boson theory. The alternative, $\{\chi_1', \chi_2\} = 0$ makes the current vanish identically (since $\tilde{a}^\mu = -a^\mu$) and so it does not give rise to an electromagnetic interaction. While no inconsistency arises with this choice, the charge interpretation cannot be made; one has two electrically neutral fields, which may perhaps interact in other ways according to the structure of $\mathcal{H}_I$. (Conceivably, the choice $\{\chi_1', \chi_2\} = 0$ may be required for a particular $\mathcal{H}_I$ not to vanish.)[21] This would not represent a breakdown of the spin-statistics connection, since the two fields cannot then be combined into a single anticommuting one possessing a particle interpretation.

We now show that the extension to charge space according to Eq. (5.2), which can only lead to Fermi quantization since the product space A^μ is

[17] For the first-order form of electrodynamics in the radiation gauge, see for example J. Schwinger, Phys. Rev. **115**, 721 (1959).

[18] The spin-2, zero-mass field may also be put into first-order form in terms of two transverse degrees of freedom [R. Arnowitt and S. Deser, Phys. Rev. **113**, 745 (1959)] so that here too Bose quantization may be deduced directly.

[19] That the extension to charge space can always be put into form (5.1) or (5.2) may be seen as follows. The independent 2×2 matrices are I, σ_i. The matrix I is symmetric and is the choice (5.1), while σ_2 is antisymmetric and yields (5.2). The other two matrices σ_1, σ_3 are symmetric; by a linear transformation in the 2-dimensional charge space, σ_1 can be reduced to σ_3 and hence does not yield an independent representation. Further, by means of a linear transformation in the original vector spaces of the Hermitian χ_1 and χ_2, one may show that the σ_3 representation is equivalent to that generated by I, that is, there exists a transformation T such that $\tilde{T}A^0T = -A^0$. (T simply interchanges canonical coordinates and momenta, e.g., sends $\phi \to \phi^0$, $\phi^0 \to -\phi$ for the scalar field.) Hence in the product space

$$\begin{bmatrix} 1 & 0 \\ 0 & \tilde{T} \end{bmatrix} (\sigma_3 \otimes A^0) \begin{bmatrix} 1 & 0 \\ 0 & T \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & \tilde{T} \end{bmatrix} \begin{bmatrix} A^0 & 0 \\ 0 & -A^0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & T \end{bmatrix} = \begin{bmatrix} A^0 & 0 \\ 0 & A^0 \end{bmatrix} = I \otimes A^0$$

and σ_3 has been made equivalent to I. Finally, we note that unless the same charge matrix is used for both A^μ and B, the resulting charged fields, ϕ, ϕ^+, will not obey the appropriate field equation for the spin in question.

[20] Note the difference between the form (5.3b) for the charged field and that given in footnote 15 for a neutral field. Form (5.3b) clearly contributes only for anticommuting variations. The existence of the antisymmetric matrix B of (5.2) prevents the contradiction $\partial_\mu \phi = 0$ of the neutral case.

[21] A simple, if artificial, example is provided by $\mathcal{H}_I \sim \chi_2 s \chi_1 - \chi_1 s \chi_2$, $s = s$. It should be noted, however, that in general, one need not necessarily specify the relation between χ_1 and χ_2 if $\mathcal{H}_I$ does not require such a specification. For a discussion of this question, see for example G. Luders, Z. Naturforsch, **13a**, 254 (1958).

symmetric, is inconsistent. The generator here reads

$$G_\chi = \tfrac{1}{2}\int d^3r(\chi A^0\,\bar\delta\chi - \bar\delta\chi A^0\chi)$$
$$= \frac{i}{2}\int d^3r(\chi_1 a^0\,\bar\delta\chi_2 - \chi_2 a^0\,\bar\delta\chi_1 - \bar\delta\chi_1 a^0\chi_2 + \bar\delta\chi_2 a^0\chi_1). \quad (5.4)$$

Since $\tilde a^0 = -a^0$, either $\{\bar\delta\chi_{1,2}, \chi_{2,1}\} = 0$ holds or else G_χ vanishes identically. Alternately, these anticommutation relations are necessary in order to obtain any field equations from $\mathfrak{L}$. Applying Eq. (2.3) with $\chi = \chi_1$ and $\bar\delta\chi_1 = 0$ one finds

$$\{(a^0\chi_1')_\alpha, (a^0\chi_1)_\beta\} = 0 \quad (5.5)$$

and similarly for the independent variables $a^0\chi_2$ of the second field. This means, however, that the independent components of $\chi_{1,2}$, i.e., $(a^0\chi_{1,2})_\alpha$ square to zero, and being Hermitian, necessarily vanish. Thus, the usual Bose quantization of charged integral fields is alone permitted.

VI. CHARGED SPIN $\frac{1}{2}$

In this section, we apply our methods to the charged spin $\frac{1}{2}$ field and show that, just as in the massless neutral spin $\frac{1}{2}$ case, Bose quantization cannot be forbidden without further assumptions. In building up the charged field from two Majorana systems, there are the same two possibilities as in the integral spin case, namely, the usual direct product $I \otimes A^\mu$ (with mass term $mI \otimes B$) and the representation $\sigma_2 \otimes A^\mu$ (with mass term $m\sigma_2 \otimes B$). We begin with the case of nonvanishing mass, and establish first that A^μ must be symmetric and B antisymmetric, just as in the neutral case. This follows from the requirement that the equation for the charged field $\psi \equiv \chi_1 - i\chi_2$ have as its free particle term the Dirac form $(-i\gamma^\mu\,\partial_\mu + m)\psi$. Since, as can easily be checked, the $\chi_{1,2}$ have free particle parts $(A^\mu\,\partial_\mu + mB)\chi_{1,2}$ in either charge representation, the relation $\gamma^\mu = iB^{-1}A^\mu$ still holds. The symmetry properties $A^\mu = s^\mu$, $B = B^a$ (s^μ and B nonsingular) then follow from the Dirac algebra, as in the neutral case.

It is clear now that the choice $I \otimes s^\mu$ leads only to the usual correct Fermi quantization of the charged field (just as (5.1) did in the normal Bose case). The remaining possibility, $\sigma_2 \otimes s^\mu$, however, leads only to Bose statistics, since $\sigma_2 \otimes s^\mu$ is antisymmetric. In contrast to its analog (5.2), however, no algebraic inconsistencies arise from the "wrong" statistics but *TCP* is violated. This may seem surprising in view of the derivation in Sec. III for the massed Majorana case, where it was shown that the Dirac matrices γ^μ could not be built up from antisymmetric A^μ. That proof, however, depended on the fact that there are only sixteen 4×4 matrices available for a^μ, while the $\sigma_2 \otimes s^\mu$ space is now 8×8. The generator G_χ is here

$$G_\chi = i/2\int d^3r[\chi_1 s^0\,\bar\delta\chi_2 + \bar\delta\chi_2 s^0\chi_1 - \chi_2 s^0\,\bar\delta\chi_1 - \bar\delta\chi_1 s^0\chi_2] \quad (6.1)$$

so that the symmetry of s^0 forces $[\bar\delta\chi_{1,2}, \chi_{1,2}] = 0$ to prevent the vanishing of G_χ and of the Lagrange equations. Applying Eq. (2.3) with $\chi = \chi_1$ and $\bar\delta\chi_1 = 0$ we find $[\chi_1', \chi_1] = 0$ and similarly for χ_2. Next, taking $\chi = \chi_1$, $\bar\delta\chi_2 = 0$ in Eq. (2.3), we find

$$\int d^3r'[\chi_1, \chi_2' s^0\,\bar\delta\chi_1'] = \tfrac{1}{2}\,\bar\delta\chi_1. \quad (6.2)$$

In order to move $\bar\delta\chi_1$ to the same side in each term of the commutator in (6.2), we need the commutation relation of $\bar\delta\chi_1$ with χ_1 itself. The first possibility, $\{\bar\delta\chi_1, \chi_1\} = 0$, leads to an anticommutation relation between χ_1 and χ_2. However, at least the free part of the Hamiltonian,

$$H_0 = -i\int d^3r[\chi_1 s^i\,\partial_i\chi_2 - \chi_2 s^i\,\partial_i\chi_1] + im\int d^3r(\chi_1 b\chi_2 - \chi_2 b\chi_1), \quad (6.3)$$

then vanishes (to within a c number), making the Heisenberg equations inconsistent with the Lagrange equations. The other possibility, $[\bar\delta\chi_1, \chi_1'] = 0$ leads to

$$[\chi_1, \chi_2'] = -i/2\;\delta^3(\mathbf{r} - \mathbf{r}') \quad (6.4)$$

where we have chosen the representation $S^0 = iI$, since $\tilde s^0 = s^0$, $s^{0*} = -s^0$. In terms of the charged field, $\psi = \chi_1 - i\chi_2$, $\psi^+ = \chi_1 + i\chi_2$, we find

$$[\psi, \psi'] = 0 = [\psi^+, \psi^{+\prime}],$$
$$[\psi^+, \psi'] = \delta^3(\mathbf{r} - \mathbf{r}'). \quad (6.5)$$

Here $\gamma^\mu = iB^{-1}s^\mu$ and $B^{-1} = \gamma^0$, which is a possible realization of the γ^μ as discussed in Sec. III. Further, the Heisenberg equations based on Eqs. (6.5) are consistent with the Lagrange equations.

The preceding discussion has thus led to Bose quantization of *this* representation of the charged spin $\frac{1}{2}$ field, free of any purely algebraic inconsistencies.[22] In the free case it is clearly sufficient

[22] It should be mentioned here that the Weyl 2-component neutrino theory may be quantized with Fermi or Bose statistics since it may be viewed as a special case of the Dirac neutrino obtained by projecting with $(1 + i\gamma^5)$. Alternately, one may see this from the fact that one may choose the Weyl Lagrangian in the form $[\psi^+, \sigma^\mu\partial_\mu\psi]$ or $i\{\psi^+, \sigma^\mu\partial_\mu\psi\}$. These two symmetrizations correspond to use of I or σ_2 in generating the charge space.

to demand positive energy, as Pauli[1] was led to do. More generally, however, this case may be eliminated in the interacting situation by invoking *TCP* invariance. For, it is easily seen that the free particle part of the Lagrangian is not invariant under *TCP*, since it is invariant under T and P but not under C. Also, as is well known, the current operator has a positive-definite charge density when Bose quantization is used, whereas it must change sign under *TCP* if *TCP* invariance is to hold.

Finally, we investigate the massless charged cases, which are obtained from the massless Majorana examples by use of I or σ_2. If A^μ is symmetric (so that an antisymmetric b must be used to form $\gamma^\mu = ib^{-1}A^\mu$) we have merely the $m = 0$ limit of the massed cases treated above and the same conclusions apply. If, however, we consider A^μ antisymmetric [and hence b antisymmetric by Eq. (3.7)] we get the same sort of situation as in the non-*TCP* invariant neutral massless field, so that we are here building up the massless charged field from such massless neutral ones. Of the two possible charge representations, the σ_2 choice leads to a null theory: In this case (which is *TCP* invariant) the combined symmetry of $\sigma_2 \otimes \gamma^5\gamma^0\gamma^\mu$ implies Fermi quantization. We obtain then, the anticommutation relations $\{\psi'_\alpha, \psi^+_\beta\} = i/2\gamma^5_{\alpha\beta}\,\delta^3(\mathbf{r} - \mathbf{r}')$ for $\psi \equiv \chi_1 - i\chi_2$, whose spin trace implies that $\chi_1^2 + \chi_2^2 = 0$ and so that $\chi_1 = 0 = \chi_2$. On the other hand, the direct product $I \otimes \gamma^5\gamma^0\gamma^\mu$ clearly behaves just like the neutral case, namely we have $\mathfrak{L} = \frac{1}{2}\{\psi^+, \gamma^5\gamma^0\gamma^\mu\partial\psi\}$, $[\psi', \psi^+] = -i/2\gamma^5\,\delta^3(\mathbf{r} - \mathbf{r}')$. Both *TCP* and positive-definiteness of the free Hamiltonian are violated, Bose quantization is not inconsistent, and so again the requirement of *TCP* invariance may be used to exclude this final case.

VII. COMMUTATION RELATIONS BETWEEN INDEPENDENT FIELDS

We discuss briefly the commutation relations among different Hermitian fields. For kinematically independent fields, i.e., systems which can be characterized by

$$\mathfrak{L} = \tfrac{1}{2}\sum_i \{\chi_{(i)}A^\mu_{(i)}\,\partial_\mu\chi_{(i)} - \partial_\mu\chi_{(i)}A^\mu_{(i)}\chi_{(i)}\} - \mathfrak{IC},$$

the generator G_χ is also a sum of independent terms. Hence Eq. (2.5) yields no information about the relations between $\chi_{(i)}$ and $\chi_{(\gamma)}$ $(i \neq j)$; that is, from

$$\int d^3r[\chi'_i, \chi_j A^0_j\,\bar{\delta}\chi_j] = \bar{\delta}_i\chi'_i = 0, \qquad (i \neq j) \quad (7.1)$$

alone, one can obtain either $[\chi_i, \chi_j] = 0$ or $\{\chi_i, \chi'_i\} = 0$ depending on the choice of relations between $\bar{\delta}\chi_i$ and χ_j. The mechanism which imposes further restrictions in this formulation is the consistency requirement between G_χ and $\mathfrak{IC}$, i.e., between Heisenberg and Lagrange equations. For example, consider two uncoupled Bose hermitian fields χ_1, χ_2 so that $\mathfrak{IC} = \mathfrak{IC}_{01}(\chi_1) + \mathfrak{IC}_{02}(\chi_2) + \chi_2^3$. The Lagrange equations for χ_1 require no knowledge of the $\bar{\delta}\chi_1$, χ_2 relations. On the other hand, the Heisenberg equations, $\dot{\chi}_1 = i[H, \chi_1]$ involve $[\chi_1, \chi_2^3]$ so that χ_1 must commute with χ_2 for consistency. For general $\mathfrak{IC}$, the consistency requirement leads to commutation conditions identical to those previously given by Luders.[23]

For some interactions, the above consistency conditions do not restrict the relations between kinematically independent fields. However, making one or another choice (when either is allowed) can alter the physical interpretation of the theory. An example was given in the charged integral spin case, where it was seen that $\{\chi_1, \chi'_2\} = 0$ implied the vanishing of the current operator, and so that $\chi_{1,2}$ were two neutral Bose fields (possibly interacting), rather than the components of a single charged field $\phi = 2^{-1/2}(\chi_1 - i\chi_2)$. In this connection, one might note that Burgoyne's proof[3] led automatically to $[\phi', \phi^+] = 0$, $(\mathbf{r} \neq \mathbf{r}')$, whereas in our analysis this result holds only if the choice $[\chi_1, \chi'_2] = 0$ is taken. The difference lies in Burgoyne's interpretation of his assumption that all quantities either commute or anticommute: he treats the charged fields ϕ, ϕ^+ (rather than χ_1 and χ_2) as the entities which obey only one of the two possibilities (as is indeed the characteristic of the correct charged field). In our analysis, the relations between χ_1 and χ_2 may, *a priori*, differ from those of each field with itself, which makes the noncharged possibility also available.

VIII. CONCLUSIONS

In the present derivation of the spin-statistics connection, emphasis was put on the separation of the usual *TCP* invariance and vacuum state $(E \geq 0)$ requirements from the conventional assumptions of local field theory. It was found that for all spin 0 and 1 fields and the neutral massed spin $\frac{1}{2}$ fields, the correct connection could be established purely from the algebraic form of the free particle part

[23] Luders[21] started from the requirement of local Heisenberg equations. We demand consistency between these and the (by assumption local) Lagrange equations, so that the two approaches are essentially the same. Within the framework of the axiomatic method, H. Araki [J. Math. Phys. **2**, 267 (1961)] has obtained the independent field relations.

of the Lagrangian, without recourse to the *TCP* or $E \geq 0$ requirements. Actually for these fields, it is the case that the *free parts* do satisfy *TCP* and $E \geq 0$. For the remaining spin $\frac{1}{2}$ cases, it was found that both connections occurred (as was the case in Pauli's original free field derivation). However, those cases for which the wrong statistics held were characterized by lack both of *TCP* invariance and of $E \geq 0$ in the free particle part of the Lagrangian. Hence these cases could be eliminated either by requiring *TCP* invariance or $E \geq 0$ *solely* for the free particle parts.

It is curious that the energy requirement is to be imposed on the free particle part of the energy rather than on the more physically meaningful total energy. This might appear more understandable if, as has been suggested, the sign of the total energy is always the sign of the kinetic energy. The significance of the alternative requirement, *TCP* invariance, seems somewhat more puzzling from the present approach. Invariance of a local field under *TCP* is a consequence of both proper Lorentz invariance and the *assumption* that the correct connection holds.[24] Consequently, *a priori* acceptance of *TCP* invariance is not so straightforward from the present point of view, and may be regarded as an empirical question. On the other hand, the fact that one need impose *TCP* only on the *free* particle part of the Lagrangian is reasonable. For, as we have seen, the latter requirement yields the correct connection which then implies *TCP* invariance for the total Lagrangian.[24]

ACKNOWLEDGMENTS

The authors are indebted to J. Schwinger for pointing out the necessity of dealing with both possible representations of charged fields. Some of the results derived here had been obtained previously by him on a somewhat different basis (unpublished lecture notes). We also wish to thank G. Feinberg and S. S. Schweber for stimulating discussions.

APPENDIX

A discussion of the postulates needed in the Schwinger action principle[4] is given here. We begin with some definitions and notations concerning unitary transformations on the basis vectors $|a'\rangle$ of a complete set of operators $\{A_i\}$ $(A_i \mid a'\rangle = a_i' \mid a'\rangle)$ in Hilbert space. Let an infinitesimal unitary transformation be $\mathfrak{U} \equiv 1 + iG$ where $G^+ = G$. If we denote the transformed ket by $\mid \bar{a}'\rangle \equiv \mathfrak{U}^{-1} \mid a'\rangle$, then the change in the ket due to the unitary transformation, $\delta \mid a'\rangle \equiv \mid \bar{a}'\rangle - \mid a'\rangle$, is given by $\delta \mid a'\rangle = -iG \mid a'\rangle$. The matrix elements of any operator B then change according to

$$\delta\langle a' \mid B \mid a''\rangle \equiv \langle \bar{a}' \mid B \mid \bar{a}''\rangle - \langle a' \mid B \mid a''\rangle.$$

If we define the operator $\delta_0 B$ by $\delta\langle a' \mid B \mid a''\rangle \equiv \langle a' \mid \delta_0 B \mid a''\rangle$, one has $\delta_0 B = -i[B, G]$. Thus the change of the matrix elements of an operator due to a change of basis can equivalently be represented by a change of the operators in the old basis.

If $\{A_i\}$ represents the complete set for the original basis $\mid a'\rangle$, then the complete set, $\{\bar{A}_i\}$ for the transformed basis $(\bar{A}_i \mid \bar{a}'\rangle = a_i' \mid \bar{a}'\rangle)$ is related to A by $\bar{A} = \mathfrak{U}^{-1} A \mathfrak{U} = A - \delta_0 A$. In accordance with the conventional physical interpretation of Hilbert space, one must associate, at any time t, a complete set of Hermitian operators $\{A_i(t)\}$ to a complete set of compatible observables. The simultaneous eigenkets of $A_i(t)$, i.e. $\mid a't\rangle$, form a basis which moves in time.[6] According to the probability interpretation, the bases at different times must be related by a unitary transformation. (The assumption of a positive-definite Hilbert-space metric is used here. This assumption is also used explicitly in some of the derivations given in text.) For the transformation representing an infinitesimal time translation (denoted by $G = G_t$) one has then that $\bar{A} = A(t + \delta t) = A(t) + \dot{A}\, \delta t$. Hence,

$$\dot{A}(t)\, \delta t = i[A(t), G_t(t)] \qquad \text{(A1a)}$$

since $\delta_0 A = -i[A, G_t]$ for this case. Equation (A1a) is the Heisenberg equation of motion. The corresponding basis vector equations of motion read: $\delta t(d \mid a't\rangle \mid dt) = -iG_t(t) \mid a't\rangle$.

The general variation of the transformation function to be considered here, $\delta\langle a_1 t_1 \mid a_2 t_2\rangle$, consists of changes of the bases due to their time motion plus variations due to changes of the complete set at a fixed time. Both variations are generated by unitary transformations so that

$$\delta\langle a_1 t_1 \mid a_2 t_2\rangle = i\langle a_1 t_1 \mid G_1(t_1) - G_2(t_2) \mid a_2 t_2\rangle. \qquad \text{(A2)}$$

Here $G(t)$ consists of two parts: one, the time translation generator G_t moving the system in time (keeping the same compatible set of measurables but at the displaced time), and a second part, G', generating the changes of bases possible at a fixed time (where the measurables are changed but the time is fixed). The change of the complete set generated by G' (i.e., $A \to \bar{A} = A - \delta_0 A$) is clearly given by

[24] See for example, G. Lüders, Ann. Phys. **2**, 1 (1957).

$$\delta_0 A = -i[A, G']. \tag{A1b}$$

For a field system, any complete set, $\{A_i\}$, and hence the function G relating two complete sets, must depend only on the field variables χ_a. We now invoke the condition that we are dealing with a local field theory. This implies that complete sets $A(t)$, at time t, can be constructed from the field operators $\chi_a(t)$ at that t [i.e. $G'(t)$ depends only on $\chi_a(t)$]. Similarly, the future behavior of the kets and operators must be determined by the field variables at time t, i.e., $G_t(t)$ depends only on $\chi_a(t)$ (so that the dynamical laws be local in time). In general then, $G(t) = G[\chi(t)]$ is a local function, in time, of $\chi_a(t)$.

We now define the Hermitian operator δW_{12} by the equation

$$\delta\langle a_1 t_1 \mid a_2 t_2\rangle \equiv i\langle a_1 t_1 \mid \delta W_{12} \mid a_2 t_2\rangle \tag{A3}$$

At this stage, δW_{12} depends only on variations at the end point times t_1 and t_2, according to Eq. (A2). However, we may divide the time interval (t_1, t_2) into many subintervals; the transformation function can then be represented by products of functions between the subintervals:

$$\langle a_1 t_1 \mid a_2 t_2\rangle = \sum \langle a_1 t_1 \mid a_3 t_3\rangle \times \langle a_3 t_3 \mid a_4 t_4\rangle \cdots \langle a_n t_n \mid a_2 t_2\rangle. \tag{A4}$$

In varying $\langle a_1 t_1 \mid a_2 t_2\rangle$ as expressed by the right-hand side of Eq. (A4), we can clearly make *arbitrary* unitary transformations on the bras and kets at the intermediate times, since such effects cancel out in the sum. In particular, one may consider interior variations in conflict with the actual time development. (For example, these intermediate variations may be generated by a G_t not proportional to the correct Hamiltonian.) On the other hand, the variation of Eq. (A4) leads to a sum of terms of the form

$$\sum_{a_m, a_{m+1}} \langle a_1 t_1 \mid a_m t_m\rangle \times \delta(\langle a_m t_m \mid a_{m+1} t_{m+1}\rangle)\langle a_{m+1} t_{m+1} \mid a_2 t_2\rangle = i\langle a_1 t_1 \mid \delta W_{m,m+1} \mid a_2 t_2\rangle \tag{A5}$$

according to the definition (A3). This shows that δW_{12} may also be viewed as a sum of terms involving variations at the intermediate times, or, in the limit as the subintervals become infinitesimal in size, δW_{12} becomes a time integral between t_2 and t_1.

We now make the basic postulate that δW_{12} is the variation of a finite operator W_{12}, i.e., that the variations of the transformation function that we are considering are to be obtained by making appropriate variations[25] of W_{12}. (The nature of these variations will be found below). Since W_{12} is a time integral, we may write

$$W_{12} = \int_{t_2}^{t_1} d^4x\, \mathfrak{L}(x), \tag{A6}$$

where $\mathfrak{L}$ must be a Hermitian operator. Comparing Eq. (A2) with Eq. (A3) gives

$$\delta \int_{t_2}^{t_1} d^4x\, \mathfrak{L}(x) = G_1(t_1) - G_2(t_2). \tag{A7}$$

In Eq. (A7), variations of W_{12} can be made in the interior as well as at the end points, since arbitrary unitary transformations are allowed at interior time when varying the right-hand side of Eq. (A4) [as expressed in Eq. (A5)]. As $G(t)$ is a local function of $\chi_a(t)$ in time, Eq. (A7) represents a quantum Hamilton's principle for obtaining Lagrange equations of motion (this will be shown below).

We now make the further postulate that $\mathfrak{L}(x)$ has the form

$$\mathfrak{L}(x) = \tfrac{1}{2}(\chi A^\mu \partial_\mu \chi - \partial_\mu \chi A^\mu \chi) - \mathfrak{H}(\chi) + \partial_\mu W^\mu(\chi), \tag{A8}$$

where χ is a column symbol whose components, $\chi_a(x)$, are Hermitian field operators and A^μ are constant square matrices in the χ space. Hermiticity of $\mathfrak{L}$ is obtained by requiring that $A^{\mu+} = -A^\mu$, $\mathfrak{H}^+ = \mathfrak{H}$, $W^{\mu+} = W^\mu$. In order to have Lorentz invariance, we require that $\mathfrak{H}$ be a scalar and W^μ a four-vector. Note that for invariance under the inhomogeneous Lorentz group to hold, neither $\mathfrak{H}$ nor W^μ can depend explicitly on x^μ. The assumption of the form (A8) for $\mathfrak{L}(x)$ stems from the fact that the equations of motion for any Lorentz covariant field system involving fields of definite spin in local interaction may be obtained by varying a Lagrangian of the above "Kemmer-Dirac" type. The $\partial_\mu W^\mu$ term represents the usual freedom available of adding an arbitrary divergence to a Lagrangian without changing the equations of motion.

The action principle (A7) becomes well defined when the variations of W_{12} to be taken are specified. As discussed above, the variations of the transformation function $\langle a_1 t_1 \mid a_2 t_2\rangle$ under consideration involve changes in all the variables on which it

[25] We have restricted the analysis to variations corresponding to unitary transformations. The postulate can be shown to be also valid for certain other changes, such as source variations.[4]

depends, namely, the time[26] ($t \to t + \delta t(t)$) and the basis vectors (through the change in the complete set of operators). In general, irrespective of whether it is the variation of a finite operator W_{12}, the operator δW_{12} may be written as $\sum_m \delta W_{m,m+1}$ in the notation of Eq. (A5). Further, the infinitesimal operator $\delta W_{m,m+1}$ must have the form

$$\delta W_{m,m+1} = \int_{t_{m+1}}^{t_m} [Z_1(t; \chi(t))\ \delta t + Z_2(t; \chi(t))\ \Delta\chi]\, dt. \qquad \text{(A9a)}$$

The right-hand side expresses the most general form possible for the operator δW to yield the class of unitary transformations in the Hilbert space contained in the variation of the transformation function. Thus the term proportional to δt must be present, since changes $t \to t + \delta t$ are being considered, while the term proportional to the as yet undefined parameter $\Delta\chi$ has been included to account for the change of basis at fixed time (when δt is zero), i.e., for the change of the complete sets of operators. [In fact, even with a pure time translation, there is necessarily associated a change of the complete set $\{A(t)\} \to \{\bar{A} = A(t + \delta t)\}$, so that for this special case, we will see that $\Delta\chi$ is proportional to δt itself.] We now invoke the integrability postulate that there exists a $W_{12}(t; \chi)$, depending on the time and the field variables, so that its general variation consists in changing these arguments. Thus to have the form (A9a) arise from varying W_{12} we will assume that[27]

$$\delta W_{12} = W(t + \delta t, \chi + \delta'\chi) - W(t, \chi). \qquad \text{(A9b)}$$

The remaining problem, then, lies in finding the form of $\delta'\chi$ for a given variation of the transformation function $\langle a_1 t_1 \mid a_2 t_2 \rangle$, (i.e., for a given infinitesimal unitary transformation).

The explicit variation of W_{12} reads

$$\delta W_{12} = \int d^4x[\{\delta\chi \cdot A^\mu\, \partial_\mu\chi - \partial_\mu\chi A^\mu\, \delta\chi - \delta\chi\, \partial\mathfrak{H}/\partial\chi\} + \delta t(t)\, dT^{00}/dt] + \int d^3r[\tfrac{1}{2}(\chi A^0\, \delta\chi - \delta\chi A^0\chi) + \delta\chi(\partial W^0/\partial\chi) - \delta t(t) T^{00}]_{t_2}^{t_1} \qquad \text{(A9c)}$$

where $\delta\chi \equiv \delta'\chi + \dot{\chi}\, \delta t$ and

$$T^{00} = \tfrac{1}{2}(\partial_i\chi A^i\chi - \chi A^i\, \partial_i\chi) + \mathfrak{H}.$$

The symbol $\delta\chi\, \partial\mathfrak{H}/\partial\chi$ means $\mathfrak{H}(\chi + \delta\chi) - \mathfrak{H}(\chi)$. The condition that δW_{12} depend only upon end-point variations implies the vanishing of the first integral:

$$\int d^4x[\delta\chi A^\mu\, \partial_\mu\chi - \partial_\mu\chi A^\mu\, \delta\chi - \delta\chi\, \partial\mathfrak{H}/\partial\chi] + \int dt\, \delta t \frac{d}{dt} H = 0, \qquad \text{(A10)}$$

where $H \equiv \int T^{00}\, d^3r$. The generator $G(t)$, which is obtained from the end-point terms of Eq. (A9c), according to (A7), is

$$G(t) = \int d^3r[\tfrac{1}{2}(\chi A^0\, \delta\chi - \delta\chi A^0\chi) + \delta\chi\, \partial w^0/\partial\chi] + \delta t H. \qquad \text{(A11)}$$

We consider first the case of no time motion, $\delta t = 0$, and obtain the generator

$$G'(t) = \int d^3r[\tfrac{1}{2}(\chi A^0\, \bar{\delta}\chi - \bar{\delta}\chi A^0\chi) + \bar{\delta}\chi\, \partial w^0/\partial\chi], \qquad \text{(A12)}$$

where $\bar{\delta}\chi$ denotes the value of $\delta'\chi$ for $\delta t = 0$. The generator G' must give rise to all possible fixed-time infinitesimal cannonical transformations. The form of G' clearly changes by changing W^0, so that W^0 must be regarded as an arbitrary function which generates the various possible bases. Further, the variation $\bar{\delta}\chi$ must necessarily be arbitrary at every space-time point.[28] This will allow one to have the freedom of generating different cannonical transformations in each of the independent mutually spacelike degrees of freedom of the field at time t. The Lagrange equations may now be obtained from Eq. (A10) by setting $\delta t = 0$. One has then that

$$(\bar{\delta}\chi A^\mu\, \partial_\mu\chi - \partial_\mu\chi A^\mu\, \bar{\delta}\chi) - \bar{\delta}\chi\, \partial\mathfrak{H}/\partial\chi = 0. \qquad \text{(A13)}$$

In order to obtain explicit Lagrange equations of motion, some condition on the operator properties of $\bar{\delta}\chi$ is required since $\bar{\delta}\chi$ need not be a c number in a quantum theory. We postulate that $\bar{\delta}\chi_a$ either commutes or anticommutes with the field operators χ_b. With this assumption, one may move all the $\bar{\delta}\chi$ either to the left or to the right side and equate the coefficient of $\bar{\delta}\chi_a$ to zero. The condition of

[26] For simplicity, we are not varying the spatial coordinates x^i. Their variation would lead to the spatial translation operators (field momenta).

[27] Strictly speaking, the general form of δW is $W(t + \Delta t; \chi + \bar{\delta}\chi) - W(t; \chi)$ where Δt has a nonvanishing part even for $\delta t \to 0$, i.e., $\Delta t = \alpha(t)\delta t + \delta\beta$. The assumption made in (A 9b) is that $\delta\beta = 0$. Invariance under time translations implies that α is a constant, which may be set to unity by a choice of units.

[28] Unless $\bar{\delta}\chi$ is arbitrary in its time dependence [so that we may choose it proportional to $\delta(t)$], one would obtain from Eq. (A 10) a set of equations of motion nonlocal in time, instead of the local Lagrange equation (A 13). The postulate of time-locality of the dynamics forbids this. Lorentz-invariance then requires that $\bar{\delta}\chi$ also be arbitrary in its spatial dependence. Note also that $\delta'\chi$ is the $\Delta\chi$ of the general discussion of Eq. (8.9a) when Eq. (8.9b) has been postulated.

commutation or anticommutation on $\bar{\delta}\chi$ eventually leads to either commutation or anticommutation relations between the field operators themselves, and is an assumption conventionally made also in other derivations of the spin-statistics connection.[29]

Let us now ask for the part of G that generates pure time translations (with no change of the complete set at time t), i.e., G_t. This means that we must restrict our variations $\delta\chi$ to those appropriate for a time translation. The generator G_t gives rise to the Heisenberg equations of motion (A1a) which should give a well-defined statement of the future dynamical motion of the system. On the other hand, the arbitrary function W^0 enters in Eq. (A12) (while the dynamics described by the Lagrange equations (A13) is independent of W^μ). We conclude, therefore, that $\delta\chi$ must vanish for pure time motion, i.e.,

$$\delta'\chi = -\dot{\chi}\,\delta t = \delta_0\chi.$$

This leads to

$$G_t = -H\,\delta t \tag{A14}$$

and from (A1a) the usual Heisenberg equations of motion $\dot{A} = -i[A, H]$ where H is the conventional field Hamiltonian. Also, we note that imposing the the condition $\delta\chi = 0$ on Eq. (A10) gives rise to the consistent result[30] $dH/dt = 0$.

We turn now to the determination of the properties of G'. More precisely we will find the unitary transformations which G' generates and this information will yield the field commutation relations. In fact, a knowledge of the transformation generated by $G_\chi \equiv G'(W^0 = 0)$ is adequate to determine the properties of the general case. Thus, in the notation of Eq. (A1b), we write generally, $\delta_0\chi = \frac{1}{2}f\bar{\delta}\chi$ for the changes generated by G_χ, i.e.,

$$[\chi(\mathbf{r})G_\chi] = \tfrac{1}{2}if\,\bar{\delta}\chi(\mathbf{r}) \tag{A15a}$$

where f is at present unknown,[31] and may even be an operator function of χ. Thus G_χ generates the transformation

$$\chi \to \bar{\chi} = \chi - \tfrac{1}{2}f\,\delta\chi. \tag{A15b}$$

The commutation relations then follow from[32] Eq. (A15a), using the fact that $\bar{\delta}\chi(\mathbf{r})$ is an arbitrary function which either commutes or anticommutes with $\chi(\mathbf{r})$:

$$[\chi(\mathbf{r}), \chi(\mathbf{r}')A^0]_\mp = \tfrac{1}{2}if\,\delta^3(\mathbf{r} - \mathbf{r}') \tag{A16}$$

As discussed in Sec. II, the bracket in Eq. (A16) is a commutator (anticommutator) when A^0 is antisymmetric (symmetric). These complete commutation relations enable us to evaluate the commutator of χ with any function. In particular, one can find the analog of Eq. (A15b) for an arbitrary

$$G' = G_\chi + \int d^3r\;\bar{\delta}\chi(\partial W^0/\partial\chi)$$

where $\bar{\delta}\chi(\partial W^0/\partial\chi)$ is shorthand for $W^0(\chi + \bar{\delta}\chi) - W^0(\chi)$. In the discussion below, it will be convenient to consider W^0 as a function of $A^0\chi$ [i.e., $W^0 = W^0(A^0\chi)$] since these are the independent field variables (even when constraints exist). We restrict ourselves to W^0's which are even in the anticommuting field variables.[33] In this case, one finds by direct computation [using Eq. (A16)] that G' generates the change

$$\chi(\mathbf{r}) \to \bar{\chi}(\mathbf{r}) = \chi - \tfrac{1}{2}f\,\bar{\delta}\chi - \tfrac{1}{2}\,\partial^2W^0/\partial(A^0\chi)^2A^0f\,\bar{\delta}\chi, \tag{A17}$$

where $\partial^2W^0/\partial(A^0\chi)^2\frac{1}{2}iA^0f\bar{\delta}\chi$ stands for

$$\eta^{-1}[W^0(A^0\chi' + A^0\,\delta\chi' + \tfrac{1}{2}iA^0\eta f) - W^0(A^0\chi' + A^0\,\bar{\delta}\chi') - W^0(A^0\chi' + \tfrac{1}{2}iA^0\eta f) + W^0(A^0\chi)] \tag{A18}$$

The η appearing in (A18) is a new infinitesimal which commutes or anticommutes with χ and $\bar{\delta}\chi$ [according to the sign in (A16)] and is to be moved to the left and canceled to obtain the explicit form. The operations in (A18) reduce to the usual definition of second derivative in the commuting situation, and provided the correct rule in the other case as well.

Information on the allowed type of f comes from the requirement that G' and G_t be mutually consistent, and that G_t be consistent with the time development as given by the Lagrange equations. To see this, we first consider $[H, G_\chi]$ which by (A15b) is

---

[29] Note that this restriction on $\bar{\delta}\chi$ means that in our discussion of G', there was no possibility of obtaining the infinity of different transformations by different choices of $\bar{\delta}\chi$ (due to its simple "c number" nature). The generality of different W^0's is therefore indeed necessary.

[30] With a more complete treatment involving space like surfaces and general coordinate variations (4) $\delta\chi^\mu$, this term gives rise to the local conservation laws $\partial_\nu T^{\mu\nu} = 0$.

[31] The operator f is necessarily coordinate independent: Translational invariance requires that $f(\mathbf{r}, \mathbf{r}') = f(\mathbf{r} - \mathbf{r}')$, while its $\delta^3(\mathbf{r} - \mathbf{r}')$ coefficient shows that only $f(0)$ enters. [See Eq. (A16) below.]

[32] We have assumed that the matrix A^0 is nonsingular (i.e., no constraints are present in the theory). When A^0 is singular, the discussion still follows in terms of the independent variables $A^0\chi$, using the result $[A^0\chi, A^0\chi']_\mp = (i/2)\,A^0f\delta^3(1 - \mathbf{r}')$.

[33] As noted above, the anticommuting case arises when A^0 is symmetric and so, as discussed in text, only for spin $\frac{1}{2}$. Lorentz invariance then requires that tensor quantities such as W^μ contain even powers of χ.

$$[H, G_\chi] = -i(H[\bar{\chi}] - H[\chi])$$
$$= -i(H[\chi - f/2\ \bar{\delta}\chi] - H[\chi])$$
$$= H\left[\chi + \frac{i}{2} f\ \bar{\delta}\chi\right] - H[\chi]. \quad \text{(A19)}$$

On the other hand, this evaluation must agree with the fact that H generates time translations. Thus

$$[H, G_\chi] = \int [H, \chi A^0]\ \bar{\delta}\chi\ d^3r + \int \chi A^0 [H,\ \bar{\delta}\chi]\ d^3r. \quad \text{(A20)}$$

The commutator $[H,\ \bar{\delta}\chi]$ must vanish for all $\bar{\delta}\chi$. For, in the anticommuting case, one obtains from it a contribution due to parts of H which are odd in the anticommuting fields. However, such parts of H would also yield nonlocal contributions to the Heisenberg equations of motion (A19), and so contradict the local Lagrange equations. No such terms may appear in H, then.[33] The remaining term on the right in (A20) involves $\dot{\chi}A^0$ by Eq. (A1a), and so by the Lagrange equations (A13), we get

$$[H, G\chi] = H[\chi + \tfrac{1}{2}i\ \bar{\delta}\chi] - H[\chi]. \quad \text{(A21)}$$

Though the required consistency between Eqs. (A21) and (A19) strongly restricts the form of f, it does *not* permit one to conclude that $f = 1$ since H is not an arbitrary function of χ (and $\bar{\delta}\chi$ is not sufficiently arbitrary, since it only commutes or anticommutes with χ). Indeed, as has been noted by Wigner,[34] the consistency of the Lagrange and Heisenberg equations does not uniquely determine f for the simple case of the one-dimensional harmonic oscillator, but restricts it in the form $f = 1 + (2E_0 - 1) \exp[i\pi(H - E_0)]$ in units where $\hbar\omega = 1$. Here E_0 is the ground-state energy which is now arbitrary. In fact, in the one-dimensional case at least, such an indeterminacy exists in a large number of cases (for example, $V = \frac{1}{2}x^2 + \lambda x^3$). Further, $f = 1$ cannot in general be forced even when the same consistency requirement is imposed on the full generator G'. Thus, calculating $[H, G']$ as was done for $[H, G_\chi]$, using the results (A17, 18), one finds the analog of (A19) to be

$$[H, G'] = H[\chi + \tfrac{1}{2}if\ \bar{\delta}\chi + \tfrac{1}{2}i\ \partial^2 W^0/\partial(A^0\chi)^2 A^0 f\ \bar{\delta}\chi] - H[\chi] \quad \text{(A22)}$$

while (A21) is replaced by[35]

$$[H, G'] = \epsilon^{-1} \int d^3r [W(A^0\chi + A^0\ \delta\chi + \tfrac{1}{2}\epsilon\ \delta H/\bar{\delta}\chi) - W(A^0\chi + A^0\ \delta\chi) - W(A^0\chi + \tfrac{1}{2}\epsilon\ \delta H/\bar{\delta}\chi) + W(A^0\chi)]. \quad \text{(A23)}$$

where ϵ is a c number infinitesimal, and $\frac{1}{2}\delta H/\bar{\delta}\chi$ has been inserted for $A^0\dot{\chi}$ by the Lagrange equations. The symbol $\delta H/\bar{\delta}\chi$ is defined by

$$\delta H \equiv H(\chi + \bar{\delta}\chi) - H(\chi) = \int d^3r\ \bar{\delta}\chi(x)\ \delta H/\bar{\delta}\chi(x).$$

For all classical fields, the right sides of Eqs. (A22) and (A23) are trivially seen to be consistent only is the choice $f = 1$ is made. However, for quantum fields, it is not even obvious that these equations are consistent with $f = 1$. This is due to the fact that the order of operators in each equation is entirely different, so that the comparison can only be made after a large number of operator reorderings has been carried out [using the field commutation relations (A16)]. However, a somewhat tedious calculation (considering the general power series terms of H and W^0) establishes that, for $f = 1$, the results are indeed consistent.[36] Thus, the full generator G' is consistent with the possibility originally allowed by G_χ alone, that $f = 1$. The above results do not of course establish the necessity of $f = 1$. The Wigner example for f in the harmonic oscillator case turns to still yield consistent results between (A22) and (A23), although we have not investigated whether $f \neq 1$ is still possible for other potentials. Since it is impossible to deduce that f must be 1 for all systems,[9a] we add the consistent postulate that $f = 1$. Only in this way does the value of the fundamental commutator remain unchanged under change of basis. It is to be noted, however, that if one just assumes f to be a c number, the consistency between (A19) and (A21) is adequate to ensure that $f = 1$.

[34] E. P. Wigner, Phys. Rev. 77, 711 (1950).

[35] Equation (A23) is obtained by taking the time derivative of δW^0, and using the Lagrange equations to replace $A^0\dot{\chi}$ by $\delta H/\bar{\delta}\chi$.

[36] That is, all extra commutators arising in reordering one of the equations into the other's form cancel.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Mass Singularities of Feynman Amplitudes*

TOICHIRO KINOSHITA
Laboratory of Nuclear Studies, Cornell University, Ithaca, New York
(Received January 4, 1962)

Feynman amplitudes, regarded as functions of masses, exhibit various singularities when masses of internal and external lines are allowed to go to zero. In this paper, properties of these mass singularities, which may be defined as pathological solutions of the Landau condition, are studied in detail. A general method is developed that enables us to determine the degree of divergence of unrenormalized Feynman amplitudes at such singularities. It is also applied to the determination of mass dependence of a total transition probability. It is found that, although partial transition probabilities may have divergences associated with the vanishing of masses of particles in the final state, they always cancel each other in the calculation of total probability. However, this cancellation is partially destroyed if the charge renormalization is performed in a conventional manner. This is related to the fact that interacting particles lose their identity when their masses vanish. A new description of state and a new approach to the problem of renormalization seem to be required for a consistent treatment of this limit.

1. INTRODUCTION

FOR any Feynman diagram, the corresponding transition amplitude is a function of scalar products of external momenta and masses of various internal lines. Analyticity of amplitudes regarded as functions of external momenta has been clarified considerably in the last few years. In these considerations, masses of internal lines are usually treated as parameters fixed to their observed values. However, it has been noted that some aspects of analyticity may be understood more clearly if the amplitude is continued analytically with respect to its masses.[1] For a complete characterization of Feynman amplitudes, it will be necessary to treat both external momenta and internal masses as (complex) variables. In this paper, we should like to see what happens to the Feynman amplitude when the domain of mass variable is extended along the real axis. In particular, we are interested in the singularities of amplitudes which are encountered at the origin of mass variables and at essentially arbitrary values of external momenta. To distinguish these singularities from the usual poles and branch points in the complex plane of energy or momentum transfer, let us call them *mass singularities*. We want to find out all possible mass singularities of Feynman amplitudes and determine in particular whether or not the amplitudes are divergent at the mass singularity.

One of the familiar examples of mass singularity is the so-called infrared divergence that appears in connection with the vanishing of the photon mass λ. The divergence of the total cross section for Coulomb scattering is also an example of this sort. There is another mass divergence that is even more common than the infrared divergence but is rarely referred to as such. It is the logarithmic divergence associated, for instance, with the vanishing of the electron mass m in quantum electrodynamics. Of course the observed electron mass is different from zero. Nevertheless it will be useful to consider the zero-mass limit since the behavior of Feynman amplitude for small m or high energy is determined to a large extent by its mass singularities.

These mass singularities have a remarkable property that divergences of *partial* transition probabilities associated with vanishing masses often cancel each other when they are summed into a *total* transition probability. This behavior is, of course, well known for the infrared divergence.[2] For other cases, however, the cancellation is more subtle and incomplete in general. Thus, it was not recognized clearly as a general property of Feynman amplitudes until a few years ago when a detailed calculation was carried out on radiative corrections to weak interactions, such as the μ-e decay, β decay, and π-μ (or π-e) decay.[3] In these calculations, it was found that the (unrenormalized) total decay probability does not contain any divergent term like $\ln \lambda$ or

* Supported in part by the joint program of the Office of Naval Research and the U.S. Atomic Energy Commission.

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

[2] We quote here only two papers that appeared most recently: D. R. Yennie, S. C. Frautschi, and H. Suura, Ann. Phys. **13**, 379 (1961); K. E. Eriksson, Nuovo cimento **19**, 1010 (1961). We note however that the present work should be regarded as an extension of earlier works quoted in references 24 and 28.

[3] T. Kinoshita and A. Sirlin, Phys. Rev. **113**, 1652 (1959); S. Berman, *ibid.* **112**, 267 (1958); T. Kinoshita, Phys. Rev. Letters **2**, 477 (1959); S. Berman, *ibid.* **1**, 468 (1958).

ln m, although each partial probability has such terms. Thus, if we consider a hypothetical problem in which the electron mass is much smaller than its actual value, the radiative correction to a partial probability becomes very large and in fact diverges logarithmically in the limit $m = 0$. This means, of course, that the perturbation theory breaks down in this limit. In spite of this, all these infinities cancel each other when they are added together and integrated over the final states. Thus we encounter the interesting situation that the range of validity of perturbation theory for small m seems to depend on what we want to measure. Although complete cancellation of ln m terms was actually proved only in the lowest order radiative corrections of decay processes, the way the cancellation takes place suggests strongly that it would happen to all orders of perturbation theory. To see whether or not this conjecture is valid is one of the motivations of this work.

If we look for similar cancellation of ln m terms in other processes, such as collision of two particles, no such cancellation seems to be operating at first sight. Actually, cancellation of ln m terms occurs in collision processes too, although it is not as complete as in the decay problems. This leads us to a new question: Which ln m terms cancel and which do not in the calculation of the total collision cross section? In fact, the same problem of cancellation will be found not only in the decay and collision probabilities but in arbitrary Feynman amplitudes. Thus we shall be concerned with the mass singularity of general Feynman amplitudes in most of this paper.

Closely related to these problems is the fact that the unrenormalized electron propagator $S_F'(=Z_2 S_{FC})$ does not diverge for $m \to 0$, whereas both Z_2 and S_{FC} have logarithmic divergences at $m = 0$. This was first pointed out by Gell-Mann and Low and used as a basis of their work on the interaction at very small distances.[4] A study of mass singularity will help us obtain a better understanding of this and related problems.

In analyzing the mass singularity of Feynman amplitude, we have found it convenient to parametrize the amplitude according to a variant of Feynman's method in which two sets of parameters are introduced in succession instead of one set as is usually done. This parametrization is discussed in Sec. 2 and Appendix B. A general definition of mass singularity as a pathological solution of Landau condition is given in Sec. 3. In Sec. 4, mass singularities of simplest Feynman amplitudes consisting of one closed loop are analyzed in detail. More complicated amplitudes are treated in Secs. 5 through 8. It will be shown in Sec. 9 how to treat amplitudes containing self-energy insertions. Results of our analysis are summarized in Sec. 10. Its application to the mass singularity of total transition probability is discussed in Sec. 11. The results of these considerations apply only to unrenormalized amplitudes. Some of the problems which may arise in connection with charge renormalization are discussed in Sec. 12. Appendix A contains a qualitative argument which may clarify the mechanism of cancellation of mass divergences.

2. DOUBLE PARAMETRIC REPRESENTATION OF FEYNMAN AMPLITUDES

We want to consider an arbitrary Feynman diagram containing n internal lines. Unless it is necessary, we shall make no particular restriction on the kind of internal lines, external lines, and vertices. In particular, masses of different lines will, in general, be treated as independent of each other.

To each internal boson line i corresponds a propagator

$$D_i = [(k_i + q_i)^2 - m_i^2 + i\epsilon]^{-1}, \tag{2.1}$$

where k_i, q_i, and m_i represent variable and fixed momenta and (renormalized) mass. We can always choose k_i and q_i in such a way that conservation of four-momentum holds at each vertex for k and q separately. We shall assume this in the following. The propagator for an internal fermion line is obtained by applying the operator[5]

$$\frac{1}{2}\int_{m_i^2}^{\infty} dm_i^2 \gamma^\mu \frac{\partial}{\partial q_i^\mu} + m_i \tag{2.2}$$

to (2.1). If a diagram is separated into two (disconnected) parts by cutting an internal line and replacing it by two external lines, any such line will be called *fixed*. Propagators for fixed lines do not contain any variable momenta. We put $k_i = 0$ in such a case.

The transition amplitude may be written as

$$\int \prod_{i=1}^{r} d^4 k_i \tilde{F} \prod_{j=1}^{n} D_j, \tag{2.3}$$

where r is the number of independent variables k_i. $\tilde{F}$ represents the contribution from vertices as well as from fermion operators (2.2), and is assumed to

[4] M. Gell-Mann and F. E. Low, Phys. Rev. 95, 1300 (1954).

[5] R. Karplus and N. M. Kroll, Phys. Rev. 77, 536 (1950).

be a regular function of coupling constants and momenta. If the diagram contains fixed lines, (2.3) is reduced to a product of subintegrals which have no common integration variable. Thus its behavior is determined by that of the subintegrals.

Our problem may be studied starting from the integral (2.3) itself, or by parametrizing (2.3) according to Feynman's method and carrying out the k integration first. We shall take the second approach here. However, we have found it convenient to employ a variant of Feynman's method, in which two sets of parameters are introduced in successive steps instead of one set as is usually done. For this purpose, let us introduce the notion of a *chain*. A chain α is defined as the largest set of internal lines having the same momentum variable k_α. Thus, two internal lines belong to the same chain if the diagram is separated into two pieces when these lines are cut. Of course this does not apply when a chain consists of just one line. Any unfixed internal line belongs to one and only one chain. Fixed lines do not belong to any chain.

A vertex (part) is called *internal* if it is connected to the rest of the diagram by chains only. It will be called *external* otherwise. A new diagram is obtained from a Feynman diagram by omitting all external lines and corresponding vertices. This will be called a *chain diagram*. To any chain diagram there corresponds a class of all Feynman diagrams that can be constructed by inserting an arbitrary number of external lines in various parts of the chain diagram in accordance with the rules of a given theory.

Each chain α contributes a factor

$$\prod_i \frac{1}{(k_\alpha \pm q_i)^2 - m_i^2 + i\epsilon}, \qquad i \in \alpha, \tag{2.4}$$

to the integrand of (2.3), where the use of $\pm$ depends on the relative sign of k_α and k_i. Our first step is to parametrize (2.4) [and not the entire denominators of (2.3)], making use of the Feynman formula

$$\frac{1}{a_1 a_2 \cdots a_n} = (n-1)! \times \int \frac{\delta(1 - x_1 - \cdots - x_n)\, dx_1\, dx_2 \cdots dx_n}{(x_1 a_1 + x_2 a_2 + \cdots + x_n a_n)^n}, \qquad x_1, x_2, \cdots, x_n \geq 0. \tag{2.5}$$

As is easily seen, (2.4) can then be written as

$$(n_\alpha - 1)! \int \frac{dx(\alpha)}{[(k_\alpha + q_\alpha)^2 - V_\alpha(x) + i\epsilon]^{n_\alpha}}, \tag{2.6}$$

where n_α is the length (namely, the number of lines) of the chain α, and

$$\begin{aligned} dx(\alpha) &= \delta\Big(1 - \sum_i^\alpha x_i\Big) \prod_j^\alpha dx_j, \\ q_\alpha &= \sum_i^\alpha \pm x_i q_i, \\ V_\alpha(x) &= \sum_i^\alpha x_i m_i^2 - \sum_{i<j}^\alpha x_i x_j (q_i \mp q_j)^2 \\ &= \frac{1}{2} \sum_{i,j}^\alpha x_i x_j [m_i^2 + m_j^2 - (q_i \mp q_j)^2]. \end{aligned} \tag{2.7}$$

Note that (2.6) has a structure similar to the ordinary propagator, q_α and V_α being the "momentum" and "mass" of the chain α. The integrand of (2.3) is a product of factors (2.6) arising from all chains. We may now perform the second parametrization by combining them into a single denominator using a formula similar to (2.5). The resulting expression can be integrated with respect to r variables k_α, as has been done by many authors,[6] and (2.3) is converted to the parametric form

$$\text{const} \times \int \frac{F\, \delta(1 - \sum z_\alpha) \prod dz(\alpha)}{U^2(z)[V(x,z) - i\epsilon]^{n-2r}}, \qquad \text{all } z_\alpha \geq 0, \tag{2.8}$$

where $n = \sum n_\alpha$,

$$dz(\alpha) = z_\alpha^{n_\alpha - 1}\, dz_\alpha\, dx(\alpha), \tag{2.9}$$

and F is derived from $\tilde{F}$ of (2.3) and is a polynomial in x and z. The function $U(z)$ is defined by

$$U(z) = \sum z_{\alpha_1} z_{\alpha_2} \cdots z_{\alpha_r}, \tag{2.10}$$

where the summation is over all sets $(\alpha_1, \alpha_2, \cdots, \alpha_r)$ such that $k_{\alpha_1}, k_{\alpha_2}, \cdots, k_{\alpha_r}$ are independent. The function $V(x, z)$ is given by

$$V(x,z) = \sum z_\alpha V_\alpha(x) + v(x,z) \tag{2.11}$$

with

$$v(x,z) = -U(z)^{-1} \times \sum_{\alpha,\beta,\cdots,\delta} \Big\{ z_\alpha \Big(z_\beta \frac{\partial}{\partial z_\beta}\Big) \cdots \Big(z_\delta \frac{\partial}{\partial z_\delta}\Big) U(z_\alpha = 0) \Big\} \times (q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2, \tag{2.12}$$

where the summation in (2.12) is over all sets $\{\alpha, \beta, \cdots, \delta\}$ with the property that the chain diagram is separated into exactly two parts if all chains $\alpha, \beta, \cdots, \delta$ are cut into two parts.[7] Signs

[6] R. Chisholm, Proc. Cambridge Phil. Soc. **48**, 300 (1952); Y. Nambu, Nuovo cimento **6**, 1064 (1957); N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **17**, 401 (1957).

[7] The function $(\partial/\partial z_\beta) \cdots (\partial/\partial z_\delta) U(z_\alpha = 0)$ of the formula (2.12) is essentially a product of two U factors, corresponding to the two parts of the chain diagram separated by the C set $\{\alpha, \beta, \cdots, \delta\}$. This was noted by Symanzik, reference 9. I would like to thank Dr. Robert B. Marr who called my attention to this point.

of $q_\alpha, q_\beta, \cdots, q_\delta$ must be so chosen that $(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2$ is invariant under any transformation of the form

$$q_{i(C)} \to q_{i(C)} + q^C, \tag{2.13}$$

where $i(C)$ runs over all internal lines of an arbitrary closed loop C and q^C is an arbitrary fixed four-vector. Because of this property, $\{\alpha, \beta, \cdots, \delta\}$ and $(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2$ will be called a *C set* and a *C invariant*, respectively. The formula (2.12) is derived in Appendix B from one of the known formulas of $V(x, z)$.

Before going into consideration of singularities, we shall make a few general remarks about the above formulas:

(a) As is obvious from (2.10), $U(z) \geq 0$ for any $z \geq 0$. The equality holds only when all z belonging to a closed loop vanish. In general $U(z)$ has a zero of mth order if all z belonging to m independent closed loops vanish simultaneously. We also note that $V(x, z)$, although it contains $U(z)^{-1}$, may be regarded as continuous in the entire domain of integration of (2.8).

(b) In general, the amplitude (2.3) contains ultraviolet divergences. In most cases this may be handled by introducing a suitable cutoff such as Feynman's.[8] In this method, $[(k_i + q_i)^2 - m_i^2 + i\epsilon]^{-1}$ is replaced by $[(k_i + q_i)^2 - L_i^2 + i\epsilon]^{-2}$ and the k integration is carried out first in (2.3). The resulting parametric integral is similar in structure to (2.8) except that m_i is replaced by L_i and the power of the denominator $n' - 2r$ is larger than $n - 2r$. Of course the actual amplitude is obtained by integrating it with respect to L_i^2 from m_i^2 to Λ_i^2 ($\Lambda_i^2 \gg m_i^2$). The final expression is given by (2.8) with additional terms containing Λ_i^2. When $n - 2r$ is not positive, the denominator of (2.8) must be interpreted as a positive power term $V^{|n-2r|}$ times a logarithmic function of V as is determined by the L_i^2 integration. With this much precaution, we may use (2.8) for our consideration of mass singularity whether or not it contains divergent contributions from high-energy end.

(c) If the line i of chain α is a fermion line, F of (2.8) contains the operator (2.2) as a factor. Carrying out the indicated operation, we find that F contains a factor

$$\gamma_\mu Q_{i\alpha}^\mu + m_i \tag{2.14}$$

with

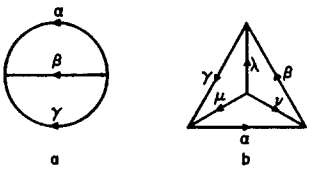

FIG. 1. Simple chain diagrams.

$$Q_{i\alpha} = q_i - q_\alpha + \frac{1}{U(z)} \sum \left\{ \left(z_\beta \frac{\partial}{\partial z_\beta} \right) \cdots \left(z_\delta \frac{\partial}{\partial z_\delta} \right) U(z_\alpha = 0) \right\} (q_\alpha \pm q_\beta \pm \cdots q_\delta), \tag{2.15}$$

where the summation is over all C sets containing α. (For simplicity it is assumed that $q_\alpha = \sum_\alpha x_i q_i$.) If the line j of chain β is also a fermion line, F consists of two terms, one being proportional to the product of $\gamma_\mu Q_{i\alpha}^\mu + m_i$ and $\gamma_\nu Q_{j\beta}^\nu + m_j$, the other to $V(x, z)(\partial Q_{i\alpha}/\partial q_j)$. Since $V(x, z)$ in the second term cancels with one $V(x, z)$ in the denominator, its contribution to the integral (2.8) is actually less singular than the first term. Thus it may be disregarded in most of our considerations. Obviously the momentum dependence of F arising from derivative-coupling vertices can be handled similarly.

(d) The formula (2.12) gives the most concise expression of the parametrized amplitude. However, it is often useful to write it in a less compressed form. For this purpose we may introduce a *subchain* as an arbitrary subset of a chain. Thus a chain α may consist of subchains $\alpha_1, \alpha_2, \cdots, \alpha_s$. Each subchain has its set of x variables and the corresponding $V_{\alpha_i}(x)$ of the form (2.7). A new $U(z)$ is obtained by the substitution

$$z_\alpha \to z_{\alpha_1} + z_{\alpha_2} + \cdots + z_{\alpha_s}.$$

It is then easy to see that $v(x, z)$ and the integral are again given by (2.12) and (2.8). However, $v(x, z)$ now contains C invariants of the form $(q_{\alpha_i} - q_{\alpha_j})^2$ besides familiar ones where α_i and α_j are subchains of α. In particular, if all subchains consist of only one internal line, all x take the value 1 and the formula (2.8) is reduced to the usual expression parametrized by a single set of variables z. C sets are still characterized by the property that the Feynman diagram is separated into two parts if all subchains of a C set are cut in the middle. This property was first noted by Symanzik.[9]

It will be appropriate here to illustrate the general formula (2.8) by applying it to some simple cases. First consider the chain diagram of Fig. 1(a). We find from (2.10) that

[8] In some cases it is necessary to use other methods such as that of Pauli and Villars. See W. Pauli and F. Villars, Revs. Modern Phys. **21**, 434 (1949).

[9] K. Symanzik, Progr. Theoret. Phys. (Kyoto) **20**, 690 (1958).

$$U(z) = z_\alpha z_\beta + z_\alpha z_\gamma + z_\beta z_\gamma. \tag{2.16}$$

Since this diagram cannot be divided into two parts unless all chains α, β, γ are bisected, we obtain only one C set. Thus

$$V(x, z) = z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma - [z_\alpha z_\beta z_\gamma / U(z)](q_\alpha + q_\beta + q_\gamma)^2. \tag{2.17}$$

A somewhat more complicated case is shown in Fig. 1(b). In this case, $U(z)$ is given by

$$\begin{aligned} U(z) = {} & z_\alpha z_\beta z_\gamma + z_\alpha z_\beta z_\lambda + z_\alpha z_\beta z_\mu + z_\alpha z_\gamma z_\lambda \\ & + z_\alpha z_\gamma z_\nu + z_\alpha z_\lambda z_\mu + z_\alpha z_\lambda z_\nu + z_\alpha z_\mu z_\nu \\ & + z_\beta z_\gamma z_\mu + z_\beta z_\gamma z_\nu + z_\beta z_\lambda z_\mu + z_\beta z_\lambda z_\nu \\ & + z_\beta z_\mu z_\nu + z_\gamma z_\lambda z_\mu + z_\gamma z_\lambda z_\nu + z_\gamma z_\mu z_\nu. \end{aligned} \tag{2.18}$$

The function $v(x, z)$ consists of seven C invariants, each corresponding one to one to the way the diagram of Fig. 1(b) is cut into two parts. Its coefficients can be calculated easily from (2.12). We find that

$$\begin{aligned} -U(z)v(x, z) = {} & z_\beta z_\gamma z_\lambda (z_\alpha + z_\mu + z_\nu)(q_\beta - q_\gamma + q_\lambda)^2 \\ & + z_\alpha z_\gamma z_\mu (z_\beta + z_\lambda + z_\nu)(q_\alpha - q_\gamma - q_\mu)^2 \\ & + z_\alpha z_\beta z_\nu (z_\gamma + z_\lambda + z_\mu)(q_\alpha - q_\beta + q_\nu)^2 \\ & + z_\lambda z_\mu z_\nu (z_\alpha + z_\beta + z_\gamma)(q_\lambda + q_\mu + q_\nu)^2 \\ & + z_\alpha z_\beta z_\lambda z_\mu (q_\alpha - q_\beta - q_\lambda - q_\mu)^2 \\ & + z_\alpha z_\gamma z_\lambda z_\nu (q_\alpha - q_\gamma + q_\lambda + q_\nu)^2 \\ & + z_\beta z_\gamma z_\mu z_\nu (q_\beta - q_\gamma - q_\mu - q_\nu)^2. \end{aligned} \tag{2.19}$$

3. DEFINITION OF MASS SINGULARITY

We shall now come to the question of how to define the mass singularity of an arbitrary Feynman amplitude. As is well known, when the internal masses are fixed, the integral (2.8) can be regarded as an analytic function of scalar products of external momenta with various threshold singularities. In our problem, however, we would rather fix the external momenta and treat (2.8) as a function of masses. In spite of this difference, singularities in mass variables can be determined by the same condition as that for the threshold singularities. We shall therefore start from the consideration of threshold singularities in general.

As is well known, a threshold is characterized by the condition that[10]

$$V(x, z) = 0 \tag{3.1}$$

is a minimum of $V(x, z)$ with respect to x and z. Note that $V(x, z)$ is continuous in x, $z \geq 0$, including the boundary. At the minimum some z_β will take the boundary value $z_\beta = 0$. The set of all β with nonvanishing z_β will be called A. The set of β with $z_\beta = 0$ will be denoted by $\bar{A}$. As is seen from (2.11), the threshold does not depend on $x_{i\beta}$ if $\beta \in \bar{A}$. For each $\alpha \in A$, $x_{i\alpha}$ is further classified into sets G_α and $\bar{G}_\alpha$, depending on whether $x_{i\alpha} \neq 0$ or $= 0$ at the threshold. Nonvanishing x and z are now determined by minimizing the expression

$$V(x, z) - \lambda \sum_A z_\alpha - \sum_A \lambda_\alpha \sum_{G_\alpha} x_{i\alpha} \tag{3.2}$$

with respect to $x_{i\alpha}$ and z_α, where λ and λ_α are Lagrange multipliers. This leads us to the Landau conditions[10]

$$\begin{aligned} \partial V/\partial x_{i\alpha} - \lambda_\alpha = 0, \quad i\alpha \in G_\alpha, \\ \partial V/\partial z_\alpha - \lambda = 0, \quad \alpha \in A. \end{aligned} \tag{3.3}$$

It is easy to see that $\lambda = 0$ and $\sum_A \lambda_\alpha = 0$. This follows from the homogeneity of $V(x, z)$ in z and in *all* x [where the second form of (2.7) is assumed for $V_\alpha(x)$].

In general, two kinds of situations, normal and pathological, may arise with respect to solutions of (3.3). In the first case, (3.3) yields normalized solutions $x_{i\alpha}$, z_α which are implicit functions of the masses and external momenta. These solutions exist when these variables satisfy certain restrictions. The simplest, and most common, situation is the restriction of the square of the total energy in some channel to a particular value, which therefore marks the energy at which the integral begins to have an absorptive part. When this value corresponds to the mass of some intermediate state, it is called a normal threshold. All other cases are referred to as anomalous threshold singularities, and have been the subject of some rather extensive study in the last few years.[11] However, none of these singularities is what we are looking for, since they are for the most part branch point singularities, and do not lead to divergence of the transition amplitude.

The other kind of situation arises when the masses and external momenta take on values such that Eq. (3.3) possesses arbitrary solutions $x_{i\alpha}$ and z_α subject only to the normalization conditions. Examples are the infrared and ln m divergences discussed in Sec. 1. Singularities of this type usually involve the vanishing of certain masses, and may

[10] L. D. Landau, Nuclear Phys. **13**, 181 (1959); N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **22**, 128 (1959); J. D. Bjorken, "Spectral representation of Green's functions in perturbation theory," Stanford University preprint (1959).

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

or may not exist for arbitrary external momenta. In what follows, we shall be most interested in those cases in which the external momenta are essentially arbitrary, the singularity arising solely as a consequence of the mass values assumed.[12] Such cases will be called *mass singularities*.

It is easily seen that (3.3) gives arbitrary z if and only if

$$V_\alpha(x) = 0, \quad \partial V_\alpha(x)/\partial x_{i\alpha} = 0, \quad x_{j\alpha} = 0, \qquad (3.4a)$$
$$i\alpha \in G_\alpha, \quad j\alpha \in \bar{G}_\alpha, \quad \alpha \in A,$$

and

$$(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2 = 0,$$
$$q_{i\xi}(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta) = 0, \quad i\xi \in G_\xi, \qquad (3.4b)$$
$$\xi = \alpha, \beta, \cdots, \delta, \quad \xi \in A,$$

where (3.4b) is needed only when $\{\alpha, \beta, \cdots, \delta\}$ is a C set. Thus, in order to find a mass singularity, we have only to look for the case where the solution x of (3.4a, b) is arbitrary. Note that (3.4a) holds for each chain separately. The role of (3.4b) is then to restrict possible solutions of (3.4a) when A contains C sets.

This consideration shows that finding the location of the mass singularity itself is much easier than finding the general thresholds for external momenta. What is not so simple, and this is our real problem, is to determine how strongly the integral (2.8) diverges at the mass singularity. This is complicated not only by the complex structure of the set A but also by the fact that the set $\bar{A}$, which does not participate in the determination of the location of the mass singularities, may influence the actual strength of the mass singularity.

4. FEYNMAN DIAGRAMS WITH ONE CLOSED LOOP

We shall first examine how mass singularities of simple Feynman diagrams consisting of only one closed loop are determined by the condition (3.4). Although the heavy machinery developed in the last two sections is actually unnecessary for this case, we shall apply it to such a diagram since it will serve as a prototype for the analysis of the more complicated diagrams.

Since our diagram consists of one chain, (2.8) is reduced to the simple form

$$\int \frac{F\, dx_1 \cdots dx_n\, \delta(1 - x_1 - \cdots - x_n)}{[V(x) - i\epsilon]^{n-2}}. \qquad (4.1)$$

If we introduce the v matrix by

$$v_{ij} = 1/2[m_i^2 + m_j^2 - (q_i - q_j)^2], \qquad (4.2)$$

we may express $V(x)$ as

$$V(x) = \sum v_{ij} x_i x_j, \qquad (4.3)$$

as is seen from (2.7). For simplicity, all q are chosen in the same direction along the loop. The threshold condition (3.4) becomes

$$\partial V/\partial x_i = 0, \quad x_j = 0; \quad i \in G, \quad j \in \bar{G}. \qquad (4.4)$$

The integral (4.1) will exhibit a mass singularity only when (4.4) gives *arbitrary* values of x_i as solutions. This takes place only if

$$v_{ij} = 0 \quad \text{for} \quad i, j \in G. \qquad (4.5)$$

For $i = j$, (4.5) means that $m_i = 0$. Thus our definition of mass singularity in fact gives a singularity at zero mass. Since any subset of G also satisfies (4.4), some mass singularities may arise from overlapping domains of integration. There is a further complication that some v_{ij} besides those of (4.5) may vanish independently of the threshold condition.

To analyze these problems, it is convenient to regard the closed loop as a sum of three subchains α, β, and γ, some of which may be empty. According to (d) of Sec. 2, (4.1) may then be expressed in the alternative form

$$\int \frac{F\, \delta(1 - z_\alpha - z_\beta - z_\gamma)\, dz(\alpha)\, dz(\beta)\, dz(\gamma)}{[z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma - z_\alpha z_\beta (q_\alpha - q_\beta)^2 - z_\alpha z_\gamma (q_\alpha - q_\gamma)^2 - z_\beta z_\gamma (q_\beta - q_\gamma)^2 - i\epsilon]^{n-2}}, \qquad (4.6)$$

where $n = n_\alpha + n_\beta + n_\gamma$ and $dz(\alpha)$, etc., are given by (2.9). We shall choose α so that $V_\alpha(x)$ vanishes for arbitrary values of $x \in \alpha$. This means that α is a subset of G. We are interested in the behavior of (4.6) in the limit $z_\beta = z_\gamma = 0$. In this neighborhood, the denominator may be written as

$$z_\alpha V_\alpha + z_\beta V_\beta' + z_\gamma V_\gamma' + z_\beta^2 (q_\alpha - q_\beta)^2 + \cdots, \qquad (4.7)$$

where

$$V_\xi' = V_\xi - (q_\xi - q_\alpha)^2$$
$$= \sum_{i \in \xi} x_i [m_i^2 - (q_i - q_\alpha)^2], \quad \xi = \beta, \gamma. \qquad (4.8)$$

[12] By arbitrary we mean the arbitrariness of relative orientation of external momenta. The magnitude of each external momentum may (or may not) be fixed by the mass shell condition. For some particular orientation of external momenta, we may find pathological singularities which are not pure mass singularities. Usually they are not very interesting since their contribution to the total transition probability is negligible because of the phase space factor. An exception is the forward scattering amplitude. How to treat this case is discussed in Secs. 4 and 11.

Noting that (4.8) is linear in $x_i \in \xi$, we shall choose β as the largest set of i such that $m_i^2 - (q_i - q_\alpha)^2 = 0$ for arbitrary values of $x \in \alpha$. Then, as a function of $x \in \alpha$ and $x \in \gamma$, V_γ' does not vanish except possibly at points of measure zero. Because of the δ function, (4.6) can easily be integrated with respect to z_α. Next, integrating by parts around $z_\gamma = 0$, we obtain

$$\int \frac{F'' z_\beta^{n_\beta + f_\beta - 1}\, dz_\beta\, dx(\alpha)\, dx(\beta)\, dx(\gamma)}{[V_\gamma']^{n_\gamma + f_\gamma}[(1 - z_\beta)V_\alpha + z_\beta V_\beta' + z_\beta^2(q_\alpha - q_\beta)^2 - i\epsilon]^{n_\alpha + n_\beta - f_\gamma - 2}} \tag{4.9}$$

$$+ \text{ less singular terms,}$$

where we have assumed that $F \sim F' z_\gamma^{f_\gamma}$ for small z_γ and $F' \sim F'' z_\beta^{f_\beta}$ for small z_β. If the set β is empty, the integral (4.9) will diverge whenever $n_\alpha - f_\gamma \geq 2$. This is obvious since $V_\alpha(x)$ vanishes for arbitrary values of $x \in \alpha$.

Let us assume next that β is not empty. At the mass singularity, V_β' vanishes for arbitrary values of $x \in \alpha + \beta$. On the other hand, it is certainly possible to choose a large enough α that $(q_\alpha - q_\beta)^2 \neq 0$ holds for almost all values of $x \in \alpha + \beta$. In fact, if $(q_\alpha - q_\beta)^2$ vanishes there, $\alpha + \beta$ is a subset of G by definition. We can, therefore, without loss of generality, regard it as a new α and repeat the above consideration. Assuming that $(q_\alpha - q_\beta)^2 \neq 0$, we may now integrate (4.9) at $z_\beta = 0$ and obtain

$$\int \frac{dx(\alpha)}{[V_\alpha(x) - i\epsilon]^d} \int \frac{F''\, dx(\beta)\, dx(\gamma)}{[V_\gamma']^{n_\gamma + f_\gamma}[(q_\alpha - q_\beta)^2]^{(n_\beta + f_\beta)/2}} \tag{4.10}$$

$$+ \text{ less singular terms,}$$

where $d = n_\alpha - f_\gamma - 2 + (n_\beta - f_\beta)/2$. Since $V_\alpha(x)$ vanishes everywhere, (4.10) will certainly diverge unless $d < 0$. On the other hand, $d < 0$ is not sufficient for convergence, since V_γ' and $(q_\alpha - q_\beta)^2$ may vanish for some values of $x \in \alpha$. To check this point, we have only to examine whether V_γ' vanishes for arbitrary values of x of a subset α' of α. No new divergence arises if such α' is not found. If α' is not empty, we repeat the above consideration for α' and see whether new d is less than zero or not. Since $V(x)$ is quadratic in x, all possible divergences will be detected by this procedure.

Going back from the subchain to the chain picture, we can easily translate the conditions $V_\alpha = 0$, $V_\beta' = 0$ to a property of the v matrix, namely,

$$v_{ij} = 0, v_{ik} = 0 \text{ for } i, j \in \alpha, k \in \beta, \alpha \subset G. \tag{4.11}$$

A necessary and sufficient condition for the convergence of the integral (4.1) at G is given by

$$d = n_\alpha - f_\gamma - 2 + \max\,[0, (n_\beta - f_\beta)/2] < 0, \tag{4.12}$$

where α, β are all possible sets of internal lines satisfying (4.11) and max (a, b) means the largest of its arguments.[12a]

We shall show next that possible values of n_α and n_β are strongly restricted by the requirement that mass singularities should take place for an arbitrary choice of external momenta. We mean by this that the relative orientation of the external momenta is arbitrary although their magnitudes may (or may not) be fixed by the mass shell condition. Hence $v_{ij} = 0$ may be satisfied for an arbitrary orientation of the external momenta if and only if it depends on at most one external momentum. This takes place only when i and j are direct neighbors, since $q_i - q_j$ is equal to the sum of the external momenta that enter the loop at the vertices between the lines i and j. Noting that each line has at most two neighbors, we see that not more than two of the off-diagonal elements in any row or column of the v matrix may vanish simultaneously. We therefore find that the only possible values of n_α and n_β are $n_\alpha = 1$, $n_\beta = 0, 1, 2$; and $n_\alpha = 2$, $n_\beta = 0$. In the following, mass singularities for $n_\alpha = 1$ and 2 will be called λ and m singularities, respectively. It should be emphasized that this restriction on n_α and n_β is valid only for general external momenta. If the external momenta are subject to some restrictions, some v_{ij} may vanish even if i and j are not direct neighbors. An important example of this sort is the forward scattering amplitude, which will be considered at the end of this section.

Equation (4.12) shows that the degree of divergence d depends on n_α and n_β but not on $n = n_\alpha + n_\beta + n_\gamma$ as a whole. Since $n_\alpha + n_\beta$ does not exceed 3 in general, this means that all types of mass singularities for $n > 3$ are already included in the case $n = 3$. However, the behavior of the latter is not difficult to analyze since the x integrations can be carried out explicitly in this case[13].

[12a] An estimate of divergence similar to (4.12) is made in reference 28 for the case of infrared divergence.

[13] Complete integration of an integral, which has the same structure as (4.1) except that the power of denominator

In the following, we shall therefore discuss some features of λ and m singularities for arbitrary n, making use of information obtained from the study of a diagram with $n = 3$.

(*a*) λ *singularity.* As is seen from (4.12), the integral (4.1) is finite at a λ singularity when $n_\beta = 0$ or 1. It may diverge only if $n_\beta = 2$. This means that the conditions

$$m_j = 0,\ m_i^2 = (q_i - q_j)^2,\ m_k^2 = (q_j - q_k)^2 \qquad (4.13)$$

are satisfied for successive lines i, j, and k. In other words, the λ singularity at $m_j = 0$ (or $x_j = 1$) may become divergent only when it is enhanced by the last two conditions of (4.13). Even then, it does not diverge if $f_\beta + f_\gamma \neq 0$. This happens in particular when j is a fermion line. In this case F contains a factor

$$\sum_k x_k \gamma^\mu (q_{j\mu} - q_{k\mu}) + m_j, \qquad (4.14)$$

as is seen from (2.14). It vanishes at $x_j = 1$ (and $m_j = 0$) and hence $f_\beta + f_\gamma = 1$. This explains why vanishing of the neutrino mass does not lead to an infrared-like divergence. In general, $f_\beta + f_\gamma$ will be zero if j is a boson line. The integral (4.1) is then divergent at $m_j = 0$ and behaves as $\ln m_j$ when this limit is approached in the order $[(q_i - q_j)^2 \to m_i^2,\ (q_j - q_k)^2 \to m_k^2]$, and finally $m_j \to 0$.[14] This λ divergence takes place only in association with vanishing boson mass. This and the necessity for strong enhancement is the characteristic feature of λ divergence. When $n \leq 2$, not all conditions of (4.13) can be satisfied. The integral (4.1) has therefore no λ divergence in these cases.

(*b*) m *singularity.* In this case the neighboring lines i and j satisfy

$$m_i^2 = m_j^2 = (q_i - q_j)^2 = 0. \qquad (4.15)$$

This means that all three lines meeting at a three-vertex must have zero mass. There is no suppression ($f_\beta + f_\gamma = 0$) in this case, since F cannot vanish identically at the (one-dimensional) singularity $x_i + x_j = 1$. Let us first assume that neither v_{ik} or v_{jl} vanishes, where k and l are direct neighbors of i and j, respectively. Then we find an m divergence of the form $\ln m_i$ at (4.15), if the limit is taken in the order $(q_i - q_j)^2 \to m_i^2$, $m_j \to 0$, and finally $m_i \to 0$.[14] No m divergence occurs if $n = 1$.

(*c*) *Overlapping of mass singularities.* If $v_{jl} = 0$ in addition to (4.15), masses and momenta of successive lines i, j, l satisfy (4.13) also. Thus, the integral (4.1) diverges at both $x_j = 1$ and $x_i + x_j = 1$. When j is a boson line, (4.1) shows simultaneous m and λ divergences of the form $(\ln m_i)(\ln m_j + c \ln m_i)$ if the limit is taken in the order $[(q_i - q_j)^2 \to m_i^2,\ (q_j - q_l)^2 \to m_l^2]$, $m_j \to 0$, and then $m_i \to 0$, where c is a numerical constant.[15] If $v_{ik} = 0$ holds in addition and if i is a boson line, (4.1) will also diverge as $(\ln m_j)(\ln m_i + c' \ln m_j)$ in an appropriate limit. Since singularities at $x_i = 1$ and $x_j = 1$ cannot occur simultaneously, these two divergences will appear as separate additive terms of (4.1). For the same reason, (4.1) does not contain a term of the form $(\ln m_i)(\ln m_l)$. When some of the lines i, j are fermion lines, the corresponding λ divergences will disappear from the above expressions.

In general, if we consider all mass singularities of (4.1), the condition (4.11) will be satisfied by various overlapping and nonoverlapping sets α. If we consider all α with $n_\alpha = 1$ (or $n_\alpha = 2$), the integration domain of (4.1) may be divided into subdomains so that each subdomain contributes at most one λ singularity (or m singularity) to the integral. When they are considered together, some α with $n_\alpha = 1$ may be contained in other α with $n_\alpha = 2$ and thus λ and m divergences will appear as a product. Consequently, (4.1) can always be written as a sum of terms, each of which is a product of at most two logarithmic factors.

Strictly speaking, the strength of mass singularity determined above is that of the dispersive part of the integral (4.1). Mass singularity of the absorptive part, which is obtained by replacing the denominator V^{-n+2} by $\delta^{(n-3)}(V)$, will in general be weaker than that of the dispersive part because of the stronger restriction on the domain of integration.

We have mentioned previously that more than two off-diagonal elements of the v matrix may vanish simultaneously if the external momenta are not completely arbitrary. In such a case, we may expect to find a stronger divergence than those considered above. Of these, an important case from the practical point of view is the amplitude for forward scattering, which is related to the total cross section by unitarity. Although it is not a pure mass singularity,[12] it may be related to the latter as is shown in the following. For this purpose, let us examine the property of the fourth-order diagram of Fig. 2(a), which describes the forward scattering of particles

is raised by one, has been carried out by G. Källen and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-fys. Skrifter **1**, No. 6 (1958), for the case $n = 3$. For our purpose, however, their result must be integrated further with respect to some mass.

[14] The value of the integral (4.1) depends on how the limit is approached in the space of n mass variables.

[15] The value of c depends on the limiting procedure.

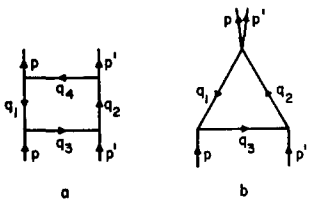

FIG. 2. The fourth-order Feynman diagram (a) describes the forward scattering of particles with momenta p and p'. Diagram (b) is obtained from (a) by contracting the line 4.

of momenta p and p'. Noting that q_3 and q_4 are equal, we find that $v_{34} = (m_3^2 + m_4^2)/2$. This has an anomalous feature in that it vanishes automatically when v_{33} and v_{44} vanish. Because of this, in addition to the mass singularities considered already, we find new singularities characterized by

(i) $m_3 = m_4 = 0$,

(ii) $m_3 = m_4 = 0$, $m_1^2 = p^2$ or $m_2^2 = (p')^2$, and

(iii) $m_3 = m_4 = 0$, $m_1^2 = p^2$, $m_2^2 = (p')^2$.

We may also consider the cases where m_1 and/or m_2 go to zero.

The behavior of these singularities may be seen clearly in the following manner: Since we are concerned with the limit $m_3 = m_4 = 0$, let us assume for simplicity that $m_3 = m_4$ even before the limit is taken. We shall also assume that 3 and 4 are boson lines. It is then easy to see that (4.1) is reduced to

$$-\frac{\partial}{\partial m_3^2}\int \frac{F\,dx_1\,dx_2\,dx_3\,\delta(1-x_1-x_2-x_3)}{\sum_{i,j}^{1,2,3} v_{ij}x_ix_j - i\epsilon}, \quad (4.16)$$

where the integral over x_1, x_2, x_3 is the amplitude for the process shown in Fig. 2(b). Thus, all singularities of Fig. 2(a) (with $m_3 = m_4$) may be obtained by differentiation from those of Fig. 2(b). But the latter can be determined from our general consideration, since Fig. 2(b) has no identical propagators. For instance, if Fig. 2(b) has a mass singularity of the form $\ln m_3$, that of Fig. 2(a) is proportional to $1/m_3^2$. This takes place in the case (iii). A physical example is the total cross section for Rutherford scattering. Because of the relation (4.16), we may include such "Coulomb" type singularities in our consideration of mass singularity. When the lines 3 and 4 are fermion lines, this argument must be slightly modified, since F vanishes at the singularity.

5. REDUCTION OF FEYNMAN AMPLITUDES

We shall now extend the consideration of the last section to general Feynman diagrams which contain several closed loops. According to Sec. 3, chains of a given diagram are classified into the sets A and $\bar{A}$ at a mass singularity. For the sake of convenience, we shall treat $\bar{A}$ as an assembly of subchains. Accordingly, $U(z)$ and $V(x, z)$ are generalized following the remark (d) of Sec. 2. A diagram obtained by contracting some (sub-) chains of the original diagram will be called a *reduced diagram*. In particular we obtain a reduced diagram A if all subchains of $\bar{A}$ are contracted. We shall show that the mass singularity of the integral (2.8) is the same as that of an appropriately reduced diagram B which contains A as a part.

Let us define $U_A(z)$ and $V_A(x, z)$ by applying (2.10) and (2.12) to the reduced diagram A. We shall also introduce $\bar{U}(z)$ by

$$U(z) = U_A(z)\bar{U}(z) + \text{higher-order terms of } z \text{ in } \bar{A}. \quad (5.1)$$

Let r_A be the number of independent closed loops of the reduced diagram A. As is easily seen, $\bar{U}(z) = 1$ if $r_A = r$. The mass singularity occurs at the origin of z_β, where β goes over all elements of $\bar{A}$. In this neighborhood, $V(x, z)$ may be written as

$$V(x, z) = V_A(x, z) + \sum_{\bar{A}} z_\beta V'_\beta + Q + \cdots, \quad (5.2)$$

where Q represents the sum of all terms quadratic in $z \in \bar{A}$ and

$$V'_\beta = \sum{}' U_A^{-1}(z)\left(z_\alpha \frac{\partial}{\partial z_\alpha}\right)\cdots\left(z_{\alpha'} \frac{\partial}{\partial z_{\alpha'}}\right)U_A(z) \times [V_\beta - (q_\beta \pm q_\alpha \pm \cdots \pm q_{\alpha'})^2]. \quad (5.3)$$

The summation $\sum'$ is over all elements $\alpha, \cdots, \alpha'$ of A such that $\{\beta, \alpha, \cdots, \alpha'\}$ is a C set of the original diagram. In deriving (5.3), we have used the formula (B11) (See Appendix B). Actually (5.2) contains another term linear in $z \in \bar{A}$. It was not written explicitly since it vanishes at the mass singularity and does not play any role in the following consideration.

As is seen from (5.2), the property of the Feynman amplitude (2.8) depends not only on $V_A(x, z)$, which vanishes at the mass singularity, but also on the behavior of V'_β. To examine this dependence, let us note that V'_β vanishes for arbitrary values of $z_\alpha, \cdots, z_{\alpha'} \in A$ if and only if

$$V_\beta - (q_\beta \pm q_\alpha \pm \cdots \pm q_{\alpha'})^2 = 0 \quad (5.4)$$

holds for all values of $x \in \alpha, \cdots, \alpha'$ and for all possible C-sets $\{\beta, \alpha, \cdots, \alpha'\}$. In general, (5.4) may not hold for a full chain β. However, noting that (5.4) can be written as

$$\sum_{i\in\beta} x_i[m_i^2 - (q_i \pm q_\alpha \pm \cdots \pm q_{\alpha'})^2] = 0, \qquad (5.5)$$

which is linear in $x_i \in \beta$, it is always possible to choose a largest subchain of β which satisfies (5.4). We shall define B as the sum of all chains of A and all subchains of $\bar{A}$ satisfying (5.4). Then the set $\bar{B}$ consists of all subchains β of $\bar{A}$ such that V'_β does not vanish identically for arbitrary values of x, $z \in A$.

If we define $U_B(z)$ and $V_B(x, z)$ for the reduced diagram B according to the general rule, we find that

$$\begin{aligned} U(z) &= U_B(z)\bar{U}(z) + \cdots, \\ V(x, z) &= V_B(x, z) + \textstyle\sum_{\bar{B}} z_\beta V'_\beta + \cdots, \end{aligned} \qquad (5.6)$$

where $\bar{U}(z)$ is the same as before, because it does not contain any $z_\beta \in \bar{A}$ such that $\{\beta, \alpha, \cdots, \alpha'\}$ is a C set of the original diagram and $\alpha, \cdots, \alpha' \in A$. If we introduce non-negative variables ρ and z' by

$$\begin{aligned} z_\alpha &= (1-\rho)z'_\alpha, \quad \textstyle\sum z'_\alpha = 1, \quad \alpha \in B, \\ z_\beta &= \rho z'_\beta, \quad \textstyle\sum z'_\beta = 1, \quad \beta \in \bar{B}, \end{aligned} \qquad (5.7)$$

the mass singularity will be found at $\rho = 0$. In this neighborhood, the integral (2.8) may be expressed in the form

$$\int \frac{F\, dz'[B]\, dz'[\bar{B}]\rho^{(n_{\bar{B}}-2r+2r_B-1)}\, d\rho}{U_B^2(z')\bar{U}^2(z')[(1-\rho)V_B(x, z') + \rho \sum_{\bar{B}} z'_\beta V'_\beta - i\epsilon]^{n-2r}} + \text{less singular terms}, \qquad (5.8)$$

where

$$dz'[B] = \delta(1 - \textstyle\sum_B z'_\alpha) \prod_B dz'(\alpha). \qquad (5.9)$$

Taking into account that $V'_\beta \neq 0$ for $\beta \in \bar{B}$, we can now carry out the integration of (5.8) in the neighborhood of $\rho = 0$ and obtain the formula

$$\int \frac{G_B\, dz[B]}{U_B^2(z)[V_B(x, z) - i\epsilon]^{n_B-2r_B}} + \text{less singular terms}, \qquad (5.10)$$

where

$$G_B = \int \frac{F\, dz[\bar{B}]}{\bar{U}^2(z)[\sum_{\bar{B}} z_\beta V'_\beta]^{n_{\bar{B}}-2r+2r_B}}. \qquad (5.11)$$

For simplicity we have replaced z' by z. It should be noted that $\sum_{\bar{B}} z_\beta V'_\beta$ may vanish for some values of x, $z \in A$. This may lead to a singularity of (2.8) stronger than that indicated by the denominator of (5.10). As was discussed in Sec. 4 for a similar situation, however, it corresponds to an enhancement of mass singularity arising from a subset of A,

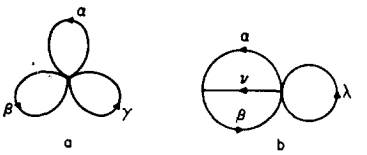

FIG. 3. Reduced diagrams of the chain diagram (b) of Fig. 1.

which can be treated as a separate problem. We may therefore regard G_B as a finite function of variables of the reduced diagram B. In this sense all information about the mass singularity of the Feynman amplitude is contained in the denominator of the first term of (5.10).

In general the diagram B may be decomposable in the sense that it consists of two components B' and B'' which share only a vertex of the *reduced* diagram B. Of course B' and B'' may be decomposed further into their components. If a component is no longer decomposable, it will be called *irreducible* with respect to this decomposition. Let us assume that B is decomposed into irreducible components $B_1, B_2, \cdots$. Similarly, A can be decomposed into its components A_i. We shall choose them so that A_i is a reduction of B_i for each i. Note that A_i is not necessarily irreducible. Corresponding to this decomposition, we find that

$$\begin{aligned} U_B(z) &= \prod_i U_{B_i}(z), \\ V_B(x, z) &= \sum_i V_{B_i}(x, z). \end{aligned} \qquad (5.12)$$

Making use of Feynman's formula (2.5) in a reverse fashion, it is now possible to show that the integral (5.10) can be factorized into a product of simpler integrals of the form

$$\int \frac{F_{B_i}\, \delta(1 - \sum_{B_i} z_\alpha) \prod_{B_i} dz(\alpha)}{[U_{B_i}(z)]^2[V_{B_i}(x, z) - i\epsilon]^{n_{Bi}-2r_{Bi}}}. \qquad (5.13)$$

Here F_{B_i} is the part of G_B which is relevant to the evaluation of the mass singularity arising from the component B_i. Thus, in order to determine the mass singularity of the Feynman amplitude (2.8), we have only to examine each irreducible component. Furthermore, the integral (5.13) can be written down immediately by applying the general formula (2.8) to the irreducible component B_i. This is a nice feature of the above reduction.

As an illustration, we shall show how the integral for the chain diagram of Fig. 1(b) may be reduced in the following cases:

(i) $B = \{\alpha, \beta, \gamma\}$, $\bar{B} = \{\lambda, \mu, \nu\}$. The reduced diagram B is shown in Fig. 3(a). It consists of three irreducible components $\{\alpha\}$, $\{\beta\}$, and $\{\gamma\}$. For

$z_\lambda, z_\mu, z_\nu \to 0$, $U(z)$ and $V(x, z)$ approach $z_\alpha z_\beta z_\gamma$ and $z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma$, respectively, and the integral (2.8) is approximated by

$$\int \frac{\delta(1 - z_\alpha - z_\beta - z_\gamma)\, dz(\alpha)\, dz(\beta)\, dz(\gamma)}{(z_\alpha z_\beta z_\gamma)^2 [z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma - i\epsilon]^{n_\alpha + n_\beta + n_\gamma - 6}}, \tag{5.14}$$

corresponding to (5.10). (Put $F = 1$ for simplicity.) We can now perform the z integration completely and obtain

$$\int \frac{dx(\alpha)}{[V_\alpha(x) - i\epsilon]^{n_\alpha - 2}} \int \frac{dx(\beta)}{[V_\beta(x) - i\epsilon]^{n_\beta - 2}} \times \int \frac{dx(\gamma)}{[V_\gamma(x) - i\epsilon]^{n_\gamma - 2}}, \tag{5.15}$$

which is a product of integrals for single closed loops.

(ii) $B = \{\alpha, \beta, \lambda, \nu\}$, $\bar{B} = \{\gamma, \mu\}$. The reduced diagram B is shown in Fig. 3(b). We find two irreducible components $\{\lambda\}$ and $\{\alpha, \beta, \nu\}$. Thus, in the neighborhood of $z_\gamma = z_\mu = 0$, the integral (2.8) is reduced to a product of

$$\int \frac{dx(\lambda)}{[V_\lambda(x) - i\epsilon]^{n_\lambda - 2}} \tag{5.16}$$

and

$$\int \frac{\delta(1 - z_\alpha - z_\beta - z_\nu)\, dz(\alpha)\, dz(\beta)\, dz(\nu)}{U_B^2(z)[z_\alpha V_\alpha + z_\beta V_\beta + z_\nu V_\nu - (z_\alpha z_\beta z_\nu / U_B)(q_\alpha - q_\beta + q_\nu)^2 - i\epsilon]^{n_\alpha + n_\beta + n_\nu - 4}}, \tag{5.17}$$

where

$$U_B(z) = z_\alpha z_\beta + z_\alpha z_\nu + z_\beta z_\nu. \tag{5.18}$$

6. CRITERION FOR THE DIVERGENCE OF MASS SINGULARITIES

We have found that an arbitrary Feynman amplitude can be reduced to a product of irreducible components in the neighborhood of a mass singularity. Thus, in order to determine the properties of a mass singularity of the whole amplitude, we have only to examine each irreducible component (5.13) separately. For this purpose it is convenient to regard each chain ξ of the set A_i as a sum of three subchains ξ_G, ξ_H, and ξ_K, where ξ_H and/or ξ_K may be empty, and define G, H, and K as assemblies of all ξ_G, ξ_H, and ξ_K, respectively. (We shall drop the suffix i in the following.) A new $U_A(z)$ is obtained from the old one by the substitution $z_\xi \to z_{\xi G} + z_{\xi H} + z_{\xi K}$ for all $\xi \in A$. Let us identify ξ_G with G_ξ of (3.4). We are thus interested in the behavior of the integral (5.13) in the limit where all z's of the sets H, K, and $D = B - A$ go to zero. In this neighborhood, $V_B(x, z)$ may be written as

$$V_B(x, z) = V_A(x, z) + \sum_D z_\delta V'_\delta + Q^0 + \cdots, \tag{6.1}$$

where

$$V_A(x, z) = V_G(x, z) + \sum_H z_\beta \tilde{V}_\beta + \sum_K z_\gamma \tilde{V}_\gamma + R + \cdots, \tag{6.2}$$

and $\tilde{V}_\beta$, $\tilde{V}_\gamma$ are defined by formulas similar to (5.3). Q^0 is the value of Q [defined by (5.2)] at $z_\beta = 0$, $\beta \in \bar{B}$, and R represents the sum of all terms quadratic in $z \in H + K$. We shall choose H as the largest set for which $\sum_H z_\beta \tilde{V}_\beta$ vanishes for all values of x, $z \in G + H$. Thus, we may regard $\tilde{V}_\gamma$ as non-vanishing insofar as the singularity at G is concerned, although it may vanish for some values of x, $z \in G$.

Making use of the method of Sec. 5, we can now carry out the integration of (5.13) with respect to the variable $\sum_K z_\gamma$. The most singular term of the resulting expression is given by

$$\int \frac{F'_B\, \delta(1 - \sum_{B-K} z_\alpha) \prod_{B-K} dz(\alpha)}{U_{B0}^2(z)[V_G(x, z) + \sum_H z_\beta \tilde{V}_\beta + \sum_D z_\delta V'_\delta + R_0 + Q_0^0 - i\epsilon]^{n_B - n_K - f_K - 2r_B}}, \tag{6.3}$$

where U_{B0}, R_0, and Q_0^0 are the values of U_B, R, and Q^0 at $z_\beta = 0$, $\beta \in K$. It is assumed that F_B behaves as $F'_B (\sum_K z_\gamma)^{f_K}$ for small $\sum_K z_\gamma$.

If non-negative variables λ, μ, and new z are introduced by

$$\begin{aligned} z_\alpha &\to (1 - \lambda - \mu) z_\alpha, & \textstyle\sum_G z_\alpha &= 1, \\ z_\beta &\to \lambda z_\beta, & \textstyle\sum_H z_\beta &= 1, \\ z_\gamma &\to \mu z_\gamma, & \textstyle\sum_D z_\gamma &= 1, \end{aligned} \tag{6.4}$$

(6.3) can be written in the neighborhood of $\lambda = \mu = 0$ as a sum of terms of the form

$$\int \frac{\lambda^{n_H + l - 1}\, d\lambda \mu^{n_D + m - 1}\, d\mu\, dz[G]\, dz[H]\, dz[D]}{U_G^2(z)[(1 - \lambda - \mu) V_G(x, z) + \lambda \sum_H z_\beta \tilde{V}_\beta + \mu \sum_D z_\gamma V'_\gamma + \lambda^2 R_1 + \mu^2 Q_1 - i\epsilon]^{n_B - n_K - f_K - 2r_B}}, \tag{6.5}$$

where $dz[G]$, etc., are defined by (5.9), $U_G(z)$ is obtained from $U_{B0}(z)$ by putting $z_\beta = 0$ for all $\beta \in H + D$, and $\lambda^2 R_1$ and $\mu^2 Q_1$ are essentially equal to R_0 and Q_0^0. We have also assumed that F' can be written as $\sum F_{lm}\lambda^l\mu^m$ for small λ and μ. The values of l and m such that $l + m$ is a minimum will be denoted as f_H and f_D in the following.

We shall now examine under what condition the integral (6.5) is finite at the mass singularity. Let us first consider the simplest case where both H and D are empty. In this case all terms of the denominator except for the first one are absent in (6.5). Furthermore, $n_B - n_K = n_G$ and $f_H = f_D = 0$. Since $V_G(x, z)$ vanishes for all values of x, $z \in G$, we find that (6.5) diverges unless

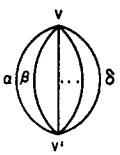

FIG. 4. A diagram representing an irreducible component which consists of one C set.

$$n_G - f_K - 2r_B < 0 \tag{6.6a}$$

is satisfied. Next consider the case where H is empty but D is not. Then, as is seen from the definition of D, Q_1 is nonvanishing for the values of x, $z \in G$ for which $\sum_D x_\gamma V'_\gamma = 0$ holds. Carrying out the μ integration at $\mu = 0$, we can reduce (6.5) to the form

$$\int \frac{dz[G]\, dz[D]}{U_G^2(z) Q_1^{(f_D+n_D)/2} [V_G(x, z) - i\epsilon]^{n_B - n_K - f_K - 2r_B - (n_D+f_D)/2}} + \cdots . \tag{6.7}$$

The integral (6.5) therefore diverges unless

$$n_G - f_K - 2r_B + (n_D - f_D)/2 < 0. \tag{6.6b}$$

In the same fashion, when D is empty but H contains some elements, we obtain the inequality

$$n_G - f_K - 2r_B + (n_H - f_H)/2 < 0. \tag{6.6c}$$

Finally, when both H and D are not empty, we find the additional condition

$$n_G - f_K - 2r_B + (n_D + n_H - f_D - f_H)/2 < 0. \tag{6.6d}$$

These inequalities may be combined into a single formula as follows:

$$\begin{aligned} d = n_G - f_K - 2r_G \\ + \max [0, \tfrac{1}{2}(n_D - f_D), \tfrac{1}{2}(n_H - f_H), \\ \times \tfrac{1}{2}(n_D + n_H - f_D - f_H)] < 0, \end{aligned} \tag{6.8}$$

where max $[a, b, c, d]$ means the largest number of its arguments. We have used the fact that $r_B = r_G$.

Of course, even if (6.8) is satisfied, the integral (5.13) may still diverge at singularities which arise from some subset G' of G. To find whether or not this is the case, we have only to repeat our analysis for all possible G'. It is important to note here that, although G is irreducible, G' may sometimes be regarded as reducible. In such a case, we have to apply the above consideration to each irreducible subset of G'.

7. IRREDUCIBLE COMPONENTS CONSISTING OF ONE C SET

We are now ready to examine how an irreducible component behaves in the vicinity of its mass singularity. Since we have discussed the case of one closed loop already, here we shall consider irreducible components with more than one closed loop. For simplicity, we shall assume in this and the next sections that all propagators in the diagram are different. The case where some lines are identical is treated in Sec. 9.

From our point of view, the simplest irreducible component B is the one which consists of just one C set $\{\alpha, \beta, \cdots, \delta\}$. The corresponding diagram has a structure such that all chains end at the two vertices v and v', as is shown in Fig. 4. In this section we shall examine this particular case in detail. The general case will be considered in Sec. 8. As is easily seen, $U_B(z)$ for this diagram can be written as

$$U_B(z) = z_\alpha z_\beta \cdots z_\delta (z_\alpha^{-1} + z_\beta^{-1} + \cdots + z_\delta^{-1}). \tag{7.1}$$

Since each chain consists of subchains, we have to make the substitution $z_\xi \to z_{\xi G} + z_{\xi H} + z_{\xi K}$ in (7.1) for all $\xi \in A$. The function $V_B(x, z)$ is given by (6.1) and (6.2) near the singularity, where $V_G(x, z)$, V'_δ and $\tilde{V}_\beta$ vanish at the mass singularity for all x and z. For general external momenta, this is possible only if each coefficient of these quantities does not depend on more than one external momentum. As is shown in the following, this puts a very strong restriction on the possible types of mass singularity. It is convenient to consider separately the cases where (I) all chains belong to A, and where (II) one of the chains does not belong to A. Cases where more than one chain does not belong to A will not be considered, since they reduce to the case of single closed loops, whose behavior has been examined already in Sec. 4.

In the case (I), the set D is empty and thus $n_D = 0$. According to the condition (3.4),

$$V_{\xi G}(x) = \partial V_{\xi G}(x)/\partial x_i = 0;\ i \in \xi_G,\ \xi \in A, \tag{7.2a}$$

and

$$\begin{aligned}&(q_{\alpha G} + q_{\beta G} + \cdots + q_{\delta G})^2 = 0,\\ &q_i(q_{\alpha G} + q_{\beta G} + \cdots + q_{\delta G}) = 0;\ i \in \xi_G,\ \xi \in A.\end{aligned} \tag{7.2b}$$

They must hold for arbitrary values of $x \in G$. The first equation tells us that each subchain ξ_G consists of at most two elements, as was shown in Sec. 4. The requirement that (7.2b) should not depend on the relative orientation of external momenta means that not more than one *external* vertex may exist between any pair of internal lines of the set G. [Equation (7.2b) gives no restriction on the number of internal vertices.] Thus, if we denote internal lines of each chain ξ (not subchain) by 1ξ, 2ξ, $\cdots$, successively, starting from the vertex v, we find that (7.2b) may hold only in the following cases: (a) Each subchain ξ_G consists of one element 1ξ, (b) $\xi_G = \{1\xi\}$ except that $\alpha_G = \{2\alpha\}$ for a suitably chosen $\alpha \in A$, (c) $\xi_G = \{1\xi\}$ except that $\alpha_G = \{1\alpha, 2\alpha\}$. In other words, (7.2b) localizes the set G to a neighborhood of the vertex v. Of course, this may hold just as well with respect to the vertex v'. However, we shall define hereafter the vertex v by the property that the *canonical form* (a), (b), or (c) holds with respect to v. The chains as well as the C set will then be called *oriented*.

Without loss of generality, we shall choose $q_{1\beta} = \cdots = q_{1\delta} = 0$. Then the remaining q's are completely determined as linear combinations of external momenta, and are nonvanishing in general, except that $q_{1\alpha}$ vanishes if the vertex v is internal.[16] Equation (7.2b) is then reduced to the simple form

$$q_{\alpha G}^2 = 0, \qquad q_i q_{\alpha G} = 0, \qquad i \in \alpha_G \tag{7.3}$$

in all three cases mentioned above.

According to (5.3), both

$$V_{\xi H} - (q_{\xi H} - q_{\xi G})^2 = 0$$

and

$$V_{\xi H} - (q_{\xi H} + \sum_{\eta \neq \xi} q_{\eta G})^2 = 0$$

must hold in order that the enhancement of the mass singularity takes place. For our particular choice of q, these equations are equivalent to

$$\begin{aligned}&m_j^2 - q_j^2 = 0, \qquad q_i q_{\alpha G} = 0;\\ &\qquad j \in \xi_H, \quad \xi = \alpha, \beta, \cdots, \delta.\end{aligned} \tag{7.4}$$

[16] Here we use the words "internal" and "external" relative to the set $\{\alpha, \beta, \cdots, \delta\}$. Thus they are different from those used elsewhere.

The second equation shows that no enhancement is possible unless $q_{\alpha G} = 0$ or $q_i = 0$. We are now ready to discuss the degree of divergence of mass singularities.

Case (a). This takes place when

$$m_{1\xi} = 0 \quad \text{for all} \quad \xi, \quad \text{and} \quad q_{\alpha G}^2 = q_{1\alpha}^2 = 0. \tag{7.5}$$

We find that $n_G = r + 1$ for $r \geq 2$. If the vertex v is not internal and thus $q_{1\alpha}(= q_{\alpha G})$ is different from zero, we find that $n_H = 0$ according to (7.4). We then obtain $d = 1 - r - f_K$ from (6.8). Since no subset of G gives divergence in this case, our integral is finite for any r.

When the vertex v is internal or $q_{1\alpha} = 0$, we find $n_H \leq r + 1$, which leads to $d \leq \max[1 - r - f_K, (3 - r - f_H - 2f_K)/2]$. Thus the irreducible component may diverge only when $r \leq 3$. If some of $1\alpha, 1\beta, \cdots, 1\delta$ are fermion lines, the mass singularity is finite for any r since $f_H + f_K \geq 2$ holds as is seen from (2.14). Another case in which the strength of mass singularity may be suppressed is the interaction of charged scalar meson with photon. If a virtual photon 1α is emitted from the three-vertex v, we find $r = 2$, $f_H + f_K \geq 1$, and thus $d \leq 0$, taking into account that F contains a factor $Q_{1\beta} - Q_{1\gamma}$, where the Q's are defined by (2.15) and vanish at the mass singularity. We may therefore obtain a logarithmic mass divergence in this case. If two photons are emitted at the four-vertex v, we obtain $r = 3$, $f_H + f_K = 0$, since F does not vanish now. We may thus find a logarithmic divergence again.

Case (b). This singularity occurs when

$$\begin{aligned}&m_{2\alpha} = m_{1\beta} = \cdots = m_{1\delta} = 0 \qquad (m_{1\alpha} \neq 0),\\ &\qquad\qquad q_{\alpha G}^2 = q_{2\alpha}^2 = 0.\end{aligned} \tag{7.6}$$

In general, we cannot choose $q_{2\alpha} = q_{\alpha G} = 0$ because it demands a specific relation between the external momentum attached to the vertex v and the one which appears between the lines 1α and 2α. An exception arises when they are the only external momenta. But we find $d < 0$ in this case since there is no enhancement. In all other cases, v must be internal and thus $q_{1\alpha} = 0$ and $q_{2\alpha} \neq 0$. This means that $n_H \leq 1$ and hence $d \leq 1 - r - f_K + \max(0, (1 - f_H)/2) < 0$. Thus the integral is again finite for any value of $r(\geq 2)$.

Case (c). This occurs only when the vertex v is internal and

$$\begin{aligned}&m_{1\alpha} = m_{2\alpha} = 0, \quad m_{1\beta} = \cdots = m_{1\delta} = 0,\\ &\qquad q_{1\alpha} = 0, \qquad\qquad q_{2\alpha}^2 = 0.\end{aligned} \tag{7.7}$$

We find that $n_G = r + 2$ $(r \geq 2)$ while $n_H = 0$

because $q_{\alpha G} \neq 0$. Thus we obtain $d = 2 - r - f_K$ for $r \geq 2$. We might therefore conclude that the integral is finite for all r except $r = 2$. However, since an arbitrary subset G' of G containing $\{1\alpha, 2\alpha\}$ is always reducible to single closed loops, the integral actually contains a logarithmic m-divergence for any r. Since this singularity is not sensitive to the value of $m_{1\beta}, \cdots, m_{1\delta}$ (even when $r = 2$), it may be regarded as an m-divergence arising from the chain α.

We shall now consider the case (II). If one assumes that $D = \{\alpha\}$, $A = \{\beta, \cdots, \delta\}$, the threshold condition and enhancement condition are given by

$$V_{\xi G}(x) = \partial V_{\xi G}(x)/\partial x_i = 0; \; i \in \xi_G, \; \xi \in A, \quad (7.8)$$

and

$$V_\alpha(x) - (q_\alpha + q_{\beta G} + \cdots + q_{\delta G})^2 = 0,$$
$$V_{\xi H}(x) - (q_{\xi H} - q_{\xi G})^2 = 0; \quad \xi \in A, \quad (7.9)$$

respectively. These equations can be treated in a manner similar to (7.2) and (7.4). We find that a mass singularity occurs only in the following cases:

Case (d). $\xi_G = \{1\xi\}$ for all $\xi \in A$, or

$$m_{1\xi} = 0, \quad \xi \in A, \quad m_i^2 = q_i^2 \neq 0, \quad i \in \alpha, \quad (7.10)$$

where we have chosen $q_{1\beta} = \cdots = q_{1\delta} = 0$ as before. This shows that the subchain α, being irreducible, consists of either one or two elements. The second possibility is found only if both v and v' are external and $\xi = \xi_G$ for all $\xi \in A$. This leads us to $n_G = r$, $n_D = 2$, $n_H = 0$, and $d = \max[-r - f_K, 1 - r - f_K - (f_D/2)] < 0$. In all other cases, we obtain $n_D = 1$, $n_H \leq r$. This also gives $d < 0$. Thus the integral is finite for any mass singularity of the type (7.10).

Case (e). $\xi_G = \{1\xi\}$ for all $\xi \in A$ except that $\beta_G = \{2\beta\}$, or

$$m_{2\beta} = m_{1\gamma} = \cdots = m_{1\delta} = 0,$$
$$m_i^2 = (q_i + q_{2\beta})^2 \neq 0, \quad i \in \alpha, \quad (7.11)$$

which may be satisfied only if the number of external vertices between the line 2β and line $i(\in \alpha)$ does not exceed one. Thus, the vertex where different chains meet must be internal. This may be satisfied at either or both of v and v'. In the first case, we obtain $n_D = 1$, $n_H \leq r + 1$, and hence $d \leq -r - f_K + \max[0, (1 - f_D)/2, (1 + r - f_H)/2, (2 + r - f_D - f_H)/2]$. The second case occurs only if $\alpha = \{1\alpha, 2\alpha\}$, $\beta = \{1\beta, 2\beta, 3\beta\}$, and $\xi = \{1\xi\}$ for the rest. This gives $d = -r - f_K + \max[0, (2 - f_D)/2, (2 - f_H)/2, (4 - f_D - f_H)/2]$. Both cases give $d = 0$ if $r = 2$ and $f_K = f_D = f_H = 0$. We find $d < 0$ otherwise. However, we may not conclude that the integral is finite for all r except $r = 2$. This is because an arbitrary subset G' of G containing $\{2\beta\}$ is reducible to single closed loops, and the loop β gives a λ divergence independent of other loops (namely, for any r) if the mass singularity at $m_{2\beta} = 0$ is enhanced by the lines 1β and 3β. Note that this enhancement is also a necessary condition for the divergence of the irreducible component as a whole. More careful analysis may reveal that it is in fact an ordinary λ divergence of the chain β.

In summary, the mass singularity of an irreducible component consisting of one C set is localized to the immediate neighborhood of the vertex v. If there is no enhancement, no new divergence is found at the mass singularity. (We may find an m divergence. But it arises from one of the chains rather than the C set as a whole.) Since the integral may diverge when some of the lines 1ξ are fixed on the mass shell, this result shows that these divergences disappear when such a restriction is removed. When there is strong enough enhancement, the integral may exhibit new kind of mass divergence besides that of the λ type. However, it becomes finite if some of the lines $1\alpha, 1\beta, \cdots, 1\delta$ are fermion lines. In the case of photon-meson interaction, the strength of mass singularity may also be suppressed but it still may diverge logarithmically in some cases.

8. GENERAL IRREDUCIBLE COMPONENTS

Let us now consider a general irreducible component which consists of several overlapping C sets. This integral can be expressed by the general formula (2.8), if one assumes that every chain of the diagram belongs to the set B. The mass singularity of the integral is determined by the requirement that all equations of (3.4) are satisfied simultaneously. To examine its properties, let us first assume that every chain of the diagram is connected and thus has just two ends. We also assume that $B = A$.

Let us consider a C set $\{\alpha, \beta, \cdots, \delta\}$. Then one can cut all chains $\alpha, \beta, \cdots, \delta$ in such a way that the subchains $\alpha_G, \beta_G, \cdots, \delta_G$ are found on the same half of the divided diagram. These chains will satisfy Eq. (3.4b) only if at most one external line is attached to this portion of the divided diagram. From this it follows that the C set $\{\alpha, \beta, \cdots, \delta\}$ must be of the canonical form (a), (b), or (c) of Sec. 7, if the chains $\alpha, \beta, \cdots, \delta$ are oriented according to the convention of the last section. When all C sets are oriented in this manner, we find that all chains of the diagram are of the type $\xi_G = \{1\xi\}$ except that one chain, say, α, may be any of the types $\alpha_G = \{1\alpha\}$, $\{2\alpha\}$, or $\{1\alpha, 2\alpha\}$.

Actually, Eq. (3.4b) is far more restrictive than the above result. This follows from the fact that, when the irreducible component is cut in the manner described above, one part of the divided diagram must not contain more than one external line for *any* C set of the diagram. To see this more explicitly, let us note that each chain in general belongs to several C sets and thus is oriented as many times as these C sets are oriented. Frequently we find that a chain is oriented in different directions by different C sets. Such a chain will be called *unoriented.* We shall now try to see how some chains are unoriented while others are not. For this purpose, let us choose a chain α whose orientation is unambiguously fixed by (3.4). A chain β may then be oriented if and only if the lines 1α and 1β are on the same half of the divided diagram for arbitrary choice of the C set which contains both α and β. The chain β has such a property, for instance, if an end of 1β touches an end of 1α.[17] The set of all these β will be denoted as $\Gamma(\alpha)$. All other chains are unoriented, since their orientation depends on how the diagram is cut into two parts.

However, not all chains of $\Gamma(\alpha)$ may be oriented simultaneously. In fact, if α' is oriented in addition to α, the set $\Gamma(\alpha')$ must contain α but not necessarily the whole $\Gamma(\alpha)$. If we denote by Γ the set of all chains that belong to both $\Gamma(\alpha)$ and $\Gamma(\alpha')$, we find that Γ is either (i) a C set consisting of all chains which share an internal vertex with α and α', or (ii) a C set consisting of three chains of which two are α and α'. In any case, oriented chains can be found only in a particular C set. All other chains are unoriented and therefore must have the property that $\xi = \xi_G$. Thus, these chains, except possibly one, must not only be of the type $\xi_G = \{1\xi\}$, but they must not contain subchains ξ_H which enhance the singularity. Enhancement of this kind may then arise only from chains of the set Γ. The exception mentioned above is found when all ξ of Γ are of the type $\xi_G = \{1\xi\}$ and some of them have no enhancement. In this case one of the unoriented chains may be of the form $\xi_G = \{1\xi, 2\xi\}$, or $\xi_G = \{2\xi\}$, $\xi_H = \{1\xi\}$. Diagrams which do not fit with these specifications cannot have mass singularity of the type being discussed.

A criterion for the convergence of the integral (2.8) at a mass singularity is given by (6.8). We shall first show that (6.8) is indeed satisfied by all irreducible components which do not contain any three-vertices. It is enough to prove it for the diagrams consisting of four-vertices only, since the integral is even more strongly convergent in other cases. From the absence of three-vertices, it follows easily that the only arbitrary solution of (3.4) is of the type $\xi_G = \{1\xi\}$ and $\xi_H =$ empty for all ξ. Thus n_H is zero and n_G is equal to the number of lines labeled 1ξ which in turn is the total number of chains. Since the latter is $2r - 2 + w$, where $2w$ is the number of three-vertices of the *chain* diagram, we find that the degree of divergence d is equal to $(2r - 2 + w) - 2r = -2 + w$. However, (3.4) permits in general $w \leq 1$ only. Thus, the integral (2.8) is finite at the mass singularity, insofar as the numerator F is nonsingular.

We shall now consider the diagrams which consist of three-vertices only. In this case, the number of chains is equal to $3r - 3$. Thus the integral (2.8) may diverge for sufficiently large r, unless the numerator F vanishes strongly at the same time. In the following, we shall examine three typical cases of this sort.

The first example is an amplitude resulting from trilinear direct couplings of scalar fields. Since the interaction does not depend on the derivatives of field operators, we obtain $f_K = f_D = f_H = 0$ and thus $d \geq r - 3$, as is seen from (6.8). This means that as r increases, an unlimited number of diagrams can be found with stronger and stronger mass divergences. As is easily seen, the coupling constants for such interactions have the dimension of mass. To the extent that they are just constants with no relation to the masses of particles involved, however, they do not compensate the mass divergences which arise from the dynamical singularities of propagators. Remembering that field theories with such couplings have only three Feynman diagrams which show divergence due to high energy virtual quanta,[18] we may recognize a certain complementarity between the ultraviolet divergence and mass divergence. Since no such interaction has been found thus far, and since there seems to be some theoretical reason for its absence,[19] we shall ignore interactions of this type in the following.

Next we shall consider a diagram in which each vertex describes emission of a photon by a charged scalar particle. Following the considerations at the beginning of this section, we find that the only

[17] A chain β may also be oriented by α if centers of the chains α and β can be connected by a continuous line which does not intersect with any other chain when the diagram is properly drawn on a sheet of paper. However, we have not found an unambiguous definition of "proper."

[18] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), p. 352.

[19] G. Baym, Phys. Rev. **117**, 886 (1960). However, Baym's argument seems to be not conclusive.

possible cases are (i) $\xi_G = \{1\xi\}$ for all ξ, (ii) $\xi_G = \{1\xi\}$ except that $\alpha_G = \{2\alpha\}$, where 1α is a scalar particle line, (iii) the same as (ii) but 1α is a photon line, (iv) $\xi_G = \{1\xi\}$ except that $\alpha_G = \{1\alpha, 2\alpha\}$, where 1α is a scalar particle line, and (v) the same as (iv) but 1α is a photon line. Obviously, n_G is equal to $3r - 3$ for the first three cases and $3r - 2$ for the rest. To evaluate the degree of divergence d for the case (i), let $n_{i,j}$ be the number of vertices where i boson lines labeled 1ξ and j photon lines labeled 1ξ meet. Then the total numbers of such boson lines and photon lines are $(2n_{2,1} + 2n_{2,0} + n_{1,1} + n_{1,0})/2$ and $(n_{2,1} + n_{1,1} + n_{0,1})/2$, respectively. Since their sum is equal to $3r - 3$, we obtain $(3n_{2,1} + 2n_{2,0} + 2n_{1,1} + n_{1,0} + n_{0,1}) = 6r - 6$. Making use of $n_{1,1} + n_{2,0} = 0$ or $= n_{1,0} + n_{0,1}$ and $f_H + f_K \geq n_{2,1} + n_{2,0}$,[20] we find that

$$f_H + f_K \geq 2r - 2 - \max\,[n_{1,1}, (n_{1,0} + n_{0,1})/3].$$

This leads us to the inequality

$$d \leq -1 - r + \max\,[n_{1,1}, (n_{1,0} + n_{0,1})/3] + \max\,[f_H, (n_H + f_H)/2]. \qquad (8.1)$$

In general, we find that $n_{1,0} + n_{0,1} \leq 3$, $f_H \leq 2$, $n_H \leq 3$, and $n_{1,1} \leq 2$. Thus we obtain $d < 0$ for any $r > 3$. When $r = 3$, f_H may take the value 2 only if $n_{1,1} = 0$. This leads us to $d < 0$. We may obtain $d = 0$ only for $r = 3$, $f_H = 1$, and $n_{1,1} = 2$. This is the only case in (i) where our integral may diverge at the mass singularity. In the cases (iii) and (v), the lower bound for $f_H + f_K$ is given by the same inequality as above. For (ii) and (iv), on the other hand, this bound must be lowered by one. We also find that $n_H = f_H = 0$ for (iv) and (v), while $n_H \leq 1$ for the cases (ii) and (iii). In all of these four cases, the integral (2.8) is found to be finite at the mass singularity. It is to be noted however that an m divergence may arise from the chain α in the cases (iv) and (v).

Finally, we consider the case where each vertex is surrounded by two fermion lines and one boson line. As is easily seen, the consideration of the last paragraph applies to this case if $n_{i,j}$ is now regarded as the number of vertices where i fermion lines labeled 1ξ and j boson lines labeled 1ξ meet. We have only to note that a lower bound for $f_H + f_K$ should now be raised by $(n_{1,1} + n_{1,0})/2$. The right-hand side of (8.1) is then reduced by $(n_{1,1} + n_{1,0})/2$. Thus, a diagram of this type gives no new mass

[20] $n_{1,1} + n_{2,0}$ is equal to 0 or $n_{1,0} + n_{0,1}$ depending on whether the tail ends of oriented chains meet at a vertex or a vertex part. The inequality $f_H + f_K \geq n_{2,1} + n_{2,0}$ follows from the remark (c) of Sec. 2.

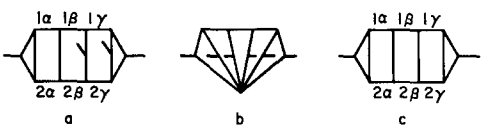

FIG. 5. The diagram (a) contains chains disconnected by vertex parts. If these chains are all of the type $\xi_G = \{1\xi\}$, its mass singularity cannot be stronger than that of the diagram (b). The diagram (c) contains chains of the type $\xi_G = \{1\xi, 2\xi\}$.

divergence except possibly an m divergence arising from one of its chains.

We shall not examine explicitly the more general case where vertices of various kinds are mixed. Since three-vertices give the strongest mass singularity, however, it is unlikely that any new divergent mass singularity would be found besides those discussed already.

We shall now remove the assumption that each chain be connected. A chain may be disconnected, for instance, if self-energy parts are inserted in the diagram. Since it gives rise to identical propagators, however, it will be discussed in Sec. 9. Thus, we consider here only the cases where some of the chains consist of two lines separated from each other by vertex parts. An example is given by the chains α, β, γ of Fig. 5(a). Note that the pair 1α, 2α or 1β, 2β has only one external line in between, while 1γ and 2γ are separated by more than one external line. This means that γ_G consists of 1γ only and that 2γ cannot enhance it. Thus mass singularity of Fig. 5(a) is not altered even if the line 2γ is contracted. In fact, if all chains α, β, γ are of the type $\xi_G = \{1\xi\}$, its mass singularity cannot be stronger than that of Fig. 5(b) which has no disconnected chain. This argument can be generalized to arbitrary irreducible components. Thus, in such a case, the problem can be reduced to the one which was solved in the above.

If some of the chains α, β are of the type $\xi_G = \{1\xi, 2\xi\}$ in Fig. 5(a), we find that no enhancement of mass singularity is allowed. Its mass singularity cannot be stronger than that of Fig. 5(c), which has only two external lines and in which α, β, γ are all of the type $\xi_G = \{1\xi, 2\xi\}$. Essentially the same situation is found for any irreducible component A. If we contract all lines 2ξ of $\xi_G = \{1\xi, 2\xi\}$ in A, we obtain a new diagram A', which can be analyzed by our method. Since n_G for A is greater than that of A', A may in general have a stronger mass singularity than A'. However, in the important cases such as the photon-meson and boson-fermion interactions, the value of $n_G - f_K$ is found to be the same for both A and A'. Thus, even if we remove

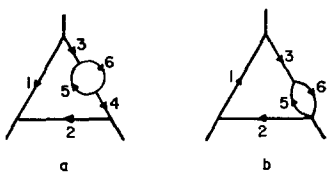

FIG. 6. The diagram (a) contains identical propagators caused by insertion of a self-energy part. The diagram (b) is obtained from (a) by contracting the line 4.

the restriction on chains, we may not encounter with any mass singularity stronger than those found already.

We shall not discuss here the case where $B - A$ is not empty, since it will, in general, give a mass singularity not stronger than the case $B = A$.

9. SELF-ENERGY PARTS

Thus far we have considered Feynman amplitudes which do not contain any identical propagators. In order to determine the mass singularity of a general amplitude, we have to know how our considerations of the last sections are modified when identical propagators are involved. Such a situation arises, for instance, when self-energy parts are inserted in arbitrary lines of a given diagram. We should therefore like to know whether or not it induces any new mass divergence or enhances the singularities considered previously.

To find an answer to this question, we shall first examine the property of a simple Feynman diagram of Fig. 6(a) which consists of three chains $\alpha = \{1, 2, 3, 4\}$, $\beta = \{5\}$, $\gamma = \{6\}$. We are interested in the case where chains β and γ form a self-energy part of the second order. We shall therefore assume that $q_3 = q_4$ and $m_3 = m_4$. A straightforward application of the formula (2.8) gives the amplitude

$$\int \frac{F z_\alpha^3\, dz_\alpha\, dz_\beta\, dz_\gamma\, \delta(1 - z_\alpha - z_\beta - z_\gamma)\, dx(\alpha)\, dx(\beta)\, dx(\gamma)}{U^2(z)[z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma - (z_\alpha z_\beta z_\gamma / U)(q_\alpha + q_\beta + q_\gamma)^2 - i\epsilon]^2}. \tag{9.1}$$

The self-energy term is not yet separated in this expression. In order to perform the separation, it is actually more convenient to parametrize the amplitude in two steps, first for lines 5 and 6 and then for the entire diagram. After the first parametrization is made, the contribution of the lines 5 and 6 can be written as

$$\int_{m_6^2}^{\Lambda^2} dL^2 \int \frac{F_{56}\, dx_5 x_6\, dx_6\, \delta(1 - x_5 - x_6)}{x_5 m_5^2 + x_6 L^2 - x_5 x_6 (k_\alpha + q_4)^2 - i\epsilon}, \tag{9.2}$$

where Λ is an ultraviolet cutoff, k_α is the variable of integration of the chain α, and F_{56} is the part of F arising from lines 5 and 6. Noting that F_{56} can be written as[21]

$$F_{56} = a m_4^2 + b[(k_\alpha + q_4)^2 - m_4^2], \tag{9.3}$$

where a is a finite constant and b is an invariant function of $(k_\alpha + q_4)^2$ finite at $(k_\alpha + q_4)^2 - m_4^2 = 0$ (for definiteness, the line 4 is assumed to be a boson line), one finds easily that (9.2) becomes

$$[(k_\alpha + q_4)^2 - m_4^2] \int_{m_6^2}^{\Lambda^2} dL^2 \int \frac{dx_5 x_6\, dx_6\, \delta(1 - x_5 - x_6)}{x_5 m_5^2 + x_6 L^2 - x_5 x_6 (k_\alpha + q_4)^2 - i\epsilon} \left[b - \frac{a m_4^2}{m_4^2 - m_5^2/x_6 - L^2/x_5 + i\epsilon} \right] \tag{9.4}$$

after the mass m_4 is renormalized. Since $[(k_\alpha + q_4)^2 - m_5^2/x_6 - L^2/x_5 + i\epsilon]^{-1}$ has the form of a propagator as a function of $k_\alpha + q_4$, the integration of the whole amplitude with respect to k_α can be carried out easily. Taking account of the fact that $(k_\alpha + q_4)^2 - m_4^2$ of (9.4) cancels one of the propagators corresponding to the lines 3 and 4, we find that the integral (9.1) should be replaced by

$$\int_{V_\gamma}^{\Lambda^2} dV_\gamma' \int \frac{\delta(1 - z_\alpha - z_\beta - z_\gamma) z_\alpha^2\, dz_\alpha\, dz_\beta z_\gamma\, dz_\gamma\, dx(\alpha)\, dx(\beta)\, dx(\gamma)}{U^2(z)[z_\alpha V_\alpha + z_\beta V_\beta + z_\gamma V_\gamma' - (z_\alpha z_\beta z_\gamma / U)(q_\alpha + q_\beta + q_\gamma)^2 - i\epsilon]^2} \cdot \left[b' - \frac{a' m_4^2}{m_4^2 - (z_\beta + z_\gamma)(V_\beta / z_\gamma + V_\gamma'/z_\beta) + i\epsilon} \right], \tag{9.5}$$

where a' and b' are derived from a and b of (9.4) and α now stands for the chain $\{1, 2, 3\}$ rather than $\{1, 2, 3, 4\}$.

If $m_4 < m_5 + m_6$, the last factor of (9.5) is finite throughout the domain of integration. Thus, (9.5) has essentially the same structure as the amplitude

[21] In the general case, the boundedness of a and b is a restriction on the kind of interaction.

for the diagram of Fig. 6(b), which does not contain identical propagators. In particular, the degree of divergence of the mass singularity of the former is the same as that of the latter. Thus our problem can be reduced to the case to which the general method developed in the last sections applies. This reduction may be extended to the general case. Thus the appearance of identical propagators due to self-energy insertion in the internal lines will not give rise to any new mass divergence. As a matter of fact, this result is more or less what one would expect insofar as one follows the usual procedure of mass renormalization of internal lines.

When $m_4 > m_5 + m_6$, the a' term of (9.5) becomes singular inside of the domain of integration. We may write it as a sum of principal value and δ-function parts. Then, mass singularity of the amplitude containing the first part can be treated by the method discussed above, since this singularity is not directly related to a mass singularity. The second part appears only when the lines 5 and 6 are on the mass shell. To find mass singularity of the corresponding amplitude, it is convenient to make use of the fact that the amplitude (9.1) is a derivative of the amplitude for the diagram of Fig. 6(b) with respect to m_3^2. In this case, we should not regard the lines 5 and 6 as self-energy insertions. Note also that this mass singularity arises from the chain α and not from the chain β or γ.

10. MASS SINGULARITY OF GENERAL FEYNMAN AMPLITUDES

In the preceding sections, we have seen how proper Feynman amplitudes will behave in the neighborhood of its mass singularities. We have found in particular that each mass singularity arises from a small neighborhood of a vertex and is not very sensitive to the over-all structure of the Feynman diagram. Physically, however, we should like to characterize an amplitude by its external lines rather than the details of its internal structure. We shall therefore reconsider our result from this point of view, classifying Feynman diagrams according to the number N of their external lines. In particular, it is important to find which diagram has the strongest mass singularity among those of given order and given initial and final states. (Of course, we shall exclude trilinear direct couplings of scalar fields from our consideration.) For the moment, let us consider proper diagrams only.

$N = 0$. Since such a diagram has no external line, there is no enhancement of mass singularity either. Hence, the mass singularity of any diagram

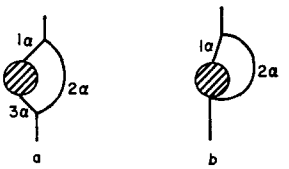

FIG. 7. General structure of diagrams in which two external lines are attached to the same chain.

which describes a vacuum-to-vacuum transition is always finite insofar as φ^3 interactions of scalar fields are excluded.

$N = 2$. The external lines are attached either to the same chain or to different chains. In the first case, the diagram must look like Fig. 7(a), where the shaded part is a structure consisting of internal lines only. If the lines 1α and 3α are identical, we may examine the mass singularity of Fig. 7(b) instead, as was shown in Sec. 9. For any irreducible component of Fig. 7(b), we find that $n_H + n_K \leq 1$. Thus, as is seen from Sec. 8, no mass divergence may arise from any multi-loop irreducible component. The only mass singularity which may diverge is the one that comes from the chain α when the condition

$$m_{1\alpha}^2 = m_{2\alpha}^2 = p^2 = 0 \tag{10.1}$$

is satisfied, where $p = q_{1\alpha} - q_{2\alpha}$ is the external momentum. Thus, if $p^2 \neq 0$, no mass divergence of the m type appears in such an amplitude. Even if p^2 vanishes, the amplitude is still finite, since it is proportional to p^2 [or $(p^2)^{1/2}$] on dimensional grounds.[22] Of course, we also find a λ singularity of the chain α when

$$m_{2\alpha} = 0, \qquad m_{1\alpha}^2 = p^2;$$
or
$$m_{1\alpha} = 0, \qquad m_{2\alpha}^2 = p^2. \tag{10.2}$$

This does not diverge as far as the leading term in the expansion of the amplitude in powers of $p^2 - M^2$ is concerned (M is the mass of the external line). However, it may give rise to a λ divergence in the second term which contributes to the wave function renormalization.

Let us now consider the second case. Again no mass divergence may arise from multi-loop irreducible components, even when $n_H = 2$. However, the amplitude may now contain several single-loop irreducible components which satisfy the conditions of the type (10.1) simultaneously. Obviously, the singularity will be the stronger, the more single-loop components belong to this type. As is easily seen, in a given order of perturbation theory, the largest number of such loops will be found in a

[22] Strictly speaking, this is true only if the dimension of coupling constants is not equal to a positive power of mass.

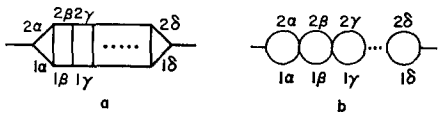

FIG. 8. The diagram (a) is an example of diagrams with two external lines which have the strongest mass singularity. The reduced diagram (b) is obtained by contracting all vertical lines of (a).

diagram of the type shown in Fig. 8. When all chains represented by the vertical lines are contracted, we obtain the diagram of Fig. 8(b), which consists of single closed loops only. Let the number of closed loops of Fig. 8(a) be s. Then the integral behaves something like m or m^2 times $(\ln m)^s$ in the neighborhood of (10.1), where m stands for the masses $m_{1\alpha}$, $m_{1\beta}$, etc.[22] As far as the λ singularity at (10.2) is concerned, it is finite for the leading term of the expansion in $p^2 - M^2$, but may give one power of logarithmic λ divergence to the second term. Incidentally, the case $N = 1$ may be treated as a special case of Fig. 7(a) in which the two external lines come out from the same vertex.

$N = 3$. There are three cases depending on whether (i) all three external lines belong to the same chain, (ii) two of the external lines are attached to one chain, or (iii) all external lines are attached to different chains. In the cases (i) and (ii), mass singularities arising from internal vertices are non-divergent. Only in the case (iii), divergent mass singularities other than λ or m divergences may be found. However, they are at most logarithmically divergent. Aside from them, mass divergences may arise only from irreducible components of single closed loop type. In a given order of perturbation theory, diagrams with strongest mass singularity are those which may be reduced at the mass singularity to the form shown in Fig. 9 by contracting smallest number of chains. The leading term of the corresponding amplitudes would look like $(a \ln \lambda + b \ln m)(\ln m)^s$ since each loop may contribute one m divergence while only one loop with three internal lines may contribute a λ divergence. Here s is the number of closed loops in Fig. 9, a, b are numerical constants, and $\ln \lambda$ and $\ln m$ stand for λ and m divergences.

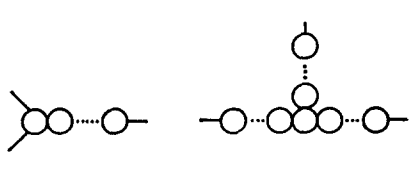

FIG. 9. Examples of reduced diagrams for $N = 3$ which have the strongest mass singularity.

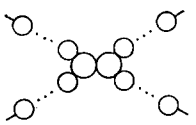

FIG. 10. An example of reduced diagrams for $N = 4$ with the strongest mass singularity.

$N \geq 4$. Analyzing diagrams with four external lines in the same manner, we find that strongest mass singularity may be obtained when the diagram is reducible with minimum contraction of chains to the form shown in Fig. 10 which consists of two 3-vertex parts touching with each other at one point. At the mass singularity, this amplitude will behave as $(a \ln \lambda + b \ln m)^2(\ln m)^s$. It should be noted that many other diagrams also behave as $(\ln \lambda)^2$ but are not as singular as this one at $m = 0$. Similarly some diagrams depend on m as $(\ln m)^s$ but behave as $(\ln \lambda)^1$ or $(\ln \lambda)^0$ at $\lambda = 0$. These considerations may be extended easily to arbitrary N. It will be not difficult to see that the amplitude will behave at most as $(a \ln \lambda + b \ln m)^{[N/2]} (\ln m)^s$, where $[N/2]$ is an integer not exceeding $N/2$ $(N \geq 3)$.

It is easy to remove the restriction that Feynman diagrams are proper. Then, a Feynman diagram will, in general, consist of several proper parts connected with each other by fixed internal lines. Mass singularities of the whole amplitude may be written down by multiplying contributions of all proper parts. We have only to note that singularities of those fixed internal lines which connect self-energy corrections of external lines to the main body of the diagram are removed when the mass renormalization is carried out. This procedure introduces λ divergences in the amplitude, as was mentioned already.

We shall now summarize the results obtained. At a mass singularity defined by (3.4), a Feynman amplitude can be reduced to a product of irreducible components. The behavior of the whole amplitude in the neighborhood of mass singularity is thus determined by that of irreducible components. In general, mass singularity arising from an irreducible component with s-ple loops is less singular than that of a product of s single closed loops. Thus, of all Feynman diagrams of given order that have no identical propagators except those caused by self-energy insertion, the strongest mass singularity is found in the diagrams which can be reduced to single loops of two or three elements by contracting as few internal lines as possible. Loops with two lines contribute primarily a $\ln m$ factor to the amplitude. If it is a self-energy insertion to an internal line, however, it may give a contribution of the type $a \ln \lambda + b \ln m$ instead. Each loop with three lines

may contribute a factor $\ln m(a' \ln \lambda + b' \ln m)$.

Finally we shall discuss the mass singularity of the absorptive part of a Feynman amplitude, which is obtained from (2.8) by replacing the denominator $(V)^{-n+2r}$ by $\delta^{(n-2r-1)}(V)$. Of course, the absorptive part exists only when the total energy in some channel is larger than its threshold value.[22a] At a mass singularity, it can be reduced to irreducible components following the procedure of Sec. 5. The same result is also obtained if one takes the absorptive part of the fully reduced amplitude (5.10). Thus, if $d < 0$ holds for the Feynman amplitude or its dispersive part, the absorptive part will also be finite at the mass singularity. As a matter of fact, because of stronger restriction on the domain of integration, mass singularity of the absorptive part may not be as strong as that of the dispersive part insofar as the diagram does not contain identical propagators other than those caused by self-energy insertions.[23]

To appreciate this result, let us note that taking the absorptive part means putting on the mass shell several internal lines which appear in a real intermediate state. In general, the absorptive part is a sum of several such terms. When a line is put on the mass shell, it may be regarded as a pair of external lines. Thus, a diagram with some of its lines fixed on the mass shell may have a mass singularity stronger than that of the Feynman amplitude itself. In spite of this, the absorptive part is no more singular than the dispersive part. We may therefore conclude that, when $d < 0$, mass singularities of various terms of the absorptive part cancel each other because of the specific relative phase of these terms. In fact, if some Feynman propagators (2.1) are replaced by other kind of Green's functions in (2.3), such a cancellation would no longer be possible, and the corresponding amplitude would have a mass singularity stronger than that of (2.3).

11. TOTAL TRANSITION PROBABILITY

We shall now come back to the cancellation of mass divergences in the total transition probability, which was the starting point of our investigation. This problem may be best handled by introducing cut diagrams.[24]

Let us consider transitions from a given initial state to all possible final states. If we denote the corresponding Feynman amplitudes as $T_1, T_2, \cdots$, the total transition probability is proportional to $\sum_{i,j} T_i^\dagger T_j$, where the summation is over those i, j whose final states are identical. The matrix multiplication in $T_i^\dagger T_j$ means integration over the entire phase space of the final state of T_j. If $T_i^\dagger$ is expressed by a diagram obtained from T_i by interchanging past and future, $T_i^\dagger T_j$ may be represented by a diagram which is obtained by connecting final state lines of T_j with corresponding initial state lines of $T_i^\dagger$. To distinguish it from a Feynman diagram, let us draw a continuous (horizontal) line that goes through all points of connection without intersecting any internal line of $T_i^\dagger$ and T_j. Let us denote by T^d a diagonal Feynman amplitude which is obtained by removing the cut line from $T_i^\dagger T_j$, and consider a set of all diagrams which reduce to T^d when the cut line is removed. Then, by optical theorem, the sum of corresponding $T_i^\dagger T_j$ is equal to the absorptive part A^d of T^d. This sum will be called a *cut diagram.*[24a] The cut lines represent various real intermediate states of A^d. Obviously the total transition probability may be regarded as a sum of cut diagrams. Thus, mass singularity of the former can be found by examining that of the latter. Results obtained in the last sections may not be applied directly to cut diagrams, however, since they are diagonal in the sense that for each incoming external line of T^d we can find an outgoing external line which carries the same momentum. As a consequence, we may find identical propagators in some A^d.

If some internal lines of A^d are identical, A^d may have a mass singularity of Coulomb type in addition to the usual one.[25] It is easy to see that this may occur only to lines belonging to the same chain. An example of this situation is shown in Fig. 11(a). Generalizing the method discussed at the end of Sec. 4, we find that its mass singularity is obtained

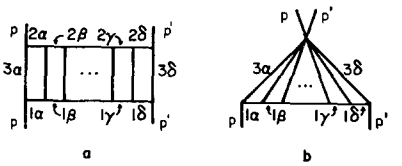

FIG. 11. The diagram (a) contains identical propagators if $m_{1\xi} = m_{2\xi}$ for some $\xi = \alpha, \beta, \cdots, \gamma, \delta$. The diagram (b) is obtained from (a) by contracting the lines $2\alpha, 2\beta, \cdots, 2\delta$.

[22a] See the footnote 24a.

[23] As was mentioned at the end of Sec. 9, some self-energy-like diagrams must be treated differently from the usual self-energy diagram. The absorptive part of the former may have a mass singularity stronger than that of the latter. It is not included in the case $N = 2$ discussed above.

[24] T. Kinoshita, Progr. Theoret. Phys. (Kyoto) 5, 1045 (1950).

[24a] A cut diagram is not the same as the sum of all topologically possible cuts of T^d unless the energy is sufficiently large. However, this does not affect our argument since cancellation of mass divergences occurs only among those intermediate states that can be opened up at the same energy.

[25] For the definition of Coulomb type see the last paragraph of Sec. 4.

from that of Fig. 11(b), which does not contain any identical propagator, by differentiation with respect to the masses $m_{1\alpha}^2$, $m_{1\beta}^2$, $\cdots$, $m_{1\gamma}^2$, $m_{1\delta}^2$. Singularities at the zero of $m_{1\beta}$, $\cdots$, $m_{1\gamma}$ may be logarithmic at most. However, singularities at $m_{1\alpha}=0$ and $m_{1\delta}=0$ may be stronger because of enhancement due to the lines 3α and 3δ. In this manner cut diagrams with identical propagators can always be reduced to those in which all propagators are different.

When A^d has no identical propagator, the result of the last section can be applied without qualification. Thus, it is free from mass divergences which may arise from purely internal vertices of A^d. The only mass divergences are λ and m divergences that arise from the self-energy and vertex parts attached to the external lines of A^d and λ divergence due to an internal line which is emitted by an initial-state line and absorbed by a final-state line of A^d.[26]

In order to examine mass divergences caused by particles of the initial state, it is convenient to consider a diagram in which initial-state lines of A^d are connected with the final-state lines carrying the same momentum. These lines may be distinguished from others making use of a second cut line representing the initial state of A^d. For each diagram like this, we can find a Feynman diagram T_v, describing a vacuum-to-vacuum transition, which looks identical with it except for the cut lines. Let us consider a set Δ of all such diagrams which give the same T_v when the cut lines are removed. We shall call Δ a *double cut diagram*.[27] Then the total transition probability is a sum of several Δ's, and each Δ is a sum of several A^d. We shall emphasize here the difference of Δ and T_v. Firstly, in closing the external lines of a cut diagram, we do not integrate over the momenta of initial-state particles, whereas T_v is fully integrated with respect to the corresponding momenta. Secondly, initial-state lines of Δ and the corresponding lines of T_v propagate energy in opposite time directions.

If a double cut diagram Δ consists of several A^d, the initial-state line of one A^d corresponds to an internal line of another A^d. Thus, in Δ as a whole, no line is fixed on the mass shell. We may then expect that mass singularity of Δ is weaker than that of individual A^d. In fact, as is discussed in Appendix A, λ divergences caused by emission and absorption of a quantum by the same external line or by different external lines carrying the same momentum cancel each other completely in this manner.[28] We find, however, that m divergences do not cancel in Δ. If Δ consists of one A^d, there is of course no cancellation of m divergence. However, such a Δ is free from λ divergence since it does not satisfy the condition for the appearance of λ divergence.

In summary, when the total transition probability is expressed as a sum of Δ's, not only the total probability but each Δ is free from λ divergence. We also find in Δ a complete cancellation of m divergences, insofar as they arise from the final-state lines of the transition amplitude. However, Δ has, in general, residual m divergences arising from self-energy and vertex parts attached to the initial-state lines. Thus, the only mass divergences to be found in the total transition probability are the m divergences caused by the vanishing of masses of particles in the initial-state and the Coulomb-type divergences that arise when *one* zero-mass particle is exchanged between the colliding particles.[29]

In particular, in the case of decay of an unstable particle, there is no singularity of Coulomb type. The only mass divergence of the total decay probability may arise from the vanishing of the mass M of the decay particle itself. To the extent that we are interested in the nonzero M only, there is therefore no mass divergence at all in the decay probability. We have thus arrived at a general proof of our conjecture that *mass divergences of partial decay probabilities cancel each other completely in all orders of perturbation theory when they are put together to form the total decay probability.*[30]

It seems to be reasonable to give the following physical interpretation for the cancellation of m divergences. (For definiteness, let us consider the emission of a photon by an electron of zero mass.) If the electron mass is zero, it emits photons so easily that, when the interaction is switched on, it becomes physically impossible to distinguish an electron travelling alone from one with a cloud of photons and pairs emitted in precisely the same direction insofar as the charge and four-momentum of the former are the same as the *total* charge and four-momentum of the latter. In other words, while measurement of the total energy of such an undifferentiated flux of electrons and photons is physically meaningful, measurement of the energy of an indi-

[26] This λ divergence belongs to the type discussed at the end of Sec. 9 and in reference 23.

[27] Double cut diagrams were first considered by T. Kinoshita (1950, unpublished). See also N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **19**, 159 (1958).

[28] N. Nakanishi, reference 27.

[29] Here the exchanged particle may emit several particles before it is absorbed. Thus, Fig. 11(a) may be regarded as a term contributing to the cross section for such a collision.

[30] It should be noted however that this result holds only for unrenormalized probabilities.

vidual particle of the flux is impossible. Consequently, we may find a sensible answer to the former question even in the limit of zero mass and in perturbation expansion, but a divergent and meaningless result for the latter. This reminds us of a similar situation in the case of the infrared divergence.

This interpretation will apply equally well to the cancellation of m divergences which arise from the final states of a collision process. If we accept it, we may also understand why the total cross section should have m divergences due to the initial state particles. It is simply because the usual definition of the initial state in terms of an incoming particle with definite energy and momentum becomes unrealistic in the limit where m divergences appear. We should instead define the initial state in terms of a flux that is an eigenstate of the total energy-momentum operator and is a certain superposition of state vectors which describe real zero-mass particles travelling in the same direction. If we consider a collision cross section of these fluxes, we find that the cut line at $t = -\infty$ may be applied more freely than the previous case to a double cut diagram Δ because of the enlarged initial state. If the cut at $t = -\infty$ can be applied as freely as the cut at $t = +\infty$, Δ will behave very much like the vacuum-to-vacuum diagram T_v with respect to its mass singularity. But, the latter is free from mass divergence as was shown in Sec. 10. Thus, there will be a complete cancellation of m divergences in Δ that are associated with the initial-state particles. Although we have not examined this problem in full detail, it seems very likely from our consideration of Appendix A that we can in fact form an initial-state vector which has such a property.

Insofar as we stick to the observed mass values of the known elementary particles, there is of course no fault in the conventional definition of initial state. The general consideration of mass singularity developed in this paper simply tells us how to find the leading terms of the cross section when the masses of particles are negligibly small compared with the total energy involved.

We have not explored the possibility that mass divergences of unrelated diagrams may cancel in the total cross section. Such a situation might be encountered when the interaction has a special property such as the invariance under the gauge transformation. Even then, we may not find a complete cancellation in general.

Finally, we shall emphasize that the conclusions of this and preceding sections hold only for unrenormalized amplitudes. How it will be modified by the renormalization will be discussed in the following section.

12. DISCUSSION

In this paper we have investigated the behavior of arbitrary Feynman amplitudes at the mass singularity. Since we have not restricted the form of interaction except that it should not give a singular numerator in any Feynman amplitude, our method and result will be applicable to various cases of physical interest. However, the usefulness of our theory is limited at present in several respects.

In the first place, we have formulated it in terms of perturbation theory. Thus, the results may not hold for an exact amplitude, although some features, being found in all orders of perturbation expansion, may hold even in the exact theory. Secondly, we have not fully exploited the analyticity of a Feynman amplitude as a function of mass variables. We have simply defined a mass singularity as a pathological solution of the Landau condition, and then estimated the strength of mass singularity by a somewhat crude integration. We might be able to learn a lot more by treating a Feynman amplitude as an analytic function of both external momenta and mass variables. Finally, we have thus far not considered the "charge" renormalization although we have assumed that the mass is renormalized and that external lines are properly normalized. Our theory therefore gives the mass singularity of unrenormalized amplitudes. In fact, it holds for any interaction, whether it is renormalizable or not, insofar as everything is made finite by ultraviolet cutoff.

For physical application of our result, we have to combine all Feynman diagrams contributing to a given transition and carry out the charge renormalization. Since we have not done this yet, let us simply outline what kind of problems we may encounter in such an attempt. If we take quantum electrodynamics as an example, the renormalization of an arbitrary amplitude is achieved by substituting the unrenormalized electron propagator S_F', photon propagator D_F', and vertex part Γ' by the corresponding renormalized quantities according to

$$\begin{aligned} &S_F'(e_0^2) = Z_2 S_{FC}(e_1^2), \quad D_F'(e_0^2) = Z_3 D_{FC}(e_1^2), \\ &\Gamma'(e_0^2) = Z_1^{-1}\Gamma_C(e_1^2), \quad e_1^2 = Z_1^{-2}Z_2^2 Z_3 e_0^2, \end{aligned} \tag{12.1}$$

where e_0 is the bare charge and Z_1, Z_2, Z_3 are the renormalization constants. Mass singularities of unrenormalized Green's functions can be examined by the method developed in this paper. For instance,

it is seen immediately from our consideration of Sec. 10 that $S_F'(p)$ or $D_F'(k)$ has no mass divergence if $p^2 \neq m^2$ or $k^2 \neq 0$. On the other hand, renormalization constants have mass divergences since they are defined with respect to free-particle states. Properties of these mass divergences may be determined by our method, or by a dimensional consideration making use of the knowledge of high-energy cutoff. Mass divergences of renormalized propagators may then be found from the fact that S_F' and D_F' are free from mass divergence. This property of propagators was first noted by Gell-Mann and Low.[4]

When we renormalize an amplitude using (12.1), parts of mass divergences are absorbed by the observed charge e_1, leaving uncompensated mass divergences of renormalized Green's functions in the observable amplitude. As far as the infrared divergence is concerned, however, the renormalization does not affect the exact amplitude since $Z_1 = Z_2$ by Ward's identity and Z_3 has no infrared divergence at all.[28] Specifically, the complete cancellation of infrared divergence in the total transition probability is preserved even after it is renormalized. Thus only m divergences will be affected by the charge renormalization. It is not difficult to see that this does not alter the degree of divergence of the strongest mass singularity in each order of perturbation expansion. For practical applications, however, it is of course important to know the explicit mass dependence of renormalized amplitudes which we have not found yet. A similar remark applies also to the pseudoscalar interaction of the pion-nucleon system, except that no simplifying relation like Ward's identity is found in this case.

As examples of nonrenormalizable cases, we shall now examine the mass singularity of radiative corrections to various weak decay processes. These processes contain a weak vertex in the lowest order besides the usual electromagnetic vertices. We know of no rule for renormalizing the weak vertices. In fact, in most cases,[3] no attempt is made to renormalize the theory except for the mass. The resulting radiative corrections are therefore not well defined and even infinite in general. In the universal $V - A$ theory of Feynman and Gell-Mann,[31] in which, however, the universality of weak interactions is assumed for the bare coupling constants, such radiative corrections must be made before theoretical prediction is compared with observation. If we make the radiative correction finite by introducing a Feynman cutoff in photon propagators, we find a few percent discrepancy between the observed and calculated ratios of lifetimes of μ-e and β decays.[32] This seems to imply that such an assumption is too naive and we must perhaps take account of electromagnetic and weak vertex form factors.[33] However, for any reasonable choice of form factors, these modifications of theory will affect only the contribution of high-energy virtual quanta and thus mass dependence of the total decay rate will not be altered.

On the other hand, if one tries to renormalize such a theory in one way or another, one finds that m divergences appear in the total probability, although the λ divergence may be avoided.[34] This is because the renormalization does modify not only the contribution of high-energy quanta, but also that of low-energy ones in contrast to the form factors discussed above. One source of m divergence is that renormalization constants corresponding to Z_1 and Z_2 do not cancel each other because there is no relation like Ward's identity in this case. Furthermore, the Z_3 factor which comes from radiative corrections to the innerbremsstrahlung processes accompanying decay contains an m divergence. This is a purely electromagnetic effect, and will be found for the first time in the fourth-order radiative correction. Since this effect is common to both $\mu - e$ and β decays, it will cancel out if one takes the ratio of total decay rates. Besides the appearance of m divergence, this theory is different from that of Feynman and Gell-Mann in that it is not possible to state the universality of weak interactions in terms of bare coupling constants.

Similar consideration applies to the radiative corrections of the branching ratio of pion decay. As is well known, if a pion is coupled with equal strength to the (parity violating) weak vector currents of lepton pairs (μ, ν) and (e, ν), the π-e to π-μ branching ratio R^0 is given by

$$R^0 = (m_e^0/m_\mu^0)^2 R_n^0, \tag{12.2}$$

when no radiative correction is included. R_n^0 is the branching ratio if lepton pairs are assumed to form a scalar instead. When radiative corrections of all

[31] R. P. Feynman and M. Gell-Mann, Phys. Rev. **109**, 193 (1958).

[32] R. K. Bardin, C. A. Barnes, W. A. Fowler, and P. A., Seeger, Phys. Rev. Letters **5**, 323 (1960); R. A. Reiter, T. A. Romanowski, R. B. Sutton, and B. G. Chidley, *ibid.* **5**, 22 (1960); V. L. Telegdi, R. A. Swanson, R. A. Lundby, and D. D. Yovanovitch, quoted in reference 34; R. J. Blin-Stoyle and J. Le Tourneux, Phys. Rev. **123**, 627 (1961); A. Altman and W. M. MacDonald, Univ. of Maryland, preprint, (1962).

[33] This possibility has been examined in detail by C. R. Schumacher (private communication).

[34] Dispersion theoretical approach by L. Durand, L. F. Landovitz, and R. B. Marr, Phys. Rev. Letters **4**, 620 (1960) belongs to this category.

orders are included, (12.2) is replaced by

$$R = (m_e^0/m_\mu^0)^2 R_n, \tag{12.3}$$

where m_e^0 and m_μ^0 are still the same *bare masses* of electron and muon, respectively.[35] If we make no charge renormalization, R_n does not have any mass divergence at $m_e = 0$ or $m_\mu = 0$. This means that R_n/R_n^0 is finite at $m_e = 0$ and will be very close to 1, if the high energy contribution is cut off properly. On the other hand, m_e^0/m_e may contain terms like $\ln m_e$. Accordingly, m_e^0/m_μ^0 may not be very close to m_e/m_μ. Thus, a deviation of the observed ratio R from $(m_e/m_\mu)^2 R_n \approx (m_e/m_\mu)^2 R_n^0$ will give an estimate of the quantity $[(m_e^0/m_e)/(m_\mu^0/m_\mu)]^2$. If one assumes that the electron and muon interact with the photon in the same manner, this mass ratio differs from 1 by about 4% in the second order which improves considerably the agreement with the observation.[36] On the other hand, if the theory is renormalized, there will no longer be any good reason why R_n should be close to R_n^0, and thus, the above argument will break down. Of course the formula (12.3) is valid only if the pion is coupled to leptons through a pure axial vector baryon current and not complicated by the pion form factor. It would therefore be not surprising if (12.3) and the following argument are modified substantially in a more rigorous treatment. Nevertheless, the good agreement of theory and experiment suggests that there may be some physical reason why the above consideration is approximately correct.

The discussion given above reveals a rather puzzling feature in that while an unrenormalized theory suffers from ultraviolet divergences but behaves in a reasonable way at m singularities, the situation is reversed in a renormalized theory. Of course this is no problem from a practical point of view since m divergences do not occur for the observed values of masses. However, if we can resolve this problem, it would be of considerable interest theoretically, since it will enable us to formulate a quantum electrodynamics which describes interaction of photons with zero-mass electrons.[37] In view of various divergences which may be found everywhere in such a theory, it has not been obvious whether it may be formulated in a consistent way at all. However, our consideration of Sec. 11 and Appendix A indicates that we may be able to eliminate m divergences from the theory if we realize that when m divergences occur, it is no longer possible to identify a line of Feynman diagram with an observed particle, and that we therefore have to characterize the initial and final states in terms of flux of particles travelling with the same velocity rather than in terms of a single particle. Unfortunately, this result seems to hold only if charge renormalization is disregarded. When the theory is renormalized, the cancellation of m divergences seems to be broken because of the Z_3 factor. Thus we still do not have a consistent formulation of quantum electrodynamics for zero-mass electron. It should be pointed out, however, that this is based on the assumption that the renormalization method used in the case $m \neq 0$ can also be applied to the case $m = 0$. This may not be correct since, as mentioned above, one particle state cannot be defined for $m = 0$ because of the degeneracy of the eigenstates of the total energy-momentum operator of interacting fields. It would be extremely interesting if this point can be clarified. Since the perturbation theory is certainly not the most satisfactory tool for such a consideration, it is desirable that it is treated by a method somewhat similar to the Bloch-Nordsieck method of the infrared problem that does not rely on perturbation expansion.

Finally, let us argue why it may be useful to study quantum electrodynamics of the zero-mass electron. It is primarily because it may lead us to the understanding of the age-old question of whether or not all of the electron mass is of electromagnetic origin. If the electron mass is zero, the theory will be invariant under the transformation $\psi \rightarrow \gamma_5 \psi$ no matter how often the electron interacts with real or virtual photons. On the other hand, the electron mass is not invariant under such a transformation. Thus, it seems that the observed electron mass cannot be produced if there is no mass to start with. In discussing related problems, however, several authors have recently pointed out the possibility that such symmetry argument may fail if we go beyond perturbation theory.[38] Thus, if we succeed in formulating a theory of quantum electrodynamics for the zero-mass electron without use of perturbation theory, we might be able to tell whether electron mass can be explained within the scheme of pure electrodynamics or if it has to be determined by interaction with other particles such as the proton.

[35] M. A. Ruderman and W. K. R. Watson, Bull. Am. Phys. Soc. 1, 383 (1956); R. Gatto and M. A. Ruderman, Nuovo cimento 8, 775 (1958).

[36] H. L. Anderson, T. Fujii, R. H. Miller, and L. Tau, Phys. Rev. 119, 2050 (1961).

[37] Importance of the study of quantum electrodynamics for the zero-mass electron has been emphasized recently by R. P. Feynman, "Report of the Solvay Conference 1961" (to be published).

[38] W. Heisenberg, Z. Naturforsch. 14, 441 (1959); Y. Nambu and G. Jona-Lasinio, Phys. Rev. 122, 345 (1961).

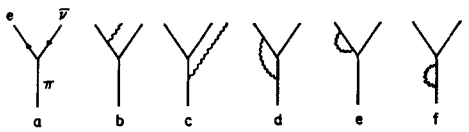

FIG. 12. Feynman diagrams describing radiative corrections to the π-e decay. Diagrams with $-\delta m$ vertex are not included for simplicity.

What is perhaps even more important is that the study of this problem will lead us to a better understanding of closely related topics such as the chirality invariance, conservation of axial vector current in weak interactions,[39] and various symmetry laws of strong interactions,[40] which are supposed to hold exactly only in the limit of zero mass or zero mass difference. It is my feeling that the method of mass singularity, although it has not been completely explored, will be one of the effective approaches to these fundamental questions of elementary particle physics.

ACKNOWLEDGMENTS

The author would like to thank Dr. S. A. Goudsmit, Dr. M. Goldhaber, Dr. G. C. Wick, and the other members of the Brookhaven National Laboratory for the hospitality extended to him during the summer of 1960 when part of this work was carried out. He also wishes to thank Professor W. E. Brittin and the Summer Institute of Theoretical Physics at the University of Colorado for their hospitality during the summer of 1961.

APPENDIX A. QUALITATIVE DISCUSSION OF CANCELLATION OF MASS DIVERGENCES

Logarithmic divergences associated with vanishing masses of partial transition probabilities cancel each other when they are summed into a total transition probability. We shall give here a qualitative argument of how such a cancellation takes place. For this purpose, let us take as an example the radiative corrections to π-e decay, given by Fig. 12, where the weak interaction vertex is assumed to be nonderivative for simplicity and two-photon vertices are omitted being irrelevant to our consideration. According to Sec. 11, the total decay probability can be expressed as a sum of cut diagrams of Fig. 13. Each cut diagram represents a sum of several partial transition probabilities. Letters on the ends of each cut line show from which diagrams of Fig. 12 the cut diagram in question is made.

To find out how cancellation of mass divergences takes place among the partial probabilities belonging to each cut diagram of Fig. 13, let us first take Fig. 13(c) and look at the internal lines k, q, p, which are the only lines that depend on how the graph is cut. Noting that a line q on the mass shell is represented by $-2\pi i\,\delta_P(q^2-m^2)$, we see that these lines contribute a factor

$$\frac{\delta_P(k^2)\,\delta_P(p^2-m^2)}{q^2-m^2}+\tfrac{1}{2}\,\delta_P(q^2-m^2)\times\left[\frac{\delta(k^2)}{p^2-m^2}+\frac{\delta(p^2-m^2)}{k^2}\right],\quad q=p+k \tag{A1}$$

to the absorptive part, where $\delta_P(k^2)=\theta(k)\,\delta(k^2)$ and $\theta(k)$ is equal to 1 or 0 depending on whether $k_0>0$ or <0. The first term corresponds to the inner bremsstrahlung cut and the second to the virtual photon cut. The denominators are principal value parts.

If we write $\delta(k^2)$ as a sum of $\delta_P(k^2)$ and $\delta_P[(-k)^2]$ in (A1), we find that these two terms give equal contributions to the amplitude when the k integration is performed. Thus (A1) is equivalent to

$$\frac{\delta_P(k^2)\,\delta_P(p^2-m^2)}{q^2-m^2}+\frac{\delta_P(k^2)\,\delta_P(q^2-m^2)}{p^2-m^2}+\frac{\delta_P(p^2-m)\,\delta_P(q^2-m^2)}{k^2}. \tag{A2}$$

We note that the same result may be obtained directly from a Feynman diagram [Fig. 13(c) less cut lines] by putting a given number (three in this case) of internal lines on the mass shell in all possible ways and applying Cutkosky's rule[41] to these lines irrespective of whether they correspond to the unitarity cut or not. From this point of view, it is because the propagators for the lines k, p, q appear as a product

$$\frac{1}{k^2+i\epsilon}\;\frac{1}{p^2-m^2+i\epsilon}\;\frac{1}{q^2-m^2+i\epsilon} \tag{A3}$$

in the Feynman amplitude that (A2) is symmetric in these lines.

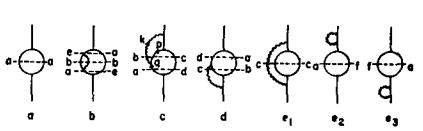

FIG. 13. Cut diagrams representing the total probability of π-e decay with radiative corrections.

[39] S. Bludman, Nuovo cimento **9**, 433 (1958); M. Gell-Mann and M. Lévy, *ibid.* **16**, 705 (1960); J. Bernstein, S. Fubini, M. Gell-Mann, and W. Thirring, *ibid.* **17**, 757 (1960); Y. Nambu, Phys. Rev. Letters **4**, 380 (1960).

[40] J. Schwinger, Ann. Phys. **2**, 407 (1957); J. J. Sakurai, *ibid.* **11**, 1 (1960); M. Gell-Mann and F. Zachariasen, Phys. Rev. **123**, 1065 (1961); M. Gell-Mann, *ibid.* **125**, 1067 (1962).

[41] See reference 1.

Let us first see how the singularities of the amplitude may arise. Consider the first term of (A2). Making use of conservation of four-momentum and the δ functions in the numerator, its denominator can be written as

$$\frac{1}{q^2 - m^2} = \frac{1}{2pk} = \frac{1}{2k_0(p_0 - |\mathbf{p}| \cos\theta)}, \qquad \text{(A4)}$$

where k_0 is the photon energy and θ is the angle between $\mathbf{p}$ and $\mathbf{k}$. Thus the propagator (A4) has a pole at $k_0 = 0$, which leads to the infrared singularity. For $m \neq 0$, $p_0 - |\mathbf{p}| \cos\theta$ is always different from zero for physical values of p, and no singularity arises from the second factor in the denominator of (A4). However, in the limit $m = 0$, $p_0 - |\mathbf{p}| \cos\theta$ may vanish and give another pole at $\theta = 0$. This is the origin of the ln m term in the amplitude. Physically, this divergence is associated with the possibility that a particle of zero mass (momentum q) can emit a photon of *arbitrary* momentum k in the forward direction (insofar as $k_0 < q_0$) without violating the energy-momentum conservation. Similar singularities are found in the remaining terms of (A2).

Since it is well known how infrared divergences cancel each other,[24,28] we shall concentrate ourselves to the cancellation of ln m terms in (A2). For simplicity, let us put $m = 0$ from the beginning. Then (A2) may be rewritten as

$$(1/2pk)\{\delta_P(k^2)\,\delta_P(p^2) - \delta_P(k^2)\,\delta_P[(p+k)^2] - \delta_P(p^2)\,\delta_P[(p+k)^2]\}. \qquad \text{(A5)}$$

Since $k_0, p_0 > 0$, this may be transformed to

$$\frac{1}{8pk}\left[\frac{\delta(p_0 - P)\,\delta(k_0 - K)}{PK} - \frac{\delta(k_0 - K)\,\delta(p_0 + k_0 - Q)}{KQ} - \frac{\delta(p_0 - P)\,\delta(p_0 + k_0 - Q)}{PQ}\right], \qquad \text{(A6)}$$

where $P = |\mathbf{p}|$, $K = |\mathbf{k}|$, and $Q = |\mathbf{p} + \mathbf{k}|$. The divergence at $m = 0$ occurs when $\mathbf{p}$ and $\mathbf{k}$ are parallel, or $Q = P + K$. Thus, as $\theta \to 0$, (A6) approaches

$$\frac{\delta(p_0 - P)\,\delta(k_0 - K)}{8pk} \times \left[\frac{1}{PK} - \frac{1}{K(P+K)} - \frac{1}{P(P+K)}\right]. \qquad \text{(A7)}$$

It is now easy to see that the three terms cancel each other for arbitrary P and K. Thus singularities of (A2) cancel out and become less singular. The formula (A7) also shows how the infrared divergences

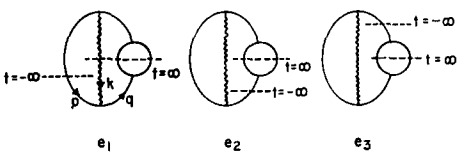

FIG. 14. Double cut diagram corresponding to the cut diagrams e_1, e_2, e_3 of Fig. 13.

at $K = 0$ cancel each other.[42] It is interesting to note that whereas cancellation of infrared divergences occur between the first two terms only, all three terms must participate in the cancellation of m divergences.

In the case of Fig. 13(b), we have to examine the absorptive part corresponding to the product

$$\frac{1}{q^2 - m^2 + i\epsilon}\frac{1}{p^2 - m^2 + i\epsilon} \times \frac{1}{k^2 + i\epsilon}\frac{1}{q^2 - m^2 + i\epsilon}, \qquad q = p + k. \qquad \text{(A8)}$$

When the self-energy effect of the electron is separated, we find again that complete cancellation of λ and m divergences occur between the three cuts of Fig. 13(b).

Let us now consider the three diagrams of Fig. 13(e). All of them give infrared divergences, but they are found to cancel each other when put together. This situation may be understood by regarding the cut diagrams as a double cut diagram shown in Fig. 14. They are obtained by cutting pion lines of a vacuum-to-vacuum type diagram in various ways. The only part of these amplitudes which differ from each other may be written as[43]

$$\frac{2\,\delta_P(q^2 - \mu^2)\,\delta_P(p^2 - \mu^2)}{k^2(q^2 - \mu^2)} + \frac{2\,\delta_P(q^2 - \mu^2)\,\delta_P(k^2)}{(p^2 - \mu^2)(q^2 - \mu^2)} + \frac{\delta_P[(-k)^2]\,\delta_P(p^2 - \mu^2)}{(q^2 - \mu^2)^2}. \qquad \text{(A9)}$$

As is easily seen, if $\delta_P[(-k)^2]$ is replaced by $\delta_P(k^2)$ in the last term, (A9) may be regarded as the absorptive part of an amplitude that has the same structure as (A8). Thus, if one writes $\delta_P[(-k)^2]$ as $\delta_P(k^2) + \{\delta_P[(-k)^2] - \delta_P(k^2)\}$, the first term will make (A9) free from λ and m divergences. The second term may be written as

$$(1/2\,|\mathbf{k}|)\,\delta(k_0 + |\mathbf{k}|) - (1/2\,|\mathbf{k}|)\,\delta(k_0 - |\mathbf{k}|). \qquad \text{(A10)}$$

These two terms tend to cancel each other in the limit $\mathbf{k} = 0$. Thus, it does not give rise to an infrared divergence either. This argument can be extended

[42] See reference 24.

[43] The formula (A9) is not well defined because of the vanishing denominators. It is given here only for comparison with (A8).

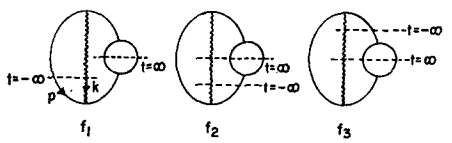

FIG. 15. Additional double cut diagrams which are introduced to take account of the degeneracy of the initial state.

to prove the cancellation of infrared divergences in the general case.[28]

It should be emphasized, however, that this mechanism of cancellation will not work for the m singularity associated with $\mu = 0$ in (A9).[44] This is because an m divergence arises from a domain where $|\mathbf{k}|$ may take arbitrary value and thus the two terms of (A10) may not cancel each other any more. As is discussed in Sec. 11, however, when $\mu = 0$, the initial state will not be defined properly if only the presence of one meson is taken into account. We must instead consider a flux of mesons and photons that travel in the same direction with a given total energy momentum. In the lowest order perturbation theory, such a state will be described as a certain linear combination of a state with one meson and a state with one meson and one photon present. The contribution of the first state to the total cross section has been discussed above. The second state will give many other terms, some of which are shown by the diagrams of Fig. 15. Now it will not be difficult to see that not only λ divergences but also m divergences cancel each other among Figs. 14(e_1), 15(f_2), and 15(f_3). Similarly for the set Figs. 14(e_2), 14(e_3), and 15(f_1). Thus, all mass divergences in fact disappear in such a treatment. Incidentally, we also find that the diagram e_1 of Fig. 14 is not really a natural counterpart of diagrams e_2 and e_3. It would look somewhat accidental that this did not destroy the cancellation of infrared divergences of Fig. 14.

APPENDIX B. DERIVATION OF FORMULA (2.8)

There is no particular rule how to parametrize a Feynman amplitude. Thus, several alternative methods of parametrization have been used in the past. Many more may be written down according to one's need. Of these, our formula (2.8) has a distinctive feature that it is perhaps the most economical expression and that it has a very close correspondence with the Feynman diagram itself. But, of course, all these formulas are equivalent and we should be able to derive one from another by transformation of parameters. Instead of proving (2.8) directly, we shall therefore take one of the published formulas and show how it is transformed to our formula by a simple rearrangement of terms of $V(x, z)$.

If we follow Nakanishi's method,[6] except that we parametrize in two successive steps as is done in Sec. 2, we are led to an integral of the form (2.8), where $v(x, z)$ is given by

$$v(x, z) = -\sum_\alpha z_\alpha q_\alpha^2 + U^{-1}(z) \sum_C U_C(z) \Big(\sum_\alpha^C \pm z_\alpha q_\alpha \Big)^2 \qquad \text{(B1)}$$

rather than (2.12). Here $\sum_C$ is the sum over all possible (not necessarily independent) closed loops C, and $\sum^C$ is the sum over all chains belonging to given C. The function $U_C(z)$ is defined by

$$U_C(z) = \sum z_{\alpha_1} z_{\alpha_2} \cdots z_{\alpha_{r-1}}, \qquad \text{(B2)}$$

where the summation is over all possible sets $(\alpha_1, \alpha_2, \cdots, \alpha_{r-1})$ such that $k_{\alpha_1}, k_{\alpha_2}, \cdots, k_{\alpha_{r-1}}$ are independent and none of the $\alpha_1, \alpha_2, \cdots, \alpha_{r-1}$ coincide with any α belonging to C. $U(z)$ and q_α are defined by (2.10) and (2.7), respectively. Our problem is therefore to show that (B1) can be written as (2.12).

For this purpose, let us note that it is always possible to give arbitrary values to r of the constant momenta q_i (one from each of the independent closed loops $C_1, C_2, \cdots, C_r$) because (2.3) contains r independent variable momenta k_i whose origin may be shifted by an arbitrary amount. In the integral (2.8), $V(x, z)$ is therefore invariant under any transformation of the form

$$q_{i(C)} \to q_{i(C)} + q^C, \qquad \text{(B3)}$$

where $i(C)$ runs over all internal lines of an arbitrary closed loop C, and q^C is an arbitrary fixed four-vector. Since $V_\alpha(x)$ is already invariant under (B3), $v(x, z)$ must also be invariant, although this is not readily seen from (B1).

In order to rewrite (B1) in an explicitly invariant form, note that the Lorentz-invariant quadratic form

$$(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2 \qquad \text{(B4)}$$

is also invariant under the transformation (B3) *for any value of* x, if any closed loop contains an even number of chains out of $\alpha, \beta, \cdots, \delta$, and if

[44] Since the decay probability vanishes when the pion mass vanishes, the following argument is actually not appropriate for the decay problem. We should rather regard it as a simplified treatment of m singularity arising from one of the incident particles of a collision process.

signs of $q_\alpha, q_\beta, \cdots, q_\delta$ are chosen properly. This is easily proved making use of the transformation

$$q_{\alpha(C)} \to q_{\alpha(C)} + q^C \tag{B5}$$

derived from (B3), where $\alpha(C)$ is any chain belonging to the loop C. Obviously q_α^2 or $(q_\alpha \pm q_\beta)^2$ cannot be invariant under (B3). Indeed, the simplest invariants are of the form $(q_\alpha \pm q_\beta \pm q_\gamma)^2$, where α, β, γ are chains that meet with each other at a three-vertex (part). For any chain diagram, the total number of these invariants is finite.

Arbitrary invariants bilinear in $q_\alpha, q_\beta, \cdots$ may be expressed as linear combinations of invariants (B4). In particular, $U(z)v(x, z)$ will be written as

$$\sum_{\alpha,\beta,\cdots,\delta} A_{\alpha\beta\cdots\delta}(z)(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2, \tag{B6}$$

where $A_{\alpha\beta\cdots\delta}$ is a homogeneous polynomial of order $r + 1$ in z and is independent of x. Comparing the coefficients of q_α^2, etc., of (B1) and (B6), we find that $A_{\alpha\beta\cdots\delta}$ satisfy

$$\sum_{\beta,\cdots,\delta} A_{\alpha\beta\cdots\delta}(z) = -z_\alpha[U(z)]_{z_\alpha=0} \tag{B7}$$

for any α. Thus $A_{\alpha\beta\cdots\delta}$ must contain the product $z_\alpha z_\beta \cdots z_\delta$ as a factor.[45] From this it follows that (B6) should not contain any term of the form $(\sum q_\alpha + \sum q_\beta)^2$ if it already contains nonoverlapping invariants $(\sum q_\alpha)^2$ and $(\sum q_\beta)^2$. To prove this, note that for each term of the polynomial $z_\alpha U(z_\alpha = 0)$ it is possible to choose r independent closed loops in such a way that r of the z factors each belong to only one closed loop, whereas the last z belongs to more than one loop. In the above example, however, there are, for any choice of r closed loops, at least two factors of z, each of which belongs to more than one loop. This is impossible. Thus the summation in (B6) and (B7) must exclude invariants of this type. This permits us to solve Eq. (B7) uniquely, with the result

$$A_{\alpha\beta\cdots\delta}(z) = -z_\alpha z_\beta \cdots z_\delta(\partial/\partial z_\beta) \cdots (\partial/\partial z_\delta)U(z_\alpha = 0). \tag{B8}$$

Equations (B6) and (B8) lead us immediately to the formula (2.12), which was what we wanted. It is very much simpler than (B1) and its structure is transparent. In particular, it depends only on $U(z)$ while (B1) contains many other functions $U_C(z)$. It should be emphasized that this simplification has been achieved easily because of our two-step parametrization of Feynman amplitudes.

In Sec. 2, a set of chains $\alpha, \beta, \cdots, \delta$ is called a C set if the corresponding quantity $(q_\alpha \pm q_\beta \pm \cdots \pm q_\delta)^2$ is one of the terms of (2.12). It is interesting to point out that a C set has the property that the chain diagram is separated into exactly two parts if all chains of the C set are cut into two parts. This property can be used to find graphically all C sets of a given diagram. It follows from this that no C set can have more than $r + 1$ elements. Conversely, any set of chains with more than r elements must contain at least one C set as a subset. To see this, we have only to note that a diagram can be divided into at least two parts if $r + 1$ or more chains are cut. On the other hand, we can find a set with r chains which does not contain any C set as a subset. It will be useful to emphasize here that the formulas (2.8)–(2.12), together with this property of C sets, may be considered as a substitute for the usual Feynman-Dyson rules. In fact, in terms of these formulas, we can write down very easily the amplitude for any given Feynman diagram once we obtain $U(z)$, which is completely determined by the topological structure of the chain diagram.

Finally we find an interesting identity

$$U(z_\alpha = 0) = \sum_{\beta,\cdots,\delta} \left(z_\beta \frac{\partial}{\partial z_\beta}\right) \cdots \left(z_\delta \frac{\partial}{\partial z_\delta}\right) U(z_\alpha = 0) \tag{B9}$$

by substituting (B8) into (B7), where the summation is over all C sets $\{\alpha, \beta, \cdots, \delta\}$ that contain α. At a mass singularity, each z belongs either to the set A or $\bar{A}$. Let us assume in particular that $\alpha \in \bar{A}$. Then we obtain

$$U(z_\alpha = 0) = U_A(z)\tilde{U}(z_\alpha = 0) + \cdots, \tag{B10}$$

as is seen from (5.1). Thus, if one substitutes (B10) in (B9), divides both sides by $\tilde{U}(z_\alpha = 0)$, and takes the limit $z_\xi \to 0$, $\xi \in \bar{A}$, one finds that

$$U_A(z) = \sum_{\beta,\cdots,\delta}{}' \left(z_\beta \frac{\partial}{\partial z_\beta}\right) \cdots \left(z_\delta \frac{\partial}{\partial z_\delta}\right) U_A(z), \tag{B11}$$

where the summation $\sum'$ is over all elements $\beta, \cdots, \delta$ of A such that $\{\alpha, \beta, \cdots, \delta\}$ is a C set of the original diagram. This formula is used in deriving (5.3).

[45] To prove this, we also need the relation $\sum_{\gamma,\cdots,\delta} A_{\alpha\beta\gamma\cdots\delta}(z) = z_\alpha z_\beta \times$ (polynomial in z) for fixed α, β, which is obtained by comparing (B1) and (B6) for $\alpha \neq \beta$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

On the Mutual Coherence Function in an Inhomogeneous Medium*

GEORGE B. PARRENT, JR.,† ROBERT A. SHORE, AND THOMAS J. SKINNER.†
Air Force Cambridge Research Laboratories, L. G. Hanscom Field, Bedford, Massachusetts
(Received October 31, 1961)

The paper is concerned primarily with determining the mutual coherence function of the field produced by a plane quasi-monochromatic source in a region of variable refractive index. A scalar theory is used throughout. Section I presents a brief review of the conceptual background of coherence theory. Section II contains an outline of the mathematical formalism of coherence theory and shows that in a region of variable refractive index the mutual coherence function is propagated according to a pair of inhomogeneous scalar wave equations. In Sec. III, the pair of wave equations are solved using appropriate Green's functions to derive an expression for the mutual coherence function of a field produced by a plane quasi-monochromatic source. In Sec. IV the case of a statistically homogeneous medium is treated and an expression for the ensemble average of the mutual coherence function is obtained in terms of integrals of the two-point correlation function characterizing the medium.

I. INTRODUCTION

THE development of coherence theory has been strongly motivated by research in visual optics. The theory is concerned with the behavior of electromagnetic fields at frequencies so high that measurements consist of recording intensities averaged over periods of time, long compared to the times involved for individual fluctuations of the fields. The fields are assumed to have stationary time dependence, at least for intervals of the order of the averaging periods.

In developing an advanced theory of optical behavior, the concepts of amplitude and phase which are often useful in other branches of electromagnetic theory and in elementary optics are no longer of much help. The usefulness of these concepts breaks down not only because the high frequencies involved in optics make it impossible to measure the amplitude and phase of field components, but even more fundamentally because light as it is usually encountered (resulting from the superposition of a large number of randomly timed statistically independent pulses) is not strictly monochromatic but consists of spectra of finite widths. Indeed, under the conditions stated it cannot be analyzed even by a Fourier decomposition into strictly monochromatic components since all that can be measured are the power spectra that allow statements to be made only about the amount of power carried by a narrow band of wavelengths.

The principal difference between coherence theory and the more elementary theory is, that while the concepts of amplitude and phase presuppose a strictly monochromatic source of illumination and are not measurable quantities (at least in the optics realm), the basic quantities of coherence theory, namely, time-averaged intensities and functions which express the degree of correlation that exists between the vibrations at different points in the field, *are* measurable and do not presuppose an unrealistic *a priori* assumption as to the nature of the field. Furthermore, as limiting cases, coherence theory yields not only a strictly monochromatic theory, but also a theory of incoherent (addition of intensities) radiation and gives an accurate description of the region between these two limits—the region of partially coherent light.

II. OUTLINE OF COHERENCE THEORY

Following this brief statement of the conceptual background of coherence theory, we proceed to an outline of the mathematical formalism of the theory. Our principal concern in this section will be to introduce the fundamental entities of coherence theory—the mutual coherence function and the complex degree of coherence—and to derive the wave equations for the propagation of the mutual coherence function in an inhomogeneous medium. For the most part, theorems will be stated and the reader referred to the literature for proofs and further discussion.

We start with a real scalar function of position and time $V^r(\mathbf{P}, t)$ which in a source free, although not necessarily homogeneous, medium satisfies the scalar wave equation

$$\nabla^2 V^r(\mathbf{P}, t) = \frac{1}{C^2(\mathbf{P})} \frac{\partial^2 V^r(\mathbf{P}, t)}{\partial t^2} ,$$

where $\mathbf{P}$ is the position vector. (At times in the

* Work performed at Air Force Cambridge Research Laboratories.
† Technical Operations, Inc., formerly with Air Force Cambridge Research Laboratories.

discussion, it will be convenient to indicate position by a subscript or to omit explicit spatial dependence.)

The intensity of $V^r(t)$ averaged over an interval of time of length $2T$ is given by the expression

$$\frac{1}{2T}\int_{-T}^{T}[V^r(\mathbf{P}, t)]^2\,dt.$$

Since in the applications we have in mind, T will be extremely large in terms of time units of the order of the actual fluctuations (for example, the mean period $1/\bar{\nu}$ where $\bar{\nu}$ is the mean frequency of the disturbance), it is convenient to let $T \to \infty$ in expressions for the time-averaged intensity and for the correlation functions to be introduced below. We assume of course that the

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}[V^r(\mathbf{P}, t)]^2\,dt$$

is finite.

Next, we associate with the real function $V^r(\mathbf{P}, t)$ a complex function $V(\mathbf{P}, t)$, the analytic signal. The advantages of choosing the analytic signal as a complex representation of the disturbance have already been discussed at length.[1]

In terms of the analytic signal, the basic quantities of coherence theory, the mutual coherence function, and the complex degree of coherence may be precisely defined. The mutual coherence function $\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau) \equiv \Gamma_{12}(\tau)$ is defined to be the complex cross correlation between the analytic signal representation of the real field at the two points $\mathbf{P}_1$ and $\mathbf{P}_2$; that is,

$$\Gamma_{12}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-\infty}^{\infty}V_1(T, t+\tau)V_2^*(T, t)\,dt,$$

where $V_i(T, t) = V_i(t)$ for $|t| \leq T$ and is zero otherwise. The time average and limiting process is denoted by sharp brackets; thus

$$\Gamma_{12}(\tau) = \langle V_1(t+\tau)V_2^*(t)\rangle.$$

By using the theorem that the cross correlation of the two real functions is equal to the cross correlation of their Hilbert transforms in the same order,[2] it is readily shown that the time-averaged intensity I_m at the point P_m is given by $\frac{1}{2}\Gamma_{mm}(0)$, i.e.,

$$I_m = \langle[V_m^r(t)]^2\rangle = \tfrac{1}{2}\Gamma_{mm}(0), \qquad (m = 1, 2).$$

Also, by use of the theorem that the convolution of two analytic signals is itself an analytic signal, it can be shown[2] that $\Gamma_{12}(\tau)$ is an analytic signal. From the last result it follows[3] that $\Gamma_{12}(\tau)$ possesses a Fourier spectrum which is zero for half the frequency range. Thus

$$\Gamma_{12}(\tau) = \int_{-\infty}^{\infty}\hat{\Gamma}_{12}(\nu)e^{-2\pi i\nu\tau}\,d\nu$$

where

$$\hat{\Gamma}_{12}(\nu) = \int_{-\infty}^{\infty}\Gamma_{12}(\tau)e^{2\pi i\nu\tau}\,d\tau \equiv 0, \nu < 0.$$

It should be noted here that in formulating the solution to actual problems, we usually obtain expressions of the form $\langle V_1(t_1 + t)V_2^*(t_2 + t)\rangle$. Under the change of variables $t' = t + t_2$, we obtain $\langle V_1(t^1 + \tau)V_2^*(t^1)\rangle$ where $\tau = t_2 - t_1$. The additional assumption of stationarity of $V(t)$ (that is, that the time averages are independent of the choice of time origin, or equivalently that the time averages are a function of the *difference* in time only) is necessary to equate $\Gamma_{12}(\tau)$ with $\langle V_1(t_1 + t)V_2^*(t_2 + t)\rangle$.

A normalized form of the mutual coherence function called the complex degree of coherence and denoted by $\gamma_{12}(\tau)$ is very useful in coherence theory; $\gamma_{12}(\tau)$ is defined to be

$$\gamma_{12}(\tau) = \Gamma_{12}(\tau)/[\Gamma_{11}(0)\Gamma_{22}(0)]^{1/2}.$$

By the use of the Schwartz inequality it can be shown that $0 \leq |\gamma_{12}(\tau)| \leq 1$. The limits characterize incoherent radiation and coherent radiation, respectively.

We shall now prove the important result that $\Gamma_{12}(\tau)$ is propagated according to a pair of wave equations. Specifically, we shall show that

$$\nabla_m^2\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau) = \frac{1}{C^2(\mathbf{P}_m)}\frac{\partial^2\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)}{\partial\tau^2}, \qquad (m = 1, 2), \tag{1}$$

Here the Laplacian ∇_m^2 acts on the coordinates of the point $\mathbf{P}_m (m = 1, 2)$, and the spatial dependence of the velocity of propagation is indicated.

To prove (1), we start from the assumption that $V^r(\mathbf{P}, t)$ and hence the truncated function $V^r(\mathbf{P}; T, t)$ satisfies the scalar wave equation

$$\nabla^2 V^r(\mathbf{P}; T, t) = \frac{1}{C^2(\mathbf{P})}\frac{\partial^2 V^r(\mathbf{P}; T, t)}{\partial t^2}. \tag{2}$$

Also, as defined above

$${}_TV^i(\mathbf{P}, t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{V^r(\mathbf{P}; T, t')}{t' - t}\,dt'. \tag{3}$$

Operating on both sides of (3) with the Laplacian ∇^2, interchanging the order of operations and using

[1] G. B. Parrent, Jr., J. Opt. Soc. Am. 49, 787 (1959).

[2] G. B. Parrent, Jr., "Contribution to the Theory of Partial Coherence," AFCRC-TR-60-124.

[3] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, New York, 1959), Chap. X.

the theorem[4] that the Hilbert transform of the derivative of a function equals the derivative of the Hilbert transform of the function, we obtain

$$\nabla_T^2 V^i(\mathbf{P}, t) = \frac{1}{C^2(\mathbf{P})} \frac{\partial_T^2 V^i(\mathbf{P}, t)}{\partial t^2}. \quad (4)$$

Multiplying (4) by i and adding to (2) we obtain

$$\nabla^2 V(\mathbf{P}; T, t) = \frac{1}{C^2(\mathbf{P})} \frac{\partial^2 V(\mathbf{P}; T, t)}{\partial t^2}. \quad (5)$$

Thus, the analytic signal itself satisfies the wave equation. Now

$$\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} V(\mathbf{P}_1; T, t+\tau) V^*(\mathbf{P}_2; T, t)\, dt. \quad (6)$$

Differentiating (6) with respect to $\mathbf{P}_1$, interchanging the order of operations, and substituting (5) we obtain

$$\nabla_1^2 \Gamma_{12}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} \nabla_1^2 [V_1(T, t+\tau)] V_2^*(T, t)\, dt$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} \frac{1}{C^2(\mathbf{P}_1)} \frac{\partial^2 V_1(T, t+\tau)}{\partial \tau^2} V_2^*(T, t)\, dt$$

$$= \frac{1}{C^2(\mathbf{P}_1)} \frac{\partial^2}{\partial \tau^2} \Gamma_{12}(\tau).$$

Similarly,

$$\nabla_2^2 \Gamma_{12}(\tau) = \frac{1}{C^2(\mathbf{P}_2)} \frac{\partial^2 \Gamma_{12}(\tau)}{\partial \tau^2},$$

and the proof is complete.

III. PROPAGATION OF $\Gamma_{12}(\tau)$ FROM A PLANE SOURCE INTO AN INHOMOGENEOUS REGION

We now come to the central problem of this paper, the determination of the mutual coherence function for a field produced by an extended polychromatic source in a region of variable index of refraction. In the following discussion, S is an arbitrary surface containing an extended polychromatic source with a known distribution of mutual coherence; P_1 and P_2 are points in the illuminated field V; and S_1 and S_2 are points on the surface S.

As shown above, the propagation of the mutual coherence function in a source free but inhomogeneous medium is governed by the pair of wave equations

$$\nabla_m^2 \Gamma_{12}(\tau) = \frac{1}{C^2(\mathbf{P}_m)} \frac{\partial^2 \Gamma_{12}(\tau)}{\partial \tau^2}, \qquad m = 1, 2. \quad (7)$$

[4] Bateman Manuscript Project, *Tables of Integral Transform* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 2.

We assume that $\Gamma_{12}(\tau)$ is known for all pairs of points S_1 and S_2 on the surface S.

Let $\hat{\Gamma}_{12}(\nu)$ be the Fourier transform of $\Gamma_{12}(\tau)$. Then, as stated above, since $\Gamma_{12}(\tau)$ is an analytic signal, its Fourier spectrum contains positive frequencies only; that is,

$$\Gamma_{12}(\tau) = \int_0^{\infty} \hat{\Gamma}_{12}(\nu) e^{-2\pi i \nu \tau}\, d\nu, \quad (8)$$

where

$$\hat{\Gamma}_{12}(\nu) = \int_{-\infty}^{\infty} \Gamma_{12}(\tau) e^{2\pi i \nu \tau}\, d\tau. \quad (9)$$

Substituting from (8) and (7) and interchanging the order of integration and differentiation, we obtain

$$\int_0^{\infty} [\nabla_m^2 + k_m^2(\nu)] \hat{\Gamma}_{12}(\nu) e^{-2\pi i \nu \tau}\, d\nu = 0, (m = 1, 2). \quad (10)$$

Since (10) holds for all τ we have

$$[\nabla_1^2 + k^2(\mathbf{P}_1, \nu)] \hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2, \nu) = 0, \quad (11a)$$

$$[\nabla_2^2 + k^2(\mathbf{P}_2, \nu)] \hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2, \nu) = 0, \quad (11b)$$

Here $k(\mathbf{P}, \nu) = 2\pi\nu/C(\mathbf{P})$. Thus, each spectral component of $\Gamma_{12}(\tau)$ satisfies the pair of Helmholtz equations (11a, 11b).

Now in Eq. (11a), P_2 is a fixed parameter as far as the operator is concerned. In particular, (11a) holds if P_2 is a fixed point S_2 on a closed surface S. Equation (11a) then becomes

$$[\nabla_1^2 + k^2(\mathbf{P}_1, \nu)] \hat{\Gamma}(\mathbf{P}_1, \mathbf{S}_2, \nu) = 0. \quad (12)$$

The boundary condition for (12) is the known values of $\hat{\Gamma}(\mathbf{S}_1, \mathbf{S}_2, \nu)$. Hence, the problem is to solve the pair of equations

$$[\nabla_1^2 + k^2(\mathbf{P}_1, \nu)] \hat{\Gamma}(\mathbf{P}_1, \mathbf{S}_2, \nu) = 0, \quad (13a)$$

with $\hat{\Gamma}(\mathbf{S}_1, \mathbf{S}_2, \nu)$ known on the boundary, and

$$[\nabla_2^2 + k^2(\mathbf{P}_2, \nu)] \hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2, \nu) = 0, \quad (13b)$$

with $\hat{\Gamma}(\mathbf{P}_1, \mathbf{S}_2, \nu)$ known on the boundary as a result of solving (13a).

A formal solution to the Eq. (13a,b) can be easily obtained in terms of the Green's function $\mathcal{G}(\mathbf{P}, \mathbf{P}')$ which satisfies the equation

$$[\nabla^2 + k^2(\mathbf{P})] \mathcal{G}(\mathbf{P}, \mathbf{P}') = -\delta(\mathbf{P} - \mathbf{P}') \quad (14)$$

and which vanishes on the boundary S. In the case that S is plane, the Green's function must be chosen not only to vanish on the boundary plane but also to satisfy the radiation condition at infinity.

In exactly the same way we would proceed if k

were a constant we obtain

$$\hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2) = \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \times \frac{\partial \mathcal{G}_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'} \frac{\partial \mathcal{G}_2(\mathbf{P}_2, \mathbf{S}_2')}{\partial n_2'} \, dS_1' \, dS_2'. \quad (15)$$

Primes in (15) and in following equations are used to indicate the variables of integration, and corresponding subscripts and primes attached to the surface S also serve this purpose. Explicit dependence on the frequency ν has been omitted for conciseness. It will be shown below that $\mathcal{G}_2 = \mathcal{G}_1^*$.

An explicit form for $\mathcal{G}(\mathbf{P}, \mathbf{P}')$ can be obtained using an iterative procedure. We first rewrite (14) representing $k^2(\mathbf{P}, \nu)$ as the sum of a fixed mean value $k^2(\nu)$ and a variable part with zero mean, $k^2(\nu)\epsilon(\mathbf{P}, \nu)$. Thus,

$$(\nabla^2 + k^2)\mathcal{G}(\mathbf{P}, \mathbf{P}') = -\delta(\mathbf{P} - \mathbf{P}') - k^2\epsilon(\mathbf{P})\mathcal{G}(\mathbf{P}, \mathbf{P}'). \quad (16)$$

[Physically $\epsilon(\mathbf{P})$ can be said to be the variable part of the dielectric constant of the medium.] Now Eq. (16) can be taken formally to be an inhomogeneous constant coefficient Helmholtz equation with right-hand side as the source term. Accordingly, a solution to (16) can be obtained in terms of the Green's function $G(P, P'')$ which satisfies the constant coefficient equation.

$$(\nabla^2 + k^2)G(\mathbf{P}, \mathbf{P}'') = -\delta(\mathbf{P} - \mathbf{P}'') \quad (17)$$

and vanishes on the boundary surface. Recalling that G itself also vanishes on the surface, we obtain

$$\mathcal{G}(\mathbf{P}, \mathbf{P}') = G(\mathbf{P}, \mathbf{P}') + k^2 \int_{V''} \epsilon(\mathbf{P}'')\mathcal{G}(\mathbf{P}'', \mathbf{P}')G(\mathbf{P}, \mathbf{P}'') \, dV''. \quad (18)$$

Equation (18) can now be used as the basis for an iterative development and we obtain

$$\mathcal{G}(\mathbf{P}, \mathbf{P}') = G(\mathbf{P}, \mathbf{P}') + k^2 \int_{V''} \epsilon(\mathbf{P}'')G(\mathbf{P}'', \mathbf{P}')G(\mathbf{P}'', \mathbf{P}) \, dV'' + k^4 \iint_{V''V'''} \epsilon(\mathbf{P}'')\epsilon(\mathbf{P}''')G(\mathbf{P}''', \mathbf{P}'') \times G(\mathbf{P}'', \mathbf{P}')G(\mathbf{P}'', \mathbf{P}) \, dV''' \, dV'' + \cdots. \quad (19)$$

Substituting the iterative series (19) into (15) we obtain the following expression for the Fourier transform of the mutual coherence function:

$$\begin{aligned}
\hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2) &= \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \\
&\times \frac{\partial G_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'} \frac{\partial G_2}{\partial n_2'}(\mathbf{P}_2, \mathbf{S}_2') \, dS_1' \, dS_2' \\
&+ k^2 \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \frac{\partial G_2}{\partial n_2'}(\mathbf{P}_2, \mathbf{S}_2') \\
&\times \int_{V_1''} \epsilon(\mathbf{P}_1'')G_1(\mathbf{P}_1'', \mathbf{P}_1) \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} \, dV_1'' \, dS_1' \, dS_2' \\
&+ k^2 \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \frac{\partial G_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'} \\
&\times \int_{V_2''} \epsilon(\mathbf{P}_2'')G_2(\mathbf{P}_2'', \mathbf{P}_2) \frac{\partial G_2(\mathbf{P}_2'', \mathbf{S}_2')}{\partial n_2'} \, dV_2'' \, dS_1' \, dS_2' \\
&+ k^4 \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \frac{\partial G_2}{\partial n_2'}(\mathbf{P}_2, \mathbf{S}_2') \\
&\times \iint_{V_1'' V_1'''} \epsilon(\mathbf{P}_1'')\epsilon(\mathbf{P}_1''')G_1(\mathbf{P}_1''', \mathbf{P}_1'') \\
&\times G_1(\mathbf{P}_1'', \mathbf{P}_1) \frac{\partial G_1}{\partial n_1'} \, dV_1''' \, dV_1'' \, dS_1' \, dS_2' \\
&+ k^4 \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \\
&\times \iint_{V_1'' V_2''} \epsilon(\mathbf{P}_1'')\epsilon(\mathbf{P}_2'')G_1(\mathbf{P}_1'', \mathbf{P}_1)G_2(\mathbf{P}_2'', \mathbf{P}_2) \\
&\times \frac{\partial G_1}{\partial n_1'}(\mathbf{P}_1'', \mathbf{S}_1') \frac{\partial G_2}{\partial n_2'}(\mathbf{P}_2'', \mathbf{S}_2') \, dV_2'' \, dV_1'' \, dS_2' \, dS_1' \\
&+ k^4 \iint_{S_2' S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2') \frac{\partial G_1(P_1, S_1')}{\partial n'} \\
&\times \iint_{V_2'' V_2'''} \epsilon(\mathbf{P}_2'')\epsilon(\mathbf{P}_2''')G_2(\mathbf{P}_2''', \mathbf{P}_2'')G_2(\mathbf{P}_2'', \mathbf{P}_2) \\
&\times \frac{\partial G_2}{\partial n_2'} \, dV_2''' \, dV_2'' \, dS_1' \, dS_2' + \cdots. \quad (20)
\end{aligned}$$

The Green's function $G(P, P')$ has been determined in an earlier paper[1] for the important case in which S is a plane surface and we give the result here:

$$G(\mathbf{P}, \mathbf{P}') = \frac{e^{+ik|\mathbf{P}-\mathbf{P}'|}}{|\mathbf{P} - \mathbf{P}'|} - \frac{e^{+ik|\mathbf{P}_i-\mathbf{P}'|}}{|\mathbf{P}_i - \mathbf{P}'|}. \quad (21)$$

Here, P_i denotes the image of the point P in the plane S (see Fig. 1). The plus sign is taken where a subscript 1 appears in (20), and the minus sign is taken where a subscript 2 appears in (20) or vice versa.

It should be noted here that an identical expression for $\hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2, \nu)$ can be obtained by a procedure which iterates for the transform of the mutual coherence function itself, rather than for the Green's

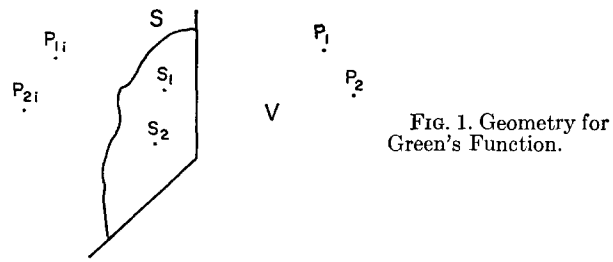

FIG. 1. Geometry for Green's Function.

function for the Helmboltz equation with a variable coefficient. This procedure begins by putting the pair of differential equations (13a, 13b) in the form of a pair of equivalent integral equations by employing Green's functions. Thus,

$$\hat{\Gamma}(\mathbf{P}_1, \mathbf{S}_2, \nu) = -\int_{S_1'} \hat{\Gamma}(\mathbf{S}_1', \mathbf{S}_2, \nu) \frac{\partial G_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'} dS_1' + k^2 \int_{V_1'} \epsilon(\mathbf{P}_1')\hat{\Gamma}(\mathbf{P}_1', \mathbf{S}_2) G_1(\mathbf{P}_1, \mathbf{P}_1')\, dV_1' \quad (22a)$$

$$\hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2, \nu) = -\int_{S_2'} \hat{\Gamma}(\mathbf{P}_1, \mathbf{S}_2') \frac{\partial G_2(\mathbf{P}_2, \mathbf{S}_2')}{\partial n_2'} dS_2' + k^2 \int_{V_2'} \epsilon(\mathbf{P}_2')\hat{\Gamma}(\mathbf{P}_1, \mathbf{P}_2') G_2(\mathbf{P}_2, \mathbf{P}_2')\, dV_2'. \quad (22b)$$

Here, the Green's function $G(\mathbf{P}, \mathbf{P}')$ is identical with the Green's function of Eq. (17), and we have again employed the representation

$$k^2(\mathbf{P}) = k^2[1 + \epsilon(\mathbf{P})].$$

The integral equations (22a, 22b) are then used as the basis for an iterative procedure which starts with the surface term as a zeroth order approximation and by successive substitutions yields the same iterative series for $\hat{\Gamma}(P_1, P_2)$ as obtained above.

The form of the iterative solution obtained [Eq. (20)] for the propagation of the mutual coherence function in a medium in which the refractive index varies is that of the uniform space solution

$$\iint_{S_2' S_1'} \hat{\Gamma}(S_1', S_2') \frac{\partial G_1(P_1, S_1')}{\partial n_1'} \frac{\partial G_2(P_2, S_2')}{\partial n_2'} dS_1'\, dS_2'$$

modified by a series of correction terms which involve volume integrals of a quantity associated with the fluctuations of the refractive index. When the refractive index is constant, the correction terms become zero and the solution reduces to the uniform space solution.

The iterative solution obtained is, of course, that for a single spectral component of $\Gamma_{12}(\tau)$. To obtain $\Gamma_{12}(\tau)$, the iterative solution $\hat{\Gamma}_{12}(\tau)$ must be substituted in Eq. (8) and the integration over ν performed. However, in the case of most practical interest, that of quasi-monochromatic fields, it is unnecessary to actually perform the integration. A quasi-monochromatic field is one for which the effective spectral range $\Delta\nu$ is small compared to the mean frequency $\bar{\nu}$; that is, $\Delta\nu/\bar{\nu} \ll 1$. For this case, provided that the time difference τ is small compared to the coherence time $1/\Delta\nu$, it is known[1,3] that the mutual coherence function is of the form

$$\Gamma_{12}(\tau) \simeq \Gamma_{12}(0) \exp(-2\pi i\bar{\nu}\tau), \qquad |\tau| \ll 1/\Delta\nu, \quad (23)$$

where $\bar{\nu}$ is the mean frequency of $\Gamma_{12}(\tau)$. Substituting from (23) into (7), we obtain

$$[\nabla_m^2 + k^2(\mathbf{P}_m, \bar{\nu})]\Gamma_{12}(0) = 0, \qquad (m = 1, 2), \quad (24)$$

where

$$k(\mathbf{P}_m, \bar{\nu}) = 2\pi\bar{\nu}/C(\mathbf{P}_m).$$

Thus, under the quasi-monochromatic approximation of narrow spectral width and small path differences, we have shown that $\Gamma_{12}(0)$ satisfies the same pair of Helmholtz equations (11) as does $\hat{\Gamma}_{12}(\nu)$ with ν now fixed at the mean frequency $\bar{\nu}$. The boundary condition becomes $\Gamma(\mathbf{S}_1, \mathbf{S}_2, 0)$ and the iterative solution yields $\Gamma_{12}(0)$ which when substituted in (23) gives us the quasi-monochromatic solution. (It should be noted that the varying propagation velocity implies that the time difference τ may be different for two paths of the same geometric length.)

The Green's function formulation of the solution in Eq. (15) leads directly to an important result which we state as a theorem.

Theorem. The field produced by a quasi-monochromatic *coherent* source extended over a surface in contact with a source-free time-invariant medium with arbitrary refractive index variation (in space) is itself coherent.

Proof. The proof of this theorem starts with the result[1,5] that a quasi-monochromatic field (source) is coherent if, and only if, the mutual intensity $\Gamma_{12}(0)$ can be represented as the product of a wave function U evaluated at P_1 with its complex conjugate U^* evaluated at P_2; that is,

$$\Gamma(\mathbf{P}_1, \mathbf{P}_2, 0) = U(\mathbf{P}_1)U^*(\mathbf{P}_2).$$

Now, in the quasi-monochromatic approximation we have from (15) and the discussion above that

$$\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau) \simeq \exp(-2\pi i\bar{\nu}\tau) \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \times \frac{\partial \mathcal{G}_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'} \frac{\partial \mathcal{G}_2(\mathbf{P}_2, \mathbf{S}_2')}{\partial n_2'} dS_1'\, dS_2', \times \left(|\tau| \ll \frac{1}{\Delta\nu}\right). \quad (25)$$

[5] L. Mandel and E. Wolf, J. Opt. Soc. Am. 51, 815 (1961).

Assuming the source to be coherent allows us to write

$$\Gamma(\mathbf{P}_1, \mathbf{P}_2, 0) = \left[\int_{S'} U(\mathbf{S}') \frac{\partial \mathcal{G}_1(\mathbf{P}_1, \mathbf{S}')}{\partial n'} dS'\right] \times \left[\int_{S'} U^*(\mathbf{S}') \frac{\partial \mathcal{G}_2(\mathbf{P}_2, \mathbf{S}')}{\partial n'} dS'\right] \quad (26)$$

where the notation has been changed slightly to make the relationship between the bracketed quantities more apparent. The proof will be complete if we can show that

$$\partial \mathcal{G}_2/\partial n' = \partial \mathcal{G}_1^*/\partial n'.$$

To prove this relationship we make use of the lemma[2] that

$$\Gamma^*(\mathbf{P}_1, \mathbf{P}_2, 0) = \Gamma(\mathbf{P}_2, \mathbf{P}_1, 0). \quad (27)$$

Substituting from (26) into (27) we have

$$\left[\int_{S'} U^*(\mathbf{S}') \frac{\partial \mathcal{G}_1^*(\mathbf{P}_1, \mathbf{S}')}{\partial n'} dS'\right] \times \left[\int_{S'} U(\mathbf{S}') \frac{\partial \mathcal{G}_2^*(\mathbf{P}_2, \mathbf{S}')}{\partial n'}\right] = \left[\int_{S'} U(\mathbf{S}') \frac{\partial \mathcal{G}_1(\mathbf{P}_2, \mathbf{S}')}{\partial n'} dS'\right] \times \left[\int_{S'} U^*(\mathbf{S}') \frac{\partial \mathcal{G}_2(\mathbf{P}_1, \mathbf{S}')}{\partial n'} dS'\right]. \quad (28)$$

Since each of the bracketed quantities is a function of one point only, we can equate the corresponding quantities obtaining

$$\int_{S'} U^*(\mathbf{S}') \frac{\partial \mathcal{G}_1^*}{\partial n'}(\mathbf{P}_1, \mathbf{S}') dS' = \int_{S'} U^*(\mathbf{S}') \frac{\partial \mathcal{G}_2}{\partial n'}(\mathbf{P}_1, \mathbf{S}') dS' \quad (29)$$

and the desired relationship follows immediately.

IV. STATISTICALLY INHOMOGENEOUS MEDIA

We now turn our attention to the important case in which the refractive index is a stationary (spatially homogeneous) isotropic random process. We assume that the fluctuations $\epsilon(\mathbf{P})$ satisfy the relation

$$\langle \epsilon(\mathbf{P}_1)\epsilon(\mathbf{P}_2)\rangle_{av} = \langle \epsilon(\mathbf{P})^2\rangle_{av} C(\rho).$$

Here $\langle \epsilon(\mathbf{P}_1)\epsilon(\mathbf{P}_2)\rangle_{av}$ denotes the average of $\epsilon(\mathbf{P}_1)\epsilon(\mathbf{P}_2)$ taken over all pairs of points $\mathbf{P}_1$ and $\mathbf{P}_2$ a fixed distance ρ apart; $\langle \epsilon(\mathbf{P})^2\rangle_{av}$ is the mean square deviation (of the dielectric constant); and $C(\rho)$ is a correlation function which depends on the separation distance only.

In practice, the fact that we are limited to a statistical knowledge of the medium implies that the most we can expect from our formulation of the propagation problem is a prediction of effects "on the average." For example, suppose that we have a plane quasi-monochromatic light source on a slab of ground glass and are interested in the distribution of the mutual coherence function on the far side of the slab. All that we can expect of our solution is that it yield a prediction of the coherence function averaged over a large number of different slabs of glass with the same statistical properties. In other words, we can predict the ensemble average distribution of the mutual coherence function but cannot accurately predict the distribution for a particular slab.

Suppose then that a series of measurements are made to determine $\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)$ corresponding to a series of independent but statistically identical samples of a medium with refractive index fluctuations. The source distribution $\Gamma(\mathbf{S}_1, \mathbf{S}_2, \tau)$ and the geometrical relations are assumed to be identical for the entire series of measurements. We are interested in the average value of $\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)$. Let us idealize the average by letting the number of experiments N become very large and asking for the limit of the average as $N \to \infty$. Thus let $\Gamma_j(\mathbf{P}_1, \mathbf{P}_2, \tau)$ denote the jth measurement; then we seek to determine the ensemble average of $\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)$, $[\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)]$ defined by

$$[\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)] = \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \Gamma_j(\mathbf{P}_1, \mathbf{P}_2, \tau).$$

Referring to Eq. (20), recalling that only ϵ varies from one measurement to another, and using the quasi-monochromatic approximation, we can write

$$[\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)] \simeq \exp(-2\pi i\bar{\nu}\tau) \times \left\{\iint_{S_2'S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1}{\partial n_1'} \frac{\partial G_2}{\partial n_2'} dS_1' dS_2' + k^2 \iint_{S_2'S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_2}{\partial n_2'} \int_{V_1''} [\epsilon(\mathbf{P}_1'')] \times G_1(\mathbf{P}_1'', \mathbf{P}_1) \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} dV_1'' dS_1' dS_2' + k^2 \iint_{S_2'S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1}{\partial n_1'} \int_{V_2''} [\epsilon(\mathbf{P}_2'')] \times G_2(\mathbf{P}_2'', \mathbf{P}_2) \frac{\partial G_2(P_2'', S_2')}{\partial n_2'} dV_2'' dS_1' dS_2' + \cdots,$$

where we have taken the ensemble average inside

the integrals. Since the samples of the medium associated with the series of experiments are assumed to be independent and statistically homogeneous, we can equate the ensemble average of $\epsilon(P_1'')$, $\epsilon(P_1'')\ \epsilon(P_2'')$, etc., (formed with *fixed* points and *different* samples of the medium) with the averages of these same quantities obtained with a particular sample and allowing the points to vary (preserving distance relationships when these enter into consideration). Hence,

$$[\epsilon(\mathbf{P}_m'')] = \langle\epsilon(\mathbf{P})\rangle_{\text{av}} = 0, \qquad (m = 1, 2),$$

$$[\epsilon(\mathbf{P}_m'')\epsilon(\mathbf{P}_m''')] = \langle\epsilon(\mathbf{P}_m'')\epsilon(\mathbf{P}_m''')\rangle_{\text{av}} = \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} C(P), \qquad (m = 1, 2),$$

etc.

Thus, to second-order terms we obtain the result that the ensemble average of the mutual coherence function (in the quasi-monochromatic approximation) is given by

$$[\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)] \simeq \exp(-2\pi i\bar{\nu}\tau)$$

$$\times \left\{ \iint_{S_1' S_2'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1}{\partial n_1'} \frac{\partial G_2}{\partial n_2'} dS_1' dS_2' \right.$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_2(\mathbf{P}_2, \mathbf{S}_2')}{\partial n_2'}$$

$$\times \iint_{V_1'' V_1'''} C(|\mathbf{P}_1''' - \mathbf{P}_1''|) G_1(\mathbf{P}_1''', \mathbf{P}_1'') G_1(\mathbf{P}_1'', \mathbf{P}_1)$$

$$\times \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} dV_1'' dV_1'' dS_1' dS_2'$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0)$$

$$\times \iint_{V_1'' V_2''} C(|\mathbf{P}_1'' - \mathbf{P}_2''|) G_1(\mathbf{P}_1'', \mathbf{P}_1) G_2(\mathbf{P}_2'', \mathbf{P}_2)$$

$$\times \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} \frac{\partial G_2(\mathbf{P}_2'', \mathbf{S}_2')}{\partial n_2'} dV_1'' dV_2'' dS_1' dS_2'$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'}$$

$$\times \iint_{V_2'' V_2'''} C(|\mathbf{P}_2''' - \mathbf{P}_2''|) G_2(\mathbf{P}_2''', \mathbf{P}_2'') G_2(\mathbf{P}_2'', \mathbf{P}_2)$$

$$\times \frac{\partial G_2(\mathbf{P}_2'', \mathbf{S}_2')}{\partial n_2'} dV_2''' dV_2'' dS_1' dS_2'.$$

The statistics of the medium enter into this expression in the mean square of the refractive index fluctuations, $\langle\epsilon(P)^2\rangle_{\text{av}}$, and the two-point correlation function $C(P)$ which must be integrated over the volume of the medium.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Schrödinger Scattering Amplitude. III*

ALEX GROSSMANN
Institute for Advanced Study, Princeton, New Jersey
AND
TAI TSUN WU†
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received February 9, 1962)

The methods of an earlier paper are used to obtain a domain of analyticity for the Schrödinger scattering amplitude minus the first Born term. The connection between the scattering integral equation and the Schrödinger equation is also studied.

1. INTRODUCTION

IN the first paper of this series[1] it is shown that the scattering integral equation may be replaced by an integral equation with square integrable kernel, which can be studied by standard methods and yields information on the analyticity properties of the scattering amplitude.

The results on analyticity, obtained in (I), refer to the total scattering amplitude. If the first Born term is subtracted from it, the remainder can be studied in a larger region. This is done in Sec. 2 of the present paper; the assumptions on the potential

* Work supported in part by the National Science Foundation.
† Alfred P. Sloan Foundation Fellow.
[1] A. Grossmann and T. T. Wu, J. Math. Phys. **2**, 710 (1961), referred to as (I).

the integrals. Since the samples of the medium associated with the series of experiments are assumed to be independent and statistically homogeneous, we can equate the ensemble average of $\epsilon(P_1'')$, $\epsilon(P_1'')\ \epsilon(P_2'')$, etc., (formed with *fixed* points and *different* samples of the medium) with the averages of these same quantities obtained with a particular sample and allowing the points to vary (preserving distance relationships when these enter into consideration). Hence,

$$[\epsilon(\mathbf{P}_m'')] = \langle\epsilon(\mathbf{P})\rangle_{\text{av}} = 0, \qquad (m = 1, 2),$$

$$[\epsilon(\mathbf{P}_m'')\epsilon(\mathbf{P}_m''')] = \langle\epsilon(\mathbf{P}_m'')\epsilon(\mathbf{P}_m''')\rangle_{\text{av}} = \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} C(P), \qquad (m = 1, 2),$$

etc.

Thus, to second-order terms we obtain the result that the ensemble average of the mutual coherence function (in the quasi-monochromatic approximation) is given by

$$[\Gamma(\mathbf{P}_1, \mathbf{P}_2, \tau)] \simeq \exp(-2\pi i\bar{\nu}\tau)$$

$$\times \left\{ \iint_{S_1' S_2'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1}{\partial n_1'} \frac{\partial G_2}{\partial n_2'} dS_1' dS_2' \right.$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_2(\mathbf{P}_2, \mathbf{S}_2')}{\partial n_2'}$$

$$\times \iint_{V_1'' V_1'''} C(|\mathbf{P}_1''' - \mathbf{P}_1''|) G_1(\mathbf{P}_1''', \mathbf{P}_1'') G_1(\mathbf{P}_1'', \mathbf{P}_1)$$

$$\times \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} dV_1'' dV_1'' dS_1' dS_2'$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0)$$

$$\times \iint_{V_1'' V_2''} C(|\mathbf{P}_1'' - \mathbf{P}_2''|) G_1(\mathbf{P}_1'', \mathbf{P}_1) G_2(\mathbf{P}_2'', \mathbf{P}_2)$$

$$\times \frac{\partial G_1(\mathbf{P}_1'', \mathbf{S}_1')}{\partial n_1'} \frac{\partial G_2(\mathbf{P}_2'', \mathbf{S}_2')}{\partial n_2'} dV_1'' dV_2'' dS_1' dS_2'$$

$$+ k^4 \langle\epsilon(\mathbf{P})^2\rangle_{\text{av}} \iint_{S_2' S_1'} \Gamma(\mathbf{S}_1', \mathbf{S}_2', 0) \frac{\partial G_1(\mathbf{P}_1, \mathbf{S}_1')}{\partial n_1'}$$

$$\times \iint_{V_2'' V_2'''} C(|\mathbf{P}_2''' - \mathbf{P}_2''|) G_2(\mathbf{P}_2''', \mathbf{P}_2'') G_2(\mathbf{P}_2'', \mathbf{P}_2)$$

$$\times \frac{\partial G_2(\mathbf{P}_2'', \mathbf{S}_2')}{\partial n_2'} dV_2''' dV_2'' dS_1' dS_2'.$$

The statistics of the medium enter into this expression in the mean square of the refractive index fluctuations, $\langle\epsilon(P)^2\rangle_{\text{av}}$, and the two-point correlation function $C(P)$ which must be integrated over the volume of the medium.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Schrödinger Scattering Amplitude. III*

ALEX GROSSMANN
Institute for Advanced Study, Princeton, New Jersey
AND
TAI TSUN WU†
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received February 9, 1962)

The methods of an earlier paper are used to obtain a domain of analyticity for the Schrödinger scattering amplitude minus the first Born term. The connection between the scattering integral equation and the Schrödinger equation is also studied.

1. INTRODUCTION

IN the first paper of this series[1] it is shown that the scattering integral equation may be replaced by an integral equation with square integrable kernel, which can be studied by standard methods and yields information on the analyticity properties of the scattering amplitude.

The results on analyticity, obtained in (I), refer to the total scattering amplitude. If the first Born term is subtracted from it, the remainder can be studied in a larger region. This is done in Sec. 2 of the present paper; the assumptions on the potential

* Work supported in part by the National Science Foundation.
† Alfred P. Sloan Foundation Fellow.
[1] A. Grossmann and T. T. Wu, J. Math. Phys. **2**, 710 (1961), referred to as (I).

are there a little more restrictive than those of (I). The domain of analyticity is described by Theorem 1. It includes the domain obtained by Hunziker[2] which, in turn, includes the domains previously obtained by many other authors. It should be noted, however, that the present results are not direct improvements of Hunziker's since the assumptions on the potential differ in some details from his.

The Hilbert space introduced in (I) is not the space of state vectors. Theorems 2 and 4 are concerned with the relationship between the two spaces, and Theorem 4 also clarifies the relationship between the scattering integral equation and the eigenvalue problem of the Hamiltonian. The Hamiltonian is defined by the method of Kato[3] which fits precisely into the general scheme of quantum mechanics and does not require the introduction of boundary conditions. Theorem 4 reduces the study of bound states to the study of an integral equation with square integrable kernel and simple symmetry properties. A few immediate consequences are mentioned at the end of Sec. 4. They are not new results[4] but are given here because the same methods of proof can also be applied to other problems which will be studied in a later paper.

2. EXTENDED DOMAIN OF ANALYTICITY

It will be assumed that $V(\mathbf{x})$ (the potential multiplied by $2m\hbar^{-2}$) is measurable and that it satisfies the following conditions:

(a) There exists a positive number α (possibly infinity) such that for every $\kappa < \alpha$, the integral

$$\int \exp(\kappa\,|\mathbf{x}|)\,|V(\mathbf{x})|\,d\mathbf{x} \tag{2.1}$$

converges and that, for every $\kappa > \alpha$, the integral (2.1) diverges.

(b) There exists a positive number $\bar{\alpha}$ (possibly infinity) such that for every $\bar{\kappa} < \bar{\alpha}$, the integral

$$\iint \exp[\bar{\kappa}(|\mathbf{x}| + |\mathbf{y}|)] \times |V(\mathbf{x})V(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y} \tag{2.2}$$

converges and that, for every $\bar{\kappa} > \bar{\alpha}$, the integral (2.2) diverges.

It will now be shown that the inequality

$$\bar{\alpha} \leq \alpha \tag{2.3}$$

holds. Let $\bar{\kappa} < \bar{\alpha}$, and write $\bar{\kappa} = \kappa + \epsilon$, where ϵ is a given positive number. There exists a constant C such that

$$\exp\{\epsilon(|\mathbf{x}| + |\mathbf{y}|)\}\,|\mathbf{x}-\mathbf{y}|^{-2} \geq C$$

for every $\mathbf{x}$ and every $\mathbf{y}$. Thus

$$\iint \exp\{(\kappa+\epsilon)(|\mathbf{x}| + |\mathbf{y}|)\}\,|\mathbf{x}-\mathbf{y}|^{-2} \times |V(\mathbf{x})V(\mathbf{y})|\,d\mathbf{x}\,d\mathbf{y} \geq C\left(\int \exp(\kappa\,|\mathbf{x}|)\,V(\mathbf{x})\,d\mathbf{x}\right)^2,$$

which shows that the integral (2.1) is convergent for every $\kappa < \bar{\alpha}$ and proves (2.3).

The amplitude $f(\mathbf{q}, \mathbf{p}; k)$ is defined in (I) as

$$\begin{aligned} f(\mathbf{q}, \mathbf{p}; k) &= (b_{-\mathbf{q}}, [1 - A(k)]^{-1} a_{\mathbf{p}}) \\ &= (b_{-\mathbf{q}}, [1 + N(k)] a_{\mathbf{p}}), \end{aligned} \tag{2.4}$$

where

$$A(\mathbf{x}, \mathbf{y}; k) = -(4\pi)^{-1} s(\mathbf{x})\,|V(\mathbf{x})V(\mathbf{y})|^{\frac{1}{2}} \times \exp(ik\,|\mathbf{x}-\mathbf{y}|)\,|\mathbf{x}-\mathbf{y}|^{-1} \tag{2.5}$$

$$s(\mathbf{x}) = \operatorname{sign} V(\mathbf{x}),$$

$$a_{\mathbf{p}}(\mathbf{x}) = (2\pi)^{-3/2} s(\mathbf{x})\,|V(\mathbf{x})|^{1/2} \exp(i\mathbf{p}\cdot\mathbf{x}), \tag{2.6}$$

$$b^*_{\mathbf{q}}(\mathbf{x}) = (2\pi)^{-3/2}\,|V(\mathbf{x})|^{1/2} \exp(i\mathbf{q}\cdot\mathbf{x}), \tag{2.7}$$

and the asterisk denotes complex conjugation. The function f is holomorphic[1] at every point of the region

$$\operatorname{Im} k > -\tfrac{1}{2}\alpha, \tag{2.8}$$

$$|\operatorname{Im} \mathbf{p}| < \tfrac{1}{2}\alpha, \tag{2.9}$$

$$|\operatorname{Im} \mathbf{q}| < \tfrac{1}{2}\alpha, \tag{2.10}$$

with the exception of points at which an eigenvalue of the completely continuous operator $A(k)$ is equal to one.

If the first Born term is subtracted from f, the remainder is

$$\begin{aligned} f^{(1)}(\mathbf{q}, \mathbf{p}; k) &= (b_{-\mathbf{q}}, N(k) a_{\mathbf{p}}) \\ &= \iint b^*_{\mathbf{q}}(\mathbf{x}) N(\mathbf{x}, \mathbf{y}; k) a_{\mathbf{p}}(\mathbf{y})\,d\mathbf{x}\,d\mathbf{y}. \end{aligned} \tag{2.11}$$

It will now be shown that the function $f^{(1)}$, defined by the r.h.s. of (2.11) has analyticity properties which can be studied in a region larger than that given by (2.8), (2.9), and (2.10).

Theorem 1. *Assume that*

$$\operatorname{Im} k > -\tfrac{1}{2}\alpha, \tag{2.12}$$

that the number one is not an eigenvalue of $A(k)$, *and*

[2] W. Hunziker, Helv. Phys. Acta **34**, 593 (1961). We would like to thank him for preprints of his paper. Dr. Hunziker has informed us that he has also extended his domain of analyticity.

[3] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951). This paper was pointed out to us by Professor V. Bargmann.

[4] T. Ikebe, Arch. Rat. Mech. Analysis. **5**, 1 (1960).

that the complex vectors $\mathbf{p}$, $\mathbf{q}$ *satisfy*

$$|\text{Im } \mathbf{p}| < \min [\tfrac{1}{2}(\alpha + \bar{\alpha}), \alpha + \text{Im } k] \tag{2.13}$$

$$|\text{Im } \mathbf{q}| < \min [\tfrac{1}{2}(\alpha + \bar{\alpha}), \alpha + \text{Im } k]. \tag{2.14}$$

Then $f^{(1)}$ *is holomorphic at* $(\mathbf{q}, \mathbf{p}; k)$.

Proof. For any real number λ, define

$$a_{\mathbf{p}}(\mathbf{x}; \lambda) = \exp(-\tfrac{1}{2}\lambda |\mathbf{x}|) a_{\mathbf{p}}(\mathbf{x}), \tag{2.15}$$

$$b_{\mathbf{q}}(\mathbf{x}; \lambda) = \exp(-\tfrac{1}{2}\lambda |\mathbf{x}|) b_{\mathbf{q}}(\mathbf{x}), \tag{2.16}$$

$$A(\mathbf{x}, \mathbf{y}; k; \lambda) = \exp\{\tfrac{1}{2}\lambda(|\mathbf{x}| + |\mathbf{y}|)\} A(\mathbf{x}, \mathbf{y}; k), \tag{2.17}$$

and

$$N(\mathbf{x}, \mathbf{y}; k; \lambda) = \exp\{\tfrac{1}{2}\lambda(|\mathbf{x}| + |\mathbf{y}|)\} N(\mathbf{x}, \mathbf{y}; k). \tag{2.18}$$

Clearly

$$\iint b^*_{\mathbf{q}}(\mathbf{x}) N(\mathbf{x}, \mathbf{y}; k) a_{\mathbf{p}}(\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}$$

$$= \iint b^*_{-\mathbf{q}}(\mathbf{x}; \lambda) N(\mathbf{x}, \mathbf{y}; k; \lambda) a_{\mathbf{p}}(\mathbf{y}; \lambda)\, d\mathbf{x}\, d\mathbf{y} \tag{2.19}$$

for every λ. The main part of the proof will consist in showing that by the assumptions of the theorem, λ can be chosen in such a way as to make square integrable every factor on the r.h.s. of (2.19). Once this is shown, the assertion of the theorem follows, as in (I), from general statements about inner products of Hilbert space elements which depend analytically on parameters.

Lemma 1.1. *If*

$$\lambda > 2 |\text{Im } \mathbf{p}| - \alpha, \tag{2.20}$$

then $a_{\mathbf{p}}(\lambda) \in L^{(2)}$.

Lemma 1.2. If

$$\lambda > 2 |\text{Im } \mathbf{q}| - \alpha, \tag{2.21}$$

then $b_{-\mathbf{q}}(\lambda) \in L^{(2)}$.

The proofs are straightforward.

Lemma 1.3. If

$$\lambda < \bar{\alpha}, \tag{2.22}$$

and

$$\lambda < \alpha + 2 \text{ Im } k, \tag{2.23}$$

then $A(x, y; k; \lambda)$ is square integrable over the whole $\mathbf{x}$, $\mathbf{y}$ space.

Proof. Define V' by

$$V'(\mathbf{x}) = \exp(\lambda |\mathbf{x}|) V(\mathbf{x}).$$

Then V' satisfies the assumptions of (I), with

$$\alpha' = \alpha - \lambda > 0,$$

and $A(\mathbf{x}, \mathbf{y}; k; \lambda) = A'(\mathbf{x}, \mathbf{y}; k)$ is square integrable for

$$\text{Im } k > -\tfrac{1}{2}\alpha' = \tfrac{1}{2}(\lambda - \alpha).$$

This proves Lemma 3.

It follows immediately that, if (2.22) and (2.23) hold, the completely continuous operator $A(k; \lambda)$ depends holomorphically on k.

Lemma 1.4. If (2.22), (2.23) are satisfied and

$$\lambda > 0, \tag{2.24}$$

then the kernel $N(\mathbf{x}, \mathbf{y}; k; \lambda)$, defined by (2.18) is square integrable over the whole $\mathbf{x}$, $\mathbf{y}$ space.

Proof. Define $A_1(\mathbf{x}, \mathbf{y}; k; \lambda)$ by

$$A_1(\mathbf{x}, \mathbf{y}; k; \lambda) = \exp\{\tfrac{1}{2}\lambda(|\mathbf{x}| - |\mathbf{y}|)\} A(\mathbf{x}, \mathbf{y}; k).$$

This kernel is majorized by $A(\mathbf{x}, \mathbf{y}; k; \lambda)$ and is consequently square integrable over the $\mathbf{x}$, $\mathbf{y}$ space. The operator

$$N_1(k; \lambda) = -1 + [1 - A_1(k; \lambda)]^{-1}$$

exists and is represented by a square integrable kernel. Otherwise the equation

$$\int \exp(\tfrac{1}{2}\lambda |\mathbf{x}|) A(\mathbf{x}, \mathbf{y}; k) \exp(-\tfrac{1}{2}\lambda |\mathbf{y}|) \phi(\mathbf{y})\, d\mathbf{y} = \phi(\mathbf{x})$$

would have a nonzero solution $\phi \in L^{(2)}$. Then, however, $\psi(\mathbf{x}) = \exp(-\tfrac{1}{2}\lambda |\mathbf{x}|)\phi(\mathbf{x})$ would also be a nonzero element of $L^{(2)}$ and would satisfy

$$\int A(\mathbf{x}, \mathbf{y}; k) \psi(\mathbf{y})\, d\mathbf{y} = \psi(\mathbf{x}).$$

This, however, is impossible, since $N(k)$ was assumed to exist.

The kernel $N(\mathbf{x}, \mathbf{y}; k; \lambda)$ satisfies

$$N(\mathbf{x}, \mathbf{y}; k; \lambda) = A(\mathbf{x}, \mathbf{y}; k; \lambda) + \int A_1(\mathbf{x}, \mathbf{z}; k; \lambda) N(\mathbf{z}, \mathbf{y}; k; \lambda)\, d\mathbf{z}.$$

This equation has, in $L^{(2)}$, the unique solution

$$N(k; \lambda) = [1 - A_1(k; \lambda)]^{-1} A(k; \lambda) = A(k; \lambda) + N_1(k; \lambda) A(k; \lambda). \tag{2.25}$$

Consequently $N(k; \lambda)$ is represented by a square integrable kernel, which proves Lemma 4.

It can be seen from (2.25) that $N(k; \lambda)$ depends holomorphically on k.

Finally, it is easy to see that (2.12), (2.13), and (2.14) are necessary and sufficient conditions for the existence of a number λ which satisfies (2.20), (2.21), (2.22), (2.23), and (2.24) simultaneously. This completes the proof of Theorem 1.

A consequence of the above is

Theorem 2. *In the region* (2.12), (2.13), (2.14) *the amplitude* $f^{(1)}$ *can have a singularity only if* k *is such that the integral equation*

$$\psi(\mathbf{x}) = -(4\pi)^{-1} \times \int \exp(ik\,|\mathbf{x}-\mathbf{y}|)\,|\mathbf{x}-\mathbf{y}|^{-1}\,V(\mathbf{y})\psi(\mathbf{y})\,d\mathbf{y} \quad (2.26)$$

has a solution $\psi(\mathbf{x})$ *for which*

$$\int |\psi(\mathbf{x})|^2\,|V(\mathbf{x})|\,d\mathbf{x} \quad \textit{exists and is positive.} \quad (2.27)$$

Proof. It follows from Theorem 1 that a singularity of $f^{(1)}$ can occur only if there exists a ϕ such that

$$A(k)\phi = \phi \qquad (\phi \in L^{(2)},\ \phi \neq 0). \quad (2.28)$$

It will now be shown that (2.28) can be satisfied if and only if it is possible to satisfy (2.26) and (2.27). Assume first that (2.28) holds. This means

$$\phi(x) = s(\mathbf{x})\,|V(\mathbf{x})|^{1/2} \times \int G_0(\mathbf{x}-\mathbf{y};k)\,|V(\mathbf{y})|^{1/2}\,\phi(\mathbf{y})\,d\mathbf{y}, \quad (2.29)$$

where

$$G_0(\mathbf{x};k) = -(4\pi)^{-1}\exp(ik\,|\mathbf{x}|)\,|\mathbf{x}|^{-1}.$$

Define $\psi(\mathbf{x})$ by

$$\psi(\mathbf{x}) = \int G_0(\mathbf{x}-\mathbf{y})\,|V(\mathbf{y})|^{1/2}\,\phi(\mathbf{y})\,d\mathbf{y}. \quad (2.30)$$

Then, by (2.28),

$$s(\mathbf{x})\,|V(\mathbf{x})|^{1/2}\,\psi(\mathbf{x}) = \phi(\mathbf{x}), \quad (2.31)$$

which shows that $\psi(\mathbf{x})$ satisfies (2.27). By substituting (2.31) into the r.h.s. of (2.30) one obtains (2.26).

Assume now, conversely, that (2.26) and (2.27) are satisfied and define $\phi(x)$ by (2.31). Then (2.26), multiplied by $s(\mathbf{x})\,|V(\mathbf{x})|^{1/2}$, becomes (2.28). Finally, (2.27) shows that $\phi \in L^{(2)}$ and $\phi \neq 0$. This proves Theorem 2.

3. OTHER ASSUMPTIONS ON POTENTIAL

The class of potentials introduced in Sec. 2 is unnecessarily wide for some applications. A convenient subclass is described by

Theorem 3. *Assume that* $V(\mathbf{x})$ *is locally square integrable and that*

$$V(\mathbf{x}) = O[\exp(-\alpha\,|\mathbf{x}|)] \qquad (|\mathbf{x}| \to \infty) \quad (3.1)$$

with $\alpha > 0$. *Then the integral* (2.1) *converges for every* $\kappa < \alpha$, *and the integral* (2.2) *converges for every* $\bar{\kappa} < \alpha$.

Proof. A function which is locally square integrable is summable over every sphere. This, together with (3.1), shows that (2.1) converges. In order to show the convergence of (2.2), write that integral as

$$\iint |V'(\mathbf{x})V'(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y}$$

with

$$V'(\mathbf{x}) = \exp(\kappa\,|\mathbf{x}|)\,V(\mathbf{x}).$$

Denote by $\int_a^b d\mathbf{x}$ the three-dimensional integral over the region $a \leq |\mathbf{x}| \leq b$.

Lemma 3.1. The integral

$$\int_0^R \int_0^R V'(\mathbf{x})V'(\mathbf{y})\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y}$$

is finite for every R.

Proof. Notice that whenever $|\mathbf{x}| \leq R$,

$$\int_0^R |\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{y} \leq 8R\pi$$

and

$$\int_0^R |V'(\mathbf{x})|^2\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y} = \int_0^R |V'(\mathbf{x})|^2 \times \left(\int_0^R |\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{y}\right) d\mathbf{x} \leq 8R\pi \int_0^R |V'(\mathbf{x})|^2\,d\mathbf{x}.$$

This means that $V'(\mathbf{x})\,|\mathbf{x}-\mathbf{y}|^{-1}$ is square integrable over the region $0 \leq |\mathbf{x}| \leq R$, $0 \leq |\mathbf{y}| \leq R$. Then, by the Schwartz inequality,

$$\int_0^R \int_0^R |V'(\mathbf{x})V'(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y} = \int_0^R \int_0^R \frac{|V'(\mathbf{x})|}{|\mathbf{x}-\mathbf{y}|}\,\frac{V'(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\,d\mathbf{x}\,d\mathbf{y} < \infty,$$

which proves the lemma.

Now choose R so that, for $|\mathbf{x}| \geq \frac{1}{2}R$,

$$|V(\mathbf{x})| \leq C\exp(-\beta\,|\mathbf{x}|),$$

where C is a constant and $\beta = \alpha - \bar{\kappa} > 0$. Then

$$\int_R^\infty \int_0^R |V'(\mathbf{x})V'(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y}$$

$$= \int_R^\infty \left(\int_0^{R/2} + \int_{R/2}^R\right) |V'(\mathbf{x})V'(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y}$$

$$\leq \frac{4C}{R^2}\left[\int_R^\infty \exp(-\beta\,|\mathbf{x}|)\,d\mathbf{x}\right]\left(\int_0^{R/2} |V(\mathbf{y})|\,d\mathbf{y}\right)$$

$$+ C^2 \int_R^\infty \int_{R/2}^R \exp[-\beta(|\mathbf{x}| + |\mathbf{y}|)] \times |\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y},$$

which is finite since $V(\mathbf{x})$ is summable over every sphere. Finally

$$\int_R^\infty \int_R^\infty |V'(\mathbf{x})V'(\mathbf{y})|\,|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y} \leq C^2$$

$$\times \int_R^\infty \int_R^\infty \exp\left[-\beta(|\mathbf{x}|+|\mathbf{y}|)\right]|\mathbf{x}-\mathbf{y}|^{-2}\,d\mathbf{x}\,d\mathbf{y} < \infty.$$

This completes the proof of Theorem 2.

An immediate consequence of Theorem 1 and Theorem 3 is

Corollary. Assume that $V(\mathbf{x})$ is locally square integrable and that the asymptotic condition (3.1) is satisfied. Assume furthermore that $\operatorname{Im} k > -\frac{1}{2}\alpha$, and that the number one is not an eigenvalue of $A(k)$. If the complex vectors $\mathbf{p}$, $\mathbf{q}$ satisfy

$$|\operatorname{Im}\mathbf{p}| < \min[\alpha, \alpha + \operatorname{Im} k]$$

$$|\operatorname{Im}\mathbf{q}| < \min[\alpha, \alpha + \operatorname{Im} k]$$

then $f^{(1)}$ is holomorphic at $(\mathbf{q}, \mathbf{p}; k)$.

4. RELATION TO THE HAMILTONIAN

In the present section it is assumed that $V(\mathbf{x})$ is locally square integrable and that it satisfies the asymptotic condition (3.1). In order to define[3] the Hamiltonian operator, it is convenient to consider, together with the Hilbert space $L^{(2)}$ of functions square integrable over the $\mathbf{x}$ space, the (isomorphic) space $\tilde{L}^{(2)}$ of Fourier-Plancherel transforms

$$\tilde{\chi}(\mathbf{s}) = (2\pi)^{-3/2}\int \exp(i\mathbf{s}\cdot\mathbf{x})\chi(\mathbf{x})\,d\mathbf{x} \quad (\chi \in L^{(2)}). \tag{4.1}$$

These are just the wave functions in momentum space. The relations (4.1) and its inverse establish a unitary correspondence between $L^{(2)}$ and $\tilde{L}^{(2)}$.

Denote by $\tilde{D}$ the set of functions $\tilde{\chi} \in \tilde{L}^{(2)}$ which are such that $|\mathbf{s}|^2\tilde{\chi}(\mathbf{s})$ also belongs to $\tilde{L}^{(2)}$. Let D be the corresponding subset[5] of $L^{(2)}$. Every function $\tilde{\chi} \in \tilde{D}$ is summable and, consequently, every $\chi \in D$ is bounded.

Let $\tilde{H}_0$ be the operator in $\tilde{L}^{(2)}$, with domain $\tilde{D}$, defined by

$$(\tilde{H}_0\tilde{\chi})(\mathbf{s}) = |\mathbf{s}|^2\,\tilde{\chi}(\mathbf{s}). \tag{4.2}$$

The corresponding operator in $L^{(2)}$ will be denoted by H_0. Its domain is D.

Denote by D_V the set of all functions $\chi \in L^{(2)}$ which are such that $V(\mathbf{x})\chi(\mathbf{x})$ also belongs to $L^{(2)}$. Define, in the obvious fashion, the operator V with domain D_V. Then $D_V \supseteq D$, since V is square integrable and every $\chi \in D$ is bounded.

[5] See reference 3. Kato's D^0 is our D. If Kato's assumptions are specialized so that (in his notation) $s = 1$, $V' = 0$, then it becomes obvious that our assumptions on V imply Kato's.

The total Hamiltonian H is defined as

$$H = H_0 + V. \tag{4.3}$$

It can be shown[3] that H is self-adjoint.

In order to study the relationship between H and the integral operator $A(k)$ defined by (2.5), consider the set D_K of all functions $\chi \in L^{(2)}$ which are such that $|V(\mathbf{x})|^{1/2}\chi(\mathbf{x})$ also belongs to $L^{(2)}$. Then $D_K \supseteq D_V$, because if $\chi \in D_V$, then $(\chi, V\chi) < \infty$, which means that $|V(\mathbf{x})|\,|\chi(\mathbf{x})|^2$ is summable or that $\chi \in D_K$. Let K_1 and K_2 be operators with domain D_K, defined by

$$(K_1\chi)(\mathbf{x}) = s(\mathbf{x})\,|V(\mathbf{x})|^{1/2}\,\chi(\mathbf{x}), \qquad (\chi \in D_K) \tag{4.4}$$

$$(K_2\chi)(\mathbf{x}) = |V(\mathbf{x})|^{1/2}\,\chi(\mathbf{x}) \qquad (\chi \in D_K). \tag{4.5}$$

Clearly

$$V = K_1K_2 = K_2K_1. \tag{4.6}$$

The equality sign between operators implies here, as throughout this paper, that the domains on both sides are equal.

The main result of this section is

Theorem 4. *If*

$$\operatorname{Im} k > 0 \tag{4.7}$$

then the eigenvalue equation

$$H\psi = k^2\psi \tag{4.8}$$

has a solution $\psi \in L^{(2)}$ $(\psi \neq 0)$ *if and only if the equation*

$$A(k)\phi = \phi \tag{4.9}$$

[*where* $A(k)$ *is the operator defined by the kernel* (2.5)] *has a solution* $\phi \in L^{(2)}$, $(\phi \neq 0)$. *These solutions are in a one-to-one correspondence, established by*

$$\phi = K_1\psi. \tag{4.10}$$

The formal validity of the equivalence between (4.8) and (4.9), (4.10) is not difficult to see. The difficulties in the proof arise from the fact that it is not obvious that every solution to (4.9) [with (4.7)] is of the form (4.10) with $\psi \in D$, whereas (4.8) clearly implies that $\psi \in D$.

In the propositions that follow, it is always assumed that $\operatorname{Im} k > 0$.

Lemma 4.1. The operator $A(k)$ is the closure of the operator

$$K_1(k^2 - H_0)^{-1}K_2. \tag{4.11}$$

Proof. Notice first that the operators $A(k)$ and (4.11) concide for all $\phi \in D_K$. Consequently,

$$||K_1(k^2 - H_0)^{-1}K_2\phi|| \leq ||A(k)||\,||\phi|| \qquad (\phi \in D_K). \tag{4.12}$$

It follows that any Cauchy sequence of elements of D_K is transformed by (4.11) into a Cauchy sequence whose limit is the image, by $A(k)$, of the limit of the original sequence. This proves the proposition. Note that the operator $A(k)$ is completely continuous and that its domain is the whole of $L^{(2)}$.

Lemma 4.2. There exist an integer $n \geq 1$ such that the range of $(A(k))^n$ [the nth power of $A(k)$] is contained in D_K.

Proof. Denote by $A_n(\mathbf{x}, \mathbf{y}; k)$ the kernel which corresponds to the operator $(A(k))^n$. Then $A_n(\mathbf{x}, \mathbf{y}; k)$ is of the form

$$A_n(\mathbf{x}, \mathbf{y}; k) = s(\mathbf{x}) \, |V(\mathbf{x}) V(\mathbf{y})|^{1/2} \, B_n(\mathbf{x}, \mathbf{y}; k),$$

where

$$B_1(\mathbf{x}, \mathbf{y}; k) = -(4\pi)^{-1} \, |\mathbf{x} - \mathbf{y}|^{-1} \exp(ik \, |\mathbf{x} - \mathbf{y}|),$$

and

$$B_{n+1}(\mathbf{x}, \mathbf{y}; k) = \int B_1(\mathbf{x}, \mathbf{z}; k) V(\mathbf{z}) B_n(\mathbf{z}, \mathbf{y}; k) \, d\mathbf{z}.$$

A direct evaluation[6] shows that $B_4(\mathbf{x}, \mathbf{y}; k)$ is bounded. Consequently, for any $\chi \in L^{(2)}$ and $\phi = [A(k)]^4 \chi$, there exist C_1 and C such that

$$|\phi(\mathbf{x})| \leq |V(\mathbf{x})|^{1/2} \left| \int B_4(\mathbf{x}, \mathbf{y}; k) \, |V(\mathbf{y})|^{1/2} \chi(\mathbf{y}) \, d\mathbf{y} \right|$$

$$\leq C_1 \, |V(\mathbf{x})|^{1/2} \int |V(\mathbf{y})|^{1/2} \, |\chi(\mathbf{y})| \, d\mathbf{y} \leq C \, |V(\mathbf{x})|^{1/2}, \quad (4.13)$$

since $|V(\mathbf{y})|^{1/2} \chi(\mathbf{y})$, being the product of two square integrable functions, is summable. Since $V(\mathbf{x})$ is square integrable, the inequality (4.13) shows that $\phi \in D_K$.

Lemma 4.3. The condition $H\psi = k^2\psi$, $(\psi \neq 0)$ is equivalent to

$$\psi = (k^2 - H_0)^{-1} V \psi \qquad (\psi \neq 0). \quad (4.14)$$

Proof. Since k^2 belongs to the resolvent set of H_0, the operator $(k^2 - H_0)^{-1}$ maps, in a one-to-one fashion, the whole of $L^{(2)}$ onto D, which is the domain of both H_0 and H. The remaining part of the proof is straightforward.

It remains to establish the equivalence between (4.14) and (4.9–10).

Lemma 4.4. If ψ satisfies (4.14), then $\phi = K_1\psi$ is not zero and satisfies (4.9).

Proof. Since $D \subseteq D_K$, the operator K_1 can be applied to (4.14), yielding

$$\phi = K_1(k^2 - H_0)^{-1} K_2 \phi$$

which is (4.9) because of Lemma 4.1. If (4.14) is written as

$$\psi = (k^2 - H_0)^{-1} K_2 K_1 \psi,$$

it is seen that $K_1\psi \neq 0$, and the proposition is proved.

Lemma 4.5. Assume that $\phi \in L^{(2)}$, $\phi \neq 0$ and that ϕ satisfies $A(k)\phi = \phi$. Then there exist a unique $\psi \in D$ such that $\phi = K_1\psi$. Moreover, this ψ satisfies $H\psi = k^2\psi$.

Proof. Notice that $(A(k))^n \phi = \phi$ for every integer $n \geq 1$. By Lemma 4.2, $\phi \in D_K$. Define a $\psi \in D$ by

$$\psi = (k^2 - H_0)^{-1} K_2 \phi. \quad (4.15)$$

Then (4.9) becomes $\phi = K_1\psi$, and (4.15) becomes (4.14), so that Lemma 4.3 applies.

This proves Theorem 4.

The results of this section make it possible to relate properties of the Hamiltonian to properties of the completely continuous operator $A(k)$ which, from a mathematical point of view, is a more elementary object. A few immediate results will be given here as an illustration.

Corollary 4.1. The operator H can have at most a finite number of negative eigenvalues.

Proof. It has been shown in (I) that the points k for which (4.9) holds are all contained in a half-plane Im $k < \gamma_0$, and that they have no finite accumulation point. Because of Theorem 4 and fact that H is self-adjoint, these points are restricted, in the half-plane Im $k > 0$, to the imaginary axis, and the assertion follows.

Corollary 4.2. Every negative eigenstate of the Hamiltonian has finite degeneracy.

This is an immediate consequence of Theorem 4 and of the fact that eigenvalues of completely continuous operators have finite multiplicities.

ACKNOWLEDGMENTS

We are indebted to Professor V. Bargmann and Professor F. J. Dyson for very helpful discussions. We would also like to thank Professor J. R. Oppenheimer for his hospitality at the Institute for Advanced Study.

[6] Reference 4, Lemma 2.2, Ikebe's $A^{(n)}$ is equal to our B_{n+1}. His assumptions in the derivation of that Lemma are: (a) square integrability of $V(\mathbf{x})$ and (b) an asymptotic condition weaker than our Eq. (3.1). Consequently, his results apply.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Spectrum of the Scattering Integral Operator in the Physical Energy Sheet*

K. MEETZ†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
and
Brookhaven National Laboratory, Upton, New York
(Received December 9, 1961)

The kernel of the integral equation for the nonrelativistic scattering of a spinless particle by a potential $-\lambda V$ can easily be symmetrized, if V has a definite sign. Under suitable conditions for the potential the symmetrized kernel is square integrable and generates a completely continuous transformation of the space L^2 that has a pure point spectrum.

The scattering kernel is a two-valued function of the energy s. For negative real values of s in the first Riemann sheet it is symmetric (Hermitian) and can be spectral decomposed in the usual way. Eigenvalues and eigenelements can be analytically continued, at least into the first s sheet and even into the finite second s sheet, if the potential decreases faster than any exponential function for $r \to \infty$. Eigenelements at two complex conjugate points s, s^* form a complete biorthogonal system of L^2. The original and the resolvent kernel can be expressed in terms of that system and the eigenvalues, for s in the first and eventually finite second sheet.

If the potential is indefinite, the scattering kernel is of polar type for negative real s and can be represented in terms of polar eigenvalues and eigenelements. These may be continued in the same way as for a definite potential.

The distribution of eigenvalues in the λ plane is studied for arbitrary complex s and related to the occurrence of bound states and resonances. It is shown that Born expansions for resolvent quantities do converge, if and only if neither λV nor $-\lambda V$ create bound states.

I. INTRODUCTION

In the modern version of scattering theory the scattering solutions of the Schrödinger equation are determined by the Lippmann-Schwinger integral equation, a formulation that originated from the interaction representation in time-dependent scattering theory. In agreement with Heisenberg's point of view, all information that a quantum mechanical description can deliver must be contained in the resolvent kernel of that integral equation. Although it is well known that bound states appear as poles in the resolvent kernel, regarded as a function of energy, there has been considerable confusion about the physical meaning of the different kinds of poles that might occur.

It seems, therefore, to be worthwhile to extract the information from the scattering integral equation in a more systematic manner. A natural starting point for such considerations is the remarkable property of the scattering kernel to be square integrable for a wide class of potentials. Square-integrable kernels generate completely continuous transformations of the space L^2 that are known to have a pure point spectrum.

We confine ourselves to the simple problem of the nonrelativistic scattering of a spinless particle by a local potential $-\lambda V$. λ may be considered as eigenvalue parameter of the scattering integral-equation. To begin with, we assume in Sec. II, potentials of definite sign that make the scattering kernel square integrable in a symmetrized form. Regarding the kernel as a two sheeted function of $k^2 = s$, it is shown to be completely continuous at least in the first sheet. It may be continued into the finite second sheet, if the potential decreases faster than any exponential function.

In Sec. III the general spectral decomposition of the scattering kernel in the first s sheet is outlined. This is an easy task for negative real values of s, where the kernel is symmetric. For complex values of s, the kernel is not even normal, that is, permutable with its adjoint. The usual spectral theory therefore breaks down. Nevertheless, a generalized spectral decomposition is obtained for s arbitrary in the first sheet by analytic continuation from the negative real axis.

The method is generalized to potentials that change sign a finite number of times in Sec. IV. The kernel is now of "polar" type for negative real values of s and can be continued in a manner completely analogous to the symmetric kernel. In Sec. V the distribution of eigenvalues in the complex λ

* Work performed partially under the auspices of the U.S. Atomic Energy Commission.
† On leave of absence from Kernforschungszentrum Karlsruhe and Technische Hochschule Karlsruhe, Karlsruhe, Germany.

plane is investigated and related to the occurrence of bound states, resonances, and second sheet poles. Finally, we discuss in Sec. VI the convergence of Born expansions for resolvent quantities into powers of λ. All known results about this subject can easily be derived from our point of view. Moreover, it is proved that under the assumed conditions for the potential the Born series converges for every value of s in the physical sheet, if and only if neither $+\lambda V$ nor $-\lambda V$ have a bound state.

II. LINEAR TRANSFORMATIONS OF THE SPACE L^2 GENERATED BY SCATTERING KERNELS

During the course of this work we shall consider the nonrelativistic scattering of a spinless particle by a potential $-\lambda V(\mathbf{x})$. The "coupling constant" λ may, in general, be complex, but $V(\mathbf{x})$ shall be a real-valued function. Further conditions to be met by the potential will be specified according to the requirements of the analysis. The Schrödinger equation then reads

$$(H - E)\psi = 0, \tag{2.1}$$

where in dimensionless units ($\hbar = 2m = 1$)$H = H_0 - \lambda V = -\Delta - \lambda V$, and $E = k^2$ (k = wave number). Out- and ingoing scattering solutions of (2.1) are determined by the Lippmann-Schwinger equations:

$$\psi_{\mathbf{k}}^{(\pm)} = \varphi_{\mathbf{k}} - \lambda[1/(E - H_0 \pm i\epsilon)]V\psi_{\mathbf{k}}^{(\pm)}. \tag{2.2}$$

These may be written as integral equations in $\mathbf{x}$ space:

$$\psi_{\mathbf{k}}^{(\pm)}(\mathbf{x}) = \varphi_{\mathbf{k}}(\mathbf{x}) - \lambda \int d\tau(\mathbf{x}') \times \left\langle \mathbf{x} \left| \frac{1}{E - H_0 \pm i\epsilon} \right| \mathbf{x}' \right\rangle V(x')\psi_{\mathbf{k}}^{(\pm)}(\mathbf{x}'), \tag{2.3}$$

where

$$\left\langle \mathbf{x} \left| \frac{1}{E - H_0 \pm i\epsilon} \right| \mathbf{x}' \right\rangle = -\frac{e^{\pm ik|\mathbf{x}-\mathbf{x}'|}}{4\pi\,|\mathbf{x} - \mathbf{x}'|} = -G^{(\pm)}(\mathbf{x}, \mathbf{x}' \mid k) \tag{2.4}$$

is the negative Green's function of $(\Delta + k^2)$. If a plane wave is chosen for $\varphi_{\mathbf{k}}(\mathbf{x})$, the solution of (2.3) satisfies (2.1) for fixed energy $E = k^2$. Square summable wave packets may then be constructed by superposition. Actually, the fact that $\varphi_{\mathbf{k}}$ may be a plane wave has very little influence on our considerations.

We shall assume, for the time being, that $V(\mathbf{x}) \geq 0$. Then the potential is attractive for $\lambda > 0$. However, we shall be able to generalize our results to a finite number of sign changes of $V(\mathbf{x})$. This can be done, following Hilbert in his theory of polar integral equations[1] (Chap. XV) that are of type (2.3). The method is pointed out in more detail in Sec. IV.

The kernel in (2.3) can then be symmetrized by multiplication with $|V|^{1/2}$:

$$|V|^{1/2}\,\psi_{\mathbf{k}}^{(\pm)} = |V|^{1/2}\,\varphi_{\mathbf{k}} + \lambda \int d\tau(\mathbf{x}')K^{(\pm)}(\mathbf{x}, \mathbf{x}' \mid k)\,|V|^{1/2}\,\psi_{\mathbf{k}}^{(\pm)} \tag{2.5}$$

with

$$K^{(\pm)}(\mathbf{x}, \mathbf{x}' \mid k) = |V(\mathbf{x})|^{1/2} \frac{e^{\pm ik|\mathbf{x}-\mathbf{x}'|}}{4\pi\,|\mathbf{x} - \mathbf{x}'|}\,|V(\mathbf{x}')|^{1/2}. \tag{2.6}$$

The kernels $K^{(\pm)}$ are symmetric, but non-Hermitian for $k \neq 0$.

We now restrict ourselves to the class of potentials that make the kernels (2.6) square integrable, that is,

$$||K^{(\pm)}||^2 = \int d\tau(\mathbf{x}) \int d\tau\,\mathbf{x}')\,|K^{(\pm)}(\mathbf{x}, \mathbf{x}' \mid k)|^2 < \infty. \tag{2.7}$$

The space of these functions is usually denoted by $\mathbf{L}^2$.[2] If (2.7) is fulfilled for real values of k, then, of course, it is also true for all values of k in the upper half k plane Im $k \geq 0$. Introducing $s = k^2$ as a parameter we then have:

$$||K(s)|| < \infty \tag{2.8}$$

in the s-plane cut from 0 to ∞. $K(s)$ is a two sheeted function of s. The upper half k plane is mapped onto the first sheet by $k = (s)^{1/2}$, the lower half plane onto the second. The kernels $K^{(\pm)}$ are the boundary values of $K(s)$ at the real axis in the first sheet, usually called the "physical" sheet:

$$K^{(\pm)} = \lim_{\epsilon \to 0} K(s \pm i\epsilon). \tag{2.9}$$

The condition (2.7) is equivalent to

$$\frac{1}{(4\pi)^2} \int d\tau(\mathbf{x}) \int d\tau(\mathbf{x}') \frac{|V(\mathbf{x})|\,|V(\mathbf{x})|}{|\mathbf{x} - \mathbf{x}'|^2} < \infty. \tag{2.10}$$

This integral exists for potentials $V(\mathbf{x})$ that

(A)
(a) vanish faster than $(|\mathbf{x}|)^{-2}$ for $|x| \to \infty$;
(b) have only a finite number of singularities of type $(|x| - |x'|)^{-\alpha}$ with $\alpha < 2$.

If the potential is spherically symmetric and has a singularity only at the origin, conditions (a) and (b)

[1] D. Hilbert, *Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen*, (Leipzig, 1912).
[2] F. Riesz, and B. Sz.-Nagy, *Functional Analysis*, (Frederick Ungar Publishing Company, New York, 1955).

guarantee that:

$$\int_0^\infty r\,|V(r)|\,dr < \infty\,. \tag{2.11}$$

Hence, for potentials of class (A), $K(s)$ always belongs to $\mathbf{L}^2$. Furthermore, the integrals

$$\int d\tau(\mathbf{x}')\;|K(\mathbf{x},\mathbf{x}'\mid s)|^2 = \int d\tau(\mathbf{x}')\;|K(\mathbf{x}',\mathbf{x}\mid s)|^2$$

exist, except for a finite number of points, and are integrable functions.

Now, if f is an arbitrary element of the space $L^2(\mathbf{x}, \mathrm{V})$ that is[2a]

$$||f||^2 = \int_{V>\sigma} d\tau(\mathbf{x})\;|f(\mathbf{x})|^2 < \infty\,, \tag{2.12}$$

then Kf belongs also to L^2, because:

$$||Kf|| \leq ||K||\;||f||. \tag{2.13}$$

Thus, the kernel $K(s)$ generates a family of linear transformations of the space L^2 depending on the complex parameter s;

$$f = g + \lambda K(s)f, \qquad f, g \in L^2, \tag{2.14}$$

where λ is to be considered fixed. The adjoint transformation is generated by the complex conjugate $K^*(s)$. $K(s)$ has the following properties:

(a) It is completely continuous on the first sheet (see reference 2, chap. VI for a definition of complete continuity). In particular, it is a completely continuous symmetric transformation on the negative real axis $-\infty < s \leq 0$.

(b) The properties on the second sheet depend on the potential. If the potential decreases as $\exp(-2\alpha r)$, $(\alpha > 0)$ for $r \to \infty$, $K(s)$ is completely continuous in a region corresponding to the strip $0 \geq \mathrm{Im}\;k > -\alpha$; if the potential decreases faster than every exponential function, $K(s)$ is completely continuous in the finite second sheet, that is, except the point ∞.

Hence, we can apply the powerful theory of completely continuous transformations to the scattering kernel at least on the physical sheet.

The scattering solutions are obtained by application of the resolvent transformation of (2.5) on the element $|V|^{1/2}\,\varphi_{\mathbf{k}}$, if the latter belongs to L^2, that is,

$$||\;|V|^{1/2}\,\varphi_{\mathbf{k}}||^2 = \int d\tau(\mathbf{x})\;|V(\mathbf{x})| < \infty \qquad \text{(B)}$$

for a plane wave $\varphi_{\mathbf{k}}$. This means that for spherically symmetric potentials

$$\int_0^\infty r^2\,|V(r)|\,dr < \infty\,. \tag{2.15}$$

Conditions (A) plus (B) also guarantee the existence of the scattering amplitude $T_{\mathbf{k}\mathbf{k}'}$:

$$|T_{\mathbf{k}\mathbf{k}'}| = |(\varphi_{\mathbf{k}},\, V\psi_{\mathbf{k}'}^{(+)})| \leq ||\;|V|^{1/2}\,\varphi_{\mathbf{k}}||\;||\;|V|^{1/2}\,\psi_{\mathbf{k}'}^{(+)}||. \tag{2.16}$$

III. SPECTRUM OF K(s) IN THE PHYSICAL SHEET

It is a well-known property of a completely continuous transformation K that its resolvent $(1 - \lambda K)^{-1}$ is a meromorphic function of the complex parameter λ, that is to say, its singular values λ_n (the "eigenvalues") form either a finite sequence that may be void, or an infinite sequence with accumulation point ∞ (see e.g., reference 2, chap. IV.). If the transformation is also Hermitian, the Hilbert-Schmidt theory gives the usual spectral decomposition.

Let us, therefore, assume that s is some negative real value $s = -\sigma < 0$. Because the kernel (2.6) is not of finite rank, we have an infinite sequence of positive eigenvalues $\lambda_n > 0$ (the kernel is positive definite) with $\lambda_n \to \infty$ for $n \to \infty$, each of finite multiplicity. The eigenvalue λ_n determines the necessary coupling strength for the formation of a bound state with energy $E = -\sigma < 0$ by the attractive potential $-\lambda_n V (V \geq 0)$. The degeneracy of such a bound state is always finite. The corresponding eigenelements satisfy

$$\phi_n(-\sigma) = \lambda_n(-\sigma)K(-\sigma)\phi_n(-\sigma), \tag{3.1}$$

where we have displayed only the dependence of all quantities on s. The kernel $K(-\sigma)$ can be expanded into the series

$$K(-\sigma) = \sum_n \frac{\phi_n(-\sigma) \otimes \phi_n(-\sigma)}{\lambda_n(-\sigma)} \tag{3.2}$$

convergent in the mean. The ϕ_n are orthogonal and belong to L^2. So we may assume them to be orthonormal: $(\phi_n, \phi_m) = \delta_{nm}$. Introducing

$$\phi_n = |V|^{1/2}\,u_n\,,\ (u_n,\, Vu_m) = \delta_{nm}, \tag{3.3}$$

it follows that $u_n \in L^2_\infty(-\sigma < 0!)$ and the series for the Green's kernel (2.4)

$$G(-\sigma) = \sum_n \frac{u_n(-\sigma) \otimes u_n(-\sigma)}{\lambda_n(-\sigma)} \tag{3.4}$$

also converges in the mean. This is due to the singular character of G for $\mathbf{x} = \mathbf{x}'$. Otherwise (3.4) would converge uniformly and absolutely. The real

[2a] The space $L^2(\mathbf{x}, \mathrm{V})$ (in the following called L^2) consists of the functions square integrable over the region where $\mathbf{V} > \sigma$.

functions u_n obey the Schrödinger equations

$$\Delta u_n - \sigma u_n + \lambda_n V u_n = 0, \tag{3.5}$$

and are bound-state wave functions for $E = -\sigma$ at coupling λ_n.

The series for the iterated kernels,

$$K^{(l)}(-\sigma) = \sum_n \frac{\phi_n(-\sigma) \otimes \phi_n(-\sigma)}{\lambda_n^l(-\sigma)};$$

$$G^{(l)}(-\sigma) = \sum_n \frac{u_n(-\sigma) \otimes u_n(-\sigma)}{\lambda_n^l(-\sigma)}$$

$$l = 2, 3 \cdots, \tag{3.6}$$

converge uniformly and absolutely according to Mercer's theorem, the first converges at all but the finite number of points where V is singular, the second converges everywhere. Hence, the same is true for the resolvent kernels

$$K_\lambda(-\sigma) = K(-\sigma) + \lambda \sum_n \frac{\phi_n(-\sigma) \otimes \phi_n(-\sigma)}{\lambda_n(-\sigma)(\lambda_n(-\sigma) - \lambda)}, \tag{3.7}$$

$$G_\lambda(-\sigma) = G(-\sigma) + \lambda \sum_n \frac{u_n(-\sigma) \otimes u_n(-\sigma)}{\lambda_n(-\sigma)(\lambda_n(-\sigma) - \lambda)},$$

where

$$[1 - \lambda K(-\sigma)]^{-1} = 1 + \lambda K_\lambda(-\sigma);$$

$$K_\lambda(-\sigma) = |V|^{1/2} G_\lambda(-\sigma) |V|^{1/2}. \tag{3.8}$$

If h is an arbitrary element of L^2, then Kh and $K_\lambda h$ can be expanded into the series

$$Kh = \sum_n \phi_n(-\sigma) \frac{(\phi_n(-\sigma), h)}{\lambda_n(-\sigma)};$$

$$K_\lambda h = \sum_n \phi_n(-\sigma) \frac{(\phi_n(-\sigma), h)}{\lambda_n(-\sigma) - \lambda}, \tag{3.9}$$

convergent absolutely and uniformly at all but the singular points of V, and, of course, $\lambda \neq \lambda_n$ for the second. Because elements of form Kh are dense in L^2, the system $[\phi_n(-\sigma)]$ is complete. Hence, in $\mathbf{L}^2$,

$$||K(-\sigma)||^2 = \sum_n \frac{1}{\lambda_n^2(-\sigma)} \tag{3.10}$$

Everything that has been said is valid for an arbitrary value of $-\sigma < 0$. Moreover, eigenvalues and eigenfunctions ϕ_n are continuous functions of σ, as follows from the structure of the kernel (2.6). The eigenvalues $\lambda_n(-\sigma)$ increase monotonically to ∞ for $\sigma \to \infty$. The limit $\sigma \to 0$ exists, because $K(0)$ still belongs to $\mathbf{L}^2$, and gives absolute minimum values for λ_n on the negative real axis. Hence, the minimum value of all λ_n in $0 \geq s > -\infty$ is $\lambda_1(0)$.

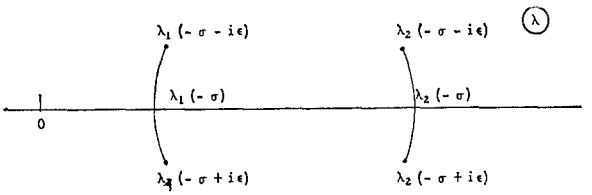

FIG. 1. Analytic continuation of eigenvalues from the negative s axis $s = -\sigma \leq 0$.

Next we proceed from some point on the negative real axis $s = -\sigma < 0$ to points $s = -\sigma \pm i\epsilon$ ($\epsilon > 0$) in the upper and lower half s plane, respectively (see Fig. 1). In doing so, we must be very careful, because the kernel is no longer Hermitian. It is not even normal, that is, permutable with its adjoint. So the system of eigenelements can no longer be orthonormal. As is well known, the normal transformations are the most general ones that can be spectral decomposed. However, the kernel $K(s)$ has the special property

$$KK^\dagger = (K^\dagger K)^*, \tag{3.11}$$

which might be called "complex normality". In our case, it allows a slight generalization of a spectral decomposition that can be obtained by analytic continuation from the negative real axis.

It is clear that, if ϕ_n is an eigenelement of K with eigenvalue λ_n, then ϕ_n^* is an eigenelement of K^* with eigenvalue λ_n^*, because K and K^* are symmetric:

$$\phi_n = \lambda_n K \phi_n, \qquad \phi_n^* = \lambda_n^* K^* \phi_n^*. \tag{3.12}$$

If ϕ_n and ϕ_m^* are eigenelements for λ_n and λ_m^*, where $\lambda_n \neq \lambda_m$, they are orthogonal. In fact,

$$(\phi_n, \phi_m^*) = \lambda_n(K\phi_n, \phi_m^*),$$

and

$$(\phi_n, \phi_m^*) = \lambda_m(\phi_n, K^*\phi_m^*) = \lambda_m(K\phi_n, \phi_m^*).$$

Hence, for $\lambda_n \neq \lambda_m$:

$$(\phi_n, \phi_m^*) = 0. \tag{3.13}$$

[We write $(f, g) = \int d\tau f g^*$.] Furthermore, it is evident that

$$\lambda_n(-\sigma - i\epsilon) = \lambda_n^*(-\sigma + i\epsilon),$$
$$\phi_n(-\sigma - i\epsilon) = \phi_n^*(-\sigma + i\epsilon). \tag{3.14}$$

We may then construct a biorthonormal system in L^2 with

$$(\phi_n, \phi_m^*) = \delta_{nm}. \tag{3.15}$$

It coincides on the negative real axis with the orthonormal system $\phi_n(-\sigma)$. To be sure, (3.15) means

$$\int d\tau\phi_n\phi_m = \int d\tau\phi_n^*\phi_m^* = \delta_{nm} \tag{3.16}$$

and is not an orthonormal system in complex L^2 space. Hence, the analytic continuation of the orthonormal system in $s = -\sigma$ is a biorthonormal system in $s = -\sigma \pm i\epsilon$. The reason is that the analytic continuation of $||\phi_n(-\sigma)||^2$ is, of course, not $||\phi_n(-\sigma \pm i\epsilon||^2 = (\phi_n(-\sigma \pm i\epsilon), \phi_n(-\sigma \mp i\epsilon))$, but $(\phi_n(-\sigma \pm i\epsilon), \phi_n^*(-\sigma \pm i\epsilon))$. Writing $\phi_n(-\sigma \pm i\epsilon) = u_n \pm iv_n$, we have from $(\phi_n, \phi_n^*) = 1$: $(u_n, v_n) = 0$ and $||u_n||^2 - ||v_n||^2 = 1$. Hence,

$$(\phi_n, \phi_n) = ||u_n||^2 + ||v_n||^2 = 1 + 2\,||v_n||^2 \geq 1. \tag{3.17}$$

The equality sign holds only on the negative real s axis, where $||v_n|| = 0$.

So we may continue the expansion (3.2) into the complex s plane. The point is that we have to replace the tensorial product in complex L^2 space $\phi_n \otimes \phi_n$ by a simpler one that does not involve the complex conjugate. We shall denote this by a dot $\phi_n \cdot \phi_n$:

$$\phi_n \otimes \phi_n = \phi_n(\mathbf{x})\phi_n^*(\mathbf{y}); \quad \phi_n \cdot \phi_n = \phi_n(\mathbf{x})\phi_n(\mathbf{y}). \tag{3.18}$$

The series (3.2), of course, still converges in the same sense as before on the negative real axis. So we have

$$K(-\sigma + i\epsilon) = \sum_n \frac{\phi_n(-\sigma + i\epsilon)\cdot\phi_n(-\sigma + i\epsilon)}{\lambda_n(-\sigma + i\epsilon)},$$

$$K(-\sigma - i\epsilon) = K^*(-\sigma + i\epsilon)$$

$$= \sum_n \frac{\phi_n^*(-\sigma + i\epsilon)\cdot\phi_n^*(-\sigma + i\epsilon)}{\lambda_n^*(-\sigma + i\epsilon)}. \tag{3.19}$$

This result can also be well understood in terms of the usual theory of integral equations or linear transformations of L^2. One may then argue as follows: To each eigenvalue of a completely continuous kernel there exists a canonical decomposition $K = R + S$, where S is of finite rank and $RS = SR$ (see reference 2, chap. IV.). In our case we have:

$$S_n = \phi_n\cdot\phi_n/\lambda_n, \qquad s = -\sigma + i\epsilon, \tag{3.20}$$

because $(\phi_n, \phi_n^*) \neq 0$; that is, the eigenelements of K and its adjoint $K^+ = K^*$ are not orthogonal. Summing all contributions we obtain

$$K = R_0 + \sum_n S_n, \tag{3.21}$$

where R_0 is orthogonal on all S_n. The series certainly converges, because of the Riesz-Fischer theorem. Then R_0 must also be completely continuous and analytic in the cut s plane. But on the negative real s axis R_0 must vanish. Hence, it is zero everywhere in the s plane and we have

$$K = \sum_n S_n. \tag{3.22}$$

The series (3.19) has in fact the property of being complex normal [see (3.11)]. We have

$$KK^* = \sum_{n,m} \frac{\phi_n\cdot\phi_m^*}{\lambda_n\lambda_m^*}(\phi_n, \phi_m)$$

$$= \left[\sum_{n,m} \frac{\phi_n^*\cdot\phi_m}{\lambda_n^*\lambda_m}(\phi_m, \phi_n)\right]^* = (K^*K)^*. \tag{3.23}$$

Furthermore, we have in L^2,

$$(K, K) = \sum_{n,m} \frac{(\phi_n, \phi_m)^2}{\lambda_n\lambda_m^*} = ||K||^2,$$

and

$$\sum_n 1/|\lambda_n|^2 \leq ||K||^2, \qquad s = -\sigma \pm i\epsilon. \tag{3.24}$$

As is well known, (3.24) holds in general for non-normal $\mathbf{L}^2$ kernels[3] (chap. VIII.). The equality sign holds, if and only if, K is normal, in our case on the negative real s axis.

Because

$$||K(-\sigma \pm i\epsilon)|| = ||K(-\sigma)||, \tag{3.25}$$

it follows from (3.24) and (3.10) that

$$\sum_n 1/|\lambda_n(-\sigma \pm i\epsilon)|^2 < \sum_n 1/\lambda_n^2(-\sigma). \tag{3.26}$$

That, of course, does not mean

$$1/|\lambda_n(-\sigma \pm i\epsilon)|^2 < 1/\lambda_n^2(-\sigma), \quad \text{all } n. \tag{3.27}$$

However, (3.27) is true for the lowest eigenvalue λ_1. In fact:

$$\frac{(\phi_1, \phi_1)}{|\lambda_1(-\sigma \pm i\epsilon)|}$$

$$= |(K(-\sigma \pm i\epsilon)\phi_1(-\sigma \pm i\epsilon), \phi_1(-\sigma \pm i\epsilon))|$$

$$\leq (K(-\sigma)\,|\phi_1|, |\phi_1|) < \frac{(\phi_1, \phi_1)}{\lambda_1(-\sigma)}, \tag{3.28}$$

because $1/\lambda_1(-\sigma)$ is the maximum of the quadratic form $(K(-\sigma)\varphi, \varphi)$ under the subsidiary condition $(\varphi, \varphi) = 1$, and $(\phi_1, \phi_1) > 1$ from (3.17). The maximum is taken only for $\varphi = \phi_1(-\sigma)$. A similar conclusion cannot be drawn for the higher eigenvalues $n \geq 2$, because $1/\lambda_n(-\sigma)(n \geq 2)$ is the

[3] F. Smithies, *Integral Equations*, (Cambridge University Press, New York, 1958).

minimum of the maxima that $(K(-\sigma)\varphi, \varphi)$ can have under the subsidiary conditions $(\varphi, \varphi) = 1$ and $(\varphi, v_l) = 0$ $(l = 1 \cdots n-1)$ for arbitrary v_l. We have indicated in Fig. 1 that the lowest eigenvalue λ_1 must increase in magnitude, if s goes off the negative real axis, but others may decrease.

In particular, we have for the boundary values on the cut $\lambda_1(s \pm i\epsilon)$ $(0 < s < \infty)$:

$$(\phi_1, \phi_1)/|\lambda_1(s \pm i\epsilon)| = |(K(s \pm i\epsilon)\phi_1, \phi_1)|$$
$$\leq (K(0)\,|\phi_1|, |\phi_1|) < (\phi_1, \phi_1)/\lambda_1(0), \qquad (3.28')$$

and in an analogous manner:

$$\frac{(\phi_n, \phi_n)}{|\lambda_n(s \pm i\epsilon)|} < \frac{(\phi_n, \phi_n)}{\lambda_1(0)} \cdot \begin{matrix}\text{all } n \\ 0 < s < \infty\end{matrix} \qquad (3.29)$$

These results will play an important role, when we discuss the convergence of the Born series in Sec. VI.

Next we want to determine the sign of Im $\lambda_n(-\sigma \pm i\epsilon)$. If s is an arbitrary point in the cut s plane, the functions $u_n(s)$, $u_n^*(s) = u_n(s^*)$ [(3.3)] are L^2_∞ and obey the Schrödinger equations

$$\begin{aligned} \Delta u_n + s u_n + \lambda_n V u_n &= 0, \\ \Delta u_n^* + s^* u_n^* + \lambda_n^* V u_n^* &= 0. \end{aligned} \qquad (3.30)$$

Hence

$$\text{Im } s \cdot (u_n, u_n) = -\text{Im } \lambda_n (u_n, Vu_n). \qquad (3.31)$$

Of course, (3.31) does not hold for the boundary values on the cut. Observing the asymptotic behavior of $u_n(s \pm i\epsilon)$ $(0 < s < \infty)$,

$$u_n(s \pm i\epsilon) \to \frac{e^{\mp i r\sqrt{s}}}{r} f_n \qquad (3.32)$$

(f_n = scattering amplitude), we obtain the familar result:

$$(s)^{1/2} \int d\Omega\, |f_n|^2 = -\text{Im } \lambda_n (u_n, Vu_n)$$
$$0 \leq s < \infty. \qquad (3.33)$$

Im λ_n has, therefore, always the opposite sign as Im s (Fig. 1), and vanishes only for $-\infty < s \leq 0$.

It is clear from what has been said that the series (3.6), (3.7), and (3.9) still converge in $s = -\sigma \pm i\epsilon$ in the same sense as in $s = -\sigma$. That means, in particular, our biorthonormal system $\{\phi_n, \phi_n^*\}$ is complete. In fact, let h be an element of L^2. Then $K(s)h$ can be expanded in $s = -\sigma$, according to (3.9). This may be continued to $s = -\sigma \pm i\epsilon$ to give a representation in terms of the elements $\{\phi_n\}$ or $\{\phi_n^*\}$. Because elements of form $K(-\sigma)h$ are dense in L^2, the biorthogonal system $\{\phi_n, \phi_n^*\}$ is complete.

For an arbitrary biorthonormal system of L^2 $\{\phi_n, \psi_n\}$ the completeness condition reads,

$$||g||^2 = \sum_n a_n b_n^* = \sum_n a_n^* b_n, \qquad (3.34)$$

where

$$a_n = (g, \phi_n); \qquad b_n = (g, \psi_n); \qquad g \in L^2.$$

In our case we have $\psi_n = \phi_n^*$. But this means neither $b_n = a_n^*$ nor $b_n = a_n$. For instance, we have in $\mathbf{L}^2$ (3.24):

$$||K||^2 = \sum_{n,m} \frac{(\phi_n, \phi_m)^2}{\lambda_n \lambda_m^*};$$
$$a_{nm} = (K, \phi_n \cdot \phi_m) = \sum_l \frac{(\phi_l, \phi_n)(\phi_l, \phi_m)}{\lambda_l} \qquad (3.35)$$
$$b_{nm} = (K, \phi_n^* \cdot \phi_m^*) = (1/\lambda_n)\, \delta_{nm}$$

(Scalar products are taken in $\mathbf{L}^2$, where $\phi_n \cdot \phi_m$, $\phi_n^* \cdot \phi_m^*$, is a complete biorthonormal system.)

We may then analytically continue the eigenvalues $\lambda_n(s)$ and eigenelements $\phi_n(s)$ into the entire s plane cut from 0 to ∞ along the positive real axis, where $k = (s)^{1/2}$ changes its sign. The system $\{\phi_n(s), \phi_n^*(s)\}$ is always complete in L^2 and the expansions (3.2), (3.4), (3.7), and (3.9) are valid in the entire cut s plane. In particular, we have the representations for the resolvent kernel $K_\lambda(s)$ and the complete Green's function $G_\lambda(s)$:

$$K_\lambda(s) = K(s) + \lambda \sum_n \frac{\phi_n(s) \cdot \phi_n(s)}{\lambda_n(s)(\lambda_n(s) - \lambda)}$$
$$G_\lambda(s) = G(s) + \lambda \sum_n \frac{u_n(s) \cdot u_n(s)}{\lambda_n(s)(\lambda_n(s) - \lambda)}. \qquad (3.36)$$

These expressions are equivalent to the Fredholm series that have been used by several authors,[4,5] but are much easier to handle than the unwieldy determinants. The Fredholm theory can, in fact, be applied to every $\mathbf{L}^2$ kernel.[6] We have shown that the "complex normal" kernel $K(s)$ allows a slight generalization of a spectral decomposition. It is determined by its eigenvalues and eigenelements that form a complete biorthonormal system with the eigenelements of the adjoint kernel K^*. It would be an interesting mathematical question, whether this holds in general for the class of transformations, characterized by (3.11).

Next, we want to generalize the results of this section to the case of a sign changing potential.

[4] R. Jost and A. Pais, Phys. Rev. **82**, 840 (1951).
[5] N. N. Khuri, Phys. Rev. **107**, 1148 (1957).
[6] F. Smithies, Duke Math. J. **8**, 107 (1941).

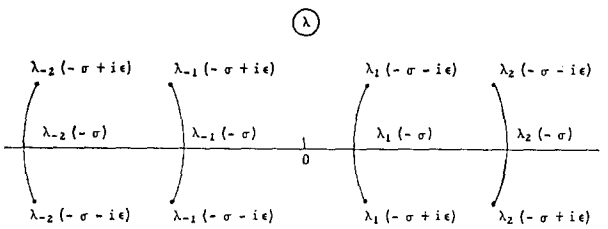

FIG. 2. Analytic continuation of polar eigenvalues from the negative real s axis $s = -\sigma \leq 0$.

IV. POLAR KERNELS

If the potential changes its sign a finite number of times, we cannot simply symmetrize the scattering kernel by multiplication with $|V|^{1/2}$ as has been done in Sec. II. However, after multiplication we arrive at a kernel, as simple on the negative real axis $s \leq 0$ as the Hilbert-Schmidt one. It has been studied by Hilbert (reference 1, chap. XV) and Garbe,[7] and is called "polar." Instead of (2.6) we get

$$K'(\mathbf{x}, \mathbf{x}' \mid k) = |V(\mathbf{x})|^{1/2} \frac{e^{ik|\mathbf{x}-\mathbf{x}'|}}{4\pi\,|\mathbf{x}-\mathbf{x}'|} |V(\mathbf{x}')|^{1/2} v(\mathbf{x}') = Kv, \quad (4.1)$$

where $v(\mathbf{x}') = \operatorname{sgn} V(\mathbf{x}')$ is the sign of V at a point $\mathbf{x}'$ and K is our previous kernel (2.6). On the negative real s axis in the physical sheet (4.1) differs from a Hilbert-Schmidt kernel by the sign function $v(\mathbf{x}')$. It is this property that is called "polarity."

If the polar kernel is of the type sign function times a definite closed Hermitian kernel as is (4.1) in $s \leq 0$, then, according to Hilbert, there is an infinite number of positive eigenvalues λ_n and negative eigenvalues λ_{-n} $(n = 1, 2, \cdots)$ with accumulation at ∞. The corresponding eigenelements $\phi_n = |V|^{1/2} u_n$ form a "polar" system that is

$$(\phi_n, v\phi_m) = \delta_{nm} \operatorname{sgn} \lambda_n; \quad n = \pm 1, \pm 2 \cdots, \quad (4.2)$$

where sgn λ_n means the sign of the eigenvalue λ_n. It can easily be seen that the spectrum of a polar kernel is of this kind. (4.1) may be considered as the sum of two kernels, the first being defined only in those domains, where $V > 0$, the second where $V < 0$. Hence, the first is positive definite, the second negative definite and the spectrum of the sum must be of the type asserted.

Everything that has been said in Sec. III about the properties of $K(-\sigma)(-\sigma < 0)$ is true also for polar kernels, if we understand all expansions in terms of polar eigenelements. We have, for instance,

$$G(-\sigma) = \sum_{\neq n} \frac{u_n(-\sigma) \otimes u_n(-\sigma)}{\lambda_n} \operatorname{sgn} \lambda_n;$$

$$K_\lambda(-\sigma) = K(-\sigma) + \lambda \sum_{\neq n} \frac{\phi_n(-\sigma) \otimes \phi_n(-\sigma)}{\lambda_n(-\sigma)(\lambda_n(-\sigma) - \lambda)} \operatorname{sgn} \lambda_n. \quad (4.3)$$

(Note that the kernels do not contain the sign function.) If $h \in L^2$, then Kvh can be expanded into the uniformly and absolutely convergent series[7]

$$f = Kvh = \sum_{\neq n} (f, v\phi_n)\phi_n \operatorname{sgn} \lambda_n. \quad (4.4)$$

Elements of form Kvh are dense in L^2, and we have the completeness condition

$$(f, vf) = \sum_{\neq n} (f, v\phi_n)^2 \operatorname{sgn} \lambda_n. \quad (4.5)$$

It is now obvious how the results of Sec. III can be transferred to the polar case. Analytic continuation off the negative real axis works in the same way. We now obtain a "bipolar" system in L^2.

The sign of Im λ_n can be read off from (3.31), (3.32) recalling

$$(u_n, Vu_n) = (\phi_n, v\phi_n). \quad (4.6)$$

The latter quantity is a continuous function of s and cannot vanish in the cut s plane. Hence,

$$\operatorname{sgn} (\phi_n(s), v\phi_n(s)) = \operatorname{sgn} (\phi_n(0), v\phi_n(0)) = \operatorname{sgn} \lambda_n(0). \quad (4.7)$$

It follows from (3.31), (3.32) that positive and negative eigenvalues of a polar kernel are moved in opposite directions, if s becomes complex (see Fig. 2).

In place of (3.10) we have:

$$||K(-\sigma)v||^2 = \sum_{\neq n} 1/\lambda_n^2(-\sigma) \leq ||K(-\sigma)||^2 = \sum_n 1/[\lambda_n^1(-\sigma)]^2 \quad (4.8)$$

The equality sign holds only for definite v.

Bounds for the eigenvalues in the complex λ plane can be derived in a similar fashion as in Sec. III. Consider $\lambda_1(s \pm i\epsilon)$ $(0 < s < \infty)$. We have

$$(\phi_1, v\phi_1)/|\lambda_1(s \pm i\epsilon)| = |(K(s \pm i\epsilon)v\phi_1, v\phi_1)| \leq (K(0)\,|\phi_1|, |\phi_1|) < (\phi_1, v\phi_1)/\lambda_1(0), \quad (4.9)$$

because $1/\lambda_1(0)$ is the maximum of $(K(0)\varphi, \varphi)$ under $(\varphi, v\varphi) = 1$. Moreover,

$$\frac{1}{|\lambda_n(s \pm i\epsilon)|} < \frac{1}{\lambda_1(0)}; \quad \begin{matrix} n = 1, 2 \cdots \\ 0 < s < \infty. \end{matrix} \quad (4.10)$$

The same is true for $\lambda_{-n}(s \pm i\epsilon)$. $-1/\lambda_{-1}(0)$ is the maximum of the positive definite form, $(K(0)\varphi, \varphi)$

[7] E. Garbe, Math. Ann., 76, 527 (1915).

under $(\varphi, v\varphi) = -1$;

$$\frac{1}{|\lambda_{-n}(s \pm i\epsilon)|} < -\frac{1}{\lambda_{-1}(0)}; \quad n = 1, 2 \cdots \quad 0 < s < \infty. \tag{4.11}$$

Corresponding results hold for $s = -\sigma \pm i\epsilon$. For the proof of (4.10) and (4.11) it is essential to know that $(\phi_n(s), v\phi_n(s))$ does not change its sign.

V. BOUND STATES AND RESONANCES

Eigenelements and eigenvalues of the scattering kernel $K(s)$ are analytic functions of s in the cut physical sheet (upper half k plane). Because $|\lambda_n(s)| \to \infty$ $(s \to \infty)$, $1/\lambda_n(s)$ can be written as a dispersion integral

$$\frac{1}{\lambda_n(s)} = -\int_0^\infty \frac{ds'}{\pi} \frac{\operatorname{Im} \lambda_n(s)}{|\lambda_n(s)|^2 (s' - s)}; \qquad n = \pm 1, \pm 2 \cdots. \tag{5.1}$$

The function $\lambda_n(s)$ maps the physical sheet onto a domain in the λ plane bounded by the image of the cut. The mapping is conform. This can be seen from the differential equations for u_n and $\partial u_n/\partial s$:

$$\begin{aligned} \Delta u_n + s u_n + \lambda_n V u_n &= 0, \\ \Delta \frac{\partial u_n}{\partial s} + s \frac{\partial u_n}{\partial s} + \lambda_n V \frac{\partial u_n}{\partial s} &= -u_n - \frac{d\lambda_n}{ds} V u_n, \end{aligned} \tag{5.2}$$

u_n and $\partial u_n/\partial s$ are L^2_∞ in the cut s plane. Hence,

$$\frac{d\lambda_n}{ds} = -\frac{(u_n, u_n^*)}{(u_n, Vu_n^*)} = -(u_n, u_n^*) \operatorname{sgn} \lambda_n(0), \tag{5.3}$$

where we have observed the normalization (4.2). (u_n, u_n^*) cannot vanish in the cut plane, because $\{u_n\}$, $\{u_n^*\}$ are complete (not orthonormal) systems in L^2_∞.

The image of the s plane cut under $\lambda_n(s)$ is a curve C_n in the λ plane that is mirrorsymmetric with respect to the real λ axis and intersects the latter in a right angle at $\lambda_n(0)$, as can be seen from (3.33). If $\lambda_n(0)$ is a positive eigenvalue, the image of the upper boundary lies in Im $\lambda < 0$ and vice versa for a negative eigenvalue $\lambda_{-n}(0)$; for $s \to \infty$ C_n approaches $\lambda = \infty$. The cut s plane is the interior of a path running around the cut in a positive sense. It is mapped onto the interior of C_n. The image of the negative real s axis is

$$\begin{aligned} \lambda_n(0) \le \lambda < \infty; &\qquad \lambda_n(0) > 0 \\ -\infty < \lambda \le \lambda_{-n}(0); &\qquad \lambda_{-n}(0) < 0. \end{aligned}$$

A simple example is the scattering of S waves by a square well potential of range one. The eigenvalues λ_n are roots of the equation

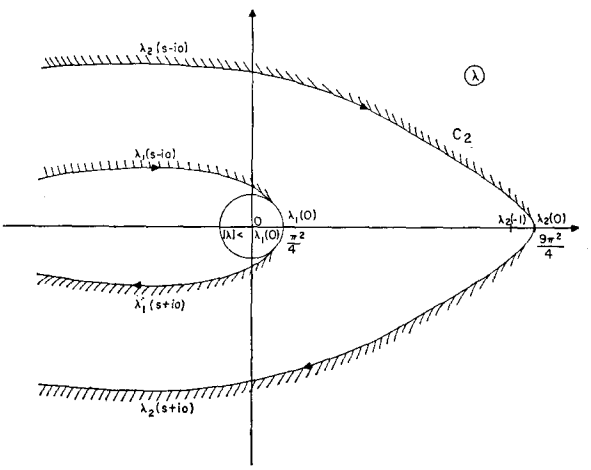

FIG. 3. Boundary values of $\lambda_{1,2}$ in the physical sheet for S-wave scattering by a square-well potential:
$V = \begin{cases} -1 & r < 1 \\ 0 & r > 1. \end{cases}$

$$(s + \lambda)^{1/2} \cot (s + \lambda)^{1/2} = i(s)^{1/2}. \tag{5.4}$$

In particular,

$$\lambda_n(0) = (2n + 1)^2 \frac{\pi^2}{4}; \qquad n = 0, 1, 2 \cdots. \tag{5.5}$$

The potential is repulsive for $\lambda < 0$ and attractive for $\lambda > 0$, according to the definition of the sign in (2.1). The curves C_1, C_2 of the lowest eigenvalues are shown in Fig. 3.

A positive or negative value of λ lies in the interior (left hand of the oriented curve) of a finite number of curves C_n or C_{-n}, because the eigenvalues $\lambda_{\pm n}(0)$ of the completely continuous kernel $K(s)$ accumulate at $\lambda = \infty$. We then have a finite number of bound states with binding energy $E_n = -\sigma_n$, determined by

$$\lambda_n(-\sigma_n) = \lambda \qquad (\lambda > 0) \tag{5.6}$$

or

$$\lambda_{-n}(-\sigma_n) = -\lambda.$$

This result is well known from the work of Bargmann[8] and, more recently, Schwinger,[9] who extended the proof to the class A of potentials.

Resonances occur at energies $s_n > 0$, if,

$$\operatorname{Re} \lambda_{\pm n}(s_n) = \lambda, \tag{5.7}$$

and Im $\lambda_{\pm n}(s_n)$ is small. To illuminate this point, we consider the total scattering cross section as a function of λ in the vicinity of a pole λ_n. The scattering amplitude is given by $(\hbar = 2m = 1)$:

[8] V. Bargmann, Proc. Natl. Acad. Sci. U.S., **38**, 961 (1952).
[9] J. Schwinger, Proc. Natl. Acad. Sci. U.S., **47**, 122 (1961).

$$f(\mathbf{k}', \mathbf{k}) = \frac{\lambda}{4\pi}(\psi_{\mathbf{k}}^{(+)}, V\varphi_{\mathbf{k}'}) = \frac{\lambda}{4\pi}(\varphi_{\mathbf{k}}, V\varphi_{\mathbf{k}'}) + \frac{\lambda^2}{4\pi}\sum_{*n}\frac{(u_n, V\varphi_{\mathbf{k}}^*)(u_n, V\varphi_{\mathbf{k}'})}{\lambda_n - \lambda}\operatorname{sgn}\lambda_n(0). \quad (5.8)$$

[Note: $(f, g) = \int d\tau f g^*$.] All quantities are to be taken at $s + i0$. Introducing the scattering amplitude f_n of $u_n(s + i0)$,

$$u_n \to [e^{ir\varkappa 1/2}/r]f_n(\Omega)\ (r \to \infty);$$

$$f_n(\Omega) = (\lambda_n/4\pi)(u_n, V\varphi_{\mathbf{k}}); \qquad \begin{matrix}\mathbf{k} = (s)^{1/2}\Omega \\ \mathbf{r} = r\Omega,\end{matrix} \quad (5.9)$$

we obtain in the neighbourhood of the pole λ_n:

$$f(\mathbf{k}', \mathbf{k}) \approx 4\pi f_n(-\Omega)f_n(\Omega')/(\lambda_n - \lambda); \qquad \mathbf{k}' = \Omega'\cdot(s)^{1/2}. \quad (5.10)$$

Assuming a spherically symmetric potential λ_n must correspond to a definite angular momentum l and $f_n(\Omega) \sim P_l(\Omega)$. Hence the total cross section is given by

$$\sigma_{\text{tot}} \approx (4\pi)(2l + 1)\int d\Omega \int d\Omega' \frac{|f_n(\Omega)|^2\,|f_n(\Omega')|^2}{|\lambda_n - \lambda|^2}. \quad (5.11)$$

With (3.33) and (5.7) this becomes a Breit-Wigner formula,

$$\sigma_{\text{tot}} \approx (2l + 1)\frac{\pi}{s}\frac{4(\operatorname{Im}\lambda_n)^2}{(\operatorname{Re}\lambda_n - \lambda)^2 + (\operatorname{Im}\lambda_n)^2}(u_n, Vu_n)^2$$

$$= (2l + 1)\frac{\pi}{s}\frac{\Gamma^2}{(s - s_n)^2 + (\Gamma/2)^2}(u_n, Vu_n)^2. \quad (5.12)$$

The resonance width Γ is

$$\frac{\Gamma}{2} = \left|\frac{\operatorname{Im}\lambda_n}{d\operatorname{Re}\lambda_n/ds}\right|_{s=s_n}. \quad (5.13)$$

The description of resonances presented here seems to be similar to a method developed by Rollnik,[10] where resonances are characterized by the condition

$$\lambda_n'(s_n) = \lambda.$$

$\lambda_n'(s)$ are the eigenvalues of the real part of the kernel $K(s)$ for a definite potential. But these are not the same as the real parts $\operatorname{Re}\lambda_n$ of the eigenvalues of $K(s)$, because $K(s)$ is not normal.

Is the occurrence of resonances related to the properties of $\lambda_n(s)$ on the second sheet? Consider a case, where a continuation into the finite second sheet is possible. We can no longer infer from (5.3) that the mapping of the second sheet onto the λ plane by $\lambda_n(s)$ is conform, because $u_n(s)$ is not L_∞^2 in contrast to $\phi_n(s)$. λ_n is real on the negative real axis in the second sheet. It follows from the maximum property of $1/\lambda_1(-1/\lambda_{-1})$ that

$$\lim_{\sigma\to\infty}\lambda_{*1}(-\sigma) = 0; \quad (5.14)$$

But, this cannot hold for all eigenvalues, because they must accumulate at ∞ for any finite σ.

Take our example: S-wave scattering by a square well. It may be seen from (5.4) that λ_1 obeys (5.14). But λ_2 arrives at a minimum value at $\sigma = 1$ and tends to ∞ for $\sigma \to \infty$. So do all higher eigenvalues. Hence,

$$d\lambda_n/ds\,|_{s\text{II}=-1} = 0; \qquad n = 2, 3\cdots \quad (5.15)$$

(s_{II} means second sheet values). $\lambda_2(s_{\text{I}})$, $\lambda_2(s_{\text{II}})$ map the negative real axis onto $\lambda_2(0) \le \lambda < \infty$ and $\lambda_2(-1) \le \lambda \le \lambda_2(0)$; $\lambda_2(-1) \le \lambda < \infty$, respectively. (See Fig. 3.) λ values below $\lambda_n(-1)$ are accepted at two complex conjugate points in the second sheet as has been shown by Nussenzweig.[11] This is true also for negative λ values (repulsive potential). In summary, $\lambda_n(s)$ is responsible for a

(a) bound state, if $\lambda > \lambda_n(0)$; $n = 2, 3, \cdots$

(b) virtual state, if $\lambda_n(0) > \lambda \ge \lambda_n(-1)$;

(c) resonance, if $\lambda_n(-1) > \lambda > -\infty$.

VI. ANALYTICITY IN s AND λ

We have proved that generalized spectral representations exist for the resolvent kernels $G_\lambda(s)$, $K_\lambda(s)$ in the case of an indefinite class A potential;

$$K_\lambda(s) = \sum_{*n}\frac{\phi_n(s)\cdot\phi_n(s)}{\lambda_n(s) - \lambda}\operatorname{sgn}\lambda_n(0) \quad (6.1)$$

and similarly for $G_\lambda(s)$ in terms of $u_n(s)$. We may infer from (6.1) and the extremal properties of the eigenvalues the following statements on analyticity in s and λ:

(a) $K_\lambda(s)$ and $G_\lambda(s)$ are meromorphic in the topological product of the cut s plane and the entire λ plane.

(b) A domain of holomorphy is the topological product of the cut s plane and the interior of the circle

$$|\lambda| = \min[\lambda_1(0), -\lambda_{-1}(0)]$$

in the λ plane.

(c) The boundary values of $K_\lambda(s)$, $G_\lambda(s)$ are continuous functions in s and have the same analytic properties in λ as stated under (a) and (b).

[10] H. Rollnik, Z. Physik. **145**, 639 (1956).

[11] H. M. Nussenzweig, Nuclear Phys. **11**, 499 (1959).

(d) The radius of the circle bounding a domain of holomorphy in the λ plane tends to ∞, if $|s| \to \infty$ in the cut s plane, inclusive on the boundaries.

$$\text{(e)} \qquad \lim_{|s|\to\infty} K_\lambda(s)\{G_\lambda(s)\} = K(s)\{G_\lambda(s)\}. \qquad (6.2)$$

The first point is also clear from the Fredholm theory,[4,5] because the latter can be generalized to $\mathbf{L}^2$ kernels[3] and $K(s)$ is $\mathbf{L}^2$ in the cut physical sheet inclusive boundary values. More important are (b) and (c). They give rise to the following theorem about the convergence of Born expansions:

Theorem. Let λV be a class A potential that has a different sign only in a finite number of domains in $\mathbf{x}$ space. Then an expansion of $K_\lambda(s)$, $G_\lambda(s)$ into powers of λ (Born expansion) converges uniformly and absolutely in $\mathbf{x}$, $\mathbf{x}'$ for all $s \geq 0$ and even arbitrary complex s in the physical sheet, if and only if neither λV nor $-\lambda V$ has a bound state.

The proof is clear. The theorem also holds for the scattering amplitude, if the potential satisfies condition B. Convergence of the Born series is sufficient for the absence of bound states of $\pm\lambda V$ only, if it is required for all s in $0 \leq s < \infty$. Davies[12] has shown that the absence of bound states is sufficient for the convergence of the Born series in the case of a definite finite range potential. On the other side we know from the work of Kohn[13] that this is not true for Born expansions of partial wave quantities of angular momentum $l \geq 1$. This is clear from the fact that the eigenvalues are not bound by the lowest one as it holds for $l = 0$ and in the three-dimensional series.

The result (d) has been obtained also by Klein and Zemach[14] and Wigner[15] by inspection of the iterated kernels for equivalent conditions on the potential as stated under A. The final point is due to the fact that $|\lambda_n(s)| \to \infty$ for $|s| \to \infty$. The resolvent kernels tend to the "free" kernels and, correspondingly, the scattering amplitude to the Born approximation. This result is well known from Khuri's work on the Fredholm series in potential scattering.[5]

ACKNOWLEDGMENTS

The author gratefully acknowledges the hospitality of the Physics Department of Princeton University and the Brookhaven National Laboratory. He wants to thank Professor F. J. Dyson and Professor M. L. Goldberger for discussions.

[12] H. Davies, Nuclear Phys. **14**, 465 (1959/60).

[13] W. Kohn, Revs. Modern Phys. **26**, 292 (1954).

[14] A. Klein and C. Zemach, Nuovo cimento **10**, 1078 (1958).

[15] E. P. Wigner (unpublished).

On Scattering of Waves by Random Distributions. I. Free-Space Scatterer Formalism*

VICTOR TWERSKY
Sylvania Electronic Defense Laboratories, Mountain View, California
(Received December 22, 1960)

Approximations are derived for the bulk parameters of the coherent multiple scattered field in a slab region of randomly distributed arbitrary scatterers. (The one-, two-, and three-dimensional cases are treated simultaneously.) The propagation number K, and, e. g., ϵ and μ, are given explicitly in terms of conventional free-space isolated scattering amplitudes; these results generalize existing special forms for monopoles, dipoles, cylinders, and spheres. Corresponding approximations are obtained for the differential-scattering cross section per unit volume (i.e., the incoherent scattering), such that the total flux (coherent plus incoherent) fulfills the energy principle explicitly. Scattering and reciprocity theorems are derived for a "multiple scattering amplitude" of a scatterer within the distribution, and these are used to trace the energy "losses" of the coherent field which "reappear" as incoherent scattering. Several applications are considered.

1. INTRODUCTION

THE earliest analytical treatment of the scattering of waves by random distributions is essentially Rayleigh's theory of the color of the sky.[1] The subject has since received much attention in the literature, but we cite only papers related to the present analysis, particularly those of Reiche,[2] Foldy,[3] Lax,[4] Ament and Urick,[5] and Twersky.[6] Additional references are listed elsewhere.[4,7]

We consider the one-, two-, and three-dimensional problems of a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ incident at an arbitrary angle on a "uniformly random" distribution of arbitrary scatterers in a region bounded by two parallel planes (e.g., the "infinite-slab region" of Fig. 1). The initial formalism is similar to Foldy's[3] and Lax's[4] in that we seek the ensemble average of the Green's function representation for the field of a fixed configuration of scatterers. Subsequent steps essentially generalize Reiche's[2] approximation procedure for a wave normally incident on a slab of dipoles. We introduce approximations in order to express the average coherent field in terms of an integral over the slab thickness and write the kernel of the integral in terms of the "average multiple-scattering amplitude" G of a scatterer within the distribution.

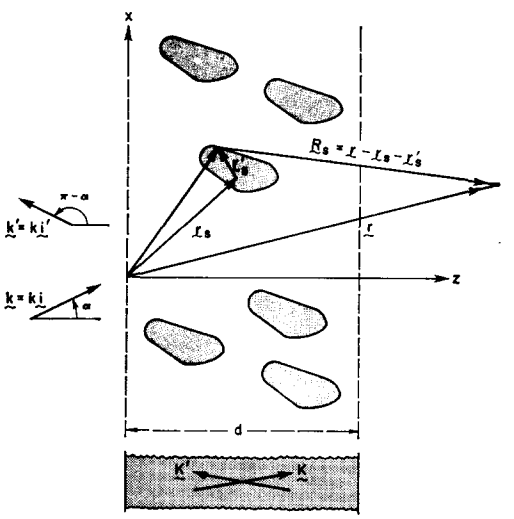

FIG. 1. Schematic for scattering of plane wave by a distribution of arbitrary identical scatterers bounded by two parallel planes (a "slab region"). Bottom part of drawing suggests the "synthetic medium" associated with the coherent scattered field.

In the present paper we express G in terms of the conventional scattering amplitude of a scatterer isolated in free space $g(\mathbf{k}_0, \mathbf{k})$ (where $\mathbf{k}$ and $\mathbf{k}_0$ are the incident and scattered propagation vectors). This leads to elementary integral equations whose solution gives a simple expression for the bulk propagation coefficient K (of the medium associated with the coherent field) explicitly in terms of special values of g. This form $K(g)$ for arbitrary scatterers and arbitrary angle of incidence is a generalization of various special forms given in the literature: If g is restricted to scatterers which are symmetrical to

* This work was partially supported by Signal Corps Contract DA 36-039 SC 75012.
[1] Lord Rayleigh, Phil. Mag. **47**, 375 (1899).
[2] F. Reiche, Ann. Physik **50**, 1 (1916); **50**, 121 (1916).
[3] L. L. Foldy, Phys. Rev. **67**, 107 (1945).
[4] M. Lax, Revs. Modern Phys. **23**, 287 (1951); Phys. Rev. **85**, 621 (1952).
[5] R. J. Urick and W. S. Ament, J. Acoust. Soc. Am. **21**, 115 (1949). This special form was rederived by P. C. Waterman and R. Truell, J. Math. Phys. **2**, 512 (1961).
[6] V. Twersky, "Multiple Scattering of Waves by a Volume Distribution of Parallel Cylinders, research report No. EM-59, Institute of Mathematical Sciences," New York University, 1953 (unpublished).
[7] V. Twersky, J. Research Natl. Bur. Standards **64**, 715 (1960).

the interfaces, then the present form reduces to one given previously by Twersky[6] (derived analytically for circular cylinders); the additional restriction to normal incidence reduces the form to one obtained by Ament and Urick[5] (derived analytically for spheres); the further restriction to monopole or dipole scatterers gives the form derived by Reiche,[2] Foldy,[3] and by Lax[4] [but Lax's is in terms of $g(\mathbf{K}, \mathbf{K})$ instead of $g(\mathbf{k}_0, \mathbf{k})$]. For the case of small scatterers all the above forms reduce to one given essentially by Rayleigh.[1]

To complete the description of the medium associated with the coherent field, we determine the average boundary conditions fulfilled at the slab interfaces; these plus $K(g)$ yield the macroscopic parameters (e.g., ϵ and μ) explicitly in terms of g. Introducing an appropriate auxiliary scattering amplitude $\mathcal{G}$ (such that for a semi-infinite slab the multiple-scattering amplitude G equals $\mathcal{G}e^{i\mathbf{K}\cdot\mathbf{r}}$), we derive its scattering and reciprocity theorems, and consider the physical basis of the propagation number K in detail. Thus, for lossless scatterers we express the coherent attenuation coefficient $2 \operatorname{Im} K$ in terms of the total scattering cross section associated with $\mathcal{G}$, etc.

Corresponding approximations are obtained for the differential scattering cross section per unit volume (i.e., the incoherent scattering), such that the resulting total flux (coherent plus incoherent) fulfills the energy principle explicitly. Using the general forms for the energy functions in terms of $\mathcal{G}$ and the scattering theorems for $\mathcal{G}$ we trace in detail the energy "losses" of the coherent field which "reappear" as incoherent scattering.

Several applications are considered; e.g., we determine the parameters for a distribution of small spheres, the behavior of the fields near grazing incidence, and the field scattered by a slab distribution backed by a ground plane.

The present paper, which neglects "hole corrections," etc., seems limited to sparse concentrations of scatterers. Subject to this restriction it provides a relatively comprehensive and explicit treatment in terms of the conventional scattering amplitude $g(\mathbf{k}_0, \mathbf{k})$. In addition, the entire introductory formalism, as well as theorems relating propagation coefficients and multiple scattered amplitudes, etc., are germane for alternative analytical procedures based on other functions than the free-space amplitude $g(\mathbf{k}_0, \mathbf{k})$. Thus, in a subsequent paper[7a] we exploit the present general results and recast the fields in terms of a "two-space scattering amplitude" $g(\mathbf{k}, \mathbf{K})$ corresponding to a scatterer excited by the coherent field $e^{i\mathbf{K}\cdot\mathbf{r}}$ but radiating into free k space.

2. NOTATION AND STATEMENT OF THE PROBLEM

We begin with a brief statement of the usual scattering problem for one object, and then proceed to configurations of objects, and an ensemble of configurations. Additional details are given elsewhere.[3,4,8]

2.1. One Scatterer

The scattering problem we consider (in one, two, or three dimensions) is specified by conditions at the scatterer's boundary A and by a wave function which in the region external to A fulfills

$$\psi = \varphi + v,$$
$$(\nabla^2 + k^2)e^{-i\omega t}\psi(r) = 0, \qquad k = 2\pi/\lambda;$$
$$\nabla^2 = \partial_z^2,\ \partial_z^2 + \partial_x^2,\ \partial_z^2 + \partial_y^2 + \partial_x^2; \qquad (2.1)$$
$$\partial_z^2 = \partial^2/\partial_z{}^2, \quad \text{etc.}$$

Here φ, the source term, is a plane wave

$$\varphi = e^{i\mathbf{k}\cdot\mathbf{r}}, \qquad \mathbf{k}\cdot\mathbf{r} = (k\mathbf{i})\cdot(r\mathbf{o})$$
$$= k(z\cos\alpha + x\sin\alpha), \qquad (2.2)$$

and v is an outgoing wave at large distances from the object ($r \to \infty$):

$$\lim_{r\to\infty} r^{(n-1)/2}(\partial_r v - ikv) = 0; \qquad n = 1, 2, 3. \qquad (2.3)$$

For brevity, we use three-dimensional terminology for all cases. Similarly, general forms indicating surface integrals are to be understood as line integrals in two dimensions, and as values at two points in one dimension (for which case r is to be read in general as $|z|\cos\alpha$, and factors involving x', x_s, etc., in following equations are redundant).

We apply Green's theorem to v and to the free space Green's functions

$$e^{ikx\sin\alpha}\left[\frac{e^{ik|z-z'|\cos\alpha}}{i2k\cos\alpha}\right], \qquad \frac{H_0^{(1)}(kR)}{i4} \equiv \frac{H_0(kR)}{i4},$$
$$\frac{e^{ikR}}{4\pi R} \equiv \frac{h_0(kR)}{i4\pi/k}, \qquad R = |\mathbf{r} - \mathbf{r}'|, \qquad (2.4)$$

where $\mathbf{r}$ and $\mathbf{r}'$ label a field point and a point on a surface inclosing the scatterer; collectively, we refer to these functions as $c\mathfrak{K}_0$, where

[7a] V. Twersky, J. Math. Phys. **3**, 724 (1962).

[8] V. Twersky, J. Acoust. Soc. Am. **29**, 209 (1957).

$\mathfrak{H}_0(kR) = e^{ikz\cos\alpha + ik|z-z'|\cos\alpha}$, $H_0(kR)$, $h_0(kR)$;

$$c = \frac{1}{i2k\cos\alpha}, \frac{1}{i4}, \frac{k}{i4\pi}. \qquad (2.4')$$

Integrating over the volume external to A, we obtain the surface integral

$$v = c\oint [\mathfrak{H}_0(k\,|\mathbf{r}-\mathbf{r}'|)\,\partial_n v(\mathbf{r}') - v(\mathbf{r}')\,\partial_n \mathfrak{H}_0(k\,|\mathbf{r}-\mathbf{r}'|)]\,dA(\mathbf{r}') \equiv \{\mathfrak{H}_0(k\,|\mathbf{r}-\mathbf{r}'|), v(\mathbf{r}', \mathbf{i})\}, \qquad (2.5)$$

where n is the outward normal, and where the operational braces include the appropriate constants. For convenience, we may add $\{\mathfrak{H}_0, \varphi\} = 0$ to the right-hand side of (2.5), and use

$$v = \{\mathfrak{H}_0(k\,|\mathbf{r}-\mathbf{r}'|), \psi(\mathbf{r}', \mathbf{i})\}. \qquad (2.5')$$

The usual boundary conditions on A, and the equation for the internal field lead in general to a pair of integral equations for specifying the scattered field. However, we do not consider single-body problems explicitly.

If $kr \gg 1$, $r \gg r'$, then

$$\mathfrak{H}_0(k\,|\mathbf{r}-\mathbf{r}'|) \sim \mathfrak{H}(kr)e^{-ik\mathbf{o}\cdot\mathbf{r}'}; \qquad \mathbf{o} = \mathbf{r}/r,$$

$$\mathfrak{H}(kr) = e^{ikz\sin\alpha + ik|z|\cos\alpha},\ (2/\pi kr)e^{ikr-i\pi/4},\ e^{ikr}/ikr.$$

The corresponding form of v may be written

$$v(r) \sim \mathfrak{H}(kr)g(\mathbf{o}, \mathbf{i}),$$

$$g(\mathbf{o}, \mathbf{i}) \equiv \{e^{-ik\mathbf{o}\cdot\mathbf{r}'}, v(\mathbf{r}', \mathbf{i})\} = \{e^{-ik\mathbf{o}\cdot\mathbf{r}'}, \psi(\mathbf{r}', \mathbf{i})\}, \qquad (2.7)$$

where the scattering amplitude g indicates the far-field response in the direction of observation $\mathbf{o}$ to plane wave excitation of direction of incidence $\mathbf{i}$.

In distinction to the usually defined scattering amplitudes (e.g., $f = g/ik$ in three dimensions) the present normalizations were chosen so that for lossless scatterers the forward amplitude theorem reads

$$-\mathrm{Re}\; g(\mathbf{i}, \mathbf{i}) = \mathfrak{M}\,|g(\mathbf{o}, \mathbf{i})|^2, \qquad (2.8)$$

where $\mathfrak{M}$ is the mean value over all directions of observation $\mathbf{o}$. Thus $\mathfrak{M}_3 = \int d\Omega/4\pi$, $\mathfrak{M}_2 = \int d\theta/2\pi$ and $\mathfrak{M}_1$ (for one-dimensional problems, i.e., for a slab scatterer) corresponds to one-half the sum of the transmitted and reflected values.

The above scalar forms for one and two dimensions also suffice for the usual electromagnetic problems. For three-dimensional electromagnetics, for present purposes, we may use $\varphi(\mathbf{i})\mathbf{e}$, $\mathbf{e} = \mathbf{y}/y$ to indicate the incident $\mathbf{E}$ or $\mathbf{H}$ field, and replace ψ, v, g by vector functions to obtain the analogs of (2.5) and (2.7) in a Cartesian representation. The corresponding analog of (2.8) is

$$-\mathrm{Re}\;\mathbf{e}\cdot\mathbf{g}(\mathbf{i}, \mathbf{i}) = \mathfrak{M}\mathbf{g}(\mathbf{o}, \mathbf{i})\cdot\mathbf{g}^*(\mathbf{o}, \mathbf{i}) = \mathfrak{M}\,|\mathbf{g}(\mathbf{o}, \mathbf{i})|^2. \qquad (2.9)$$

For simplicity, we retain the scalar formalism and limit consideration to cases such that $\mathbf{g}$ preserves the incident "polarization" $\mathbf{e}$ in the plane perpendicular to $\mathbf{e}$; specifically, we require[9]

$$\mathbf{g}(\mathbf{i}, \mathbf{i}) = \mathbf{e}g(\mathbf{i}, \mathbf{i}), \qquad \mathbf{g}(\mathbf{i}', \mathbf{i}) = \mathbf{e}g(\mathbf{i}', \mathbf{i}), \qquad (2.10)$$

where $\mathbf{i}'$ is the image of $\mathbf{i}$ in the plane $z = 0$. [Eq. (2.10) holds, for example, for spheres.] The corresponding plane wave image of φ will be written

$$\varphi' = e^{i\mathbf{k}'\cdot\mathbf{r}},$$

$$\mathbf{k}'\cdot\mathbf{r} = (k\mathbf{i}')\cdot(r\mathbf{o}) = k(-z\cos\alpha + x\sin\alpha). \qquad (2.11)$$

2.2. Configuration of Many Scatterers

A fixed configuration of N scatterers is specified by boundary conditions on surfaces $A_1, A_2, \cdots A_N$, located by fixed position vectors $\mathbf{r}_1 \cdots r_N$. We write the total field as

$$\Psi(\mathbf{r}) = \varphi(\mathbf{r}) + U(\mathbf{r}), \qquad (2.12)$$

where U has the same form as (2.5) with the surface of integration inclosing all objects. Breaking up the surface into individual portions inclosing individual scatterers (see Fig. 1 for geometry) we write

$$U(\mathbf{r}) = \sum_s u_s(\mathbf{r}-\mathbf{r}_s),$$

$$u_s(\mathbf{r}-\mathbf{r}_s) = \{\mathfrak{H}_0(kR_s), \Psi(\mathbf{r}_s + \mathbf{r}_s')\}_s, \qquad \mathbf{R}_s = \mathbf{r} - \mathbf{r}_s - \mathbf{r}_s', \qquad (2.13)$$

where $\Psi(\mathbf{r}_s + \mathbf{r}_s')$ is the total field at a point on a surface inclosing only scatterer s. Essentially, as for the single-body problem, the boundary conditions on A_s and the equations for the corresponding internal fields lead in general to a system of $2N$ equations for the $2N$ surface fields and their derivatives; however, we do not consider this system of equations explicitly.

If $kR_s \gg 1$, and $|\mathbf{r}-\mathbf{r}_s| \gg r_s'$, then analogous to (2.7) we write

$$u_s(\mathbf{r}-\mathbf{r}_s) \sim \mathfrak{H}(k\,|\mathbf{r}-\mathbf{r}_s|)\,G_s(\mathbf{o}),$$

$$G_s(\mathbf{o}) \equiv \{e^{-ik\mathbf{o}\cdot\mathbf{k}'s}, \Psi(\mathbf{r}_s + \mathbf{r}_s')\}_s, \qquad (2.14)$$

where G_s is the "multiple scattered amplitude" of scatterer s of the configuration. As the other scatterers recede to infinity, the functions u_s and G_s reduce to the corresponding functions for an isolated scatterer located at $\mathbf{r}_s$ with reference to the phase

[9] V. Twersky, IRE Trans. **AP-5**, 81 (1957).

origin $r = 0$ (i.e., to their "single-scattered" values $v_s e^{i\mathbf{k}\cdot\mathbf{r}_s}$ and $g_s e^{i\mathbf{k}\cdot\mathbf{r}_s}$).

2.3. Ensemble of Configurations

An ensemble of configurations is specified by an appropriate probability distribution function $W(\mathbf{1}, \mathbf{2}, \cdots \mathbf{N})$, where 1 stands for all properties of scatterer 1, etc. The ensemble average of U may be written as

$$\langle U\rangle = \sum_s \langle u_s\rangle = \sum_s \int \langle u_s(\mathbf{r} - \mathbf{r}_s)\rangle_s w_s(\mathbf{r}_s)\, d\mathbf{r}_s,$$

$$\langle u_s\rangle_s = \langle\{\mathfrak{C}_0(kR_s), \Psi\}_s\rangle_s = \{\mathfrak{C}_0, \langle\Psi\rangle_s\}_s, \qquad (2.15)$$

where $\langle u_s\rangle_s$ is the average of u_s over all variables except $\mathbf{r}_s$ (and is consequently independent of all distribution variables but $\mathbf{r}_s$), and where $w_s(\mathbf{r}_s)$ (the one-particle spatial-distribution function) is obtained by integrating W over the other variables. Similarly, we have

$$\langle\Psi\rangle_s = \varphi + \langle U\rangle_s$$

$$= \varphi + \langle u_s\rangle_s + \sum_t{}' \int \langle u_t\rangle_{st} w_{st}(r_s; r_t)\, d\mathbf{r}_t,$$

$$\langle u_t\rangle_{st} = \{\mathfrak{C}_0(kR_t), \langle\Psi(\mathbf{r}_t + \mathbf{r}_t')\rangle_{st}\}_t, \qquad (2.16)$$

where $\langle\ \rangle_{st}$ indicates an average with two position vectors fixed, and where w_{st} (the two-particle conditional probability density) is obtained by integrating W/w_s over all variables except $\mathbf{r}_s$ and $\mathbf{r}_t$; the prime on the sum means $t \neq s$. See Foldy,[3] Lax,[4] and Twersky[8] for additional general discussion.

2.4. Statement of the Problem

We consider an ensemble of configurations of identical similarly aligned scatterers $u_s = u$. We assume that W is symmetrical and that

$$w_s(r_s) = \rho/N, \qquad (2.17)$$

where ρ, the one-particle density function (the average number of scatterers in unit available volume) is independent of s. Substituting into (2.15), dropping unnecessary subscripts on u_s and r_s', and summing over s, we write

$$\langle U(\mathbf{r})\rangle = \rho \int \langle u(\mathbf{r} - \mathbf{r}_s; \mathbf{r}_s)\rangle_s\, d\mathbf{r}_s,$$

$$\langle u\rangle_s = \{\mathfrak{C}_0(kR_s), \langle\Psi(\mathbf{r}_s + \mathbf{r}')\rangle_s\}, \qquad (2.18)$$

where $\langle u\rangle_s$ is the average wave scattered by an object fixed at $\mathbf{r}_s$ (now a dummy variable for the volume integration). For concreteness, we take the origin of $\mathbf{r}'$ at the center of the sphere circumscribing a scatterer and restrict its end point to the scatterer's surface; thus $\langle\Psi(\mathbf{r}_s + \mathbf{r}')\rangle_s$ is the average field at a point $\mathbf{r}'$ on the scatterer fixed at $\mathbf{r}_s$.

If the "centers" of all scatterers (the end points of $\mathbf{r}_s$) lie within the "infinite-slab region" $0 < z < d$ (see Fig. 1), then from the symmetry of the distribution we require

$$\langle\Psi(\mathbf{r}_s + \mathbf{r}')\rangle_s = e^{ikx_s \sin\alpha}\langle\Psi(\mathbf{z}_s + \mathbf{r}')\rangle_s, \qquad (2.19)$$

e.g., in three dimensions, the average field at a point $\mathbf{r}'$ on a scatterer fixed at $\mathbf{r}_s = \mathbf{x}_s + \mathbf{y}_s + \mathbf{z}_s$ differs from that at the corresponding point on a scatterer fixed at $\mathbf{r}_s = \mathbf{z}_s$ only by the phase factor introduced by the incident wave (the essential feature of Snell's law). Substituting (2.19) into (2.18), we obtain

$$\langle U\rangle = \rho \int e^{ikx_s \sin\alpha}\{\mathfrak{C}_0(kR_s), \langle\Psi(\mathbf{z}_s + \mathbf{r}')\rangle_s\}\, d\mathbf{r}_s. \qquad (2.20)$$

For three and two dimensions, respectively, we introduce the plane-wave representations

$$h_0(kr) = \frac{1}{2\pi}\int_0^{2\pi} d\beta \times \int_0^{\pi/2 - i\infty} e^{ik[|z|\cos\tau + \sin\tau(x\cos\beta + y\sin\beta)]} \sin\tau\, d\tau,$$

$$H_0(kr) = \frac{1}{\pi}\int_{-\pi/2 + i\infty}^{\pi/2 - i\infty} e^{ik[|z|\cos\tau + x\sin\tau]}\, d\tau, \qquad (2.21)$$

and integrate over all coordinates but z_s [e.g., by using

$$\int_{-\infty}^{\infty} e^{ikx(t - \sin\alpha)}\, dx = (2\pi/k)\,\delta(t - \sin\alpha),$$

etc., after transforming to real limits in (2.21)]. Thus

$$\langle U\rangle = C\int_0^d \{e^{ik(x - x')\sin\alpha + ik|z - z_s - z'|\cos\alpha}, \langle\Psi(\mathbf{z}_s + \mathbf{r}')\rangle_s\}\, dz_s, \qquad (2.22)$$

$$C_1 = \rho_1, \quad C_2 = 2\rho_2/k\cos\alpha, \quad C_3 = 2\pi\rho_3/k^2\cos\alpha,$$

where the subscripts indicate the number of dimensions involved.

In analogy with (2.14), we introduce the "average multiple-scattering amplitude" of a scatterer on the z axis ($\mathbf{r}_s = \mathbf{z}_s$):

$$\langle u(\mathbf{r} - \mathbf{z}_s)\rangle_s \sim \mathfrak{C}(k\,|\mathbf{r} - \mathbf{z}_s|)\, G(z_s; \mathbf{0}),$$

$$G(z_s; \mathbf{0}) \equiv \langle G_s(z_s; \mathbf{0})\rangle_s = \{e^{-ik\mathbf{o}\cdot\mathbf{r}'}, \langle\Psi(\mathbf{z}_s + \mathbf{r}')\rangle_s\}. \qquad (2.23)$$

In terms of G, we rewrite $\langle U\rangle$ of (2.22) for

$z > (z_s + z')_{max} = d + z'_{max}$ (the forward scattered field) as

$$\langle U_> \rangle = \varphi C \int_0^d e^{-i\gamma\zeta} G(\zeta; \mathrm{i})\, d\zeta, \quad \gamma \equiv k \cos\alpha. \quad (2.24)$$

Similarly for $z < (z_s + z')_{min} = z'_{min}$ (the reflected field), we write

$$\langle U_< \rangle = \varphi' C \int_0^d e^{i\gamma\zeta} G(\zeta; \mathrm{i}')\, d\zeta, \quad (2.25)$$

where φ' is the image of φ in the slab's faces. Within the distribution ($z'_{max} < z < d - z'_{min}$), if we ignore that the ranges $z - z'_{min}$ to z, and z to $z + z'_{max}$ seem to call for special treatment as far as the definition of (2.23) and the form $\langle U \rangle$ of (2.22) are concerned, we may write

$$\langle \Psi \rangle = \varphi\left[1 + C \int_0^z e^{-i\gamma\zeta} G(\zeta; \mathrm{i})\, d\zeta\right] + \varphi' C \int_z^d e^{i\gamma\zeta} G(\zeta; i')\, d\zeta. \quad (2.26)$$

For present purposes we regard (2.26) as valid for $0 \leq z \leq d$; similarly, we use (2.24) for $z \geq d$ and (2.25) for $z \leq 0$. Thus $\langle \psi \rangle$ is continuous at $z = 0$ and $z = d$, and is fully determined by the average multiple scattering amplitude G.

In order to obtain a simple explicit representation of $\langle \Psi \rangle$, we assume that the two-particle function is a product of one-particle forms, and that the average with two variables held fixed may be approximated by that for one fixed variable (an assumption implicit in Reiche,[2] and introduced explicitly in Foldy[3]). Thus

$$\sum_t{}' w_{st} = \rho(N - 1)/N \approx \rho,$$

$$\langle \Psi(\mathbf{r}_t + \mathbf{r}') \rangle_{st} \approx \langle \Psi \rangle_t. \quad (2.27)$$

Introducing (2.27) into (2.16), we obtain

$$\langle \Psi \rangle_s \approx \langle \Psi \rangle + \langle u \rangle_s \quad (2.28)$$

which in view of (2.1), (2.18), and (2.26) is the form of the solution for a single object excited by a set of plane waves, and radiating into free space. The corresponding scattering amplitude G of $\langle u \rangle_s$ can thus be represented by a superposition of scattering amplitudes of single objects radiating into free space.

There are essentially two such representations we can obtain: one based on conventional isolated scattering amplitudes (2.7) is developed in Sec. 3; the other based on "two-space" isolated scattering amplitudes is developed in a subsequent paper.[7a]

3. FORMALISM BASED ON FREE-SPACE SCATTERERS

3.1. Average Wave Function

In the present paper we interpret (2.26) as a set of free-space plane waves $\langle \Psi \rangle = e^{i\mathbf{k}'\cdot\mathbf{r}} \mathfrak{A}(0, z) + e^{i\mathbf{k}'\cdot\mathbf{r}} \mathfrak{A}'(z, d)$, i.e., (2.26) indicates a multiple-scattering process in free space, the field at a scatterer at z consisting of the source term φ, plus the free-space forward scattered fields of the planes of scatterers in $0 < \zeta < z$, plus the free-space reflected fields of the planes $z < \zeta < d$. From this view it is convenient to rewrite (2.26) as

$$\langle \Psi \rangle = [\psi_+(0, z) + \psi_-(z, d)] e^{ikx \sin\alpha};$$

$$\psi_+(0, z) = e^{i\gamma z}\left[1 + C \int_0^z e^{-i\gamma\zeta} G(\zeta; \mathrm{i})\, d\zeta\right],$$

$$\psi_-(z, d) = e^{-i\gamma z} C \int_z^d e^{i\gamma\zeta} G(\zeta; \mathrm{i}')\, d\zeta. \quad (3.1)$$

Using (3.1) in (2.28), we superpose the appropriate isolated scattering amplitudes and write the corresponding average-scattering amplitude G of $\langle u \rangle_s$ as

$$G(\zeta; \mathrm{o}) = g(\mathrm{o}, \mathrm{i})\psi_+(0, \zeta) + g(\mathrm{o}, \mathrm{i}')\psi_-(\zeta, d). \quad (3.2)$$

Substituting into (2.26) and introducing

$$S_+ = Cg(\mathrm{i}, \mathrm{i}), \quad R_+ = Cg(\mathrm{i}, \mathrm{i}'), \quad S_- = Cg(\mathrm{i}', \mathrm{i}'), \quad R_- = Cg(\mathrm{i}', \mathrm{i}), \quad (3.3)$$

we obtain the pair of integral equations

$$\psi_+(0, z) = e^{i\gamma z} + e^{i\gamma z} \int_0^z e^{-i\gamma\zeta}[S_+\psi_+(0, \zeta) + R_+\psi_-(\zeta, d)]\, d\zeta,$$

$$\psi_-(z,d) = e^{-i\gamma z} \int_z^d e^{i\gamma\zeta}[R_-\psi_+(0, \zeta) + S_-\psi_-(\zeta, d)]\, d\zeta, \quad (3.4)$$

which hold even if the R's and S's are functions of ζ (i.e., if the average concentration of scatterers, or their sizes, etc., vary in the direction perpendicular to the slab). Differentiating with respect to z yields

$$\psi'_\pm = \pm i\gamma\psi_\pm \pm (S_\pm\psi_\pm + R_\pm\psi_\mp) = \pm T_\pm\psi_\pm \pm R_\pm\psi_\mp, \quad T_\pm \equiv i\gamma + S_\pm, \quad (3.5)$$

Differentiating again with respect to z and using (3.5), we obtain in general

$$\psi''_\pm + (T_- - T_+ - R'_\pm/R_\pm)\psi'_\pm + (R_+R_- - T_+T_- \mp T'_\pm \pm T_\pm R'_\pm/R_\pm)\psi_\pm = 0, \quad (3.6)$$

which, for coefficients independent of z, reduces to

$$\psi''_\pm + (T_- - T_+)\psi'_\pm + (R_+R_- - T_+T_-)\psi = 0. \quad (3.7)$$

For present purposes, we restrict consideration to (3.7).

The solutions of (3.7) may be written as

$$\psi_\pm = A_\pm e^{i\Gamma z} + B_\pm e^{i\Gamma' z}; \tag{3.8}$$

$\Gamma = \Delta + \Gamma_0, \quad \Gamma' = \Delta - \Gamma_0$:

$$i\Gamma_0 = [\tfrac{1}{4}(T_+ + T_-)^2 - R_+R_-]^{1/2} \equiv [T_a^2 - R_g^2]^{1/2},$$
$$\Delta = i(S_- + S_+)/2. \tag{3.9}$$

Substituting (3.8) into (3.5) we obtain

$$\begin{pmatrix} i\Gamma - T_+ & -R_+ \\ R_- & i\Gamma + T_- \end{pmatrix}\begin{pmatrix} A_+ \\ A_- \end{pmatrix} \equiv \mathbf{M}(\Gamma)\cdot\mathbf{A} = 0,$$
$$\mathbf{M}(\Gamma')\cdot\mathbf{B} = 0 \tag{3.10}$$

(such that $|\mathbf{M}| = 0$ gives (3.9)). The constants in (3.8) are determined by (3.10) plus the boundary values obtained from (3.6):

$$\psi_+(0, 0) = A_+ + B_+ = 1,$$
$$\psi_-(d, d) = A_-e^{i\Gamma d} + B_-e^{i\Gamma' d} = 0 \tag{3.11}$$

(where the first condition "extinguishes" the incident field inside the slab, and the second "extinguishes" the internal free-space reflected wave). Thus

$$A_+ = 1/[1 - QQ'e^{i(\Gamma-\Gamma')d}] \equiv D, \quad A_- = -QD,$$
$$B_+ = -Q'B_- = -Q'Qe^{i(\Gamma-\Gamma')d}\, D, \tag{3.12}$$
$$Q = R_-/(T_- + i\Gamma) = R_-/[T_a + (T_a^2 - R_g^2)^{1/2}],$$
$$Q' = R_+/(T_+ - i\Gamma') = R_+Q/R_-$$

[which also follow on substituting (3.8) into (3.4) and equating the coefficients of $e^{\mp i\gamma z}$, $e^{i\Gamma z}$, $e^{i\Gamma' z}$ to zero]. Consequently,

$$\psi_+ = e^{i\Delta z}(e^{i\Gamma_0 z} - QQ'e^{i\Gamma_0(2d-z)})\, D,$$
$$\psi_- = -Qe^{i\Delta z}(e^{i\Gamma_0 z} - e^{i\Gamma_0(2d-z)})\, D, \tag{3.13}$$

i.e., the sum of the elementary free-space forward waves in ψ_+ equals two imaged plane waves traveling in a new medium, etc. Thus the internal, reflected, and transmitted coherent fields [equal to $e^{ikx\sin\alpha}$ times $\psi_+ + \psi_-$, $\psi_-(0, d)$, and $\psi_+(0, d)$, respectively] are given by

$$\Psi_I = [(1 - Q)e^{i\mathbf{K}\cdot\mathbf{r}} + Q(1 - Q')e^{i\mathbf{K}'\cdot\mathbf{r} + i(\Gamma-\Gamma')d}]\, D,$$
$$\mathbf{K}\cdot\mathbf{x} = \mathbf{K}'\cdot\mathbf{x} = kx\sin\alpha, \tag{3.14}$$
$$\Psi_R = \langle U_< \rangle = -Q(1 - e^{i(\Gamma-\Gamma')d})\, De^{i\mathbf{k}'\cdot\mathbf{r}} = \mathfrak{R}_+\varphi',$$
$$\Psi_T = \langle U_> \rangle = (1 - QQ')\, De^{i(\Gamma-\gamma)d + i\mathbf{k}\cdot\mathbf{r}} = \mathfrak{I}_+\varphi,$$

where $\Gamma - \Gamma' = 2\Gamma_0$. These are essentially the forms for a continuous slab, with Q and Q' playing the role of generalized "single-surface" Fresnel reflection coefficients. The above scalar results also apply for electromagnetic problems subject to (2.10).

The special forms of the above for $T_+ = T_-$ and $R_+ = R_-$ were obtained originally[6] by solving the analogous integral equations for G in terms of g. The special case of dipoles ($S_+ = S_- \simeq -R_+ = -R_-$) and normal incidence ($\alpha = 0$) was treated by Reiche[2] who introduced essentially the present procedure. The integral equations of (3.4) which led to (3.14) specify "four-terminal" continuous transmission lines. The discrete forms of these equations (in which ζ is replaced by a multiple of a constant spacing between planes, and the integrals by sums) specify a "one-mode" approximation for the periodic case (provided the S's and R's are replaced by the appropriate values for an isolated lattice plane); essentially the corresponding difference equations for the case of a semi-infinite crystal were treated originally by Darwin[10] to obtain the discrete analogs of the present results.

Since the final result Ψ_I of (3.14) for the internal field in the synthetic "macroscopic medium" has the form $Ae^{i\mathbf{K}\cdot\mathbf{r}} + A'e^{i\mathbf{K}'\cdot\mathbf{r}}$ (where the A's are independent of distance), while the starting form (3.1) at the "microscopic level" was

$$\Psi = \mathfrak{A}(0, z)e^{i\mathbf{k}\cdot\mathbf{r}} + \mathfrak{A}'(z, d)e^{i\mathbf{k}'\cdot\mathbf{r}},$$

the present formalism is not "fully self-consistent." However, the first-order equations (3.5) give the essential physics and serve to make the procedure plausible. Thus if we "return from the limit" and replace $\partial_z\psi_+$ by its difference form $[\psi_+(z + \Delta z) - \psi_+(z)]/\Delta z$, and regroup the corresponding terms of (3.5), we obtain

$$\psi_+(z + \Delta z) = \psi_+(z) + \Delta z[T_+\psi_+(z) + R_+\psi_-(z)], \tag{3.5$_+$}$$

i.e., the "forward field" at $z + \Delta z$ equals that at z plus the fraction of forward field transmitted through the elementary thickness Δz, plus the fraction of "backward field" reflected at z through Δz. Similarly, for the backward field,

$$\psi_-(z) = \psi_-(z + \Delta z)$$
$$+ \Delta z[T_-\psi_-(z + \Delta z) + R_-\psi_+(z + \Delta z)]. \tag{3.5$_-$}$$

Thus (3.5) could have been written directly from elementary physical considerations. Indeed essentially such arguments were used by Darwin[10] to construct the difference-equation analogs of (3.5$_-$) for the periodic case. [However, the procedure which led us to (3.5) is closer to Ewald's[11] analysis of scattering by a periodic array of dipoles.]

[10] C. G. Darwin, Phil. Mag. **27**, 315, 675 (1914).
[11] P. P. Ewald, Ann. Physik **49**, 1, 117 (1916); **54**, 519 (1917).

[The present formalism in terms of free-space scattering amplitudes appears limited to sparse concentrations (such that $1/\rho$ is much larger than the volume of a scatterer), i.e., we have essentially the start of a perturbation procedure around the properties of free space. To extend this procedure to dense distributions requires a more general two-particle distribution function than used in (2.27). Thus if in (2.16) we use $\sum w_{st} = \rho$ only for values of $\mathbf{r}_t$ outside some volume V containing scatterer s, and use $w_{ts} = 0$ inside V, then instead of (2.28) we obtain

$$\langle \Psi(\mathbf{r}_s)\rangle_s \approx \langle \Psi(\mathbf{r}_s)\rangle + \langle u(\mathbf{r}_s - \mathbf{r}_s)\rangle_s - \rho \int_V \langle u(\mathbf{r}_s - \mathbf{r}_t)\rangle_t \, d\mathbf{r}_t, \qquad (2.28')$$

where V is the "hole" arising from the size and shape of the scatterers, etc. For some cases we would expand $\langle u(\mathbf{r}_s - \mathbf{r}_t)\rangle_t$ as a Taylor series of derivatives of $\langle u(\mathbf{r}_s - \mathbf{r}_s)\rangle_s$ and powers of $|\mathbf{r}_t - \mathbf{r}_s|$, and were the first term to suffice, we would obtain (2.28) with $\langle u\rangle_s$ multiplied by $1 - \rho V$; other simple cases arise when the integral in (2.28′) is proportional to $\langle \Psi\rangle$ or to $\langle U\rangle$. More generally, we would expand the integral as a set of plane waves and apply the superposition principle to the complete set[12]; such expansions are also useful for "liquid-state" statistics, for which we introduce an explicit approximation for the pair distribution function w_{st} in (2.16), and attempt to integrate over the total volume.[13] Analytical treatments for dense gases based essentially on free-space scatterers are given by Mazur[14] and his associates, and various more physical approaches are reviewed by Böttcher.[15] We do not consider such procedures, and therby limit the discussion of the formalism based on $g(\mathbf{k}, \mathbf{k})$ to sparse distributions.]

Specular amplitudes. The coefficients $\mathfrak{R}_+$ and $\mathfrak{T}_+$ equal the reflection and transmission amplitudes of the slab for direction of incidence $\mathbf{i}$. For a wave incident in the image direction $\mathbf{i}'$, we reflect the structure by interchanging $+$ and $-$ in the above; this corresponds to replacing Δ by $-\Delta$ and Q by Q'. Thus the ratios of the two transmitted fields, reflected fields equal

$$\mathfrak{T}_+/\mathfrak{T}_- = e^{i(\Gamma+\Gamma')d} = e^{(S_+ - S_-)d}, \qquad \mathfrak{R}_+/\mathfrak{R}_- = R_+/R_-; \qquad (3.15)$$

only the first depends on slab thickness. For normal incidence the usual reciprocity theorem gives $g(\mathbf{i}, \mathbf{i}) = g(\mathbf{i}', \mathbf{i}') = g(-\mathbf{i}, -\mathbf{i})$; consequently $S_+ = S_-$, and the first ratio reduces to unity. For scatterers symmetrical to the faces of the slab [$g(\mathbf{i}, \mathbf{i}) = g(\mathbf{i}', \mathbf{i}')$, and $g(\mathbf{i}, \mathbf{i}') = g(\mathbf{i}', \mathbf{i})$], both ratios reduce to unity. For direction of incidence $-\mathbf{i}$, the reciprocity theorem gives a transmission coefficient $\mathfrak{T}_+$ as in (3.14) and a reflection coefficient $\mathfrak{R}_-$ as in (3.15); for direction of incidence $-\mathbf{i}'$ we get $\mathfrak{R}_+$ as in (3.14), and $\mathfrak{T}_-$ as in (3.15). (See also Redheffer,[16] for additional germane results for such "four-terminal" networks.)

Propagation coefficients. The waves in the new medium are of the form

$$e^{i\mathbf{K}\cdot\mathbf{r}} = e^{i\Gamma z + ikx\sin\alpha} = e^{-\operatorname{Im}\Gamma z + i(\operatorname{Re}\Gamma z + kx\sin\alpha)}. \qquad (3.16)$$

The planes of constant amplitude are parallel to the slab, the attenuation coefficient equaling 2 Im Γ. The planes of constant phase are given by

$$z \operatorname{Re}\Gamma + kx\sin\alpha = k\eta_r(z\cos\beta + x\sin\beta),$$

$$\eta_r = \operatorname{Re}\Gamma/k\cos\beta = \sin\alpha/\sin\beta = [(\operatorname{Re}\Gamma/k)^2 + \sin^2\alpha]^{1/2}, \qquad (3.17)$$

where η_r is the real index of refraction, and β is the real angle of refraction.

The propagation coefficients of the medium associated with the coherent field, and the complex indices of refraction are

$$K_\pm^2 = k^2\eta_\pm^2 = k^2\sin^2\alpha + (\Delta \pm \Gamma_0)^2. \qquad (3.18)$$

In particular, for scatterers symmetrical to the slab faces, (3.18) reduces to

$$K^2 = k^2 - i2kS\cos\alpha - S^2 + R^2, \qquad S = S_+ = S_-, \qquad R = R_+ = R_-, \qquad (3.19)$$

a form obtained originally by separation of variables in circular cylindrical coordinates.[6] In terms of $g = g(\mathbf{i}, \mathbf{i})$ or $\mathbf{e}\cdot\mathbf{g}(\mathbf{i}, \mathbf{i})$, and $g' = g(\mathbf{i}', \mathbf{i})$ or $\mathbf{e}\cdot\mathbf{g}(\mathbf{i}', \mathbf{i})$ we have in one, two, and three dimensions

$$K_1^2 = k^2 - 2i\rho_1 kg\cos\alpha + \rho_1^2(g'^2 - g^2),$$

$$K_2^2 = k^2 - 4i\rho_2 g + (2\rho_2/k\cos\alpha)^2(g'^2 - g^2), \qquad (3.20)$$

$$K_3^2 = k^2 - 4\pi i\rho_3 g/k + (2\rho_3\pi/k^2\cos\alpha)^2(g'^2 - g^2).$$

[12] Integral equation representations for G in terms of g for arbitrary configurations are given by V. Twersky in *Electromagnetic Waves*, edited by R. E. Langer (University of Wisconsin Press, Madison, Wisconsin, 1962), pp. 361-389; J. Math. Phys. **3**, 83 (1962).

[13] See V. Twersky, IRE Trans. **AP-7**, S307 (1959) for a multiple scattering treatment of a "one-dimensional liquid" (e.g., a "random grating") based on the pair distribution function used by F. Zernike and J. A. Prins, Z. Phys. **41**, 184 (1927) in their single scattering treatment of the analogous problem.

[14] P. Mazur and M. Mandel, Physica **22**, 289, 299 (1956); L. Jansen and P. Mazur, *ibid.* **21**, 193 (1955).

[15] C. J. F. Böttcher, *Theory of Electric Polarization* (Elsevier Publishing Company, Princeton, New Jersey, 1952).

[16] R. M. Redheffer, J. Math. and Phys. **28**, 237 (1950).

To facilitate comparison with the three-dimensional results of others, we rewrite K_3 of (3.18) in terms of the conventionally defined isolated scatterer amplitude $f = g/ik$ (the amplitude in $v \sim fe^{ikr}/r$):

$$K^2 = k^2 + 4\pi\rho f - (2\pi\rho/k\cos\alpha)^2(f'^2 - f^2);$$
$$f = f(\mathbf{i}, \mathbf{i}), \quad \mathbf{e}\cdot\mathbf{f}(\mathbf{i}, \mathbf{i});$$
$$f' = f(\mathbf{i}', \mathbf{i}), \quad \mathbf{e}\cdot\mathbf{f}(\mathbf{i}', \mathbf{i}). \tag{3.21}$$

For incidence normal to the slab, (3.21) reduces to

$$K^2 = k^2 + 4\pi\rho f + \left(\frac{2\pi\rho}{k}\right)^2(f^2 - f'^2)$$
$$= \left(k + \frac{2\pi\rho f}{k}\right)^2 - \left(\frac{2\pi\rho f'}{k}\right)^2$$
$$= \left[k + \frac{2\pi\rho}{k}(f + f')\right]\left[k + \frac{2\pi\rho}{k}(f - f')\right], \tag{3.22}$$

where $f' = f(-\mathbf{i}, \mathbf{i})$; this result (with f expressed as a series of the usual scattering coefficients) was obtained originally by Urick and Ament[5] for a slab of spheres. For normal incidence on a slab of electric dipoles Reiche[2] obtained

$$K^2 = k^2 + 4\pi\rho f, \tag{3.23}$$

which follows from (3.22) by specializing to dipoles, $f = -f'$. The form was also obtained by Foldy[3] for general regions of monopoles, $f = f'$. The result obtained by Lax[4] has the form (3.23) but involves $f(\mathbf{K}, \mathbf{K})$ instead of the present $f(\mathbf{k}, \mathbf{k})$. Similarly, the results obtained by Rayleigh[1] for small scatterers and spheres may be put in the general form

$$K \approx k + 2\pi\rho f/k, \tag{3.24}$$

as obtained by expanding the square root of any of the above.

Bulk parameters. The boundary conditions fulfilled by $\langle\Psi\rangle$ and $\partial_z\langle\Psi\rangle$ at $z = 0, d$ plus the explicit representation for the propagation coefficient (3.18) enable us in general to determine the bulk parameters of the medium associated with the coherent field. From (3.14) we see that $\langle\Psi\rangle$ is continuous at the boundaries: $\varphi(0) + \Psi_R(0) = \Psi_I(0)$, and $\Psi_T(d) = \Psi_I(d)$, or, for brevity,

$$\Psi_{\text{out}} = \Psi_{\text{in}}. \tag{3.25}$$

Similarly it may be shown that

$$\partial_z\Psi_{\text{out}} = B\,\partial_z\Psi_{\text{in}} + B'\Psi_{\text{in}},$$
$$B = \frac{\gamma}{\Gamma - \Gamma'}\left(\frac{1 + Q}{1 - Q} + \frac{1 + Q'}{1 - Q'}\right)$$
$$\equiv \frac{\gamma}{\Gamma - \Gamma'}(Z + Z'),$$
$$B' = \frac{-iB}{2}(\Gamma + \Gamma') + \frac{i\gamma}{2}(Z - Z'), \tag{3.26}$$

where the Z's have the form of "impedances." Substituting for Q and Q' from (3.12) we obtain

$$B = [1 + (S_+ + S_- - R_+ - R_-)/2i\gamma]^{-1},$$
$$B' = B[R_- - R_+ - i(S_- - S_+)]/2.$$

If the scatterers are symmetrical to the face then $B' = 0$, and (3.26) reduces to

$$\partial_z\Psi_{\text{out}} = B\,\partial_z\Psi_{\text{in}}, \qquad B = \frac{\gamma}{\Gamma}\left(\frac{1 + Q}{1 - Q}\right) = \frac{\gamma Z}{\Gamma};$$
$$\frac{1}{B} = 1 + \frac{S - R}{i\gamma}. \tag{3.27}$$

To illustrate the above, we apply (3.27) to the electromagnetic case. Thus if the incident field, either $\mathbf{H}_i$ or $\mathbf{E}_i$, equals $\mathbf{e}\,e^{i\mathbf{k}\cdot\mathbf{r}}$, then (3.27) provides either the bulk permittivity ϵ or permeability μ:

$$\left.\begin{matrix}\mathbf{H}_i\\ \mathbf{E}_i\end{matrix}\right\} = \mathbf{e}e^{i\mathbf{k}\cdot\mathbf{r}}, \quad \left.\begin{matrix}\epsilon\\ \mu\end{matrix}\right\} = 1 + \frac{S - R}{ik\cos\alpha};$$
$$\mathbf{e} = \mathbf{y}/y, \quad k\cos\alpha = \mathbf{k}\cdot\mathbf{z}/z \equiv \mathbf{k}\cdot\mathbf{z}_o \tag{3.28}$$

The remaining parameter follows from (3.19) and the relation

$$K^2 = k^2\eta^2 = k^2\,\epsilon\,\mu. \tag{3.29}$$

In particular for normal incidence we have

$$\eta^2 = \left(1 - \frac{S - R}{ik}\right)\left(1 + \frac{S + R}{ik}\right), \tag{3.30}$$

and consequently

$$\mathbf{H}_i = \mathbf{e}e^{i\mathbf{k}\cdot\mathbf{r}}; \qquad \epsilon = 1/B = 1 + (S - R)/ik,$$
$$\mu = \eta^2 B = 1 + (S + R)/ik. \tag{3.31}$$

Here $S \pm R = C(g \pm g')$, with $C = \rho_1, 2\rho_2/k, 2\rho_3\pi/k^2$. In three dimensions, we write (3.31) in terms of $f = g/ik$ as

$$\epsilon = 1 + 2\pi\rho(f - f')/k^2,$$
$$\mu = 1 + 2\pi\rho(f + f')/k^2. \tag{3.32}$$

Thus, for example, for small perfectly conducting spheres of radius a, we have to lowest order $f = k^2a^3/2$ and $f' = -3k^2a^3/2$; consequently $\epsilon \approx 1 + 4\pi\rho a^3$ and $\mu \approx 1 - 2\pi\rho a^3$ in accord with alternative derivations. Similarly for an electric dipole ($f = -f'$) we obtain $\mu = 1$ and $\epsilon = 1 + 4\pi\rho f/k^2$, where $4\pi f/k^2$ is identified with the usual electric dipole moment.

As a more general illustration we consider the results for arbitrary angle of incidence on a slab of

"adjustable" electric dipoles (e.g., small dielectric spheres of radius a and permittivity ϵ', or the elementary model of an electron). For this case and $\mathbf{E}_i = \mathbf{e}\varphi$, we have $f = f' = f_0$ [where $f_0 = k^2a^3(\epsilon' - 1)/(\epsilon' + 2)$ for spheres]; on the other hand, for $\mathbf{H}_i = \mathbf{e}\varphi$, we have $f' = -f \cos 2\alpha$, $f = f_0$. Using (3.28) for either ϵ or μ, and (3.21) to obtain the other from $\eta^2 = \epsilon\mu$, we find

$$\begin{aligned}
\mathbf{E}_i = \mathbf{e}\varphi: \quad & \epsilon_e = 1 + M = 1 + 4\pi\rho f_0/k^2 = \eta_e^2, \\
& \mu_e = 1. \\
\mathbf{H}_i = \mathbf{e}\varphi: \quad & \epsilon_m = 1 + M, \\
& \eta_m^2 = 1 + M + M^2 \sin^2\alpha, \\
& \mu_m = 1 + M^2 \sin^2\alpha/(1 + M). \qquad (3.33)
\end{aligned}$$

Thus these results indicate that ϵ is independent of polarization for all angles of incidence, but that μ and η depend on polarization to the second order of a generally small quantity. (The slight polarization dependence of μ and η for this special case indicates the limitations of the present approach.) For normal incidence, these results reduce to those obtained originally by Reiche.[2]

As another illustration we consider the behavior of the various functions of this section for near grazing incidence $\alpha = (\pi/2) - \tau \to \pi/2$, and demonstrate that no singularities arise. As shown elsewhere,[8,9] as $\tau \to 0$, the sum $f + f'$ is nonvanishing while the difference $f - f'$ vanishes as τ^2. Consequently, for $\tau \to 0$, we find that $(S - R)/i\gamma \to M_0$, while $\tau^2(S + R)/i\gamma \to P_0$, such that M_0 and P_0 are nonvanishing constants. From (3.29) it follows that $\epsilon, \mu \to 1 + M_0$; similarly $\eta^2 = \mu\epsilon \to 1 + P_0(1 + M_0)$, and $\Gamma^2 \to k^2P_0(1 + M_0) \equiv N_0^2$.

In order to determine the behavior of the fields for $\alpha = (\pi/2) - \tau \to \pi/2$, we note that $Z\tau \to kP_0/N_0$, and that consequently $Q = (Z - 1)/(Z + 1) \sim 1 - 2N_0\tau/kP_0 \to 1$. Using this result in (3.14) we see that the transmitted and internal fields vanish, and that the reflected field reduces to $\Psi_R \sim -\varphi' - \varphi'(2N_0/kP_0)\tau i \cot N_0 d \to -\varphi'$. Thus independently of polarization the reflected field reduces linearly to $-\varphi'$ as the "grazing angle" τ vanishes, the limit being that of a perfect reflector governed by the boundary condition of vanishing wave function for either polarization: The incident wave is reflected with a phase change of π with no amplitude change. This is identical with the result obtained previously for planar random distributions of scatterers ("random screens"), and for random distributions of bosses on a ground plane ("rough surfaces").[8,9]

Additional applications of the present formalism (e.g., the restrictions on g required for unit permeability or unit permittivity, or for polarization independent bulk parameters) will be found elsewhere.[17]

3.2. Physical Bases of the Propagation Number

Rayleigh's approximation. The propagation number in terms of g has a simple physical interpretation provided that Cg is small enough for the higher order terms to be neglected. For this range we obtain essentially Rayleigh's result[1] generalized to arbitrary scatterers:

$$\Gamma_+ = \Delta + \Gamma \approx \gamma - iS_+ = \gamma - iCg(\mathbf{i}, \mathbf{i}) \equiv \Gamma_R, \qquad (3.34)$$

which states that the component of the propagation vector in the direction normal to the slab equals the free-space value $\gamma = k \cos\alpha$ perturbed by the effects of a single-scattering traversal of unit thickness of material. This result may be obtained directly by an elementary procedure based on asymptotic forms and a stationary phase evaluation. Thus the far scattered field of a single object at $\mathbf{r}_s$ excited by $e^{i\mathbf{k}\cdot\mathbf{r}} = \varphi$ equals $\mathfrak{C}(k\,|\mathbf{r} - \mathbf{r}_s|)g(\mathbf{o}, \mathbf{i})e^{i\mathbf{k}\cdot\mathbf{r}_s}$, and ρ times this function integrated over a plane parallel to the slab face equals the contribution of a layer of unit thickness. Evaluating the integral by the method of stationary phase gives $\varphi Cg(\mathbf{i}, \mathbf{i})$ as the forward wave scattered by an elementary layer; the total transmitted field at z is thus $\varphi(1 + Cgz) \approx e^{ikz\sin\alpha + i(\gamma - iC\rho)z}$. [For the electromagnetic case, we use $\varphi\mathbf{e}$ and $\mathfrak{C}(k\,|\mathbf{r} - \mathbf{r}_s|)\mathbf{g}(\mathbf{o}, \mathbf{i})e^{i\mathbf{k}\cdot\mathbf{r}_s}$, and obtain the same forward waves in terms of $g(\mathbf{i}, \mathbf{i}) = \mathbf{g}(\mathbf{i}, \mathbf{i})\cdot\mathbf{e}$.] The same result follows from an equivalent argument based on Fresnel zones: the size of the first Fresnel zone is $2(\lambda r)^{1/2} \sec\alpha$ or $\pi r\lambda \sec\alpha$ in two or three dimensions (provided $\lambda \ll r\cos^2\alpha$); thus φCg is proportional to the forward scattered wave of an element at the center of the first zone times the number of elements in the zone.

The real part of Γ_R of (3.34) equals the phase change normal to the slab arising from a single scattering traversal of unit thickness. Twice the imaginary part of Γ_R, the attenuation coefficient, equals

$$2 \operatorname{Im} \Gamma_R = -2C \operatorname{Re} g(\mathbf{i}, \mathbf{i}). \qquad (3.35)$$

Since the total energy cross section $p = p_a + p_s$ (absorption plus scattering cross sections) of an isolated object is given by

[17] V. Twersky, "On Scattering of Waves by a Slab Region of Randomly Distributed Objects," Report EDL-E26, Sylvania Electronic Defense Laboratories, 1958 (unpublished).

$$p = p_a + p_s = p_a + 2C_0\mathfrak{M}\,|g(\mathbf{o}, \mathbf{i})|^2$$
$$= -2C_0 \operatorname{Re} g(\mathbf{i}, \mathbf{i}),$$
$$C_0 \equiv \cos\alpha, \quad 2/k, \quad 2\pi/k^2, \tag{3.36}$$

and since $C = C_0\rho \sec\alpha$, we obtain

$$2 \operatorname{Im} \Gamma_R = \rho p \sec\alpha, \tag{3.37}$$

where $\rho \sec\alpha$ times unit volume is the number of scatterers irradiated by unit area of wave front in traversing unit thickness, and $e^{-2\operatorname{Im}\Gamma_R} = e^{-\rho p \sec\alpha}$ is the corresponding energy loss, the resultant time-averaged coherent power at $z + dz$ equaling that at z less the fraction scattered and absorbed by the included elements. From the physical meaning of the terms, one could start with

$$e^{-2\operatorname{Im}\Gamma_R(z+\Delta z)} = e^{-2\operatorname{Im}\Gamma_R z}\left(1 - \frac{\rho p}{\cos\alpha}\,\Delta z\right)$$

and proceed to the limit

$$d_z e^{-2\operatorname{Im}\Gamma_R z} = -e^{-2\operatorname{Im}\Gamma_R z}\rho p \sec\alpha$$

to obtain (3.37); essentially this procedure was used by Rayleigh[1] for lossless scatterers for which

$$2 \operatorname{Im} \Gamma_R = \rho p \sec\alpha = 2C\mathfrak{M}\,|\mathbf{g}(\mathbf{o}, \mathbf{i})|^2. \tag{3.38}$$

For the three-dimensional case of small lossless arbitrary scatterers, Rayleigh approximated g by its leading term (proportional to the volume of the scatterer divided by $i\lambda^3$), say g_R. He used g_R and essentially the stationary phase procedure to obtain $\operatorname{Re} \Gamma_R = \gamma - iCg_R$; he then used the physical argument mentioned prior to (3.38) to obtain $2 \operatorname{Im} \Gamma_R \approx 2C\mathfrak{M}\,|g_R(\mathbf{o}, \mathbf{i})|^2$. In addition, for small lossless spheres, Rayleigh used sufficient terms of the Legendre series expansion of g to obtain both the leading phase and amplitude components of Γ_R, and demonstrated that the corresponding value of $\operatorname{Im} \Gamma_R$ was identically that obtained through energy considerations (e.g., by integrating the square of the leading term of the series over the angle of observation). Thus although Rayleigh did not use the cross section theorem (3.36) explicitly, he essentially demonstrated that (2.8) held for small lossless spheres. More generally, Rayleigh's separate treatment of the phase and intensity effects made maximum use of the incomplete representation g_R for g.

The complete form of Γ. The analogous physical interpretation for the complete form of Γ follows in terms of a different scattering amplitude, say $\mathcal{G}$. Refering back to (3.2) and (3.13), we write the average multiple scattered amplitude for a scatterer fixed at z within the distribution as

$$G(z; \mathbf{o}) = \mathcal{G}(\mathbf{o}, \mathbf{i})e^{i\Gamma z} D + \mathcal{G}(\mathbf{o}, \mathbf{i}')Qe^{i\Gamma' z + i(\Gamma - \Gamma')d} D,$$
$$\mathcal{G}(\mathbf{o}, \mathbf{i}) = g(\mathbf{o}, \mathbf{i}) - g(\mathbf{o}, \mathbf{i}')Q,$$
$$\mathcal{G}(\mathbf{o}, \mathbf{i}') = g(\mathbf{o}, \mathbf{i}') - g(\mathbf{o}, \mathbf{i})Q', \tag{3.39}$$

where $\mathcal{G}(\mathbf{o}, \mathbf{i})$ is associated with $e^{i\mathbf{K}\cdot\mathbf{r}}$, etc.

If the region is semi-infinite ($d \to \infty$), then (3.39) reduces to

$$G(z; \mathbf{o}) = \mathcal{G}(\mathbf{o}, \mathbf{i})e^{i\Gamma z}, \tag{3.40}$$

which indicates that $\mathcal{G}$ plays the same role in Γ as g plays in Γ_R. More explicitly, since

$$Q = [S_+ + i(\gamma - \Gamma)]/R_+$$

[as obtained from $Q = R_-/(T_- + i\Gamma)$ of (3.12), the relation $R_-R_+ = (T_+ - i\Gamma)(T_- + i\Gamma)$ of (3.9), and the definition $T_\pm = i\gamma + S_\pm$], it follows that

$$\mathcal{G}(\mathbf{i}, \mathbf{i}) = g(\mathbf{i}, \mathbf{i}) - g(\mathbf{i}, \mathbf{i}')Q$$
$$= (S_+ - R_+Q)/C = i(\Gamma - \gamma)/C.$$

Consequently

$$\Gamma = \gamma - iC\mathcal{G}(\mathbf{i}, \mathbf{i}) \tag{3.41}$$

has the same form as $\Gamma_r(g)$ of (3.34). We now define a corresponding cross section P through

$$2 \operatorname{Im} \Gamma = 2C \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i}) \equiv \rho P \sec, \tag{3.42}$$

and complete the analogy by showing that P has the required characteristics.

In particular, for lossless scatterers, the average of $|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2$ over angles of observation equals

$$\mathfrak{M}\,|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2 = \mathfrak{M}[|g(\mathbf{o}, \mathbf{i})|^2 + |Qg(\mathbf{o}, \mathbf{i}')|^2$$
$$-2 \operatorname{Re} g^*(\mathbf{o}, \mathbf{i})g(\mathbf{o}, \mathbf{i}')Q].$$

Using the general theorem for an isolated lossless scatterer,

$$-2\mathfrak{M}g^*(\mathbf{o}, \mathbf{i})g(\mathbf{o}, \mathbf{j}) = g(\mathbf{i}, \mathbf{j}) + g^*(\mathbf{j}, \mathbf{i}), \tag{3.43}$$

we obtain

$$C\mathfrak{M}\,|\mathcal{G}|^2 = -\operatorname{Re}\,[S_+ + S_-\,|Q|^2 - (R_-^* + R_+)Q]$$
$$= (1 - |Q|^2) \operatorname{Im} \Gamma,$$

where we used $\operatorname{Re}\,(R_+Q) = \operatorname{Re} S_+ + \operatorname{Im} \Gamma$, and $\operatorname{Re}\,(R_-^*Q) = \operatorname{Re}\,(S_- - \operatorname{Im} \Gamma)\,|Q|^2$. Thus for lossless elements,

$$2 \operatorname{Im} \Gamma = -2C \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i})$$
$$= 2C\mathfrak{M}\,|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2/(1 - |Q|^2), \tag{3.44}$$

where (as discussed fully in Sec. 3.3) $1 - |Q|^2$ is the fraction of incident energy flux transmitted into

the slab region. Comparison with (3.42) gives

$$P = -2C_0 \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i}) = 2C_0 \mathfrak{M} |\mathcal{G}(\mathbf{o}, \mathbf{i})|^2/(1 - |Q|^2), \quad C_0 = C/\rho \sec \alpha. \tag{3.45}$$

Thus, essentially as for an isolated lossless scatterers, the function P is the power outflow

$$2C_0 \mathfrak{M} |\mathcal{G}(\mathbf{o}, \mathbf{i})|^2 e^{-i2\operatorname{Im}\Gamma z}$$

an element diverts from the forward traveling coherent wave (diverts to *incoherent* scattering as shown in Sec. 3.3) divided by the incident coherent forward flux per unit area $(1 - |Q|^2)e^{-2\operatorname{Im}\Gamma z}$.

Scattering theorems. In order to include absorption, and to derive additional relations for the amplitudes, we consider the elementary scattering process defined by $\mathcal{G}$, i.e., the free-space process of a single object excited simultaneously by two plane waves φ and $-Q\varphi'$. By superposition, the total field equals

$$\Phi(i) = \Phi_i + \mathcal{V}_i; \qquad \Phi_i = \varphi - Q\varphi' \equiv \varphi_i - Q_i\varphi_{i'}, \quad \mathcal{V} = v_i - Q_i v_{i'}, \tag{3.46}$$

where v_i is the field scattered by an isolated object in response to φ_i (i.e., as in Sec. 2.1), and $v_{i'}$ is the response to $\varphi_{i'}$. We have

$$\begin{aligned} \mathcal{V} &= \{\mathfrak{K}_0(k|\mathbf{r} - \mathbf{r}'|), \mathcal{V}(\mathbf{r}', \mathbf{i}')\} \sim \mathfrak{K}(kr)\mathcal{G}(\mathbf{o}, \mathbf{i}), \\ \mathcal{G}(\mathbf{o}, \mathbf{i}) &= \{\varphi_0^*(\mathbf{r}'), \mathcal{V}(\mathbf{r}', \mathbf{i})\} \\ &= \{\varphi_0^*, v_i\} - Q_i, \{\varphi_0^*, v_{i'}\} \\ &= g(\mathbf{o}, \mathbf{i}) - Q_i g(\mathbf{o}, \mathbf{i}'); \qquad \varphi_0^* = e^{-ik\mathbf{o}\cdot\mathbf{r}}. \end{aligned} \tag{3.47}$$

The total time averaged energy flow into the scatterer is proportional to $-\{\Phi^*(\mathbf{r}'), \Phi(\mathbf{r}')\}$ and the total outflow to $\{\mathcal{V}^*(\mathbf{r}'), \mathcal{V}(\mathbf{r}')\}$. We define the corresponding energy cross sections to be

$$\begin{aligned} P_a &\equiv -\{\Phi^*(\mathbf{r}'), \Phi(\mathbf{r}')\}b \\ &= -2 \operatorname{Re} \{\Phi_i^*, \mathcal{V}\}b - \{\mathcal{V}^*, \mathcal{V}\}b, \\ P_s &\equiv \{\mathcal{V}^*(\mathbf{r}'), \mathcal{V}(\mathbf{r}')\}b = 2\mathfrak{M} |\mathcal{G}(\mathbf{o}, \mathbf{i})|^2 b; \\ b &= \frac{1}{1 - |Q|^2}\left(\cos\alpha, \frac{2}{k}, \frac{2\pi}{k^2}\right) = \frac{C_0}{1 - |Q|^2}, \end{aligned} \tag{3.48}$$

where the second form of P_s follows on using Green's theorem to replace $\{\ \}$ by the surface integral at ∞, and then using the asymptotic for of $\mathcal{V}$. The total cross section is thus

$$\begin{aligned} P \equiv P_a + P_s &= -2b \operatorname{Re} \{\Phi_i^*, \mathcal{V}\} \\ &= -2b \operatorname{Re} \{(\varphi_i - Q_i\varphi_{i'})^*, \mathcal{V}(\mathbf{r}', \mathbf{i})\} \\ &= -2b \operatorname{Re} [\mathcal{G}(\mathbf{i}, \mathbf{i}) - Q_i \mathcal{G}(\mathbf{i}', \mathbf{i})]. \end{aligned} \tag{3.49}$$

From $Q_i = Q$ of (3.12) and $\mathcal{G}(\mathbf{o}, \mathbf{r})$ of (3.39) we have

$$\begin{aligned} i(\Gamma - \gamma) &= S_+ - QR_+ = C\mathcal{G}(\mathbf{i}, \mathbf{i}), \\ i(\Gamma + \gamma) &= (R/Q) - S_- = C\mathcal{G}(\mathbf{i}', \mathbf{i})/Q, \end{aligned} \tag{3.50}$$

and consequently

$$-\operatorname{Im} \Gamma/2C = \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i}) = \operatorname{Re} [\mathcal{G}(\mathbf{i}', \mathbf{i})/Q] = \operatorname{Re} [\mathcal{G}(\mathbf{i}', \mathbf{i})Q^*]/|Q|^2. \tag{3.51}$$

Eliminating $\mathcal{G}(\mathbf{i}', \mathbf{i})$ from (3.49) we obtain, in accord with (3.42),

$$P = 2b(1 - |Q|^2) \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i}) = -2C \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i})/\rho \sec \alpha. \tag{3.52}$$

[The phase component of Γ may also be interpreted on the basis of the process (3.46), the phase change of $e^{i\Gamma z}$ being determined by the free space phase factor γ plus the response of an isolated elementary layer of scatterers to $\varphi - Q\varphi'$. Both phase and amplitude effects are clear from the first-order equations (3.5); thus $\partial_z\psi_+ = (i\gamma + S_+)\psi_+ + R_+\psi_-$ states that the change in the forward free-space field ψ_+ equals the result of the transmission of the wave through an elementary layer plus the reflection of the backward free-space wave at the boundary of the layer. In terms of the field in the semi-infinite medium, we have [from (3.13)] $\psi_+ = e^{i\Gamma z}$ and $\psi_- = -Qe^{i\Gamma z}$; consequently (3.5) equals $\partial_z e^{i\Gamma z} = (i\gamma + S_+ - QR_+)e^{i\Gamma z} = [i\gamma + C\mathcal{G}(\mathbf{i}, \mathbf{i})]e^{i\Gamma z}$, which equals $i\Gamma e^{i\Gamma z}$ as required.]

The elementary process defined in (3.46) leads to additional relations that $\mathcal{G}$ fulfills. Thus if we apply Green's theorem in the external region to $\Phi^*(i)$ and $\Phi(j)$, then for lossless scatterers

$$-2\mathfrak{M}\mathcal{G}^*(\mathbf{s}, \mathbf{i})\mathcal{G}(\mathbf{s}, \mathbf{j}) = \mathcal{G}(\mathbf{i}, \mathbf{j}) - Q_i^*\mathcal{G}(\mathbf{i}', \mathbf{j}) + \mathcal{G}^*(\mathbf{j}, \mathbf{i}) - Q_j\mathcal{G}^*(\mathbf{j}', \mathbf{i}) \tag{3.53}$$

(a generalization of (3.43) which reduces to (3.43) if $Q_i = Q_j = 0$). If $j = i$, $Q_j = Q$, then (3.53) reduces to the energy theorem

$$\begin{aligned} -\mathfrak{M} |\mathcal{G}(\mathbf{o}, \mathbf{i})|^2 &= \operatorname{Re} [\mathcal{G}(\mathbf{i}, \mathbf{i}) - Q^*\mathcal{G}(\mathbf{i}', \mathbf{i})] \\ &= (1 - |Q|^2) \operatorname{Re} \mathcal{G}(\mathbf{i}, \mathbf{i}) \\ &= (1 - |Q|^2) \operatorname{Re} [\mathcal{G}(\mathbf{i}', \mathbf{i})/Q]. \end{aligned} \tag{3.54}$$

Another special case of present interest is $j = i'$, $Q_i = Q$, $Q_j = Q'$; here we use

$$\begin{aligned} -i(\Gamma' + \gamma) &= S_- - R_-Q' = C\mathcal{G}(\mathbf{i}', \mathbf{i}), \\ -i(\Gamma' - \gamma) &= (R_+/Q') - S_+ = C\mathcal{G}(\mathbf{i}, \mathbf{i}')/Q' \end{aligned} \tag{3.55}$$

together with (3.50) and (3.53) to obtain

$$-2C\mathfrak{M}\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}') = i(\Gamma^* - \Gamma')(Q' - Q^*),$$

$$C \operatorname{Re} \mathfrak{M}\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}') = \operatorname{Re} \Gamma_0 \operatorname{Im} (Q + Q') - \operatorname{Im} \Delta \operatorname{Re} (Q - Q'). \quad (3.56)$$

If the scatterers are symmetrical to the slab face, then (3.56) reduces to

$$\mathfrak{M}\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}') = -\operatorname{Re} [\mathcal{G}(\mathbf{i}, \mathbf{i}') - Q^*\mathcal{G}(\mathbf{i}, \mathbf{i})] = 2 \operatorname{Re} \Gamma \operatorname{Im} Q/C. \quad (3.57)$$

The $\mathcal{G}$'s may also be used to represent the propagation function e.g., $K^2 = k^2 - C^2\mathcal{G}(\mathbf{i}, \mathbf{i})\mathcal{G}(\mathbf{i}', \mathbf{i})/Q$, etc.

Reciprocity theorem. Green's theorem applied to $\Phi(i)$ and $\Phi(j)$ in the interior region of the scatterer gives $\{\Phi(i), \Phi(j)\} = 0$, and since $\{\Phi_i, \Phi_i\} = 0$, and $\{\mathcal{V}_i, \mathcal{V}_i\}_s = \{\mathcal{V}_i, \mathcal{V}_i\}_\infty = 0$, we obtain

$$\{(\varphi_i - Q_i\varphi_{i'})\mathcal{V}_i\} - \{(\varphi_j - Q_j\varphi_{j'})\mathcal{V}_i\} = 0.$$

Consequently

$$\mathcal{G}(-\mathbf{i}, \mathbf{j}) - Q_i\mathcal{G}(-\mathbf{i}', \mathbf{j}) = \mathcal{G}(-\mathbf{j}, \mathbf{i}) - Q_j\mathcal{G}(-\mathbf{j}', \mathbf{i}), \quad (3.58)$$

which is the extension of the reciprocity principle to a scatterer excited simultaneously by two waves. If $Q_i = Q_j = 0$, then (3.58) reduces to the usual relation for one exciting wave

$$g(-\mathbf{i}, \mathbf{j}) = g(-\mathbf{j}, \mathbf{i}). \quad (3.59)$$

Regrouping the terms of (3.58) and using the definition of $\mathcal{G}$, we obtain

$$\begin{aligned}\mathcal{G}(-\mathbf{i}, \mathbf{j}) - \mathcal{G}(-\mathbf{j}, \mathbf{i}) &= Q_i\mathcal{G}(-\mathbf{i}', \mathbf{j}) - Q_j\mathcal{G}(-\mathbf{j}', \mathbf{i}) \\ &= Q_ig(-\mathbf{i}', \mathbf{j}) - Q_jg(-\mathbf{j}', \mathbf{i}). \quad (3.60)\end{aligned}$$

Thus $\mathcal{G}$ possesses the "reciprocity in angles" shown by g in (3.59) only if the right-hand side of (3.60) vanishes, e.g., if the scatterer is symmetrical to the plane in which i and j have been imaged (i.e., $g(-\mathbf{i}', \mathbf{j}) = g(-\mathbf{i}, \mathbf{j}') = g(\mathbf{j}', -\mathbf{i})$), and provided that $Q_j = Q_i$.

For the particular values of Q_i of interest for the coherent field, we have $Q_i = Q_{-i'} = Q$, and $Q_{i'} = Q_{-i} = Q'$; in addition, $Qg(\mathbf{i}, \mathbf{i}') = Q'g(\mathbf{i}', \mathbf{i})$. Consequently for $j = -i$ in (3.60) we obtain

$$\mathcal{G}(-\mathbf{i}, -\mathbf{i}) = \mathcal{G}(\mathbf{i}, \mathbf{i}); \quad (3.61)$$

thus the case of forward scattering fulfills the elementary relation. On the other hand, the forms for $j = i'$ and $j = -i'$ do not:

$$\mathcal{G}(-\mathbf{i}, \mathbf{i}') = \mathcal{G}(-\mathbf{i}', \mathbf{i}) + Q[g(-\mathbf{i}', \mathbf{i}') - g(-\mathbf{i}, \mathbf{i})g(\mathbf{i}, \mathbf{i}')/g(\mathbf{i}', \mathbf{i})],$$

$$\mathcal{G}(-\mathbf{i}, -\mathbf{i}') = \mathcal{G}(\mathbf{i}', \mathbf{i}) + Q[g(\mathbf{i}', \mathbf{i}') - g(\mathbf{i}, \mathbf{i})]. \quad (3.62)$$

The "correction terms" in (3.62) vanish for scatterers symmetrical to the image plane, for which case [using $g(\mathbf{i}, \mathbf{i}') = g(\mathbf{i}', \mathbf{i})$], we also have

$$\mathcal{G}(\mathbf{i}', \mathbf{i}') = \mathcal{G}(\mathbf{i}, \mathbf{i}), \qquad \mathcal{G}(\mathbf{i}', \mathbf{i}) = \mathcal{G}(\mathbf{i}, \mathbf{i}'). \quad (3.63)$$

3.3. Average Intensity and Energy Flux

In order to complete the discussion of the scattering losses shown by the coherent field we derive the "energy theorem" for the total average power (the sum of the "coherent" and "incoherent" components), and show that it is satisfied by our explicit forms for the power components in terms of Γ and $\mathcal{G}$. Equivalently, our procedure demonstrates that the energy theorem for the distribution leads to the same theorems for Γ and $\mathcal{G}$ obtained by isolated scatterer considerations in Sec. 3.2 (and this is of particular interest for models[7a] not based on free-space scatterers, i.e., for alternative formalisms, the results of this section provide restrictions which help define the appropriate elementary scatterer).

Average intensity. For a single configuration, the "intensity" is $|\Psi|^2 = |\varphi + U|^2$. Its ensemble average may be written

$$\langle|\Psi|^2\rangle = |\langle\Psi\rangle|^2 + V,$$

$$V = \langle|U - \langle U\rangle|^2\rangle = \langle|U|^2\rangle - |\langle U\rangle|^2, \quad (3.64)$$

where the "incoherent intensity" V (the absolute mean square fluctuation of Ψ from its mean value $\langle\Psi\rangle$) is given by

$$\begin{aligned}V &= \iint [\sum{}' \sum \langle u_su_t^*\rangle_{st}w_sw_{st} \\ &\quad - \sum \sum \langle u_s\rangle_s\langle u_t^*\rangle_t w_sw_t] \, d\mathbf{r}_s \, d\mathbf{r}_t \\ &\quad + \sum \int \langle|u_s|^2\rangle_s w_s \, d\mathbf{r}_s \\ &= \rho^2 \iint \left[\frac{N-1}{N}\langle u_su_t^*\rangle_{st} - \langle u_s\rangle_s\langle u_t^*\rangle_t\right] d\mathbf{r}_s \, d\mathbf{r}_t \\ &\quad + \rho \int \langle|u_s|^2\rangle_s \, d\mathbf{r}_s; \quad (3.65)\end{aligned}$$

here $\langle u_s\rangle_s$ represents $\langle u(\mathbf{r} - \mathbf{r}_s)\rangle_s$, etc., N equals the number of scatterers, and we used $w_s = \rho/N$ and $w_sw_{st} = \rho^2/N^2$.

Using $(N - 1)/N \approx 1$, and

$$\langle u_su_t^*\rangle_{st} \approx \langle u_s\rangle_s\langle u_t^*\rangle_t, \quad (3.66)$$

we obtain

$$V \approx \rho \int |\langle u_s \rangle_s|^2 \, d\mathbf{r}_s \tag{3.67}$$

which is adequate for our present purposes.

The far field form of (3.67) may be written

$$V \sim \rho \int |\mathfrak{C}(k\,|\mathbf{r} - \mathbf{r}_s|)|^2 \, |G(z_s; \mathbf{s})| \, d\mathbf{r}_s,$$

$$\mathbf{s} = \frac{\mathbf{r} - \mathbf{r}_s}{|\mathbf{r} - \mathbf{r}_s|}, \tag{3.68}$$

where $|\mathfrak{C}|^2$ is given in (2.6). Introducing q as the "differential scattering cross section" of the unit volume of scatterers located at z_s, i.e.,

$$q \equiv (\cos\alpha, 2/\pi k, 1/k^2)\rho \, |G|^2, \tag{3.69}$$

and using $dx_s \, dy_s = -|\mathbf{r} - \mathbf{r}_s|^2 \sec\theta_s \, d\Omega_s$, in three dimensions, and $dx_s = -|\mathbf{r} - \mathbf{r}_s| \sec\theta_s \, d\theta_s$, in two (where $\cos\theta_s = \mathbf{s}\cdot\mathbf{z}_0$), we rewrite (3.68) inside the slab as

$$V = \left[\int_{+\frac{1}{2}} \int_0^z - \int_{-\frac{1}{2}} \int_z^d\right] q(z_s; \mathbf{s}) \frac{d\Omega_s}{\cos\theta_s} \, dz_s$$

$$\equiv V_+(0, z) + V_-(z, d), \tag{3.70}$$

where the integrals over the forward and back half-spaces of angles in three dimensions have Ω_s replaced by θ_s in two dimensions, and are to be replaced by the two values $\theta_s = \alpha, \pi - \alpha$ in one dimension. Outside the slab, we have

$$\begin{aligned} V &= V_+(0, d), \qquad z > d; \\ V &= V_-(0, d), \qquad z < d, \end{aligned} \tag{3.71}$$

which are the variances of the coherent transmitted and reflected fields, respectively. We may substitute the explicit forms for q in the above and integrate directly over z_s.

The average energy flux. For a single configuration, the total time averaged energy flux per unit area divided by the time-averaged incident flux density equals $\mathbf{J} = \mathrm{Re}\,[\Psi^* \nabla \Psi/ik]$. Using $\nabla\Psi = \nabla\varphi + \nabla \sum u_s = i\mathbf{k}\varphi + \sum \nabla_s u_s$, and $\langle \sum \nabla_s u_s \rangle = \nabla \sum \langle u_s \rangle = \nabla\langle U \rangle$, we obtain the ensemble average

$$\langle \mathbf{J} \rangle = \mathrm{Re}\,[\mathbf{i} + \varphi^*(\mathbf{i} + \nabla/ik)\langle U \rangle$$

$$+ \sum \sum \langle u_t^* \nabla_s u_s/ik \rangle] \equiv \mathbf{C} + \mathbf{I}. \tag{3.72}$$

Here $\mathbf{C}$, the "coherent flux", equals

$$\mathbf{C} = \mathrm{Re}\,[\langle \Psi \rangle^* \nabla \langle \Psi \rangle/ik], \tag{3.73}$$

and $\mathbf{I}$, the "incoherent flux", is analogous to V of (3.65).

If the scatterers are lossless, then Green's theorem applied to Ψ^* and Ψ gives $\oint \mathbf{J}\cdot d\mathbf{A} = 0$ for a fixed configuration; consequently $\oint \langle \mathbf{J} \rangle \cdot d\mathbf{A} = 0$ (the time averaged, ensemble averaged energy flow through any closed surface equals zero). It then follows from Gauss' theorem that $\nabla\cdot\langle \mathbf{J} \rangle = 0$, and since $\partial_x \langle \mathbf{J} \rangle \cdot \mathbf{x}_0$ and $\partial_y \langle \mathbf{J} \rangle \cdot \mathbf{y}_0$ vanish because of symmetry, we obtain $\partial_z \langle \mathbf{J} \rangle \cdot \mathbf{z}_0 = 0$. Consequently

$$\langle \mathbf{J} \rangle \cdot \mathbf{z}_0 = (\mathbf{C} + \mathbf{I}) \cdot \mathbf{z}_0 = J_0 = \text{const.} \tag{3.74}$$

This energy theorem (that the component of the average flux normal to a slab region of nonabsorbing scatterers does not vary along the normal) is of course a consequence of the symmetry of the infinite slab excited by a plane wave; there are no scattering energy losses in any real sense for this problem, there being no place for energy to get lost. However, it is convenient to describe the history of the "beam" whose cross section is unit area of incident wave front, and the discussion of the corresponding coherent and incoherent components of the flux facilitates identifying the "losses" of the coherent flux indicated by the attenuation coefficient 2 Im Γ with the corresponding incoherent flux. (In a practical situation involving a bounded incident beam, the scattering losses along the coherent beam also appear in other directions as the incoherent flux, and even when $\mathbf{I}$ and $\mathbf{C}$ overlap in space, they can be "separated" by appropriate measurement procedures—at least in microwave experiments.)

In order to trace the energy loss of the coherent field we use

$$\mathbf{I} \approx \sum \langle u_s \rangle^* \langle \nabla u_s/ik \rangle \sim \sum |\langle u_s \rangle|^2 \mathbf{s} \equiv V(\mathbf{s}) \tag{3.75}$$

where $V(\mathbf{s})$ indicates the insertion of

$$\mathbf{s} = (\mathbf{r} - \mathbf{r}_s)/|\mathbf{r} - \mathbf{r}_s|$$

inside the integrals of (3.68) and (3.70). The asymptotic form suffices, since the coherent power loss via scattering may radiate directly to infinity as far as the effects shown by Γ are concerned. Physically speaking, any rescattering of the incoherent radiation cannot affect the coherent field: although the incoherent flux contributes in the directions in which the coherent flux is observed, it cannot become "coherent."

Using (3.73), (3.74), and (3.75), we obtain for $z < 0$,

$$(\mathbf{C} + \mathbf{I}) \cdot \mathbf{z}_0 = \mathrm{Re}\,[\mathbf{i} + \varphi^* \langle U_< \rangle (\mathbf{i} + \mathbf{i}') + \mathbf{i}' \, |\langle U_< \rangle|^2] \cdot \mathbf{z}_0$$

$$+ V_-(\mathbf{s}\cdot\mathbf{z}_0) = (1 - |\mathcal{R}|^2)\cos\alpha + V_-(\cos\theta_s), \tag{3.76}$$

where the "reflection coefficient" $|\mathcal{R}|^2 = |\Psi_R|^2$ is given in (3.14), and where

$$V_-(\cos\theta_s) = -\int_{-1/2} \int_0^d q(z_s; \mathbf{s}) \, d\Omega \, dz_s. \tag{3.77}$$

Similarly, for $z > d$, we obtain

$$(\mathbf{C} + \mathbf{I})\cdot\mathbf{z}_0 = |\mathfrak{I}|^2 \cos\alpha + V_+(\cos\theta_s), \quad (3.78)$$

where $|\mathfrak{I}|^2 = |\Psi_T|^2$ is given in (3.14), and where

$$V_+(\cos\theta_s) = \int_{1/2}\int_0^d q(z_s; \mathbf{s})\, d\Omega\, dz_s. \quad (3.79)$$

Using (3.74), we equate (3.76) to (3.78) and obtain

$$(1 - |\mathfrak{R}|^2)\cos\alpha + V_-(\cos\theta_s) = |\mathfrak{I}|^2\cos\alpha + V_+(\cos\theta_s) = \text{const}, \quad (3.80)$$

and consequently

$$1 - |\mathfrak{R}|^2 - |\mathfrak{I}|^2 = \sec\alpha[V_+(\cos\theta_s) - V_-(\cos\theta_s)] = \sec\alpha\int_0^d \mathfrak{Q}(z_s; \mathbf{s})\, dz_s, \quad (3.81)$$

where

$$\mathfrak{Q}(\zeta) = \oint q(\zeta; \mathbf{s})\, d\Omega_s = 2C\Re\, |G(\zeta; \mathbf{s})|^2 \cos\alpha = 2C\,|D|^2\,\Re\,|\mathcal{G}(\mathbf{s}, \mathbf{i})e^{i\Gamma\zeta} + Q\mathcal{G}(\mathbf{s}, \mathbf{i}')e^{i\Gamma'\zeta + id(\Gamma - \Gamma')}|^2\cos\alpha \quad (3.82)$$

is the average total scattering cross section for a unit volume at a distance ζ within the slab. Thus the energy diverted from unit area of incident wave front by coherent reflection and transmission equals the total energy incoherently scattered by the elements irradiated by unit area of wave front. [Redefining $\mathfrak{Q}$ as the sum of absorption plus scattering cross sections, we extend (3.81) directly to lossy scatterers.]

In order to show that our explicit forms for $|\mathfrak{R}|^2$, $|\mathfrak{I}|^2$, and q fulfill (3.82), we use essentially (3.54) and (3.56) to reduce (3.82):

$$|D|^2\{(1 - |Q|^2)2\,\mathrm{Im}\,\Gamma e^{-2\,\mathrm{Im}\,\Gamma\zeta} - (1 - |Q'|^2)\,|Q|^2\,2\,\mathrm{Im}\,\Gamma' e^{-2\,\mathrm{Im}\,\Gamma'\zeta - 4\,\mathrm{Im}\,\Gamma_0 d} - 2\,\mathrm{Re}\,[i(\Gamma^* - \Gamma')Q(Q' - Q^*)e^{-i(\Gamma^* - \Gamma')\zeta + i2\Gamma_0 d}]\} = \mathfrak{Q}/\cos\alpha;$$

integrating the result over ζ gives

$$\int_0^d \frac{\mathfrak{Q}}{\cos\alpha}\, d\zeta = -|D|^2(1 - |Q|^2)[e^{-2\,\mathrm{Im}\,\Gamma d} - 1] + |D|^2(1 - |Q'|^2)\,|Q|^2[e^{-2\,\mathrm{Im}\,\Gamma d} - e^{-4\,\mathrm{Im}\,\Gamma_0 d}] + 2\,\mathrm{Re}\,|D|^2\,Q(Q' - Q^*)(e^{-2\,\mathrm{Im}\,\Gamma d} - e^{i2\Gamma_0 d})]$$

which equals $1 - |\mathfrak{R}|^2 - |\mathfrak{I}|^2$ as obtained from (3.14).

The normal flux inside the slab region also equals (3.74), i.e.,

$$\langle\mathbf{J}_{\mathrm{in}}\rangle\cdot\mathbf{z}_0 \equiv |\mathcal{g}(z)|^2\cos\alpha + V(z; \cos\theta_s) \equiv J_0, \quad (3.83)$$

and since the incoherent component is continuous at $z = 0, d$ it follows that the coherent component is also continuous:

$$(1 - |\mathfrak{R}|^2) = |\mathcal{g}(0)|^2, \qquad |\mathcal{g}(d)|^2 = |\mathfrak{I}|^2. \quad (3.84)$$

Thus from the average boundary conditions (3.26) and (3.25), we have

$$|\mathcal{g}(z)|^2 = \mathrm{Re}\,[\Psi_I^*(B\,\partial_z\Psi_I + B'\Psi_I)/ik\cos\alpha], \quad (3.85)$$

where Ψ_I is defined in (3.14), and B and B' are given in (3.27). In particular, for scatterers symmetrical to the interfaces, the internal coherent field equals

$$\mathbf{C}_{\mathrm{in}} = \mathrm{Re}\,(\Psi_I^* B\nabla\Psi_I/ik) = |\Psi_I|^2\,\mathrm{Re}\,(B\mathbf{K}/k),$$

$$\mathrm{Re}\,(B\mathbf{K}/k) = \mathbf{z}_0\,\mathrm{Re}\,Z\cos\alpha + \mathbf{x}_0\,\mathrm{Re}\,B\sin\alpha, \quad (3.85')$$

where

$$\mathrm{Re}\,Z = \mathrm{Re}\,\frac{1 + Q}{1 - Q} = \frac{1 - |Q|^2}{|1 - Q|^2}$$

and B^{-1} equals ϵ or μ. The direction of coherent power flow is defined by the unit vector

$$\mathrm{Re}\,(B\mathbf{K}/k)\cos\nu/\mathrm{Re}\,Z\cos\alpha$$

such that $\tan\nu = \mathrm{Re}\,B\tan\alpha/\mathrm{Re}\,Z$.

We use (3.83) to trace the energy loss of the internal coherent field. Writing

$$|\mathcal{g}(z_2)|^2 - |\mathcal{g}(z_1)|^2 = \Delta\,|\mathcal{g}|^2 = -\sec\alpha\Delta V(z; \cos\theta_s),$$

and using

$$V(z; \cos\theta_s) = V_+(0, z; \cos\theta_s) + V_-(z, d; \cos\theta_s),$$

etc., [where $0, z$ in V_+ and z, d in V_- indicate the limits on the integrals in (3.79) and (3.77), respectively], we obtain

$$\Delta V = V_+(z_1, z_2; \cos\theta_s) - V_-(z_1, z_2; \cos\theta_s).$$

Thus [essentially as for (3.81)] we have

$$|\mathcal{g}(z_2)|^2 - |\mathcal{g}(z_1)|^2 = -\sec\alpha\int_{z_1}^{z_2}\mathfrak{Q}(\zeta; \mathbf{i})\, d\zeta, \quad (3.86)$$

i.e., the energy loss from unit area of the coherent field in traversing the thickness of distribution $z_2 - z_1$ equals the total incoherent scattering of the scatterers in the volume $1 \times \sec\alpha \times z_2 - z_1$. If $z_1 = 0$, $z_2 = d$, then (3.86) reduces to (2.81). Essentially as for (3.81), we may verify directly that our explicit forms for $\mathcal{g}$ and $\mathfrak{Q}$ fulfill (3.86).

The explicit forms also determine the constant J_0 in (3.74), (3.80), and (3.83). Using (3.83) with the two components of $V = V_+(0, z) + V_-(z, d)$

written in the form $V_+(0, z) = V_+(0, \ ) + V_+(\ , z)$, $V_-(z, d) = V_-(\ , d) + V_-(z, \ )$ (corresponding to the terms of the evaluated integrals arising from the upper and lower limits), we see that

$$\cos\alpha\,|g(z)|^2 + V_+(\ , z) + V_-(z, \ ) = 0.$$

Thus

$$\begin{aligned} J_0 &= V_+(0, \ ) + V_-(\ , d) = |D|^2\, C\cos\alpha \\ &\times \{\mathfrak{M}_{\frac{1}{2}}[f_1 + f_2 e^{-4\,\mathrm{Im}\,\Gamma_0 d} + \mathrm{Re}\,(f_3 e^{i2\Gamma_0 d})] \\ &+ \mathfrak{M}_{-\frac{1}{2}}[f_1 + f_2 + \mathrm{Re}\,f_3]e^{-2\,\mathrm{Im}\,\Gamma d}\}, \qquad (3.87) \\ f_1 &\equiv \frac{|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2}{\mathrm{Im}\,\Gamma}, \qquad f_2 \equiv \frac{|Q\mathcal{G}(\mathbf{o}, \mathbf{i}')|^2}{\mathrm{Im}\,\Gamma'}, \\ f_3 &\equiv \frac{4\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}')Q}{i(\Gamma^* - \Gamma')}, \end{aligned}$$

where $\mathfrak{M}_{1/2}$ and $\mathfrak{M}_{-1/2}$ represent the normalized integrals over the forward and back angular regions; here we may eliminate Im Γ, etc. From (3.87) and (3.80) it follows that

$$\begin{aligned} |\mathfrak{I}|^2 &= \frac{[V_+(0, \ ) + V_-(\ , d) - V_+(0, d)]}{\cos\alpha} \\ &= \frac{[V_-(\ , d) - V_+(\ , d)]}{\cos\alpha} \\ &= |D|^2\, C\mathfrak{M}[f_1 + f_2 + \mathrm{Re}\,f_3]e^{-2\,\mathrm{Im}\,\Gamma d}, \qquad (3.88) \\ 1 - |\mathfrak{R}|^2 &= \frac{[V_+(0, \ ) + V_-(\ , d) - V_-(0, d)]}{\cos\alpha} \\ &= \frac{[V_+(0, \ ) - V_-(0, \ )]}{\cos\alpha} \\ &= |D|^2\, C\mathfrak{M}[f_1 + f_2 e^{-4\,\mathrm{Im}\,\Gamma_0 d} + \mathrm{Re}\,f_3 e^{-2\,\mathrm{Im}\,\Gamma_0 d}], \end{aligned}$$

which may be verified by using (3.54) and (3.52), and $|\mathfrak{I}|^2$ and $|\mathfrak{R}|^2$ of (3.14). Thus we have related $|\mathfrak{I}|^2$ to the total incoherent scattering of a layer at the slab face d, and $(1 - |\mathfrak{R}|^2)$ to that of a layer at the slab face 0.

For $d \to \infty$, Eq. (3.88) reduces to

$$\begin{aligned} J_0 &= (C\cos\alpha/\mathrm{Im}\,\Gamma)\mathfrak{M}_{1/2}\,|\mathcal{G}\,\mathbf{o}, \mathbf{i})|^2, \\ 1 - |\mathfrak{R}|^2 &= C\mathfrak{M}\,|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2/\mathrm{Im}\,\Gamma = 1 - |Q|^2 \\ &= |1 - Q|^2\,\mathrm{Re}\,Z. \qquad (3.89) \end{aligned}$$

3.4. Slab Distribution on a Ground Plane

Using the image technique we now apply the above to treat a slab distribution of scatterers (symmetrical to the interfaces) backed by a "ground plane."

For a plane wave φ incident on a slab region $-d \le z \le d$ of scatterers symmetrical to the interfaces [such that $g(\mathbf{o}, \mathbf{i}) = g(\mathbf{o}', \mathbf{i}')$], we see that (3.13), (3.14), and (3.39) reduce to

$$\begin{aligned} \psi_+(\mathbf{i}) &= De^{i(\Gamma-\gamma)d}[e^{i\Gamma z} - Q^2 e^{i\Gamma(2d-z)}], \\ \psi_-(\mathbf{i}) &= -De^{i(\Gamma-\gamma)d}Q[e^{i\Gamma z} - e^{i\Gamma(2d-z)}], \\ \Psi_I &= D(1-Q)e^{i(\Gamma-\gamma)}[e^{i\Gamma z} + Qe^{i\Gamma(2d-z)}], \qquad (3.90) \\ D &= [1 - Q^2 e^{i4\Gamma d}]^{-1}, \\ \Psi_R(\mathbf{i}) &= -Qe^{-i2\gamma d}(1 - e^{i4\Gamma d})\,D\varphi', \\ \Psi_T(\mathbf{i}) &= (1 - Q^2)e^{i(\Gamma-\gamma)2d}D\varphi, \\ G(z; \mathbf{o}, \mathbf{i}) &= g(\mathbf{o}, \mathbf{i})\psi_+(\mathbf{i}) + g(\mathbf{o}, \mathbf{i}')\psi_-(\mathbf{i}) \\ &= De^{i(\Gamma-\gamma)d}[\mathcal{G}(\mathbf{o}, \mathbf{i})e^{i\Gamma z} + \mathcal{G}(\mathbf{o}, \mathbf{i}')Qe^{i\Gamma(2d-z)}]. \qquad (3.91) \end{aligned}$$

Similarly for an incident wave φ' (the image of φ) we interchange z and $-z$ in the above; we indicate this by

$$\Psi(\mathbf{i}'; z, -z) = \Psi(\mathbf{i}; -z, z). \qquad (3.92)$$

The corresponding multiple scattered amplitude equals

$$\begin{aligned} G(z; \mathbf{o}, \mathbf{i}') &= g(\mathbf{o}, \mathbf{i}')\psi_+(\mathbf{i}') + g(\mathbf{o}, \mathbf{i})\psi_-(\mathbf{i}') \\ &= De^{i(\Gamma-\gamma)d}[\mathcal{G}(\mathbf{o}, \mathbf{i}')e^{-i\Gamma z} + Q\mathcal{G}(\mathbf{o}, \mathbf{i})e^{i\Gamma z + i2\Gamma d}]. \qquad (3.93) \end{aligned}$$

The sum and difference of the above sets, say $\Phi^{\pm} = \Psi(\mathbf{i}) \pm \Psi(\mathbf{i}')$, are the required functions in the region $z < 0$ for φ incident on the slab region $-d$ to 0 backed by a "ground plane" at $z = 0$. The function Φ^+ (such that $\partial_z\Phi^+ = 0$ at $z = 0$) applies for a rigid plane in acoustics, or for a perfect conductor and **E** in the plane of incidence in electromagnetics; similarly Φ^- (which vanishes at $z = 0$) applies for a free plane in acoustics, or for a perfect conductor and **E** perpendicular to the plane of incidence. Thus

$$\begin{aligned} \Phi_I^{\pm} &= \Psi_I(\mathbf{i}) \pm \Psi_I(\mathbf{i}') = (1 - Q)[e^{i\Gamma z} \pm e^{-i\Gamma z}]\,D_{\pm}, \\ D_{\pm} &\equiv [1 \mp Qe^{i2\Gamma d}]^{-1}, \qquad (3.94) \\ \Phi_R^{\pm} &= \Psi_R(\mathbf{i}) \pm \Psi_T(\mathbf{i}') = -e^{-i2\gamma d}(Q \mp e^{i2\Gamma d})\,D_{\pm}. \end{aligned}$$

The corresponding multiple scattered amplitude is

$$\begin{aligned} F_{\pm}(z; \mathbf{o}, \mathbf{i}) &= G(z; \mathbf{o}, \mathbf{i}) \pm G(z; \mathbf{o}, \mathbf{i}') \\ &= e^{i(\Gamma-\gamma)d}[\mathcal{G}(\mathbf{o}, \mathbf{i})e^{i\Gamma z} \pm \mathcal{G}(\mathbf{o}, \mathbf{i}')e^{-i\Gamma z}]\,D_{\pm}. \qquad (3.95) \end{aligned}$$

The flux corresponding to (3.70), etc., is obtained in terms of $|F|^2$ integrated from $-d$ to $+d$, i.e., we require the intensity contributions of the scatterers in the range $-d$ to 0 plus the contributions of those in the image range 0 to d. The appropriate energy theorems follow by inspection of the previous

results. In particular, the incident, coherent-reflected, and incoherent scattered fluxes for lossless scatterers are related by

$$1 - |\Phi_R^{\pm}| = \sec\alpha \int_{-1/2} d\Omega \int_{-d}^{d} q_{\pm}(\zeta; \mathbf{s})\, d\zeta,$$

$$q_{\pm}(\zeta; \mathbf{s}) = \left(\cos\alpha, \frac{2}{\pi k}, \frac{1}{k^2}\right)\rho\, |F_{\pm}|^2. \tag{3.96}$$

In order to demonstrate that this relation holds for the present explicit forms, we first integrate over ζ:

$$\int_{-1/2} d\Omega \int_{-d}^{d} q_{\pm}\, d\zeta \sec\alpha$$

$$= 2C\mathfrak{M}_{-1/2} \int_{-d}^{d} |F_{\pm}|^2\, d\zeta e^{-2\,\mathrm{Im}\,\Gamma d}\, |D_{\pm}|^2$$

$$= C\mathfrak{M}_{-1/2}\left\{\frac{[|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2 + |\mathcal{G}(\mathbf{o}, \mathbf{i}')|^2]}{\mathrm{Im}\,\Gamma}\,[1 - e^{-4\,\mathrm{Im}\,\Gamma d}]\right.$$

$$\pm \frac{\mathcal{G}(\mathbf{o}, \mathbf{i})\mathcal{G}^*(\mathbf{o}, \mathbf{i}') + \mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}')}{\mathrm{Re}\,\Gamma}$$

$$\left.\times\, 2\,\mathrm{Im}\, e^{-2i\Gamma^* d}\right\} |D_{\pm}|^2. \tag{3.97}$$

Exploiting the symmetry of the scatterers with respect to the interfaces, i.e.,

$$\mathfrak{M}_{-1/2}\, |\mathcal{G}(\mathbf{o}, \mathbf{i}')|^2 = \mathfrak{M}_{1/2}\, |\mathcal{G}(\mathbf{o}, \mathbf{i})|^2,$$

$$\mathfrak{M}_{-1/2}\mathcal{G}(\mathbf{o}, \mathbf{i})\mathcal{G}^*(\mathbf{o}, \mathbf{i}') = \mathfrak{M}_{1/2}\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}'),$$

and using (3.44) and (3.57), reduces (3.97) to

$$C\mathfrak{M}\left\{\frac{|\mathcal{G}(\mathbf{o}, \mathbf{i})|^2\,[1 - e^{-4\,\mathrm{Im}\,\Gamma d}]}{\mathrm{Im}\,\Gamma}\right.$$

$$\left.\pm \frac{\mathcal{G}^*(\mathbf{o}, \mathbf{i})\mathcal{G}(\mathbf{o}, \mathbf{i}')}{\mathrm{Re}\,\Gamma}\, 2\,\mathrm{Im}\, e^{-i2\Gamma^* d}\right\} |D_{\pm}|^2$$

$$= [(1 - |Q|^2)[1 - e^{-4\,\mathrm{Im}\,\Gamma d}]$$

$$\mp 4\,\mathrm{Im}\, Q\,\mathrm{Im}\, e^{i2\Gamma d}]\, |D_{\pm}|^2,$$

which equals $1 - |\Phi_R|^2$ [obtained from (3.94)] as required.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

On a General Class of Scattering Problems*

VICTOR TWERSKY
Sylvania Electronic Defense Laboratories, Mountain View, California
(Received December 22, 1960)

This paper considers a class of scattering problems corresponding to a wave of propagation number K_1 exciting an object characterized by K_2, and giving rise to a scattered field of propagation number K_3. In addition, the boundary conditions on the wave functions and their normal derivatives at the object involve three sets of two parameters associated with the three propagation numbers. This generalized class of scattering problems can be treated by conventional analytical procedures (as for the usual problems $K_1 = K_3$, etc.) to obtain series solutions for symmetrical shapes and Green's function representations for general shapes. The numerators of the new series coefficients differ from the usual ones, and the Green's function representation has an additional inhomogeneous term, i.e., a volume integral containing the source term in its kernel. These results are applied for special cases to obtain explicit approximations for the "two-external-space" scattering amplitude $g(K_3, K_1)$ (e.g., for small spheres with three sets of ϵ's and μ's, and for large tenuous scatterers).

1. INTRODUCTION

THE usual scattering problems of the reduced wave equations correspond to a wave of propagation number K_1 exciting an object characterized by K_2 subject to boundary conditions on the object involving two sets of two parameters, and subject to radiation conditions at infinity. The class of problems we now consider is more general in that three sets of parameters are required. (Loosely speaking, we consider scattering problems in which the incident and scattered waves travel in different all pervading "exterior spaces" K_1 and K_3.) Solutions and associated scattering amplitudes $g(\mathbf{K}_3, \mathbf{K}_1)$ are derived by conventional analytical procedures. Elementary series and Green's function representations are given for one-, two-, and three-dimensional scalar problems, and for the three-dimensional electromagnetic case.

Such "two-space" scatterers are of physical interest, for example, in analyzing multiple scattering of waves by random distributions of scatterers.[1] In such problems, the ensemble averaged field at a scatterer fixed within the distribution may be approximated as the sum of the "effective field" plus the field radiated by the fixed scatterer: the effective field which excites the scatterer travels in the "synthetic medium" (say, K_1) associated with coherent propagation, but the fixed scatterer is constrained to radiate into free space (say, K_3).

* This work was partially supported by Signal Corps Contract DA 36-039 SC-75012.

[1] The explicit results derived in the present paper are applied in detail in V. Twersky, J. Math. Phys. **3**, 724 (1962), this issue.

2. SCALAR PROBLEMS

2.1. Statement of the Problem

We seek a field which in the volume V bounded by S is specified by a function ψ such that

$$(\nabla^2 + K_2^2)\psi = 0,$$

$$\nabla^2 = \partial_x^2, \qquad \partial_x^2 + \partial_y^2, \qquad \partial_x^2 + \partial_y^2 + \partial_z^2. \tag{1}$$

External to V, the field is specified by two functions ϕ and u which fulfill

$$(\nabla^2 + K_1^2)\varphi = 0, \qquad (\nabla^2 + K_3^2)u = 0. \tag{2}$$

On S, the functions and their normal derivatives are related by boundary conditions of the form

$$\begin{aligned} &A\varphi + u = A'\psi; \quad A = A_1/A_3, \quad A' = A_2/A_3. \\ &B\,\partial_n\varphi + u = B'\,\partial_n\psi; \quad B = B_1/B_3, \quad B' = B_2/B_3. \end{aligned} \tag{3}$$

The "interior field" ψ is to be nonsingular; for convenience, we use $K_2 = K' = k\eta'$. The function φ, the source term, is taken as the plane wave

$$\begin{aligned} &\varphi = e^{i\mathbf{K}\cdot\mathbf{r}}, \qquad \operatorname{Im} K > 0; \\ &\mathbf{K} = K\mathbf{i} = k\eta\mathbf{i}, \qquad \mathbf{r} = r\mathbf{o}. \end{aligned} \tag{4}$$

The remaining function u is specified as the scattered wave by the Sommerfeld radiation condition at infinity (with respect to some interior point of V):

$$\begin{aligned} &\lim_{r\to\infty} r^{(n-1)/2}(\partial_r u - iK_3 u) = 0, \\ &K_3 = k = |k| = 2\pi/\lambda, \end{aligned} \tag{5}$$

where $n = 1, 2, 3$ gives the dimensionality of the

problem, and where we have assumed that the scatterer radiates into free space. (Although superfluous,[2] an additional condition at infinity for $n = 3, 2$ obviates detailed manipulations with surface integral forms; i.e., for $r \to \infty$, it is convenient to regard u bounded as $r^{(n-1)/2}u < M_n$, where the M's are constants.)

In the above, the wave function $\psi e^{-i\omega t}$ corresponds to such physical fields as the velocity potential in acoustics, the probability amplitude in quantum-mechanics, or the z component of **E** or **H** in one- or two-dimensional electromagnetics. The parameters A, B, K, are determined by the physical constants of the appropriate problem, and the boundary conditions follow from the requirement that the physically observable fields or certain components, be continuous (e.g., velocity, pressure, tangential components of **E**, etc.). In general, we regard the parameters having different subscripts as specifying three different physical media. Thus in acoustics $A_n = \rho_n$, $B_n = 1$, $K_n^2 = \omega^2\rho_n/(\nu_n - i\omega\xi_n)$, where ρ, ν, and ξ are, respectively, the density, compressibility, and the compressive viscosity. Similarly in one- and two-dimensional electromagnetics, $A_n = 1$, $B_{nE} = 1/\mu_n$, $B_{nH} = 1/\epsilon_n$, $K_n^2 = \omega^2\mu_n(\bar{\epsilon}_n + i\sigma_n/\omega) \equiv \omega^2\mu_n\epsilon_n$, where the subscripts E and H correspond, respectively, to $E_z = \psi$ (TM case), and $H_z = \psi$ (TE case); here μ, ϵ, and σ are, respectively, the permeability, permittivity, and conductivity. Finally in scalar wave mechanics, ψ is the probability amplitude, and $A_n = B_n = 1$.

One view of the situation governed by Eqs. (1) to (5) is implicit in the last paragraph. We may visualize an object "made up" of K_2 space excited by a field traveling in K_1 space, and scattering a wave which then travels in K_3 space. The monopole form of this type of "schizoid scatterer" is implicit in the usual Green's function representation for the conventional field scattered by a homogeneous body for the case $K_1 = K_3 = k$, $A_n = B_n = 1$: When we write the usual scattered wave as a volume integral proportional to

$$\int (e^{ik|\mathbf{r}-\mathbf{r}'|}/|\mathbf{r} - \mathbf{r}'|)\ \psi(K_2;\mathbf{r}')\ dV(\mathbf{r}')$$

and interpret the kernel as the field of a monopole located at $\mathbf{r}'$ excited by the local field at $\mathbf{r}'$, we recognize that although the exciting field $\psi(K_2;\ \mathbf{r}')$ travels in K_2 space, the monopole scatters into k space.

[2] See discussion and references on p. 192 of A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press Inc., N. Y., 1949).

In the following, we first obtain explicit solutions for elementary cases which may be treated by separating variables. Then we consider the Green's function representations for the general case. For simplicity we restrict discussion to the case where all parameters A, B, and K are constants.

2.2. Elementary Illustrations

The derivations in this section are conventional, but the explicit results for the scattering coefficients they lead to differ from those of the usual problem. The conventional steps are sketched in order to delineate where the differences arise.

Plane slab: We consider the one-dimensional problem of a plane wave $\varphi = e^{iKx}$ incident on a slab specified by K' and bounded by the planes $x = \pm a$. The scattered wave u is to travel in k space, i.e., $u \propto e^{ik|x|}$. The internal field consists of terms proportional to $e^{iK'x}$ and $e^{-iK'x}$. Thus, in the three regions of space, the field is specified by

$$\begin{aligned} x < -a: &\quad e^{iKx}, \quad S_-e^{-ikx}; \\ -a < x < a: &\quad I_+e^{iK'x} + I_-e^{-iK'x}; \\ x > a: &\quad e^{iKx}, \quad S_+e^{ikx}. \end{aligned} \tag{6}$$

The unknowns are determined by the four equations obtained on applying the boundary conditions (3) at the two interfaces $x = \pm a$, i.e., from

$$\begin{aligned} Ae^{\mp iKa} + S_\pm e^{ika} &= A'(I_+e^{\pm iK'a} + I_-e^{\mp iK'a}), \\ KBe^{\mp iKa} \pm kS_\pm e^{ika} &= K'B'(I_+e^{\pm iK'a} - I_-e^{\mp iK'a}). \end{aligned} \tag{7}$$

In terms of the "impedances"

$$\begin{aligned} Z &= KB/kA = K_1A_3B_1/K_3A_1B_3, \\ Z' &= K'B'/kA' = K_2A_3B_2/K_3A_2B_3, \end{aligned} \tag{8}$$

we solve (7) for the amplitudes I_+ and I_- of the internal field:

$$\begin{aligned} (\Delta A'/A)I_\pm &= (1 + Z')(1 \pm Z)e^{-i(K' \pm K)a} \\ &\quad - (1 - Z')(1 \mp Z)e^{i(K' \pm K)a}, \\ \Delta &\equiv (1 + Z')^2e^{-i2K'a} - (1 - Z')^2e^{i2K'a}. \end{aligned} \tag{9}$$

Similarly, the back-scattered amplitude S_- (i.e., the reflection coefficient), and the forward-scattered amplitude S_+ are given by

$$\begin{aligned} (\Delta e^{ika}/A)S_\pm &= e^{\mp iKa}(1 \pm Z)2Z' \\ &\quad + e^{\pm i(K+2K')a}(1 \mp Z')(Z - Z') \\ &\quad - e^{\pm i(K-2K')a}(1 \pm Z')(Z + Z') \end{aligned} \tag{10}$$

which reduce to the standard form if $K = k$ and $Z = 1$. The denominator Δ of $I_\pm$ and $S_\pm$ (which

represents multiple reflection within the slab K') is the same as for the standard problem; however, the numerators are significantly different. For a perfect reflector, such that $A'\psi = 0$ at $x = -a$, we have $Z' \to \infty$, and (10) reduces to $S_- = -Ae^{-i(K+k)a}$; similarly, for $B'\,\partial_n\psi = 0$, we have $Z' \to 0$ and (9) reduces to $S_- = (KB/k)e^{-i(K+k)a}$.

The above results for normal incidence may be directly extended to an arbitrary angle of incidence θ_1 (measured from the surface normal) by replacing K_n^2 by $K_n^2 \cos^2\theta_n$, and multiplying all wave functions by $e^{iK_3 y \sin\theta_3}$, such that

$$K_1 \sin\theta_1 = K_2 \sin\theta_2 = K_3 \sin\theta_3. \tag{11}$$

Equation (11) is a generalization of the usual "laws" of specular reflection, transmission, and refraction. In the present case both the reflected and forward scattered directions (although still images of each other in the slab face) are "broken" with respect to the direction of incidence, i.e., we have $\sin\theta_3 = (K_1/K_3)\sin\theta_1$, so that the reflected and incident directions are not images in the slab normal (not even for a perfect reflector). Similarly the internal field is refracted both with respect to the direction of incidence, and the directions of scattering; thus

$$\sin\theta_2 = (K_1/K_2)\sin\theta_1 = (K_3/K_2)\sin\theta_3.$$

The circular cylinder: For the circular cylinder of radius a, proceeding essentially as above, we expand the wave functions in cylindrical coordinates around the cylinder's axis:

$$\begin{aligned}\varphi(K) &= e^{iKx} = \sum_{n=-\infty}^{\infty} i^n J_n(Kr)e^{in\theta},\\ u(k) &= \sum i^n C_n H_n(kr)e^{in\theta},\\ \psi(K') &= \sum i^n D_n J_n(K'r)e^{in\theta}.\end{aligned} \tag{12}$$

The nonsingular internal and source waves involve Bessel functions J_n, the scattered wave involves Hankel functions of the first kind $H_n = H_n^{(1)}$ to satisfy the radiation condition, and n is an integer in these Fourier-Bessel series because the field must be single valued in θ.

Applying the boundary conditions (3) at $r = a$, and using the orthogonality property of the exponential to equate the coefficients of the resultant Fourier series individually to zero, we obtain

$$\begin{aligned}AJ_n(Ka) + C_nH_n(ka) &= A'\,D_nJ_n(K'a),\\ BKJ_n'(Ka) + kC_nH_n'(ka) &= B'K'J_n'(K'a),\end{aligned} \tag{13}$$

where the prime on the cylindrical functions represents differentiation with respect to argument. Thus,

$$\begin{aligned}D_n &= A[J_n(Ka)H_n'(ka) - ZJ_n'(Ka)H_n(ka)]/A'\Delta_n,\\ \Delta_n &\equiv J_n(K'a)H_n'(ka) - Z'J_n'(K'a)H_n(ka);\\ C_n &= A[Z'J_n(Ka)J_n'(K'a) - ZJ_n'(Ka)J_n(K'a)]/\Delta_n,\end{aligned} \tag{14}$$

where the Z's are defined in (8). As for the slab, only the denominators are the same as for the standard problem.

If $A'\psi = 0$ at $r = a$, then (14) reduces to

$$C_n = -AJ_n(Ka)/H_n(ka); \tag{15}$$

in particular, for the electromagnetic case $\psi = E_z$ and a perfect conductor, we have $A = 1$. Similarly if $B'\partial_n\psi = 0$ at $r = a$, then

$$C_n = -AZJ_n'(Ka)/H_n'(ka),\quad AZ = KB/k = \eta B; \tag{16}$$

for $\psi = H_z$ and a perfector conductor, we have $AZ = \eta/\epsilon = \mu/\eta$.

If $a \to 0$ in the above, then for a dielectric such that $\mu_2/\mu_3 = \mu' = 1$, we have for $\psi = E_z$,

$$\begin{aligned}C_0 &\approx \tfrac{1}{4}i\pi(K_2a)^2(1 - \epsilon_1/\epsilon_2) = \tfrac{1}{4}i\pi(ka)^2(\epsilon' - \epsilon),\\ C_1 &\propto a^4, \qquad C_n \propto a^{2n}.\end{aligned} \tag{17}$$

Similarly, for $\psi = H_z$,

$$\begin{aligned}C_1 &\approx \tfrac{1}{4}i\pi K_1K_3a^2\,\frac{\epsilon_3(\epsilon_2 - \epsilon_1)}{\epsilon_1(\epsilon_3 + \epsilon_2)} = \frac{i\pi(ka)^2\eta}{4\epsilon}\left(\frac{\epsilon' - \epsilon}{\epsilon' + 1}\right),\\ C_0 &\propto a^4, \qquad C_n \propto a^{2n}\end{aligned} \tag{18}$$

On the other hand, for a perfect conductor and $\psi = E_z$ we obtain

$$\begin{aligned}C_0 &\approx -\frac{1 - (Ka/2)^2}{1 - (ka/2)^2 - i(2/\pi)\ln(2/\gamma ka)},\\ \gamma &\approx 1.781; \qquad C_n \approx -\frac{i\pi n}{(n!)^2}(Kk)^n\left(\frac{a}{2}\right)^{2n},\end{aligned} \tag{19}$$

whose leading term $C_0 \approx -i\pi/2 \ln(2/\gamma ka)$ is the leading term for the standard problem. Similarly for $\psi = H_z$,

$$\begin{aligned}C_0 &\approx \frac{-i\pi(Ka)^2}{4\epsilon}, \qquad C_1 \approx \frac{i\pi Kka^2}{4\epsilon},\\ C_n &\approx \frac{i\pi n(Kk)^n}{(n!)^2\epsilon}\left(\frac{a}{2}\right)^{2n}.\end{aligned} \tag{20}$$

The sphere: For the sphere, we have

$$\begin{aligned}\phi &= e^{iKz} = \sum_{n=0}^{\infty} i^n(2n + 1)j_n(Kr)P_n(\cos\theta),\\ u &= \sum i^n(2n + 1)c_nh_n(kr)P_n(\cos\theta),\\ \psi &= \sum i^n(2n + 1)\,d_nj_n(K'r)P_n(\cos\theta),\end{aligned} \tag{21}$$

where j and $h = h^{(1)}$ are the usual spherical func-

tions,[2] and the P's are the Legendre polynomials. Proceeding as previously, leads to (14) with D, C, J, and H replaced by their lower case analogs.

2.3. Green's Function Representation

For an arbitrary shaped scatterer, Green's theorem applied to u and to the free space Green's function in the region outside the scatterer yields the conventional surface integral representation for u. However, the corresponding representation for u as an integral over the volume of the scatterer is significantly different. In order to introduce notation, and to facilitate derivation of the volume integral representation we begin by sketching the standard procedure for obtaining the surface integral form.

Thus, applying Gauss' theorem

$$\oint \mathbf{F}\cdot d\mathbf{S} = \oint \nabla\cdot\mathbf{F}\, dV \tag{22}$$

to $F = \mathcal{G}\,\nabla u$, $u\nabla\mathcal{G}$, and subtracting the results, we obtain Green's theorem

$$\oint (\mathcal{G}\nabla u - u\nabla\mathcal{G})\cdot\mathbf{n}\, dS = \oint (\mathcal{G}\,\partial_n u - u\,\partial_n\mathcal{G})\, dS$$

$$= \int (\mathcal{G}\nabla^2 u - u\nabla^2\mathcal{G})\, dV, \tag{23}$$

where $\mathbf{n}$ is the normal out of V. We take $\mathcal{G}$ to satisfy

$$(\nabla^2 + k^2)\mathcal{G}(|\mathbf{r} - \mathbf{r}'|) = \delta(\mathbf{r} - \mathbf{r}'), \tag{24}$$

where δ is the appropriate Dirac δ function. In one, two, and three dimensions, respectively, we have

$$\mathcal{G} = \frac{e^{ikR}}{i2k}, \qquad \frac{H_0^{(1)}(kR)}{i4}, \qquad -\frac{e^{ikR}}{4\pi R} = \frac{h_0^{(1)}(kR)}{4\pi i/k},$$

$$R = |\mathbf{r} - \mathbf{r}'|. \tag{25}$$

Using $\nabla^2 u = -k^2 u$, and (24) in the volume integral in (23) for the region external to the scatterer gives $-u(\mathbf{r})$; then using the conditions at infinity for $\mathcal{G}$ and u to eliminate the surface integral at infinity we write

$$u(kr) = \int [\mathcal{G}(k\,|\mathbf{r} - \mathbf{r}'|)\,\partial_n u(kr')$$

$$- u(kr')\,\partial_n\mathcal{G}(k\,|\mathbf{r} - \mathbf{r}'|)]\, dS(\mathbf{r}'), \tag{26}$$

where the integral is over any surface which incloses the scatterer and excludes the observation point $\mathbf{r}$, and where the normal points out of the scatterer.

Using the boundary conditions (3), and then using (22), we rewrite u of (26) as a volume integral. Thus,

$$u = \int [\mathcal{G}(B'\,\partial_n\psi - B\,\partial_n\phi) - (A'\psi - A\phi)\,\partial_n\mathcal{G}]\, dS$$

$$= \int [\mathcal{G}\psi(A'k^2 - B'K'^2) + (B' - A')\nabla\mathcal{G}\cdot\nabla\psi]\, dV$$

$$- \int [\mathcal{G}\phi(Ak^2 - BK^2) + (B - A)\nabla\mathcal{G}\cdot\nabla\phi]\, dV, \tag{27}$$

where we used the fact that $(\nabla^2 + K^2)\psi = 0$ inside the scatterer, and that $(\nabla^2 + K^2)\varphi = 0$ everywhere in space. The present u differs from that of the conventional problem in containing the additional volume integral involving φ (which is of the form of a "modified Born approximation" for an ordinary scatterer with parameters A, B, K embedded in free space). However, the present form lends itself as readily for obtaining integral equations for the surface fields and derivatives, or for constructing approximations.

If all boundary coefficients equal unity, then (27) gives the integral equation

$$\psi = \varphi - \int (k^2 - K^2)\mathcal{G}\varphi\, dV$$

$$+ \int (k^2 - K'^2)\mathcal{G}\psi\, dV \tag{28}$$

which differs from the form of the conventional result in that the inhomogeneous term contains the integral involving ϕ as an additional component.

For one- and two-dimensional electromagnetics, for $\psi = E_z$, (27) reduces to

$$u = \int [\mathcal{G}\psi k^2(1 - \epsilon') - (1 - 1/\mu')\nabla\mathcal{G}\cdot\nabla\psi]\, dV$$

$$- \int [\mathcal{G}\varphi k^2(1 - \epsilon) - (1 - 1/\mu)\nabla\mathcal{G}\cdot\nabla\varphi]\, dV. \tag{29}$$

Similarly for $\psi = H_z$, we obtain (28) with the μ's and ϵ's interchanged.

If $k|\mathbf{r} - \mathbf{r}'| \gg 1$, $r \gg r'$, then

$$\mathcal{G}(k\,|\mathbf{r} - \mathbf{r}'|) \sim c\mathcal{K}(kr)e^{-ik\mathbf{o}\cdot\mathbf{r}'};$$

$$\mathcal{K} \equiv e^{ik|z|},\quad (2/\pi kr)^{1/2}e^{ikr - i\pi/4},\quad e^{ikr}/ikr;$$

$$c = 1/i2k,\quad 1/i4,\quad k/i4\pi, \tag{30}$$

and (26) reduces to

$$u \sim \mathcal{K}(kr)g(k\mathbf{o}, K\mathbf{i}),$$

$$g(k\mathbf{o}, K\mathbf{i}) \equiv c\int [e^{-ik\mathbf{o}\cdot\mathbf{r}'}\,\partial_n u - u\,\partial_n e^{-ik\mathbf{o}\cdot\mathbf{r}'}]\, dS, \tag{31}$$

where g is the corresponding "two-exterior" scat-

tering amplitude.[2a] Essentially as for u, we have

$$g=c\int e^{-ik\mathbf{o}\cdot\mathbf{r}'}[(A'k^2-B'K'^2)\psi-ik(B'-A')\mathbf{o}\cdot\nabla\psi]\,dV$$
$$-[Ak^2-BK^2+kK(B-A)\mathbf{i}\cdot\mathbf{o}]c\int e^{i(K\mathbf{i}-k\mathbf{o})\cdot\mathbf{r}'}\,dV \quad (32)$$

Corresponding to the symmetrical one-, two-, and three-dimensional cases of Sec. 2.2, the scattering amplitudes are given by

$$g_1 = S_n, \qquad g_2 = \sum_{n=-\infty}^{\infty} C_n e^{in\theta},$$
$$g_3 = \sum_{n=0}^{\infty} C_n(2n+1)P_n(\cos\theta),$$

where g_2 is obtained directly by using $H_n(x) \sim i^{-n}e^{ix-i\pi/4}(2/\pi x)^{1/2}$ in u of (12) to obtain $\mathfrak{IC}_2 g_2$, and g_3 by using $h_n(x) \sim i^{-n}e^{ix}/ix$ in u of (21) to obtain $\mathfrak{IC}_3 g_3$.

Illustration: We apply (32) to consider scattering of an incident wave $\varphi = e^{iKz}$ by "tenuous scatterers" ($K \approx K' \approx k$) large compared to wavelength. Our approximation procedure is essentially that used for ordinary scatterer problems by Montroll and Greenberg,[3] Saxon,[4] van de Hulst,[5] and others: We neglect refraction and reflection effects and approximate the internal field $\psi(\mathbf{r}')$ by visualizing φ arriving as a straight ray modified only by a change in exponent corresponding to the thickness of the scatterer it traversed in reaching $\mathbf{r}'$, i.e., we use the leading term of the WKB approximation with the slowly varying amplitude factor replaced by unity. Thus we approximate the internal field in (32) by

$$\psi \approx e^{iKz_0+iK'(z'-z_0)} = \varphi(z_0)e^{iK'(z'-z_0)}, \quad (33)$$

where $z_0(x',y')$ is the impact point on the surface for the ray reaching $\mathbf{r}'$, and where $z' - z_0$ is the path length within the scatterer.

Using (33) in the three-dimensional form of (32) (i.e., $c = k/i4\pi$), we obtain

$$g(\theta) = M'I_1 - MI_2,$$
$$M = (ik/4\pi)[(A-B)kK\cos\theta - Ak^2 + BK^2],$$

$\cos\theta = \mathbf{r}\cdot\mathbf{z}/rz = \mathbf{o}\cdot\mathbf{z}_1$,

$$I_1 = \int e^{iK'(z'-z_0)+iKz_0-ik\mathbf{o}\cdot\mathbf{r}'}\,dV,$$
$$I_2 = \int e^{iKz'-ik\mathbf{o}\cdot\mathbf{r}'}\,dV, \quad (34)$$

where M' is obtained by replacing A, B, and K of M by primed quantities. In particular, in the forward direction ($\mathbf{o} = \mathbf{z}_1$), we have

$$I_1 = \int e^{-i(K'-K)z_0+i(K'-k)z'}\,dV,$$
$$I_2 = \int e^{i(K-k)z'}\,dV. \quad (35)$$

The integral I_2 of (34) has the form of a "modified Born approximation". In particular, for the the sphere, we obtain[4]

$$I_2 = 4\pi a^3\frac{j_1(x)}{x} = 4\pi a^3\left[\frac{\sin x}{x^3} - \frac{\cos x}{x^2}\right],$$
$$\frac{x}{a} = |K\mathbf{z}_1 - k\mathbf{o}| = \left[(K-k)^2 + 4kK\sin^2\frac{\theta}{2}\right]^{1/2}, \quad (36)$$

which reduces to $4\pi a^3/3$ in the forward direction. The corresponding forward scattered value of I_1 may be obtained directly. Thus for a sphere, the arrival point (z_0) and "departure point" (z_1) of a ray are given by $z_0 = -z_1 = (a^2 - \rho^2)^{1/2}$, where a is the radius, and $\rho^2 = x^2 + y^2$. Using

$$\int dV = \int_0^a \rho\,d\rho\int_0^{2\pi} d\varphi\int_{-z_1}^{z_1} dz'$$

in (35), we integrate to obtain the forward scattered value

$$I_1 = \frac{2\pi}{i(K'-k)}\left[e^{iq_2}\left(\frac{a}{iq} + \frac{1}{q^2}\right) - \frac{1}{q^2}\right]_{q_1}^{q_2},$$
$$q_2 = 2K' - K - k, \qquad q_1 \equiv k - K, \quad (37)$$

which is of the form $f(q_2) - f(q_1)$. Similarly, for the back-scattered direction, we replace k by $-k$ in (37). For arbitrary θ, we may construct series representations for I_1 by proceeding essentially as in references 3–5. If $(K - K')z_0$ is neglected in the exponent of (34), then

$$I_1 \approx I_2(K', k). \quad (38)$$

For arbitrary shaped scatterers we may expand the exponentials in I_1 and I_2 of (35) and obtain representations in terms of the volume moments of the original shape (say, S), and an auxiliary shape (say, S_a) generated by translating the volume elements of S parallel to the direction of incidence

[2a] *Note added in proof.* We may also express g as an integral over the surface of the scatterer in terms of the solution for the corresponding conventional problem. Thus $\mathbf{g}(\mathbf{k}, \mathbf{K}) = c\int(A\varphi\,\partial_n\psi - \psi B\,\partial_n\varphi)\,dS$, where $\varphi = \exp(i\mathbf{K}\cdot\mathbf{r})$, and where $\psi = \exp(-i\mathbf{k}\cdot\mathbf{r}) + U$ is the solution of the conventional scattering problem for the same object (K', A', B') excited by $\exp(-i\mathbf{k}\cdot\mathbf{r})$. Volume integral representations may be obtained as above.

[3] E. W. Montroll and J. M. Greenberg, in *Wave Motion and Vibration Theory*, edited by A. E. Heins (McGraw-Hill Book Company, Inc., New York, 1954), pp. 103–127.

[4] D. D. Saxon, "Lectures on Scattering of Light," Scientific Report No. 9, Dept. of Meteorology, U.C.L.A. (1955) (unpublished).

[5] H. C. van de Hulst, *Light Scattering by Small Particles* (John Wiley and Sons, Inc., New York, 1957), p. 174 ff.

(z_1) to yield a flat face at $z = 0$. In particular, if $K - k \approx 0$ and $K' - k \approx 0$, then

$$I_2 \approx V + i(K - k)\int_{z_0}^{z_1} z'\,dV \equiv V[1 + i(K - k)L'],$$

$$I_1 \approx V + i(K' - K)\int_{z_0}^{z_1}(z' - z_0)\,dV + i(K - k)\int_{z_0}^{z_1} z'\,dV \equiv V[1 + i(K' - K)L + i(K - k)L'], \quad (39)$$

where V is the volume of the scatterer. For concreteness, we visualize the origin of coordinates as the center of the longest line through the scatterer drawn parallel to the direction of incidence. The constant L' is thus the distance of the centroid of the scatterer from the "midplane" $z = 0$, and consequently vanishes for shapes (S) symmetrical to the midplane. Similarly, the other constant

$$L = \frac{1}{V}\int_{z_0}^{z_1}(z' - z_0)\,dV(x, y, z) = \frac{1}{V}\int_0^{z_1 + |z_0|}\zeta\,dV(x, y, z) \quad (40)$$

is the distance of the centroid of the auxiliary shape S_a from its flat surface facing the direction of incidence. Thus we may write

$$L = \left[\int \zeta A(\zeta)\,d\zeta\right] \Big/ \int A(\zeta)\,d\zeta, \qquad A(\zeta) = \iint dx\,dy, \quad (41)$$

where $A(\zeta)$ is the area of a cut of S_a perpendicular to z. As an example, consider that S is an ellipsoid with semiaxes c, b, a along x, y, z (so that $L' = 0$). For this case $A(\zeta) = \pi cb(1 - \zeta^2/a^2)$, and the distance of the centroid of the distorted ellipsoid S_a is given by

$$L = \int_0^{2a}\zeta(a^2 - \zeta^2)\,d\zeta \Big/ \int_0^{2a}(a^2 - \zeta^2)\,d\zeta = 3a/4. \quad (42)$$

3. THREE-DIMENSIONAL ELECTROMAGNETICS

3.1. Statement of the Problem

For three-dimensional electromagnetic problems, we replace the previous scalar functions ψ, φ, and u by vector analogs representing either **E** or **H** fields. Thus the field inside the scatterer's volume V is given by

$$\nabla\times\nabla\times\boldsymbol{\psi} - K'^2\boldsymbol{\psi} = 0, \qquad \nabla\cdot\boldsymbol{\psi} = 0, \quad (43)$$

and external to V we use

$$\nabla\times\nabla\times\boldsymbol{\varphi} - K^2\boldsymbol{\varphi} = 0, \qquad \nabla\cdot\boldsymbol{\varphi} = 0,$$
$$\nabla\times\nabla\times\mathbf{u} - k^2\mathbf{u} = 0, \qquad \nabla\cdot\mathbf{u} = 0. \quad (44)$$

The boundary conditions at the scatterer's surface are

$$\mathbf{n}\times(\boldsymbol{\varphi} + \mathbf{u}) = \mathbf{n}\times\boldsymbol{\psi};$$
$$\mathbf{n}\times(B\nabla\times\boldsymbol{\varphi} + \nabla\times\mathbf{u}) = \mathbf{n}\times(B'\nabla\times\boldsymbol{\psi}) \quad (45)$$

where $B = 1/\mu$ for **E** fields, and $B = 1/\epsilon$ for **H** fields, etc. The source is taken as

$$\boldsymbol{\varphi} = \mathbf{x}_1 e^{iKz} = \mathbf{x}_1\varphi(K), \qquad \mathbf{x}_1 = \mathbf{x}/x \quad (46)$$

and the scattered wave fulfills Silver's radiation condition

$$\lim_{r\to\infty} r[\mathbf{o}\times(\nabla\times\mathbf{u}) + ik\mathbf{u}] = 0, \qquad \mathbf{o} = \mathbf{r}/r. \quad (47)$$

From (47), it follows that for $r \to \infty$ we have asymptotically

$$\mathbf{u} \sim \mathbf{g}e^{ikr}/ikr = \mathbf{g}(\mathbf{k}_0, \mathbf{K})h_0(kr), \quad (48)$$

where the vector scattering amplitude is independent of r. Since $\nabla\cdot\mathbf{u} = 0$, we also have $\nabla\cdot\mathbf{u} \sim ik\mathbf{o}\cdot\mathbf{u} = \mathbf{o}\cdot\mathbf{g}e^{ikr}/r = 0$, and consequently $\mathbf{g}$ is perpendicular to the direction of scattering. Similarly $\mathbf{o}\times(\nabla\times\mathbf{u}) \sim -ik\mathbf{u}$.

If φ represents an **E** field, say $\mathbf{E}_i = \boldsymbol{\varphi} = \mathbf{x}_i e^{iKz}$, then from Maxwell's equations $\nabla\times\mathbf{E} - i\omega\mu\mathbf{H} = \nabla\times H + i\omega\epsilon\mathbf{E} = 0$, it follows that $\mathbf{H}_i = e^{iKz}\mathbf{y}_1\,\eta_1/\mu_1$, $\mathbf{y}_1 = \mathbf{y}/y$. Similarly, the corresponding scattered waves normalized for $\mathbf{E}_i = \mathbf{x}_1 e^{iKz}$ are $\mathbf{E}_s = \mathbf{u} \sim \mathbf{g}_e h_0$ and $\mathbf{H}_s \sim \mathbf{g}'_e h_0$ such that $\mathbf{g}'_e = \mathbf{o}\times\mathbf{g}_e(\eta_3/\mu_3)^{1/2}$. However, if we take $\boldsymbol{\varphi} = \mathbf{y}_1 e^{iKz} = \mathbf{H}_i$, then the corresponding normalized amplitude of $\mathbf{H}_s \sim \mathbf{g}_m h_0$ is related to the others by

$$\mathbf{g}_m = \mathbf{g}'_e\mu_1/\eta_1 = \mathbf{o}\times\mathbf{g}_e\mu_1\eta_3/\eta_1\mu_3 = \mathbf{o}\times\mathbf{g}_e\mu/\eta.$$

3.2. The Sphere

For $\boldsymbol{\varphi} = \mathbf{E}_i = \mathbf{x}_1 e^{iKz}$, we use the vector Mie formalism as in Stratton[6] to treat the sphere by separation of variables. Exploiting his development (pp. 564 and 565), we introduce a third set of parameters in his form for $\mathbf{E}_i = \boldsymbol{\varphi}$ and write the corresponding form $\mathbf{H}_i = \nabla\times\boldsymbol{\varphi}/i\omega\mu$. Thus we obtain

$$\mathbf{u} = \mathbf{E}_s = \sum i^n(2n + 1)[C_n\mathbf{m}^{(3)}_{o_{in}} - C'_n\mathbf{n}^{(3)}_{e_{in}}] \quad (49)$$

where the functions $\mathbf{m}^{(3)}$ and $\mathbf{n}^{(3)}$ are given on p. 564 of Stratton, and where the scattering coefficients C_n and C'_n are given by

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

$$C_n = -\frac{j_n(Ka)[K'aj_n(K'a)]'/\mu' - j_n(K'a)[Kaj_n(Ka)]'/\mu}{h_n(ka)[K'aj_n(K'a)]/\mu' - j_n(K'a)[kah_n(ka)]'} = C_n(\mu, \mu'), \qquad C_n' = \frac{\epsilon}{\eta} C_n(\epsilon, \epsilon') \tag{50}$$

in terms of the usual spherical Bessel and Hankel functions of the first kind; the primes on the brackets indicate differentiation with respect to Ka, etc. The corresponding scattering amplitude, as defined through $\mathbf{E}_s \sim \mathbf{g}h_0$ equals

$$\mathbf{g} = \sum_{n=1} \frac{2n+1}{n(n+1)} \left[C_n \left(\frac{P_n^1}{\sin\theta} \cos\varphi \boldsymbol{\theta}_1 - \frac{\partial P_n^1}{\partial\theta} \sin\varphi \boldsymbol{\varphi}_1 \right) + C_n' \left(\frac{\partial P_n^1}{\partial\theta} \cos\varphi \boldsymbol{\theta}_1 - \frac{P_n^1}{\sin\theta} \sin\varphi \boldsymbol{\varphi}_1 \right) \right], \tag{51}$$

where $P_n^1 = P_n^1(\cos\theta)$ is the associated Legendre function; $\boldsymbol{\theta}_1$ and $\boldsymbol{\varphi}_1$ indicate unit vectors. [Only in (51) and (52) is the symbol φ used for the azimuthal angle; elsewhere, it represents the source.] If $a \to 0$, then to order a^6, we have

$$2\mathbf{g} = [3(C_1 + C_1'\cos\theta) + 5(C_2\cos\theta + C_2'\cos 2\theta)]\boldsymbol{\theta}_1\cos\varphi - [3(C_1\cos\theta + C_1') + 5(C_2\cos 2\theta + C_2'\cos\theta)]\boldsymbol{\varphi}_1 \sin\varphi$$

$$C_1' = \frac{i2(ka)^3(\epsilon' - \epsilon)}{3(\epsilon' + 2)} \left\{ 1 + \frac{K^2(2\epsilon' - \epsilon)(2 + \epsilon') + 5k^2(\epsilon' - 2)(\epsilon - \epsilon') - K'^2\epsilon'(\epsilon + 2)}{10(\epsilon - \epsilon')(2 + \epsilon')} a^2 + \frac{i2}{3}(ka)^3 \frac{\epsilon' - 1}{\epsilon' + 2} \right\},$$

$$C_2' = i\eta(ka)^5(\epsilon' - \epsilon)/15(3 + 2\epsilon'), \qquad C_n = C_n'(\mu, \mu')\eta/\mu. \tag{52}$$

The corresponding results for a perfect conductor ($\mathbf{n} \times \mathbf{E} = 0$ on S) are

$$C_n = -j_n(Ka)/h_n(ka), \qquad C_n' = -[Kaj_n(Ka)]'/[kah_n(ka)]'\eta \tag{53}$$

$$C_1 = -\frac{i\eta(ka)^3}{3}\left[1 - \frac{(ka)^2}{10}(5 + \eta^2) - \frac{i(ka)^3}{3}\right], \qquad C_2 = -\frac{i\eta(ka)^5}{45},$$

$$C_1' = \frac{i2(ka)^3}{3}\left[1 + \frac{(ka)^2}{10}(5 - 2\eta^2) + \frac{i2}{3}(ka)^3\right], \qquad C_2' = \frac{i\eta(ka)^5}{30}. \tag{54}$$

3.3. Green's Function Formalism

We proceed essentially as in Sec. 2.3. We apply Gauss' theorem (21) to the vectors $(\boldsymbol{\Gamma}\cdot\mathbf{e}) \times (\nabla \times \mathbf{u})$ and $\mathbf{u} \times (\nabla \times \boldsymbol{\Gamma}\cdot\mathbf{e})$, where $\mathbf{e}$ is an arbitrary constant vector, and $\boldsymbol{\Gamma}$ is the free space dyadic Green's function

$$\boldsymbol{\Gamma}(\mathbf{r}, \mathbf{r}') = \boldsymbol{\Gamma}(\mathbf{r}', \mathbf{r}) = \left[\mathbf{I} - \frac{\nabla\nabla'}{k^2}\right] \frac{kh_0(k\,|\mathbf{r} - \mathbf{r}'|)}{4\pi i},$$

$$\nabla \times \nabla \times \boldsymbol{\Gamma} - k^2\boldsymbol{\Gamma} = -\mathbf{I}\,\delta(\mathbf{r} - \mathbf{r}') \tag{55}$$

with $\mathbf{I}$ as the unit dyadic. Thus, analogous to (23):

$$\oint [\boldsymbol{\Gamma}\cdot\mathbf{e} \times (\nabla \times \mathbf{u}) - \mathbf{u} \times (\nabla \times \boldsymbol{\Gamma}\cdot\mathbf{e})]\cdot\mathbf{n}\, dS = -\int \{(\boldsymbol{\Gamma}\cdot\mathbf{e})\cdot\nabla \times (\nabla \times \mathbf{u}) - \mathbf{u}\cdot[\nabla \times (\nabla \times \boldsymbol{\Gamma}\cdot\mathbf{e})]\}\, dV. \tag{56}$$

Using the wave equation (44) for $\mathbf{u}$ and (55) for $\boldsymbol{\Gamma}$, we apply (56) in the region external to the scatterer and obtain the analog of (26):

$$\mathbf{u}\cdot\mathbf{e} = \int [(\boldsymbol{\Gamma}\cdot\mathbf{e}) \times (\nabla \times \mathbf{u}) - \mathbf{u} \times (\nabla \times \boldsymbol{\Gamma}\cdot\mathbf{e})]\cdot\mathbf{n}\, dS. \tag{57}$$

where $\mathbf{e}$ may be dropped for brevity.

The volume integral representation for $\mathbf{u}$ of (57) follows from the boundary conditions (45), some elementary vector algebra, and Gauss' theorem. Thus

$$\mathbf{u} = \int \{\boldsymbol{\Gamma} \times [\nabla \times (B'\psi - B\varphi)] - (\psi - \varphi) \times (\nabla \times \boldsymbol{\Gamma})\}\cdot d\mathbf{S} = \int [\psi\cdot\boldsymbol{\Gamma}(k^2 - B'K'^2) + (B' - 1)(\nabla \times \psi)\cdot(\nabla \times \boldsymbol{\Gamma})]\, dV - \int [\varphi\cdot\boldsymbol{\Gamma}(k^2 - BK^2) + (B - 1)(\nabla \times \varphi)\cdot(\nabla \times \boldsymbol{\Gamma})]\, dV. \tag{58}$$

The conventional scattering problem ($B = 1, K = k$) does not contain the integral involving φ; for the conventional problem involving $\mathbf{E}$ fields with $B' = 1/\mu' = 1$, (58) reduces to the result discussed in detail by Saxon.[4]

If $k\,|\mathbf{r} - \mathbf{r}'| \gg 1$, $r \gg r'$, then

$$\mathbf{\Gamma}(\mathbf{r}, \mathbf{r}') \sim (\mathbf{I} - \mathbf{oo})h_0(kr)e^{-ik\mathbf{o}\cdot\mathbf{r}}c, \qquad c = k/i4\pi \quad (59)$$

Consequently (58) reduces to $\mathbf{u} \sim h_0(kr)\mathbf{g}$, where, e.g.,

$$\mathbf{g}(k\mathbf{o}, \mathbf{K}) = (\mathbf{I} - \mathbf{oo})\cdot\int [(\mathbf{n}\times\mathbf{u})\times\nabla e^{-ik\mathbf{o}\cdot\mathbf{r}'} - e^{-ik\mathbf{o}\cdot\mathbf{r}'}(\nabla\times\mathbf{u})\times\mathbf{n}]\,dS, \qquad \mathbf{g}\cdot\mathbf{o} = 0, \quad (60)$$

which may be recast in different form by using vector algebra. Similarly, corresponding to (58),

$$\mathbf{g} = c(\mathbf{I} - \mathbf{oo})\cdot\int e^{-ik\mathbf{o}\cdot\mathbf{r}'}[(k^2 - B'K'^2)\psi + (B' - 1)ik\mathbf{o}\times(\nabla\times\psi)]\,dV - c(\mathbf{I} - \mathbf{oo})\cdot[(k^2 - BK^2)\mathbf{x}_1 + kK(B - 1)\mathbf{o}\times(\mathbf{z}_1\times\mathbf{x}_1)]\int e^{i(K\mathbf{i}-k\mathbf{o})\cdot\mathbf{r}'}\,dV, \quad (61)$$

which is the vector analog of (32).

Illustration: We apply (62) to the analogous illustration considered for the scalar problem, i.e., to large tenuous scatterers. Thus for an incident wave $\varphi = \mathbf{x}_1 e^{iKz}$, we multiply (33) by $\mathbf{x}_1$ and substitute in (61) to obtain

$$\mathbf{g}(\theta) = \mathbf{M}'I_1 - \mathbf{M}I_2,$$

$$M = (ik/4\pi)[(1 - B)kK\mathbf{a} - (k^2 - BK^2)\mathbf{b}],$$

$$\mathbf{a} = \mathbf{y}_1\times\mathbf{o}, \qquad \mathbf{b} = \mathbf{x}_1 - \mathbf{o}(\mathbf{o}\cdot\mathbf{x}_1), \quad (62)$$

where M' has the same form as M in terms of the primed parameters, and where I, and I_2 were considered in (34) ff. If we take $\varphi = \mathbf{E}_i$, then $B = 1/\mu$ and $B' = 1/\mu'$. In particular if $\mu' = 1$, and $\mu = \eta = K/k$ (a case of particular interest for reference 1), we obtain

$$\mathbf{g} = (ik/4\pi)[(K'^2 - k^2)\mathbf{b}I_1 - k(K - k)(\mathbf{a} + \mathbf{b})I_2], \quad (63)$$

where, since $K' \approx k$, we may use $K'^2 - k^2 \approx 2k(K' - k)$.

In the forward direction ($\mathbf{a} = \mathbf{b} = \mathbf{x}_1$), for arbitrary tenuous scatterers we have

$$\mathbf{g}\cdot\mathbf{x} = g = (ik^2V/2\pi)(K' - K) \times [1 + i(K' - k)L + u(K - k)L'] \quad (64)$$

where L and L' are as in (39) ff. For a sphere, (64) reduces to

$$g = i2k^2a^3(K' - K)[1 + i(K' - k)3a/4]/3. \quad (65)$$

For other than the forward direction, to lowest order in $K' - k$ and $K - k$, we use (36) and (38) and approximate (62) by

$$\mathbf{g} \approx (ik^2a^3/3)J(\theta)[2(K' - k)\mathbf{b} - (K - k)(\mathbf{a} + \mathbf{b})], \qquad J(\theta) = 3j_1(2ka\sin\tfrac{1}{2}\theta)/2ka\sin\tfrac{1}{2}\theta. \quad (66)$$

In the plane perpendicular to the incident polarization, we have $\mathbf{b} = \mathbf{x}_1$ and $\mathbf{a} = \mathbf{x}_1\cos\theta$; thus

$$\mathbf{g}(\theta) = \mathbf{x}_1(i2k^2a^3/3)J(\theta)[K' - k - (K - k)\cos^2\tfrac{1}{2}\theta]. \quad (67)$$

The results of this section are exploited in detail elsewhere[1] to treat multiple scattering of waves by random distributions of conventional scatterers.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

On Scattering of Waves by Random Distributions. II. Two-Space Scatterer Formalism*

VICTOR TWERSKY
Sylvania Electronic Defense Laboratories, Mountain View, California
(Received December 22, 1960)

The bulk parameters (propagation number K, etc.) of the coherent multiple scattered field in a slab region of randomly distributed scatterers are expressed functionally in terms of "two-space" isolated scatterer amplitudes. These amplitudes correspond to a single object excited by a wave traveling in "K space" (the medium associated with the coherent field) but radiating into "k space" (free space). The formalism is applied to small spheres of arbitrary ϵ' and μ', and the "Lorentz-Lorenz" form for each bulk parameter ϵ and μ is obtained. It is shown that ϵ and μ are independent of the direction and of the polarization of the incident field; thus, the coherent field for the distribution defines a unique Maxwellian medium. As a second approximation for small spheres, the "loss terms" corresponding to incoherent scattering are included; these appear in the bulk ϵ and μ in the roles of electric and magnetic "conductivities." As another illustration, we consider normal incidence on a slab of tenuous scatterers (parameters close to those of free space) large compared to wavelength; for this case the bulk parameters ϵ and μ are equal to the index of refraction. We consider the incoherent as well as the coherent effects.

1. INTRODUCTION AND STATEMENT OF THE PROBLEM

IN a previous paper[1] we considered a plane wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ incident at an arbitrary angle on a uniformly random distribution of arbitrary scatterers in a region bounded by two parallel planes. (We refer to this paper[1] as I and cite its equations as $(n.m)$ to indicate the m'th equation of Sec. n, i.e., just as they are numbered in I.) In Sec. I.2, starting with the ensemble average of the Green's function surface integral representation for the field scattered by a configuration, we introduced approximations in order to express the average coherent field in terms of an integral over the slab region ($0 \leq z \leq d$). Thus, corresponding to the geometry of Fig. I.1, we obtained (2.26) of I:

$$\langle\Psi\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}\left[1 + C\int_0^z e^{-i\gamma\zeta}G(\zeta;\mathbf{k})\,d\zeta\right] + e^{i\mathbf{k}'\cdot\mathbf{r}}C\int_z^d e^{i\gamma\zeta}G(\zeta;\mathbf{k}')\,d\zeta, \tag{1}$$

$$\mathbf{k}\cdot\mathbf{r} = kz\cos\alpha + kx\sin\alpha = \gamma z + k_x x, \qquad \mathbf{k}'\cdot\mathbf{r} = -\gamma z + k_x x,$$

$$C_1 = \rho_1, \qquad C_2 = 2\rho_2/k\cos\alpha, \qquad C_3 = 2\pi\rho_3/k^2\cos\alpha,$$

where $\mathbf{k}$ and $\mathbf{k}'$ are images in the slab face $z = 0$, and where the one-particle density function ρ_n is the number of scatterers divided by the available volume in one, two, or three dimensions. The function G, the "average multiple scattering amplitude" of a scatterer fixed at $\mathbf{r}_s = \mathbf{z}_s$ within the distribution, was defined by (2.23):

$$G(z_s;\mathbf{k}_0) = \langle G(z_s;\mathbf{k}_0)\rangle_s \equiv \{e^{-i\mathbf{k}_0\cdot\mathbf{r}'}, \langle\Psi(\mathbf{z}_s + \mathbf{r}')\rangle_s\}, \qquad \mathbf{k}_0 = k\mathbf{o}, \tag{2}$$

where $\mathbf{o}$ is a unit vector in the direction of observation, $\langle\Psi\rangle_s$ is the average field at a fixed scatterer, and where the operational braces represent the usual Green's formula surface integral, e.g., in three dimensions, $\{x, y\} = (k/i4\pi)\int(x\,\partial_n y - y\,\partial_n x)\,dA$. (In distinction to I, we now use $\mathbf{k}_0 = k\mathbf{o}$ instead of merely $\mathbf{o}$ in the argument of G, etc.) The statement of the problem was completed by the approximation for the average field at a fixed scatterer (2.28):

$$\langle\Psi\rangle_s \approx \langle\Psi\rangle + \langle u(\mathbf{r} - \mathbf{r}_s)\rangle_s,$$

$$\langle u(\mathbf{r} - \mathbf{z}_s)\rangle_s \sim \mathfrak{C}(k\,|\mathbf{r} - \mathbf{z}_s|)\,G(\mathbf{z}_s;\mathbf{k}_0), \tag{3}$$

where $\mathfrak{C}(x) = e^{i|x|}$, $(2/i\pi x)^{1/2}e^{ix}$, e^{ix}/ix in one, two, or three dimensions. Equation (3) has the form of the solution for an isolated scatterer excited by a set of plane waves and radiating into free space. In the limit $\rho \to 0$ [i.e., $C \to 0$ in (1)], we have $\langle\Psi\rangle \to e^{i\mathbf{k}\cdot\mathbf{r}}$ and $\langle u\rangle_s \to v(k\,|\mathbf{r} - \mathbf{r}_s|)e^{i\mathbf{k}\cdot\mathbf{r}}$, where v is the presumably known solution for the single-body scattering problem; for large $k\,|\mathbf{r} - \mathbf{r}_s|$, we have $v \sim \mathfrak{C}g(\mathbf{k}_0, \mathbf{k})$ where $g(\mathbf{k}_0, \mathbf{k})$ is essentially the conventional scattering amplitude for an isolated object excited by a plane wave traveling along $\mathbf{k}$.

In Sec. I.3, we obtained explicit results by interpreting $\langle\Psi\rangle$ of (1) as a set of free-space plane waves $e^{i\mathbf{k}\cdot\mathbf{r}}\mathfrak{A}(0, z) + e^{i\mathbf{k}'\cdot\mathbf{r}}\mathfrak{A}'(z, d)$ representing a

* This work was partially supported by Signal Corps Contract DA 36-039 SC 78281.
[1] V. Twersky, J. Math. Phys. 3, 700 (1962), this issue.

multiple scattering process in free k space (i.e., at the "microscopic level" (1) indicates that the field at z consists of the source term $e^{i\mathbf{k}\cdot\mathbf{r}}$ plus the free-space forward-scattered fields of the planes of scatterers in $0 < \zeta < z$, plus the free-space reflected field of the planes $z < \zeta < d$). The corresponding average scattering amplitude G of $\langle u\rangle_s$, i.e.,

$$G(\zeta; \mathbf{k}_0) = g(\mathbf{k}_0, \mathbf{k})\mathfrak{A}(0, z)e^{i\gamma z} + g(\mathbf{k}_0, \mathbf{k}')\mathfrak{A}'(z, d)e^{-i\gamma z},$$

was substituted in (1) and the resulting integral equations solved to obtain the form

$$\langle\Psi\rangle = Ae^{i\mathbf{K}\cdot\mathbf{r}} + A'e^{i\mathbf{K}'\cdot\mathbf{r}},$$
$$\mathbf{K}\cdot\mathbf{r} = \Gamma z + k_x x, \qquad \mathbf{K}'\cdot\mathbf{r} = \Gamma' z + k_x x, \tag{4}$$

where A and A' are independent of z, and where $\mathbf{K}$ and $\mathbf{K}'$ are the "macroscopic" propagation vectors of the medium associated with the average coherent field. Section I.3 obtained A, K etc., explicitly in terms of the known functions of known parameters $g(\mathbf{k}_0, \mathbf{k})$ and $g(\mathbf{k}_0, \mathbf{k}')$; the results were generalizations of existing results derived elsewhere by different methods for monopoles, dipoles, small scatterers, circular cylinders, and spheres. It was pointed out that the procedure was not fully "self-consistent" (the microscopic field traveled in "k space" but the macroscopic internal field traveled in "K space"), and that these approximations (obtained by essentially perturbing around the characteristics of free space) were not appropriate for dense distributions.

Although several formalisms exist for extending such perturbation procedures to dense distributions (analytical "hole" corrections, etc., as mentioned in I), it is believed that a more rapidly convergent development may be obtained in terms of a more appropriate isolated scattering amplitude than the conventional $g(\mathbf{k}_0, \mathbf{k})$. In particular, the form (4) substituted into (3) indicates that a fully self-consistent procedure may be based on a new kind of isolated scatterer problem: The scattered wave must travel in k space [as seen from the asymptotic form of $\langle u\rangle_s$ in (3)] whereas the excitation must travel in K space [as seen from the form of $\langle\Psi\rangle$ of (4)]. In general, such scattering problems involve three wave equations (one for the interior of the scatterer, and two for its "two exteriors"), and three sets of physical parameters (e.g., ϵ's and μ's) associated with the three "media". Suppressing the space in the scatterer's interior for simplicity, we denote the corresponding "two-space" scattering amplitudes by $g(\mathbf{k}_0, \mathbf{K})$.

Thus the "average scatterer" in (3) is now regarded as excited by the two plane waves of (4) traveling in K space, but still radiating into k space as required by the asymptotic form of $\langle u\rangle_s$. By superposition, the corresponding multiple scattered amplitude is consequently

$$G(\zeta: \mathbf{k}_0) = g(\mathbf{k}_0, \mathbf{K})Ae^{i\Gamma z} + g(\mathbf{k}_0, \mathbf{K}')A'e^{i\Gamma' z}. \tag{5}$$

The generalized single-body scattering problem which yields $g(\mathbf{k}_0, \mathbf{K})$ as the scattering amplitude has been treated elsewhere.[2] Thus we assume in the present development that the forms of the two-space single-scattering amplitudes required in (5) are known. Equations (1), (4), and (5) define the present treatment of the problem of the coherent field in random distributions.

2. THE AVERAGE WAVE FUNCTION

Substituting (4) and (5) into (1), and introducing

$$S = Cg(\mathbf{k}, \mathbf{K}), \qquad S' = Cg(\mathbf{k}', \mathbf{K}'),$$
$$R = Cg(\mathbf{k}', \mathbf{K}), \qquad R' = Cg(\mathbf{k}, \mathbf{K}'), \tag{6}$$

we obtain

$$Ae^{i\Gamma z} + A'e^{i\Gamma' z} = e^{i\gamma z}\left[1 + \int_0^z e^{-i\gamma\zeta}(ASe^{i\Gamma\zeta} + A'R'e^{i\Gamma'\zeta})\,d\zeta\right] + e^{-i\gamma z}\int_z^d e^{i\gamma\zeta}(ARe^{i\Gamma\zeta} + A'S'e^{i\Gamma'\zeta})\,d\zeta. \tag{7}$$

Integrating over ζ, and equating the coefficients of $e^{i\Gamma z}, e^{i\Gamma' z}, e^{i\gamma z}$, and $e^{-i\gamma z}$ to zero gives, respectively,

$$1 = \frac{S}{i(\Gamma - \gamma)} - \frac{R}{i(\Gamma + \gamma)}, \qquad 1 = \frac{R'}{i(\Gamma' - \gamma)} - \frac{S'}{i(\Gamma' + \gamma)} \tag{8}$$

$$1 = \frac{AS}{i(\Gamma - \gamma)} + \frac{A'R'}{i(\Gamma' - \gamma)}, \qquad \frac{ARe^{i\Gamma d}}{i(\Gamma + \gamma)} + \frac{A'S'e^{i\Gamma' d}}{i(\Gamma' + \gamma)} = 0. \tag{9}$$

Equation (8) specifies the propagation coefficients. Equation (9) gives the boundary conditions on the coherent field at $z = 0$ and $z = d$. (The first condition "extinguishes" the incident wave $e^{i\mathbf{k}\cdot\mathbf{r}}$ within the slab region, and the second "extinguishes" the internal free-space reflected wave.)

Solving (9) for A and A', and using (8) to simplify the results, we obtain

$$A = (1 - Q)[1 - QQ'e^{i(\Gamma - \Gamma')d}]^{-1} \equiv (1 - Q)\,D,$$
$$A' = (1 - Q')Qe^{i(\Gamma - \Gamma')d}\,D$$
$$Q \equiv (\Gamma - \gamma)R/(\Gamma + \gamma)S, \tag{10}$$
$$Q' \equiv (\Gamma' + \gamma)R'/(\Gamma' - \gamma)S'.$$

[2] V. Twersky, J. Math. Phys. 3, 716 (1962), this issue.

Substituting these values in (4) yields the internal field Ψ_I. Similarly these values together with (1) and (5), yield the reflected field

$$\Psi_R = e^{i\mathbf{k}'\cdot\mathbf{r}} \int_0^d e^{i\gamma\zeta} G(\zeta;\mathbf{k}')\, d\zeta$$

in the range $z < 0$, and the transmitted field

$$\Psi_T = e^{i\mathbf{k}\cdot\mathbf{r}}\left[1 + C\int_0^d e^{-i\gamma\zeta} G(\zeta;\mathbf{k})\, d\zeta\right]$$

for $z > d$. In terms of the present Γ's and Q's, the forms of the coherent field in the three regions are identical with those given previously in (3.14), and fulfill the previous average boundary conditions (3.25) and (3.26).

The corresponding multiple scattering amplitude obtained from (1) and (5) equals

$$G(\zeta;\mathbf{k}_0) = (1-Q)g(\mathbf{k}_0, \mathbf{K})e^{i\Gamma z} D + Q(1-Q')g(\mathbf{k}_0, \mathbf{K}')e^{i\Gamma' z + i(\Gamma-\Gamma')d} D, \quad (11)$$

which has the form (3.39) in terms of

$$\mathcal{G}(\mathbf{k}_0, \mathbf{K}) = (1-Q)g(\mathbf{k}_0, \mathbf{K}),$$
$$\mathcal{G}(\mathbf{k}_0, \mathbf{K}') = (1-Q')g(\mathbf{k}_0, \mathbf{K}'). \quad (12)$$

The present forms for the coherent field and amplitude are identical to those obtained in Sec. I.3, and consequently much of the previous discussion is applicable. The results differ in that the Q's and Γ's are now differently expressed in terms of an isolated scattering amplitude, and that the amplitude $g(\mathbf{k}_0, \mathbf{K})$ corresponds to a new kind of single scatterer. In order to specify the new scatterer we require in addition to the relation $\Gamma(g)$ a relation between g and one of the macroscopic parameters of the medium we have synthesized. The required parameters are implicit in the boundary conditions of the coherent field (3.25) and (3.26).

For simplicity we restrict detailed consideration to scatterers which are essentially symmetrical with respect to the slab face $z = 0$. Such scatterers fulfill the symmetry relations

$$S = S', \quad g(\mathbf{k}, \mathbf{K}) = g(\mathbf{k}', \mathbf{K}') \equiv g;$$
$$R = R', \quad g(\mathbf{k}', \mathbf{K}) = g(\mathbf{k}, \mathbf{K}') \equiv g'. \quad (13)$$

Consequently (8) reduces to

$$i = S/(\Gamma-\gamma) - R/(\Gamma+\gamma);$$
$$i(\Gamma^2-\gamma^2) = \Gamma(S-R) + \gamma(S+R) \equiv C(\Gamma g_- + \gamma g_+), \quad (14)$$

where $g_- = g - g'$ and $g_+ = g + g'$ are twice the antisymmetric and symmetric components of g with respect to reflection of one direction in the plane of symmetry. The corresponding Q's of (10) are given by

$$Q = Q' = (\Gamma-\gamma)R/(\Gamma+\gamma)S = (\Gamma-\gamma)g'/(\Gamma+\gamma)g. \quad (15)$$

For this case, the coherent field has the elementary form for a uniform isotropic slab

$$\Psi_I = (1-Q)[e^{i\Gamma z} + Qe^{-i\Gamma z + i2\Gamma d}]e^{ikx\sin\alpha} D, \qquad D = [1 - Q^2 e^{i2\Gamma d}]^{-1},$$
$$\Psi_R = -Q(1-e^{i2\Gamma d})\, De^{i\mathbf{k}'\cdot\mathbf{r}} = \mathcal{R}e^{i\mathbf{k}'\cdot\mathbf{r}},$$
$$\Psi_T = (1-Q^2)\, De^{i(\Gamma-\gamma)d}e^{i\mathbf{k}\cdot\mathbf{r}} = \mathcal{T}e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (16)$$

and satisfies simple boundary conditions at $z = 0$, and d:

$$\Psi_{\text{out}} = \Psi_{\text{in}}; \qquad \partial_z\Psi_{\text{out}} = B\,\partial_z\Psi_{\text{in}} = \frac{\gamma Z}{\Gamma}\partial_z\Psi_{\text{in}};$$
$$iZ = i\frac{1+Q}{1-Q} = \frac{S}{\Gamma-\gamma} + \frac{R}{\Gamma+\gamma}. \quad (17)$$

The corresponding multiple scattered amplitude is

$$G(\zeta;\mathbf{k}_0) = \frac{(1-Q)[e^{i\Gamma z}g(\mathbf{k}_0, \mathbf{K}) + Qe^{i\Gamma(2d-z)}g(\mathbf{k}_0, \mathbf{K}')]}{1 - Q^2 e^{i2\Gamma d}}. \quad (18)$$

From (14) and (17), we obtain

$$i(\Gamma - Z\gamma) = Cg_-, \qquad i(Z\Gamma - \gamma) = Cg_+. \quad (19)$$

Equivalently in terms of Γ and B, we have

$$B = 1 - \frac{Cg_-}{i\Gamma}, \quad \frac{B\Gamma^2}{\gamma^2} = \frac{B\eta^2\cos^2\beta}{\cos^2\alpha} = 1 + \frac{Cg_+}{i\gamma}, \quad (20)$$

where η and β (such that $\eta\sin\beta = \sin\alpha$) are the complex index and angle of refraction, respectively. Thus

$$B\eta^2 = B\sin^2\alpha + B\eta^2\cos^2\beta = 1 - Cg_-\sin^2\alpha/i\Gamma + Cg_+\cos^2\alpha/i\gamma. \quad (21)$$

An alternative expression for η^2 than that obtained from (20) and (21) follows directly from (14):

$$\eta^2 = (\Gamma^2 + k^2\sin^2\alpha)/k^2 = 1 + C(\Gamma g_- + \gamma g_+)/ik^2. \quad (22)$$

In particular, for the electromagnetic case, we have

$$\eta^2 = \epsilon\mu, \qquad B^{-1} = \begin{Bmatrix}\epsilon\\ \mu\end{Bmatrix}, \qquad B\eta^2 = \begin{Bmatrix}\mu\\ \epsilon\end{Bmatrix}, \quad (23)$$

where ϵ and μ are the relative permeability and

permittivity. More explicitly, if the incident H field is normal to the plane of incidence ($y = 0$), then

$$\epsilon = 1/B, \quad \mu = B\eta^2; \quad \mathbf{H}_i = (y/y)e^{i\mathbf{k}\cdot\mathbf{r}} \equiv \mathbf{e}\varphi. \tag{24}$$

Similarly if the incident E field is normal to the plane of incidence, then

$$\mu = 1/B, \qquad \epsilon = B\eta^2: \qquad \mathbf{E}_i = \mathbf{e}\varphi. \tag{25}$$

For the three-dimensional electromagnetic case we require

$$\mathbf{g}(\mathbf{k}, \mathbf{K}) = g(\mathbf{k}, \mathbf{K})\mathbf{e}, \quad \mathbf{g}(\mathbf{k}', \mathbf{K}) = g(\mathbf{k}', \mathbf{K})\mathbf{e} \tag{26}$$

in order for the present scalar formalism to apply.

Specializing the above to the three-dimensional case, and introducing the more conventional form $f = g/ik$ [with $f = \mathbf{e}\cdot\mathbf{f}(\mathbf{k}, \mathbf{K})$, $f' = \mathbf{e}\cdot\mathbf{f}(\mathbf{k}', \mathbf{K})$] we obtain for $\mathbf{H}_i = \mathbf{e}\varphi$,

$$\begin{aligned}\epsilon^{-1} &= 1 - \frac{2\pi\rho g_-}{ik\gamma\Gamma} = 1 - \frac{2\pi\rho(f - f')}{kK\cos\beta\cos\alpha},\\ \mu &= 1 + \frac{2\pi\rho}{ik}\left(\frac{g_+}{k^2} - \frac{g_-\sin^2\alpha}{\gamma\Gamma}\right)\\ &= 1 + 2\pi\rho\left[\frac{f + f'}{k^2} - \frac{(f - f')\sin^2\alpha}{kK\cos\beta\cos\alpha}\right].\end{aligned} \tag{27}$$

Similarly for $\mathbf{E}_i = \mathbf{e}\varphi$ we interchange ϵ and μ. For normal incidence ($\alpha = \beta = 0$) (27) simplifies to

$$\epsilon^{-1} = 1 - 2\pi\rho(f - f')/\eta k^2, \qquad \mu = 1 + 2\pi\rho(f + f')/k^2. \tag{28}$$

[The present μ has the previous form (3.32), and for small f, the sum of the first two terms of the expansion of the present ϵ is of the form (3.32).]

The above forms simplify for limiting cases; we illustrate this in three dimensions. Thus if the scatterers are simple monopoles $f = f'$, Eqs. (14), (17), and (23) reduce to

$$\Gamma^2 = \gamma^2 + 4\pi\rho f, \qquad Z = \Gamma/\gamma, \qquad \eta^2 = 1 + 4\pi\rho_1/k^2; \tag{29}$$

the form for η^2 is identical with that obtained by Foldy.[3] On the other hand, for simple dipoles $f = -f'$, we obtain

$$\Gamma^2 = \gamma^2 + 4\pi\rho\Gamma f/\gamma, \qquad Z = \gamma/\Gamma, \qquad \eta^2 = 1 + 4\pi\rho\Gamma f/k^2\gamma. \tag{30}$$

Another special case of interest corresponds to $f' \ll f$. Here, for normal incidence we obtain

$$K = k + \frac{2\pi\rho g}{ik^2} = k + \frac{2\pi\rho f}{k}; \qquad \eta = \epsilon = \mu, \quad Z = 1. \tag{31}$$

[3] L. L. Foldy, Phys. Rev. **67**, 107 (1945).

For this case,

$$\varphi = e^{ikz}, \quad \Psi_I = e^{iKz}, \quad \Psi_T = e^{iKd + ikz}, \quad \Psi_R = 0. \tag{32}$$

We have thus obtained simple relations for the required parameters ϵ and μ in terms of f. However, f itself is a function of ϵ and μ. Thus, in order to solve the above equations, we must express f explicitly in terms of ϵ and μ for specific scatterers. In a following section we consider special cases in detail.

3. ENERGY CONSIDERATIONS

In Sec. I.3.2, we worked with the usual theorems for the isolated scattering amplitude $g(\mathbf{k}_0, \mathbf{k})$ in order to construct analogous theorems for $\mathcal{G}(\mathbf{k}_0, \mathbf{k}) = g(\mathbf{k}_0, \mathbf{k}) - Q(k)g(\mathbf{k}_0, \mathbf{k}')$, and used these theorems to interpret the propagation number $\Gamma(k)$ physically. Then in Sec. I.3.3, for lossless scatterers we showed that our explicit forms for the average energy flux $\langle\mathbf{J}\rangle = \mathrm{Re}\,\langle\Psi^*\nabla\Psi/ik\rangle$ in terms of $\mathcal{G}$ fulfilled the energy theorem for the distribution

$$\langle\mathbf{J}\rangle\cdot\mathbf{z}_0 = (\mathbf{C} + \mathbf{I})\cdot\mathbf{z}_0 = \text{const}, \tag{33}$$

i.e., that the sum of the coherent and incoherent flux normal to the slab is constant. In effect, Sec. I.3.3 obtained theorems for $\mathcal{G}$ by applying energy considerations to the distribution, and showed they were identical to the forms obtained in Sec. I.3.2 by considerations for an isolated scatterer.

In the present treatment the equivalent isolated scatterer problem yielding

$$\mathcal{G}(\mathbf{k}_0, \mathbf{K}) = [1 - Q(K)]g(\mathbf{k}_0, \mathbf{K})$$

appears too general to enable us to parallel the previous procedure. Instead we essentially exploit the development of Sec. I.3.3 to determine the constraints on $\mathcal{G}$ arising from applying the energy theorem to the distribution.

Thus the various forms for the coherent and incoherent intensities given in Sec. I.3.3 in terms of G and $\mathcal{G}$ apply equally in terms of the present functions. In particular, the coherent intensities outside the slab are given by

$$\mathbf{C} = |\mathfrak{R}|^2\mathbf{k}'/k = |\mathfrak{R}|^2\mathbf{i}' \quad \text{for} \quad z < 0, \qquad \mathbf{C} = |\mathfrak{I}|^2\mathbf{k}/k = |\mathfrak{I}|^2\mathbf{i} \quad \text{for} \quad z > d \tag{34}$$

in terms of the present $\mathfrak{R}$ and $\mathfrak{I}$. Similarly, to trace the energy lost from the coherent field, we may approximate the corresponding incoherent flux by

$$\mathbf{I} \sim \left[\rho_1\cos\alpha, \frac{2\rho_2}{\pi k}, \frac{\rho_3}{k^2}\right]\int_0^d d\zeta\int_{\mathfrak{h}}\frac{|G(\zeta:\mathbf{k}_s)|^2}{\cos\theta_s}\,\mathbf{s}\,d\boldsymbol{\Omega}_s, \qquad \mathbf{s} = \frac{\mathbf{r} - \mathbf{r}_s}{|\mathbf{r} - \mathbf{r}_s|}, \tag{35}$$

where the three terms in the brackets are appropriate multipliers for one, two, and three dimensions, respectively, and where $\int_{1/2} s d\Omega_s/\cos\theta_s$ ranges over the forward or back half-space of angles (i.e., 2π or π in three or two dimensions), or reduces to $\mathbf{s} = \mathbf{i}$ or $\mathbf{i}'$ in one dimension.

Applying the energy theorem (33), we obtain analogous to (3.81)

$$1 - |\mathfrak{R}|^2 - |\mathfrak{I}|^2 = 2C \int_0^d \mathfrak{M} |G(\zeta; \mathbf{k}_0)|^2 \, d\zeta, \qquad (36)$$

where $\mathfrak{M}$ is the average over all directions of observation. Substituting $|G|^2$ of (18) we integrate over ζ; then substituting $|\mathfrak{R}|^2$ and $|\mathfrak{I}|^2$ from (16), we equate corresponding terms to obtain

$$2 \operatorname{Im} \Gamma = \frac{2C\mathfrak{M} |\mathcal{S}(\mathbf{k}_0, \mathbf{K})|^2}{1 - |Q|^2} = \frac{2C\mathfrak{M} |\mathcal{S}(\mathbf{k}_0, \mathbf{K}')|^2}{1 - |Q|^2} = \frac{\rho P_s}{\cos\alpha}, \qquad (37)$$

$$2 \operatorname{Re} \Gamma = \frac{C\mathfrak{M}\mathcal{S}^*(\mathbf{k}_0 \mathbf{K})\mathcal{S}(\mathbf{k}_0, \mathbf{K}')}{\operatorname{Im} Q} = \frac{C\mathfrak{M}\mathcal{S}(\mathbf{k}_0, \mathbf{K})\mathcal{S}^*(\mathbf{k}_0, \mathbf{K}')}{\operatorname{Im} Q}, \qquad (38)$$

where P_s is the multiple scattering cross section of one element for conversion from coherent to incoherent radiation, and 2 Im Γ equals the energy lost from unit area of coherent wave front as a result of incoherent scattering.

In order to relate $\mathfrak{M} |\mathcal{S}|^2$ to Re $\mathcal{S}$, etc., we refer back to (14), (15), and (17) and write

$$i(\Gamma - \gamma) = C(1 - Q)g(\mathbf{k}, \mathbf{K}) = C\mathcal{S}(\mathbf{k}, \mathbf{K}),$$
$$i(\Gamma + \gamma) = C(1 - Q)g(\mathbf{k}, \mathbf{K}')/Q = C\mathcal{S}(\mathbf{k}, \mathbf{K}')/Q \qquad (39)$$

[cf. (3.50)]. Consequently, from (39) and (37), we obtain

$$2 \operatorname{Im} \Gamma = -2C \operatorname{Re} \mathcal{S}(\mathbf{k}, \mathbf{K}) = 2C \operatorname{Re} [\mathcal{S}(\mathbf{k}, \mathbf{K}')/Q] = \frac{-2C \operatorname{Re} [\mathcal{S}(\mathbf{k}, \mathbf{K}) - Q^*\mathcal{S}(\mathbf{k}, \mathbf{K}')]}{1 - |Q|^2} = \frac{2C\mathfrak{M} |\mathcal{S}(\mathbf{k}_0, \mathbf{K})|^2}{1 - |Q|^2}. \qquad (40)$$

Similarly, from (39) and (38),

$$2 \operatorname{Re} \Gamma = -\frac{C \operatorname{Re} [\mathcal{S}(\mathbf{k}, \mathbf{K}') - Q^*\mathcal{S}(\mathbf{k}, \mathbf{K})]}{\operatorname{Im} Q} = \frac{C\mathfrak{M}\mathcal{S}^*(\mathbf{k}_0, \mathbf{K})\mathcal{S}(\mathbf{k}_0, \mathbf{K}')}{\operatorname{Im} Q}. \qquad (41)$$

Analogous results for $\mathcal{S}(\mathbf{k}_0, \mathbf{k})$ are discussed in Sec. I.3.2.

Rewriting the $\mathcal{S}$ relations to make g explicit, we have for example

$$\operatorname{Re} [(1 - Q)g(\mathbf{k}, \mathbf{K})] = \operatorname{Re} [(1 - Q)g(\mathbf{k}, \mathbf{K}')/Q] = -\mathfrak{M} |g(\mathbf{k}_0, \mathbf{K})|^2 \, |1 - Q|^2/(1 - |Q|^2), \qquad (42)$$

or equivalently

$$2\mathfrak{M} |g(\mathbf{k}_0, \mathbf{K})|^2 = -4 \operatorname{Re} Z \operatorname{Re} [g(\mathbf{k}, \mathbf{K})/(Z + 1)] = -4 \operatorname{Re} Z \operatorname{Re} [g(\mathbf{k}, \mathbf{K}')/(Z - 1)]. \qquad (43)$$

Similarly

$$2\mathfrak{M} |g(\mathbf{k}_0, \mathbf{K})|^2 = -4 \operatorname{Re} Z \operatorname{Re} [g(\mathbf{k}, \mathbf{K})/(Z + 1)] = -\operatorname{Re} [g'(Z^* + 1) - g(Z^* - 1)]. \qquad (44)$$

4. ILLUSTRATIONS

In order to apply the present formalism in detail to specific scatterers, we require the appropriate $g(\mathbf{k}, \mathbf{K})$ explicitly in terms of K. For scalar phenomena, $g(\mathbf{k}, \mathbf{K})$ corresponds to the generalized single body scattering problem specified by three wave functions such that

$$(\nabla^2 + K^2)\varphi = 0, \qquad (\nabla^2 + K'^2)\psi = 0, \qquad (\nabla^2 + k^2)v = 0 \qquad (45)$$

where $\varphi = e^{i\mathbf{K}\cdot\mathbf{r}}$ (with Im $K > 0$) is the excitation, ψ (which is to be nonsingular) is the field inside the scatterer, and v is the scattered wave fulfilling the usual free-space radiation condition at infinity. Thus, asymptotically, $v \sim \mathfrak{K}(kr)g(\mathbf{k}_0, \mathbf{K})$, where the new scattering amplitude g has the usual Green's formula surface integral form in terms of v, i.e., $g(\mathbf{k}_0, \mathbf{K}) = \{e^{-ik_0\cdot\mathbf{r}}, v(\mathbf{r}'; \mathbf{K})\}$. The problem is made determinate by boundary conditions at the scatterer, e.g.,

$$A\varphi + v = A'\psi, \qquad B\,\partial_n\varphi + \partial_n v = B'\,\partial_n\psi. \qquad (46)$$

The volume integral forms for u and g this formulation leads to, as well as the results for the scattering coefficients in the series representations for the separable problems of the slab, cylinder, and sphere, are significantly different from those for the conventional isolated scatterer problem.[2]

For three-dimensional electromagnetics, we replace the scalar functions φ, ψ, and v by analogous vector functions having zero divergence (representing $\mathbf{E}$ or $\mathbf{H}$ fields) and use the operator $-\nabla \times \nabla \times$ instead of ∇^2. Thus

$$\nabla \times \nabla \times \mathbf{F} - \kappa^2\mathbf{F} \approx 0, \qquad \nabla \cdot \mathbf{F} = 0, \qquad (47)$$

where (F, κ) stand for $(\varphi = \mathbf{e}e^{i\mathbf{K}\cdot\mathbf{r}}, K = k\eta)$, $(\psi, K' = k\eta')$, or $(\mathbf{v} \sim \mathfrak{K}(kr)\mathbf{g}, k)$. The boundary conditions at the scatterer are

$$\mathbf{n}\times(\varphi + \mathbf{v}) = \mathbf{n}\times\psi, \qquad \mathbf{n}\times(B\nabla\times\varphi + \nabla\times\mathbf{v}) = \mathbf{n}\times(B'\nabla\times\psi), \tag{48}$$

where $B = 1/\mu$, for $\varphi = \mathbf{E}_i$, and $B = 1/\epsilon$ for $\varphi = \mathbf{H}_i$, etc. Surface integral and volume integral representations for $\mathbf{u}$ and $\mathbf{g}$, as well as scattering coefficients for the separable problem of the sphere are given elsewhere.[2]

To use the explicit results[2] for $\mathbf{g}(\mathbf{k}, \mathbf{K})$ in our present treatment of scattering by random distributions, we identify K of (45) with K of (4) and (21): thus $\mathbf{K}\cdot\mathbf{r} = \Gamma z + kx\sin\alpha$, where Γ is given in (14). Similarly we use $A = 1$ and B of (20) in $\mathbf{g}(\mathbf{k}, \mathbf{K})$.

4.1. Small Spheres

Our first illustration deals with small spheres of radius a and arbitrary ϵ' and μ'. The case where the scattering from an isolated sphere is specified by the sum of an electric plus magnetic dipole is of particular interest, since the resultant bulk parameters ϵ and μ should then correspond to those of a homogeneous "Maxwellian medium"; i.e., ϵ and μ should be independent of the polarization and of the direction of the incident field. [The shortcoming in this respect of ϵ and μ obtained from the procedure based on $g(\mathbf{k}, \mathbf{k})$ was noted in (3.33) of I.] We show that the present procedure gives satisfactory results; also, if we take the "available volume" V_a for the distribution to equal the original volume V_0 less the space occupied by the N scatterers (NV_s), then in terms of $\rho_0 = N/V_0$ we obtain the Lorentz-Lorenz[4] or "L^2" form as the first approximation.

As discussed elsewhere,[2] we may solve the generalized electromagnetic scattering problem for the sphere by separation of variables and represent the scattered wave as a Hankel-Legendre series. For an incident wave $\varphi = e^{iKr\cos\theta}\mathbf{x}/x$ the corresponding "two-space" scattering amplitude to order a^6 was shown equal to

$$\mathbf{g} = [\tfrac{3}{2}(C_1 + C_1'\cos\theta) + \tfrac{5}{2}(C_2\cos\theta + C_2'\cos 2\theta)]\cos\omega\boldsymbol{\theta}_1 - [\tfrac{3}{2}(C_1\cos\theta + C_1') + \tfrac{5}{2}(C_2\cos 2\theta + C_2'\cos\theta)]\sin\omega\boldsymbol{\omega}_1,$$

$$C_1 = \frac{i2k^2Ka^3(\mu' - \mu)}{3\mu(\mu' + 2)} \times\left\{1 + \frac{K^2(2\mu' - \mu)(2 + \mu') + 5k^2(\mu' - 2)(\mu - \mu') - K'^2\mu'(\mu + 2)}{10(\mu - \mu')(2 + \mu')}a^2 + i\tfrac{2}{3}k^3a^3\frac{(\mu' - 1)}{\mu' + 2}\right\},$$

$$C_2 = \frac{iK^2k^3a^5}{15\mu}\left[\frac{\mu' - \mu}{3 + 2\mu'}\right], \qquad C_1' = \frac{i2k^3a^3(\epsilon' - \epsilon)}{3(\epsilon' + 2)}\{\epsilon', \epsilon\}, \qquad C_2' = \frac{iKk^4a^5}{15}[\epsilon', \epsilon], \tag{49}$$

where the notation is to indicate that the braced and bracketed functions of C' differ from their analogs in C in having the μ's replaced by ϵ's. The subscript 1 indicates a unit vector, and θ and ω are essentially the usual polar and azimuthal angles.

The above amplitude is an "electric amplitude" normalized for an incident wave $\mathbf{E}_i = e^{i\mathbf{K}\cdot\mathbf{r}}\mathbf{x}_1$. The corresponding "magnetic amplitude" normalized for an incident field $\mathbf{H} = e^{i\mathbf{K}\cdot\mathbf{r}}\mathbf{y}_1$ is related to (49) through

$$\mathbf{g}_m = \mathbf{o}\times\mathbf{g}\eta/\epsilon,\ \eta = (\mu\epsilon)^{1/2},\ \mathbf{o} = \mathbf{r}/r;\ \mathbf{H}_i = e^{i\mathbf{K}\cdot\mathbf{r}}\mathbf{y}_1. \tag{50}$$

We work with g_m when we treat the slab distribution excited by a field polarized in the plane of incidence. On the other hand, for the case of polarization perpendicular to the plane of incidence, $\mathbf{E} = e^{i\mathbf{K}\cdot\mathbf{r}}\mathbf{y}_1$, we rotate (49) through 90°; i.e., we replace ω by $\omega + \pi/2$ and indicate the corresponding amplitude by

$$\mathbf{g}_e = \mathbf{g}(\omega + \pi/2); \qquad \mathbf{E}_i = e^{i\mathbf{k}\cdot\mathbf{r}}\mathbf{y}_1. \tag{51}$$

Equations (50) and (51) correspond to (24) and (25), respectively; the scalar formalism suffices for the coherent field since the sphere fulfills (26).

We use (50) and (51) for our present applications. In these forms, θ is the angle between the directions of incidence and observation, such that

$$\mathbf{K}\cdot\mathbf{r} = Kr\cos\theta = Kr[\cos\theta_K\cos\theta_r + \sin\theta_K\sin\theta_r\cos(\omega_K - \omega_r)]. \tag{52}$$

[4] For brevity, we label as the "L^2 form" of any parameter, the form $(1 + 2\rho_0 x)/(1 - \rho_0 x)$ obtained by Mossotti (1847), Clausius (1897), Maxwell (1873), Lorentz (1880), and Lorenz (1880); see V. Twersky, J. Research Natl. Bur. Standards. **64D**, 715-730 (1960) for citations to the early literature, as well as for citations to other work.

The angles θ_K and θ_r are measured from the slab normal ($\mathbf{z}_1$), and the ω's counterclockwise from the plane of incidence ($y = 0$).

If we retain only the dipole terms of $\mathbf{g}$, and restrict attention to the plane of incidence then

$$\mathbf{g}_m = (3/2)[C_1 + C_1' \cos(\theta_K - \theta_r)](\eta/\epsilon)\mathbf{y}_1 \equiv g_m \mathbf{y}_1, \tag{53}$$

where we used $\omega = 0$ in (49), and $\mathbf{r}_1 \times \theta_1 = \mathbf{y}_1$. Similarly

$$\mathbf{g}_e = (3/2)[C_1 \cos(\theta_K - \theta_r) + C_1']\mathbf{y}_1 = g_e \mathbf{y}_1, \tag{54}$$

where we used $\omega = \pi/2$ and $\omega_1 = -\mathbf{y}_1$ in (49). From (53) we obtain twice the symmetric and antisymmetric components with respect to reflection of either θ_K or θ_r in the slab face:

$$\begin{aligned} g_m + g_m' &= (3\eta/\epsilon)(C_1 + C_1' \sin\theta_K \sin\theta_r), \\ g_m - g_m' &= (3\eta/\epsilon)C_1' \cos\theta_K \cos\theta_r. \end{aligned} \tag{55}$$

In particular, in the specular directions $\theta_r = \theta_k \equiv \alpha$ [and reverting to $\theta_K \equiv \beta$ as in (20)], we have to order a^3

$$\begin{aligned} g_m + g_m' = g_+ &= i2k^3a^3\left[\frac{\mu' - \mu}{\mu' + 2} + \frac{\eta}{\epsilon}\left(\frac{\epsilon' - \epsilon}{\epsilon' + 2}\right)\sin\beta \sin\alpha\right] \\ &= i2(ka)^3\left[\frac{\mu' - \mu}{\mu' + 2} + \frac{1}{\epsilon}\left(\frac{\epsilon' - \epsilon}{\epsilon' + 2}\right)\sin^2\alpha\right], \end{aligned} \tag{56}$$

$$g_m - g_m' = g_- = \frac{i2(ka)^3}{\epsilon}\eta\left(\frac{\epsilon' - \epsilon}{\epsilon' + 2}\right)\cos\beta\cos\alpha = \frac{i2a^3k\Gamma\gamma}{\epsilon}\left(\frac{\epsilon' - \epsilon}{\epsilon' + 2}\right),$$

where we used $k\eta \cos\beta = K \cos\beta = \Gamma$ and $\eta \sin\beta = \sin\alpha$. Similarly for g_e we merely interchange ϵ's and μ's in the above; we indicate this by

$$\begin{aligned} g_e + g_e' &= g_+' = g_+(\epsilon \leftrightarrow \mu), \\ g_e - g_e' &= g_-' = g_-(\epsilon \leftrightarrow \mu). \end{aligned} \tag{56'}$$

Thus for $\mathbf{H}_i = \mathbf{e}e^{i\mathbf{k}\cdot\mathbf{r}}$, we substitute g_- of (56) into ϵ of (27) to obtain

$$\epsilon^{-1} = 1 - \frac{2\pi\rho g_-}{ik\gamma\Gamma} = 1 - \frac{3w}{\epsilon}\left(\frac{\epsilon' - \epsilon}{\epsilon' + 2}\right), \tag{57}$$

where

$$w \equiv \rho 4\pi a^3/3 = \rho V_s = NV_s/V_a \equiv w_0/(1 - w_0), \tag{58}$$

is the volume of the scatterers divided by the available volume. Thus

$$\epsilon = \frac{1 + 3w\epsilon'/(\epsilon' + 2)}{1 + 3w/(\epsilon' + 2)} = \frac{1 + 2\mathfrak{E}}{1 - \mathfrak{E}} = 1 + \frac{3\mathfrak{E}}{1 - \mathfrak{E}},$$

$$\mathfrak{E} \equiv w_0\left(\frac{\epsilon' - 1}{\epsilon' + 2}\right), \qquad w_0 = \frac{w}{1 + w}. \tag{59}$$

If we assume $V_a = V_0 - NV_s$, then $w_0 = \rho_0 4\pi a^3/3 = \rho_0 V_s = NV_s/V_0$ is the fractional volume (the fraction of unit volume of distribution occupied by scatterers) and (59) is the usual L^2 form. Similarly, substituting g_+ and g_- of (56) into μ of (28) we obtain

$$\mu = 1 + \frac{2\pi\rho}{ik}\left(\frac{g_+}{k^2} - \frac{g_- \sin^2\alpha}{\gamma\Gamma}\right) = 1 + 3w\frac{\mu' - \mu}{\mu' + 2}. \tag{60}$$

Solving for μ we obtain the same form as (59):

$$\mu = \frac{1 + 3w\mu'/(\mu' + 2)}{1 + 3w/(\mu' + 2)} = 1 + \frac{3\mathfrak{M}}{1 - \mathfrak{M}}, \qquad \mathfrak{M} = w_0\left(\frac{\mu' - 1}{\mu' + 2}\right). \tag{61}$$

On the other hand for $\mathbf{E}_i = \mathbf{e}e^{i\mathbf{k}\cdot\mathbf{r}}$, we use the forms of (56') in the forms obtained by interchanging ϵ and μ in (27). Thus solving for μ in

$$\mu^{-1} = 1 - \frac{2\pi\rho g_-'}{ik\gamma\Gamma} = 1 - \frac{3w}{\mu}\left(\frac{\mu' - \mu}{\mu' + 2}\right), \tag{62}$$

we again obtain (61). Similarly

$$\epsilon = 1 + \frac{2\pi\rho}{ik}\left(\frac{g_-'}{k^2} - \frac{g_-' \sin^2\alpha}{\gamma\Gamma}\right) = 1 + 3w\frac{\epsilon' - \epsilon}{\epsilon' - 2} \tag{63}$$

leads directly to (59).

For all cases we may write the symmetric form

$$X = w_0 X' = \frac{w}{1 + w}X', \qquad X = \frac{x - 1}{x + 2}, \qquad X' = \frac{x' - 1}{x' + 2}, \tag{64}$$

where $x \to x'$ as $w_0 \to 1$, and $x \to 1$ as either $w_0 \to 0$ or $x' \to 1$.

Thus the present multiple scattering formalism in terms of "schizoid scatterers" $\mathbf{g}(\mathbf{k}, \mathbf{K})$ has led to satisfactory results for small spheres of arbitrary ϵ' and μ' (i.e., for scatterers possessing both electric and dipole moments); (1) we obtained the L^2 form as the first approximation for both ϵ and μ; (2) the results for ϵ and μ are independent of the polarization and of the direction of the incident field (i.e., the parameters associated with the coherent field specify a unique Maxwellian medium); (3) the behavior of

the bulk parameters parallels that of the parameters of the isolated scatterers (e.g., if $\mu' \to 1$ so that the scatterers become simple electric dipoles, then $\mu \to 1$ and the medium shows no magnetic effects). In all three of the above respects, the formalism of I in terms of conventional scatterers $g(\mathbf{k}, \mathbf{k})$ is inadequate (unless hole corrections, etc., are computed). Thus (3.33) for electric dipoles ($\mu' = 1$) gave $\mu = 1$ only for polarization $\mathbf{E}_i$; for the other case, expect for normal incidence, μ differed slightly from unity. For both cases, (3.33) gave only $\epsilon = 1 + 3w(\epsilon' - 1)/(\epsilon + 2)$. For sparse concentrations, the results of I and of the present section reduce to the same forms.

The corresponding index of refraction for all cases is

$$\eta^2 = \epsilon\mu = 1 + 3(\mathfrak{E} + \mathfrak{M} + \mathfrak{E}\mathfrak{M})/(1 - \mathfrak{E} - \mathfrak{M} + \mathfrak{E}\mathfrak{M}), \quad (65)$$

which reduces to the L^2 form if either $\mathfrak{M} = 0$ (i.e., $\mu' = 1$, and $\eta^2 = \epsilon$) or $\mathfrak{E} = 0$ (i.e., $\epsilon' = 1$, and $\eta^2 = \mu$).

For the limiting case of a perfect conductor, we let $\epsilon' \to \infty$ and $\mu' = \eta'^2/\epsilon' \to 0$ in the above to obtain

$$\epsilon = 1 + 3w = \frac{1 + 2w_0}{1 - w_0}, \qquad \mu = \frac{1}{(1 + 3w/2)} = \frac{1 - w_0}{1 + w_0/2}, \qquad \eta^2 = \frac{1 + 2w_0}{1 + w_0/2}. \quad (66)$$

The above results based on retaining only the a^3 terms of g neglect the effects of incoherent scattering on the coherent field. In order to take into account such "losses," we keep in addition the leading real terms of g [i.e., the a^6 terms of (49)].

Thus in (56), we now multiply the terms of the form $(x' - x)/(x' + 2)$ by

$$L(x') \equiv 1 + iAX', \qquad A \equiv 2(ka)^3/3, \qquad X' = (x' - 1)/(x' + 2). \quad (67)$$

The net effect is to replace w in ϵ of (57) and (59) by $wL(\epsilon')$, and w in μ of (60) and (61) by $wL(\mu')$. Thus the symmetric form $X = w_0X' = wX'/(1 + w)$ of (64) goes over to

$$X = \frac{w(1 + iAX')X'}{1 + w(1 + iAX')} = \frac{w_0(1 + iAX')X'}{1 + iw_0AX'}. \quad (68)$$

Equivalently, ϵ and μ are given by

$$x = 1 + \frac{3w_0X'(1 + iAX')}{1 - w_0X'[1 - i3A/(\epsilon' + 2)]} \approx 1 + \frac{3w_0X'}{1 - w_0X'}\left[1 + \frac{i2(ka)^3}{3}\frac{(1 - w_0)X'}{1 - w_0X'}\right], \quad (69)$$

where the final form retains only terms to the order k^3. For x equal ϵ and μ, the imaginary terms proportional to $(ka)^3$ play the role of electric and magnetic "conductivities," respectively.

Similarly for the limiting case of a perfect conductor, we have $X'(\epsilon') = 1$ and $X'(\mu') = -\frac{1}{2}$. Thus

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{w_0[1 + i2(ka)^3/3]}{1 + iw_02(ka)^3/3}, \qquad \frac{\mu - 1}{\mu + 2} = \frac{-w_0[1 - i(ka)^3/3]/2}{1 - i(ka)^3/3}, \quad (70)$$

$$\epsilon = \frac{1 + 2w_0[1 + i(ka)^3]}{1 - w_0}, \qquad \mu = \frac{1 - w_0}{(1 + w_0/2)[1 - i(ka)^3]} \approx \frac{1 - w_0}{(1 + w_0/2)}\left[1 + \frac{i(ka)^3}{(1 + w_0/2)}\right]. \quad (71)$$

From the above we may construct $\eta^2 = \epsilon\mu$, and its real and imaginary parts. Thus writing $x = x_r + ix_i$, we obtain in general

$$\eta_r^2 - \eta_i^2 = \epsilon_r\mu_r - \epsilon_i\mu_i \equiv \mathfrak{R},$$

$$2\eta_r\eta_i = \epsilon_i\mu_r + \epsilon_r\mu_i \equiv \mathfrak{g}, \quad \left.\begin{matrix}\eta_r \\ \eta_i\end{matrix}\right\} = \left\{\frac{\mathfrak{R}}{2}\left[\left(1 + \frac{\mathfrak{g}^2}{\mathfrak{R}^2}\right)^{1/2} \pm 1\right]\right\}^{1/2}. \quad (72)$$

Since $\mathfrak{g}/\mathfrak{R} \ll 1$, we may use

$$\eta_r \approx \sqrt{\mathfrak{R}}\,(1 + \mathfrak{g}^2/8\mathfrak{R}^2) \approx (\epsilon_r\mu_r)^{1/2}\left[1 + \frac{1}{8}\left(\frac{\epsilon_i}{\epsilon_r} + \frac{\mu_i}{\mu_r}\right)^2\right],$$

$$\eta_i \approx \sqrt{\mathfrak{R}}\,(\mathfrak{g}/\mathfrak{R}2) \approx \frac{(\epsilon_r\mu_r)^{1/2}}{2}\left(\frac{\epsilon_i}{\epsilon_r} + \frac{\mu_i}{\mu_r}\right) = \frac{\epsilon_i}{2}(\mu_r/\epsilon_r)^{1/2} + \frac{\mu_i}{2}(\epsilon_r/\mu_r)^{1/2}. \quad (73)$$

If only one dipole is present (e.g., $\mu_r = 1$ and $\mu_i = 0$), then $\eta_r^2 - \eta_i^2 = x_r$, $2\eta_r\eta_i = x_i$, and

$$\eta_r \approx \sqrt{x_r}\left[1 + \frac{1}{8}\left(\frac{x_i}{x_r}\right)^2\right] \approx \left[1 + \frac{3w_0X'}{1 - w_0X'}\right]^{1/2} \cdot \left\{1 + \frac{1}{2}\left[\frac{(ka)^3w_0(1 - w_0)X'^2}{(1 - w_0X')(1 + 2w_0X')}\right]^2\right\}, \quad (74)$$

$$\eta_i \approx \frac{x_i}{2\sqrt{x_r}} \approx \frac{(ka)^3w_0(1 - w_0)X'^2}{(1 - w_0X')^2[1 + 3w_0X'/(1 - w_0X')]^{1/2}}.$$

The scattering amplitude g_m of (53) including loss terms equals

$$g_m = \frac{i\eta}{\epsilon}(ka)^3\left\{\frac{\eta}{\mu}\left(\frac{\mu'-\mu}{\mu'-2}\right)L(\mu') + \left(\frac{\epsilon'-\epsilon}{\epsilon'+2}\right)\cos(\theta_K-\theta_r)L(\epsilon')\right\},$$

$$L(x) = 1 + \frac{i2k^3a^3}{3}X, \qquad X = \left(\frac{x-1}{x+2}\right). \tag{75}$$

Substituting (69) for ϵ and μ, we obtain

$$g_m = (i\eta/\epsilon)(ka)^3\{(\eta/\mu)M(\mu')[1 + (i2/3)(ka)^3M(\mu')] + M(\epsilon')[1 + (i2/3)(ka)^3M(\epsilon')]\cos(\theta_K-\theta_r)\},$$

$$M(x) = (1-w_0)X/(1-w_0X) \tag{76}$$

where terms involving $(ka)^9$ were neglected.

For real ϵ' and μ', we may show that (76) fulfills the single-body relation (43) obtained by "working backwards" from the theorems for the distributions, i.e.,

$$\mathfrak{M}\,|\mathbf{g}(\mathbf{k}_0, \mathbf{K})|^2 = \frac{1}{4\pi}\int |\mathbf{g}|^2\,d\Omega = -2\,\mathrm{Re}\,Z\cdot\mathrm{Re}\left[\frac{g(\mathbf{k},\mathbf{K})}{1+Z}\right]. \tag{77}$$

For simplicity, we consider only the case $\mu' = \mu = 1$, $\alpha = \beta = 0$.

For $\mathbf{E} = \mathbf{e}\varphi$, we obtain

$$\mathbf{g}_e = g_e(\cos\theta\sin\omega\boldsymbol{\theta}_1 - \cos\omega\boldsymbol{\omega}_1),$$

$$g_e = i(ka)^3M(\epsilon')[1 + (i2/3)(ka)^3M(\epsilon')] \equiv iT(1 + i\tfrac{2}{3}T), \tag{78}$$

which corresponds to the "monopole form" (29). From (29), we have $Z = \eta$, and we must show that

$$\mathfrak{M}\,|\mathbf{g}|^2 = -2\,\mathrm{Re}\,\eta\cdot\mathrm{Re}\,\frac{g}{1+\eta} = -\frac{2\eta_r[g_r(1+\eta_r) + g_i\eta_i]}{(1+\eta_r)^2+\eta_i^2}. \tag{79}$$

Now $\mathbf{g}_e$ itself satisfies the theorem for a conventional dipole, i.e., $\mathfrak{M}\,|\mathbf{g}_e|^2 = \frac{2}{3}\,|g_e|^2 = -\mathrm{Re}\,g_e = -g_r$. Substituting this in (79) and clearing the fraction gives

$$g_r(1 + \eta_i^2 - \eta_r^2) = 2g_i\eta_i\eta_r. \tag{80}$$

Equivalently, we require

$$g_r(1-\epsilon_r) = g_i\epsilon_i, \quad \mathrm{Re}\,g\epsilon^* = \mathrm{Re}\,g = -\mathfrak{M}\,|\mathbf{g}|^2. \tag{81}$$

Using (69) for ϵ, and T as defined in (78), we have $\mathrm{Im}\,\epsilon = \epsilon_i = -(1-\epsilon_r)2T/3$; thus since $g_i = T$ and $g_r = -2T^2/3$ we see that (79) fulfills (77) and (81).

For the other polarization, $\mathbf{H} = \mathbf{e}\varphi$, we obtain

$$\mathbf{g}_m = \frac{iT}{\eta}(1 + i\tfrac{2}{3}T)(\cos\theta\cos\omega\boldsymbol{\omega}_1 + \sin\omega\boldsymbol{\theta}_1),$$

$$g_m = \mathbf{e}\cdot\mathbf{g}_m(\omega = 0) = \frac{iT}{\eta}(1 + i\tfrac{2}{3}T)\cos\theta \equiv g(\theta), \tag{82}$$

which corresponds to the dipole form (30). From (30) we have $Z = 1/\eta$, and we must show that

$$\mathfrak{M}\,|\mathbf{g}|^2 = -2\,\mathrm{Re}\,\frac{1}{\eta}\,\mathrm{Re}\,\frac{g(0)}{1+1/\eta} = -\frac{2}{|\eta|^2}\,\mathrm{Re}\eta\cdot\mathrm{Re}\,\frac{g(0)\eta}{1+\eta}, \tag{83}$$

Now, $\eta\mathbf{g}_m$ of (82) fulfills the theorem for a conventional dipole, i.e., $\mathfrak{M}\,|\eta\mathbf{g}_m|^2 = -\mathrm{Re}\,[\eta g(0)]$. Thus if we multiply (83) through by $|\eta|^2$ and note that the present $\eta g(0)$ is identically g_e of (79), we see that we have reduced the present case to the previous. Thus $\mathbf{g}_m$ of (82) fulfills (77), and also the additional relation

$$\mathrm{Re}\,[\eta\epsilon^* g(0)] = \mathrm{Re}\,[\eta g(0)] = -\mathfrak{M}\,|\eta\mathbf{g}|^2. \tag{84}$$

Equivalently, since $\eta\epsilon^* = |\eta|^2\,\eta^*$, we have

$$\mathrm{Re}\,[\eta^* g(0)] = -\mathfrak{M}\,|\mathbf{g}|^2. \tag{84'}$$

The effects of the quadrupole terms of the sphere on ϵ_i and μ_i may be obtained from (49), and more complete results can be obtained from the separable series representation for the arbitrary sphere given in reference 2. The quadrupole terms if significant, would introduce factors of $(ka)^2$ in the real parts of the parameters.

4.2. Large Tenuous Scatterers

Another case of particular interest is that of "tenuous scatterers" ($\mu' = 1$, and $K' \approx k \approx K$) large compared to wavelength. For such scatterers, we approximate the volume integral representation of $\mathbf{g}(\mathbf{k}, \mathbf{K})$ by a "modified WKB technique".[2] Thus for $\mathbf{E}_i = \mathbf{x}_1 e^{iKz}$, and corresponding to $\mu = \eta$ of (31),

$$\mathbf{g} \approx (ik^2/4\pi)[2(K'-k)\mathbf{b}I_1 - (K-k)(\mathbf{a}+\mathbf{b})I_2],$$

$$I_1 = \int e^{iK'(z'-z_0)+iKz_0-ik\mathbf{o}\cdot\mathbf{r}'}\,dV(\mathbf{r}'),$$

$$I_2 = \int e^{iKz-ik\mathbf{o}\cdot\mathbf{r}'}\,dV(\mathbf{r}'), \tag{85}$$

where $z'(x', y')$ ranges over a line $z_1 - z_0$ within the scatterer, and where $\mathbf{a} \equiv \mathbf{y}_1 \times \mathbf{o}$ and $\mathbf{b} \equiv \mathbf{x}_1 - (\mathbf{o}\cdot\mathbf{x}_1)\mathbf{o}$.

In the forward direction ($\mathbf{o} = z_1$), we retain only

the first powers of $K' - k$ and $K - k$ in the integrand and obtain

$$\begin{aligned} I_1 &= V_s[1 + i(K' - K)L + i(K - k)L'], \\ I_2 &= V_s[1 + i(K - k)L']; \\ L' &\equiv \frac{1}{V}\int_{[z_0]}^{[z_1]} z'\, dV(x', y', z'), \\ L &\equiv \frac{1}{V}\int_{[z_0]}^{[z_1]} (z' - z_0)\, dV(x', y', z') \end{aligned} \tag{86}$$

where we may take the origin as the midpoint of the longest line through the scatterer drawn parallel to the direction of incidence. The constant L' is the distance of the centroid of the scatterer from the "midplane" $z = 0$; consequently it vanishes for shapes symmetrical to that plane. On the other hand, if we generate a new body by translating the volume elements of the original body parallel to z to yield a flat surface facing the direction of incidence, then L is the distance of the centroid of the new shape from its flat face. Thus for an ellipsoid of semiaxis a along z, we have[2]

$$L = \tfrac{3}{4}a, \qquad L' = 0, \tag{87}$$

where L is twice the distance of the centroid of a hemi-ellipsoid with respect to its flat face. Similarly the next terms of such expansions as series of volume moments involve the radii of gyration, etc.

Substituting (86) into the forward value of (85) (i.e., $\mathbf{a} = \mathbf{b} = \mathbf{x}_1$) we obtain

$$\begin{aligned} g = \mathbf{x}_1 \cdot \mathbf{g}(k\mathbf{z}_1, K\mathbf{z}_1) &= (ik^2 V_s/2\pi)(K' - K) \\ &\cdot[1 + i(K' - k)L + i(K - k)L']. \end{aligned} \tag{88}$$

For scatterers symmetrical to a plane perpendicular to the direction of incidence (88) reduces to

$$g = (ik^2 V_s/2\pi)(K' - K)[1 + i(K' - k)L]. \tag{89}$$

In the back scattered direction ($\mathbf{o} = -\mathbf{z}_1$), we have $g' = ik^2(K' - k)I_1(-k)/2\pi$, where I_1 is, in general, small. For example, for spheres of radius a, I_1 is proportional to a/k^2, and the ratio of back to forward scattering fulfills

$$|g'/g| \propto (ka)^{-2} \ll 1. \tag{90}$$

(See reference 2 for complete results for the forward and back scattered values for spheres.)

The above results suffice to determine the coherent field for a plane wave normally incident on a slab region of tenuous scatterers. [For nonnormal incidence, i.e., $\alpha \neq \beta \neq 0$, we require the values for $\mathbf{o}\cdot\mathbf{z}_1 = \cos(\beta - \alpha) \approx 1$, and $\mathbf{o}\cdot\mathbf{z}_1 = -\cos(\beta + \alpha) \approx -\cos 2\alpha$.] Corresponding results for the incoherent scattering are obtained from simple approximations of $\mathbf{g}(k\mathbf{o}, K\mathbf{z}_1)$. Thus the integral I_2, for arbitrary angles of observation, has the standard form of a modified Born approximation, say,

$$I_2 = V_s J(K, k), \tag{91}$$

where, e.g., for a sphere,

$$J = 3[\sin(x)/x^3 - \cos(x)/x^2],$$

$$x/a = |K\mathbf{z}_1 - k\mathbf{o}| = [(K-k)^2 + 4kK\sin^2(\theta/2)]^{1/2}. \tag{92}$$

Similarly, if we neglect $(K - K')z_0$ in I_1, we obtain

$$I_1 \approx V_s J(K', k). \tag{93}$$

To lowest order in $K' - k$ and $K - k$ we use $J(k, k) \equiv J_B$ in (91) and (93). Thus

$$\begin{aligned} g(k\mathbf{o}, Kz_1) &\approx (ik^2 V_s/4\pi) \\ &\cdot[2(K' - k)\mathbf{b} - (K - k)(\mathbf{a} + \mathbf{b})]J_B, \end{aligned} \tag{94}$$

where J_B is the usual Born approximation (e.g., for spheres the argument of (92) is to be taken as $2ka\sin(\theta/2)$) such that $J_B \to 1$ as $\theta \to 0$.

Substituting (89) into (31), we obtain

$$\begin{aligned} K &= k + (2\pi\rho/ik^2)g \\ &= k + w(K' - K)[1 + i(K' - k)L], \end{aligned} \tag{95}$$

where $w = \rho V_s = w_0/(1 - w_0)$. Consequently,

$$\begin{aligned} K &= k + \frac{(K' - k)w_0[1 + i(K' - k)L]}{1 + iw_0(K' - k)L} \\ &\approx k + (K' - k)w_0[1 + i(1 - w_0)(K' - k)L], \end{aligned} \tag{96}$$

which gives the real and imaginary parts of K to lowest order in $K' - k$. Thus for lossless scatterers the phase change Re $K - k$ increases linearly with w_0, and the corresponding attenuation (conversion to incoherent scattering) has a parabolic distribution around $w_0 = \frac{1}{2}$. The function K is symmetrical in the parameters in that interchanging k and K', and simultaneously interchanging w_0 and $1 - w_0$ leaves K unaltered. (Thus, essentially as for the case of small spheres, it is plausible to take $V_s = V_0 - NV_s$ in order to interpret w_0 as the fractional volume.)

Substituting (96) back into g we obtain

$$\begin{aligned} g(k, K) &\approx (iV_s/2\pi)k^2(K' - k)(1 - w_0) \\ &\cdot[1 + i(K' - k)(1 - w_0)L]. \end{aligned} \tag{97}$$

(Thus as the fractional volume w_0 approaches unity, the scatterer represents less of a discontinuity in its environment.) The corresponding approximation for $|\mathbf{g}(\theta)|^2$ of (94) [obtained on using (96)] may be taken as

$$|\mathbf{g}(k\mathbf{o}, K\mathbf{z})|^2 \approx$$
$$|(k^2 V_s/4\pi) J_B|^2 |2(K'-k)\mathbf{b} - (K'-k) w_0(\mathbf{a}+\mathbf{b})|^2$$
$$= |[k^2(K' - k) V_s/2\pi] J_B|^2$$
$$\times \left[\left(1 - w_0 \cos^2 \frac{\theta}{2}\right)^2 - (1-w_0) \sin^2 \theta \cos^2 \omega\right], \quad (98)$$

which, for $\theta = 0$ reduces to $|g|^2$ of (97) to order $(K' - k)^2$. In general, $|J_B(\theta)|^2$ is peaked sharply around $\theta = 0$ and only a narrow cone of angles need be considered for computing the total cross section (e.g., for a sphere and $ka \approx 20$, a cone of half-angle 10° around $\theta = 0$ receives about 95% of the energy); consequently, we may use

$$|\mathbf{g}(k\mathbf{o}, K\mathbf{z}_1)|^2 \approx (1 - w_0)^2 |[k^2(K' - k) V_s/2\pi] J|^2$$
$$= (1 - w_0)^2 |\mathbf{g}_B(k\mathbf{o}, k\mathbf{z}_1)|^2, \quad (99)$$

where $\mathbf{g}_B(k\mathbf{o}, k\mathbf{z})$ is the Born approximation for the ordinary isolated scatterer problem.

From (32), the coherent power transmitted through a slab of thickness d is given by

$$C = |\mathfrak{I}|^2 = e^{-2 \operatorname{Im} Kd} \equiv e^{-2\tau},$$
$$2 \operatorname{Im} K = 2L(K' - k)^2 w_0(1 - w_0). \quad (100)$$

For the present case $Q = 0$ (i.e., $Z = 1$), theorem (40) reduces to

$$2 \operatorname{Im} K = \frac{-4\pi\rho}{k^2} \operatorname{Re} g(\mathbf{k}, \mathbf{K}) = \frac{4\pi\rho}{k^2} \mathfrak{M} |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2$$
$$= \frac{\rho}{k^2} \oint |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2 \, d\Omega. \quad (101)$$

Thus $\mathbf{g}(\mathbf{k}, \mathbf{K})$ fulfills the scattering theorem for ordinary scatterers. In view of the behavior of $|J_B|^2$ as a function of angles, we may restrict the integration over Ω to the forward half-space, i.e.,

$$2 \operatorname{Im} K \approx \frac{\rho}{k^2} \int_{(1/2)} |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2 \, d\Omega. \quad (102)$$

Using $|\mathbf{G}|^2 = |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2 e^{-2 \operatorname{Im} K\zeta}$ in (35) we obtain the incoherent power received by a cone of half-angle δ around the direction of incidence:

$$\mathbf{I}\cdot\mathbf{z}_0 = I = \frac{\rho}{k^2} \int_0^d d\zeta \int_\delta |\mathbf{g}|^2 e^{-2 \operatorname{Im} K\zeta} \, d\Omega$$
$$= \frac{\rho}{k^2 2 \operatorname{Im} K} (1 - e^{-2\tau}) \int_\delta |\mathbf{g}|^2 \, d\Omega$$
$$= (1 - e^{-2\tau}) q(\delta), \quad (103)$$

where

$$q(\delta) = \frac{\int_\delta |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2 \, d\Omega}{\oint |\mathbf{g}(k\mathbf{o}, \mathbf{K})|^2 \, d\Omega} \approx \frac{\int_\delta |\mathbf{g}(k\mathbf{o}, \mathbf{k})|^2 \, d\Omega}{\int_{\frac{1}{2}} |\mathbf{g}(k\mathbf{o}, \mathbf{k})|^2 \, d\Omega}, \quad (104)$$

i.e., q is the fraction of the total scattering cross section of one scatterer in free space received by the cone δ, such that $q \to 1$ as $\delta \to \pi/2$. The corresponding total normalized average power in the forward direction is thus

$$\mathbf{J}\cdot\mathbf{z}_0 = C + I = e^{-2\tau} + (1 - e^{-2\tau}) q(\delta). \quad (105)$$

Since the coherent reflected flux and the total incoherent scattering into the back half-space are negligible for this distribution, the energy relation exhibits itself in that $\mathbf{J}\cdot\mathbf{z}_0 = 1$ for $\delta = \pi/2$, i.e., the coherent transmitted power plus the total incoherent power in the forward half-space equals that incident.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

A Symmetric Representation for Three-Body Problems. I. Motion in a Plane*

FELIX T. SMITH
Stanford Research Institute, Menlo Park, California
(Received January 4, 1962)

A symmetric representation is sought for the motion of three particles in the limit of weak interaction. Operators with the desired symmetry can be obtained from the 6-dimensional generalized angular momentum tensor [F. T. Smith, Phys. Rev. **120**, 1058 (1960)]. Here, the 4-dimensional problem of motion in a plane is worked out. Symmetric angular coordinates are found, operators and eigenfunctions are constructed, and the coupling coefficients connecting this with more familiar representations are discovered. Formally, the eigenfunctions are similar to the symmetric rotor functions, but with different arguments.

INTRODUCTION

GENERALIZED orbital angular momentum[1] has recently proved to be a common feature implicit in a number of treatments of quantal problems involving three or more particles.[2] It has likewise arisen in the problem of the n-dimensional harmonic oscillator.[3] Bargmann and Moshinsky,[4] treating the collective motion of N oscillators in a common 3-dimensional harmonic potential, also introduce some generalized angular momentum operators of a vector which they term the "pseudospin." Most of the representations previously given are not completely symmetrical in the particles of the system, but there are problems in which such symmetry would be a convenience.

In this series of papers the possibility of constructing a symmetrical representation for a system of three particles is explored. This representation will be based on operators derived from the generalized angular-momentum tensor $\mathbf{\Lambda}$, all elements of which commute with the total-kinetic-energy operator and with $\Lambda^2 = \frac{1}{2}\sum_{ij}(\Lambda_{ij})^2$. These operators are connected with the total ordinary orbital angular momentum $\mathbf{L}$ and with the symmetric tensor $\mathbf{\Sigma}$ which can be called the *togetherness tensor* since it determines (along with $\mathbf{L}$) the probability of a collision bringing the three particles together.

Symmetric representations of some 3-body problems have also been investigated by Möbius.[5] His attention is directed primarily to the construction of collective coordinates which simplify the solution of the problem, whereas in this paper the primary focus is on the operators.

Ultimately, one would like to know both the complete set of commuting operators and the coordinates in which their eigenfunctions can be most simply expressed. The preferable form for these eigenfunctions is a simple product of one-argument functions. In addition to such a complete expression for the functions, it is desirable to know how they can be expressed as a linear combination of the eigenfunctions in a more familiar representation.

In this paper we shall pursue this program for the case of three particles confined to motion in a plane. This problem has considerable interest in its own right and is a prototype for the more complicated problem involving motion in 3-dimensional space.

In a very recent paper Biedenharn[6] has discussed extensively the properties of the 4-dimensional rotation group. His treatment was motivated by physical problems involving the Coulomb central field. The problem to be treated here may also be

* Supported principally by National Aeronautics and Space Agency and in part by National Science Foundation.

[1] F. T. Smith, Phys. Rev. **120**, 1058 (1960). The results and notation of that paper will be assumed here; references to its equations will be preceded by the numerals I or II, indicating the relevant section. The following corrections should be made: read "$-Y$" in Eq. (I-38); read "$3n$" instead of "n" in Eqs. (II-25) and (II-27).

[2] Helium atom: T. H. Gronwall, Phys. Rev. **51**, 655 (1937); J. H. Bartlett, *ibid.* 51, 661 (1937); V. Fock, Izvest. Akad. Nauk S.S.S.R. Ser. Fiz. **18**, 161 (1954) [translation: Kgl. Norske Videnskab. Selskabs Forh. **31**, 138, 145 (1958)]; P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1730. Nuclear 3-body problem (triton): R. E. Clapp, Phys. Rev. **76**, 873 (1949), R. E. Clapp, Ann. Phys. **13**, 187 (1961). 3-body collisions or dissociation: G. H. Wannier, Phys. Rev. **90**, 817 (1953); L. M. Delves, Nuclear Phys. **9**, 391 (1958–1959); L. M. Delves, *ibid.* **20**, 275 (1960).

[3] G. A. Gallup, J. Mol. Spectroscopy **3**, 673 (1959), J. D. Louck and W. H. Schaffer, *ibid.* **4**, 285, 298 (1960). Louck has now generalized this further to give the n-dimensional analog of spin angular momentum as well [Los Alamos Report LA-2451, 1960 (unpublished)].

[4] V. Bargmann and M. Moshinsky, Nuclear Phys. **18**, 697 (1960).

[5] P. Möbius, Nuclear Phys. **16**, 278 (1960); **18**, 224 (1960).

[6] L. C. Biedenharn, J. Math. Phys. **2**, 433 (1961).

considered as an exercise in the same group theoretical area, but we shall not discuss it from that point of view. However, it may be noted that the full 4-dimensional rotation group is applicable only to problems in which there is no interaction between the particles. Here we shall be principally interested in a subgroup which is constructed in such a way that some of the symmetries will remain unchanged when the interaction is introduced.

In what follows, attention is confined to the case of structureless, spinless particles. Once the principles are worked out in this case, they can be extended to other cases as needed. Möbius[5] and Clapp[2] have given some examples of problems in which the spin and symmetry of the particles must be considered.

This paper is divided into two main sections. In the first, the coordinate systems which are most useful for the 3-body collision problem are described, and the relationships between them explicitly set forth. Many of these relations are applicable to the problem of three particles in space, as well as the restricted problem where they are confined to a plane. In the second part, the operators, eigenfunctions, and coupling coefficients are worked out for the special case of three bodies in a plane.

In a subsequent paper we hope to discuss the symmetric representation for 3-body problems in 3-dimensional space.

A. COORDINATE SYSTEMS

1. Rectangular Coordinates

Thinking about three-body interactions is aided if we use a normalized center-of-mass coordinate system. This has the following properties: (a) The kinetic energy is a diagonal form with a common reduced mass for all the coordinates except those of the center of mass; and (b) the transformation between these coordinates and the laboratory coordinates of the particles is a linear one with constant coefficients that leaves the volume element unchanged and is invariant under rotations in ordinary space. The most general transformation with these properties leads to a family of coordinate systems that depend upon a single parameter, of which we shall mostly need only three special members. Except for a normalizing factor, these special coordinates represent the vector between two of the particles and the vector from the center of mass of this pair to the third.

Basic to this representation is the 3-body reduced mass μ. It is useful to know its connection with the 2-body reduced masses, μ_{ij} for the pair ij and μ'_k for the relative motion of k and the center of mass of ij. Closely related to these are the normalizing constants d_k. Definitions and identities are

$$\begin{aligned} M &= m_1 + m_2 + m_3, \\ \mu_{ij} &= m_i m_j/(m_i + m_j), \\ \mu'_k &= m_k(m_i + m_j)/M, \\ \mu^2 &= m_1 m_2 m_3/M = \mu_{ij}\mu'_k, \\ d_k^2 &= (m_k/\mu)(1 - m_k/M) = \mu/\mu_{ij} = \mu'_k/\mu. \end{aligned} \tag{1}$$

The masses also define certain angles. If (ijk) is an even permutation of (123), β_{ij} is an obtuse angle with the properties

$$\begin{aligned} \beta_{ij} &= -\beta_{ji}, \qquad \beta_{ij} = 0, \\ \tan\beta_{ij} &= -m_k/\mu, \qquad d_i\, d_j \sin\beta_{ij} = 1, \\ & \qquad d_i\, d_j m_k \cos\beta_{ij} = -\mu, \end{aligned} \tag{2a}$$

which lead to the identities

$$\begin{aligned} \beta_{12} + \beta_{23} + \beta_{31} &= 2\pi, \\ \sum_k (1 - m_k/M)\sin^2\beta_{jk} & \\ = \sum_k (1 - m_k/M)\cos^2\beta_{jk} &= 1, \\ \sum_k (1 - m_k/M)\cos 2\beta_{jk} & \\ = \sum_k (1 - m_k/M)\sin 2\beta_{jk} &= 0. \end{aligned} \tag{2b}$$

With these relations we can now define the normalized relative coordinates $\{{}_m\xi\}$. Here the prefix m labels the order in which the particles are paired. If x_i^k is the ith Cartesian component of the position vector $\mathbf{x}^k$ of the kth particle, the transformation is

$$\begin{aligned} {}_m\xi_i^1 &= \sum_{k=1}^{3} d_k x_i^k \sin\beta_{km} = d_m^{-1}(-x_i^{m+1} + x_i^{m+2}), \\ {}_m\xi_i^2 &= \sum_{k=1}^{3} d_k x_i^k \cos\beta_{km} = d_m[x_i^m - (m_{m+1} + m_{m+2})^{-1} \\ & \qquad \times (m_{m+1}x_i^{m+1} + m_{m+2}x_i^{m+2})], \\ X_i &= \sum_{k=1}^{3} \frac{m_k}{M} x_i^k. \end{aligned} \tag{3}$$

[If $m = 1$, Eq. (3) reduces to Eq. (I.2) of reference 1, except for a different convention as to signs.] Similarly the momenta become

$${}_m\pi_i^1 = \mu \sum_{k=1}^{3} \frac{d_k}{m_k} p_i^k \sin\beta_{km} = \frac{\mu}{d_m}\left(-\frac{p_i^{m+1}}{m_{m+1}} + \frac{p_i^{m+2}}{m_{m+2}}\right),$$

$$ {}_m\pi_i^2 = \mu \sum_{k=1}^{3} \frac{d_k}{m_k} p_i^k \cos \beta_{km} $$

$$ = \mu\, d_m \left(\frac{p_i^m}{m_m} - \frac{p_i^{m+1} + p_i^{m+2}}{m_{m+1} + m_{m+2}} \right), $$

$$ P_i = \sum_{k=1}^{3} p_i^k. \tag{4} $$

The volume elements are invariant:

$$ dx_i^1\, dx_i^2\, dx_i^3 = d_m\xi_i^1\, d_m\xi_i^2\, dX_i, $$

$$ dp_i^1\, dp_i^2\, dp_i^3 = d_m\pi_i^1\, d_m\pi_i^2\, dP_i. \tag{5} $$

The kinetic energy as transformed is

$$ K = \frac{1}{2\mu} \sum_{i=1}^{6} ({}_m\pi_i)^2 + \frac{1}{2M} \sum_{i=1}^{3} P_i^2. \tag{6} $$

In Eq. (6), the superscript is suppressed on the π's and the subscript runs from 1 to 6; this convention will often be used, always in such a way that $\{\xi_i^1\} \to \{\xi_1, \xi_2, \xi_3\}$ and $\{\xi_i^2\} \to \{\xi_4, \xi_5, \xi_6\}$. The position X_i and momentum P_i of the center of mass will usually be ignored.

Physically—except for a scale factor—${}_k\boldsymbol{\xi}^1$ represents the vector from particle j to i, and ${}_k\boldsymbol{\xi}^2$ represents the vector from particle k to the center of mass of the pair (ij). If (ijk) run cyclically,

$$ \mathbf{x}^{ij} = -\mathbf{x}^i + \mathbf{x}^j = d_k\, {}_k\boldsymbol{\xi}^1, $$

$$ \mathbf{y}^k = \mathbf{x}^k - (m_i + m_j)^{-1}(m_i\mathbf{x}^i + m_j\mathbf{x}^j) = d_k^{-1}\, {}_k\boldsymbol{\xi}^2. \tag{7} $$

Another useful quantity is the vector from the center of mass of all three particles to one of them,

$$ \mathbf{z}^k = \left(1 - \frac{m_k}{M}\right)\mathbf{y}^k = \frac{\mu}{m_k}\, d_k\, {}_k\boldsymbol{\xi}^2. \tag{8} $$

The identities obviously follow:

$$ x_i^{12} + x_i^{23} + x_i^{31} = \sum_k d_k\, {}_k\xi_i^1 = 0 $$

and

$$ \sum_k m_k z_i^k = \sum_k d_k\, {}_k\xi_i^2 = 0. \tag{9} $$

Clearly, the coordinate system $\{{}_k\xi\}$ is the one to use when particles i and j are close together and k is far away, so that the potential is a function of ${}_k\boldsymbol{\xi}^1$ only. It is a consequence of Eq. (3) that the transformation between the coordinates $\{{}_k\xi\}$ and an equivalent system $\{{}_i\xi\}$ is a simple orthogonal transformation, the kinematic rotation through the angle β_{ki},

$$ {}_i\xi_i^1 = {}_k\xi_i^1 \cos\beta_{ki} + {}_k\xi_i^2 \sin\beta_{ki}, $$

$$ {}_i\xi_i^2 = -{}_k\xi_i^1 \sin\beta_{ki} + {}_k\xi_i^2 \cos\beta_{ki}; \tag{10} $$

it should be noted that this is not a rotation in ordinary space. A more general coordinate system $\{{}_\phi\xi\}$ is obtained by a rotation through the angle ϕ. The momenta transform in just the same way.

2. Hyperspherical Polar Coordinates

a. The Hyperradius

The rectangular coordinates ${}_k\xi_i$ in the six-dimensional interaction space are useful for the formulation of generalized relations, but it is also valuable to introduce a hyperspherical polar-coordinate system in this space. This involves one distance coordinate, the hyperradius ρ, and five angles. The hyperradius is of fundamental importance, as may be seen from the identities which follow from its definition:

$$ \rho^2 = \sum_{i=1}^{6} ({}_k\xi_i)^2 \qquad \text{(for any } k\text{)} $$

$$ = \sum_{k=1}^{3} \left(1 - \frac{m_k}{M}\right) \sum_{i=1}^{3} ({}_k\xi_i^a)^2 \qquad \text{(for } a = 1 \text{ or } 2\text{)} $$

$$ = \mu^{-1} \sum_{k=1}^{3} m_k\, |\boldsymbol{z}^k|^2 $$

$$ = (2\mu M)^{-1} \sum_{i,j=1}^{3} m_i m_j\, |\boldsymbol{x}^i - \boldsymbol{x}^j|^2 $$

$$ = \mu \sum_{\text{cyc}} m_i^{-1}\, |\boldsymbol{x}^i - \boldsymbol{x}^k|^2. \tag{11} $$

Among other properties, the identities of Eq. (11) show that $\mu\rho^2$ is the moment of inertia of the three-body system about the axis perpendicular to the plane in which they lie.

The five angles can be chosen in various ways. We shall be principally interested in two, one that treats the particles symmetrically and a second suited for the representation involving two uncoupled angular momenta.

b. The Symmetric Coordinate System

To start with, one can separate the external from the internal coordinates of the configuration. For two of the external coordinates the obvious choice is the pair of angles locating the orientation of the plane of the three bodies. The third external coordinate fixes in some manner the orientation of the figure with respect to some axis in that plane. One way of doing this is by the angle in the plane from a fixed axis to one of the principal axes of inertia of the figure. In this case, the three external angles can be taken as the Euler angles orienting the principal figure axes with respect to a set of fixed axes. For motion confined to a plane only one external angle is needed, and it can again be defined

by the orientation φ of the principal axes in the plane.

One of the internal coordinates is the hyperradius ρ. It is related to a moment of inertia, and it is natural to look to the other principal moments of inertia for additional coordinates defined symmetrically with respect to the three particles.

If the system lies in the xy plane oriented so that its principal axes coincide with the Cartesian coordinate axes, the moments of inertia are

$$\begin{aligned} I_3 &= \mu\rho^2 \\ I_2 &= \sum_k m_k(z_2^k)^2 = \mu \sum_k (1 - m_k/M)({}_k\xi_1^2)^2 \\ &= \mu[({}_j\xi_1^1)^2 + ({}_j\xi_1^2)^2] \quad \text{(any } j\text{)}, \\ I_1 &= \sum_k m_k(z_2^k)^2 = \mu \sum_k (1 - m_k/M)({}_k\xi_2^2)^2 \\ &= \mu[({}_j\xi_2^1)^2 + ({}_j\xi_2^2)^2]. \end{aligned} \tag{12}$$

Since I_1, I_2, I_3, are the principal moments, there is a subsidiary condition,

$$\begin{aligned} I_{12}/\mu = (1/\mu) \sum_k m_k z_1^k z_2^k &= \sum_k (1 - m_k/M)\, {}_k\xi_1^2\, {}_k\xi_2^2 \\ &= {}_i\xi_1^1\, {}_i\xi_2^1 + {}_i\xi_1^2\, {}_i\xi_2^2 = 0. \end{aligned} \tag{13}$$

The moments are not all independent, because $I_1 + I_2 = I_3$ (this is a well-known result for any plane lamina). One is therefore led to define the difference $I_0 = I_1 - I_2$,

$$\begin{aligned} I_0/\mu = (I_1 - I_2)/\mu &= \sum_k (1 - m_k/M)[({}_k\xi_2^2)^2 - ({}_k\xi_1^2)^2] \\ &= ({}_i\xi_2^1)^2 + ({}_i\xi_2^2)^2 - ({}_i\xi_1^1)^2 - ({}_i\xi_1^2)^2. \end{aligned} \tag{14}$$

This is related to the area A of the triangle by

$$I_0^2 + (4\mu A)^2 = I_3^2. \tag{15}$$

We can then define a new angular coordinate Θ:

$$I_0/\mu = \rho^2 \cos 2\Theta, \qquad 4A = \rho^2 \sin 2\Theta. \tag{16}$$

The quantity I_0/μ is closely related to some internal coordinates used by Clapp.[2] He introduces three scalar coordinates related to

$${}_i v = 2\, {}_i\boldsymbol{\xi}^1 \cdot {}_i\boldsymbol{\xi}^2 = 2 \sum_i {}_i\xi_i^1\, {}_i\xi_i^2; \tag{17}$$

and to $({}_k r_1)^2$ and $({}_k r_2)^2$. The sum of the latter two is ρ^2, and their difference can be written

$${}_i u = ({}_i r_1)^2 - ({}_i r_2)^2. \tag{18}$$

The pair $({}_i u, {}_i v)$ transform under the kinematic rotation of Eq. (10) just like a vector rotated through an angle $2\beta_{ki}$. The magnitude of this vector is an invariant, and it proves to be just

$$({}_i u^2 + {}_i v^2)^{\frac{1}{2}} = I_0/\mu = \rho^2 \cos 2\Theta. \tag{19}$$

It is appropriate to choose as the third internal coordinate an angle ${}_i\Phi$ such that

$$\begin{aligned} {}_i u = (I_0/\mu) \cos 2\, {}_i\Phi &= \rho^2 \cos 2\Theta \cos 2\, {}_i\Phi \\ &= ({}_i\xi_1)^2 + ({}_i\xi_2)^2 - ({}_i\xi_4)^2 - ({}_i\xi_5)^2, \\ {}_i v = (I_0/\mu) \sin 2\, {}_i\Phi &= \rho^2 \cos 2\Theta \sin 2\, {}_i\Phi \\ &= 2({}_i\xi_1\, {}_i\xi_4 + {}_i\xi_2\, {}_i\xi_5). \end{aligned} \tag{20}$$

These equations should be supplemented by

$$4A = \rho^2 \sin 2\Theta = 2({}_i\xi_1\, {}_i\xi_5 - {}_i\xi_2\, {}_i\xi_4). \tag{21}$$

The kinematic rotation to a new basis, say $({}_k u, {}_k v)$, is accomplished by replacing ${}_i\Phi$ by

$${}_k\Phi = \beta_{ki} + {}_i\Phi. \tag{22}$$

This coordinate is not completely independent of the numbering of the particles, but the dependence is the trifling one of a shift of origin. The quantities $({}_i u, {}_i v, A)$ and the coordinates $(\rho, \Theta, {}_i\Phi)$ are invariant to ordinary rotations in 3-dimensional Cartesian space.

The internal coordinate ρ, Θ, ${}_i\Phi$ will prove to be the appropriate ones for the symmetric representation of the three-body problem. For the planar problem they are naturally associated with the external coordinate φ which represents the orientation of one of the principal axes in the plane. In terms of the rectangular coordinates in the xy plane $({}_i\xi_1, {}_i\xi_2, {}_i\xi_4, {}_i\xi_5)$, we can write

$$\begin{aligned} s = \rho^2 \cos 2\Theta \cos 2\varphi &= ({}_i\xi_1)^2 + ({}_i\xi_4)^2 - ({}_i\xi_2)^2 - ({}_i\xi_5)^2 \\ t = \rho^2 \cos 2\Theta \sin 2\varphi &= 2({}_i\xi_1\, {}_i\xi_2 + {}_i\xi_4\, {}_i\xi_5). \end{aligned} \tag{23}$$

There is an interesting symmetry in the equations defining φ and ${}_i\Phi$, Eqs. (20) and (23). φ is invariant to a kinematic rotation (changing j), but it changes by an additive constant under a rotation of the basic Cartesian axes.

The potential for field-free problems is dependent only on the internal coordinates. It is convenient to represent it by equipotential surfaces in a spherical polar coordinate system $(\rho, \Theta, {}_i\Phi)$ such as

$$\begin{aligned} z &= \rho \sin\Theta \\ {}_i x &= \rho \cos\Theta \cos {}_i\Phi \\ {}_i y &= \rho \cos\Theta \sin {}_i\Phi. \end{aligned} \tag{24}$$

Note that z is just a function of the area A. In the plane $z = 0$, $\Theta = 0$, A vanishes and the particles are in a collinear configuration. The remaining coordinates in the xy plane are then just those that have previously been found useful for the

description of the motion of three particles on a line.[1,7]

The symmetries of the potential $V(\rho, \Theta, \Phi)$ are important. They can be obtained from Eqs. (16) and (20) by noting that the potential is completely determined by $(A, {}_iu, {}_iv)$ and that it is invariant to reflection of the figure in a line, which replaces A by $-A$:

$$V(A, {}_iu, {}_iv) = V(-A, {}_iu, {}_iv). \tag{25}$$

Obviously Θ and ${}_i\Phi$ can each be changed by a integral multiple of π without affecting $(A, {}_iu, {}_iv)$ and A alone changes sign as a result of the replacements $\Theta \leftrightarrow -\Theta$ and $(\Theta, {}_i\Phi) \leftrightarrow (\Theta + \frac{1}{2}\pi, {}_i\Phi + \frac{1}{2}\pi$. Thus we have, if m and n are integers,

$$V(\rho, \Theta, {}_i\Phi) = V(\rho, \pm\Theta + m\pi, {}_i\Phi + n\pi)$$
$$= V[\rho, \pm\Theta + (m + \tfrac{1}{2})\pi, {}_i\Phi + (n + \tfrac{1}{2})\pi]. \tag{26}$$

c. Other Angular Coordinates

When considering a representation in which two angular momenta are uncoupled, the natural way of introducing a set of five angles is to express the vectors ${}_k\xi^a$ in the spherical polar coordinates $({}_kr_a, {}_k\theta_a, {}_k\varphi_a)$ and then define a new angle ${}_k\chi$ by the equations

$$({}_kr_a)^2 = \sum_{i=1}^{3} ({}_k\xi_i^a)^2, \quad {}_kr_1 = \rho \cos {}_k\chi, \quad {}_kr_2 = \rho \sin {}_k\chi. \tag{27}$$

Since we shall here be interested only in motion in the plane, from here on we can consider just the special case ${}_k\theta_1 = {}_k\theta_2 = \pi/2$. Then,

$${}_k\xi_1 = \rho \cos {}_k\chi \cos {}_k\varphi_1, \quad {}_k\xi_2 = \rho \cos {}_k\chi \sin {}_k\varphi_2,$$
$${}_k\xi_4 = \rho \sin {}_k\chi \cos {}_k\varphi_2, \quad {}_k\xi_5 = \rho \sin {}_k\chi \sin {}_k\varphi_2. \tag{28}$$

In order to separate internal and external coordinates, let us define

$${}_k\varphi_+ = \tfrac{1}{2}({}_k\varphi_1 + {}_k\varphi_2), \qquad {}_k\varphi_- = \tfrac{1}{2}({}_k\varphi_1 - {}_k\varphi_2). \tag{29}$$

The external coordinate is ${}_k\varphi_+$; the internal ones are $(\rho, {}_k\chi, {}_k\varphi_-)$. The transformation between the coordinates $(\Theta, {}_k\Phi)$ and $({}_k\chi, {}_k\varphi_-)$ appears in a most symmetric form if we define

$${}_k\chi' = \pi/4 - {}_k\chi, \qquad {}_k\varphi_-' = \pi/4 + {}_k\varphi_-. \tag{30}$$

Then

$$4A/\rho^2 = \sin 2\Theta = \cos 2\,{}_k\chi' \cos 2\,{}_k\varphi_-',$$
$${}_ku/\rho^2 = \cos 2\Theta \cos 2\,{}_k\Phi = \sin 2\,{}_k\chi',$$
$${}_kv/\rho = \cos 2\Theta \sin 2\,{}_k\Phi = \cos 2\,{}_k\chi' \sin 2\,{}_k\varphi_-'. \tag{31}$$

From these follow also

$$\tan 2\,{}_k\Phi = \cot 2\,{}_k\chi' \sin 2\,{}_k\varphi_-',$$
$$\tan 2\,{}_k\varphi_-' = \cot 2\Theta \sin 2\,{}_k\Phi. \tag{32}$$

The relation between φ and ${}_k\varphi_+$ can be found from

$$s/\rho^2 = \cos 2\Theta \cos 2\varphi = \sin 2\,{}_k\varphi_-' \cdot \cos 2\,{}_k\varphi_+$$
$$+ \sin 2\,{}_k\chi' \cos 2\,{}_k\varphi_-' \cdot \sin 2\,{}_k\varphi_+,$$
$$t/\rho^2 = \cos 2\Theta \sin 2\varphi = -\sin 2\,{}_k\chi' \cos 2\,{}_k\varphi_-'$$
$$\times \cos 2\,{}_k\varphi_+ + \sin 2\,{}_k\varphi_-' \cdot \sin 2\,{}_k\varphi_+. \tag{33}$$

${}_k\varphi_+$ and φ differ by an additive constant that depends on the internal configuration,

$${}_k\varphi_+ = \varphi + {}_k\varphi_+^0, \tag{34}$$

where

$$\cos 2\Theta \cos 2\,{}_k\varphi_+^0 = \sin 2\,{}_k\varphi_-'$$

and

$$\cos 2\Theta \sin 2\,{}_k\varphi_+^0 = \sin 2\,{}_k\chi' \cos 2\,{}_k\varphi_-', \tag{35}$$

or

$$\tan 2\,{}_k\varphi_+^0 = \cot 2\,{}_k\varphi_-' \sin 2\,{}_k\chi'$$
$$= \cot 2\,{}_k\Phi \sin 2\Theta. \tag{36}$$

The parallelism between the coordinate sets $(\Theta, {}_k\Phi, \varphi)$ and $({}_k\chi', {}_k\varphi_-, {}_k\varphi_+)$ will ultimately be exploited in the discussion of the eigenfunctions.

d. Domain of Integration

The domain of variation of the coordinates $({}_k\chi, {}_k\varphi_1, {}_k\varphi_2)$ is $0 \leq {}_k\chi \leq \pi/2$, $0 \leq {}_k\varphi_1, {}_k\varphi_2 \leq 2\pi$, and the angular element of integration is

$$d^3\omega = \tfrac{1}{2} \sin 2\,{}_k\chi \, d_k\chi \, d_k\varphi_1 \, d_k\varphi_2. \tag{37}$$

It is appropriate to use the same domain and angular element when $({}_k\varphi_+, {}_k\varphi_-)$ replace $({}_k\varphi_1, {}_k\varphi_2)$. As for the symmetric coordinates, similar conditions also apply for φ and ${}_k\Phi$, for which we can take $0 \leq {}_k\Phi$, $\varphi \leq 2\pi$, but the range of Θ must be $-\pi/4 \leq \Theta \leq \pi/4$ in order to span positive and negative values of A (see Eq. 21). The angular element is

$$d^3\Omega = \tfrac{1}{2} \cos 2\Theta \, d\Theta \, d_k\Phi \, d\varphi. \tag{38}$$

An integration over the whole domain covers the hypersphere in the space of $(\xi_1, \xi_2, \xi_4, \xi_5,)$. The angular part of the eigenfunctions will be normalized to unity over the hypersphere.

[7] F. T. Smith, J. Chem. Phys. **31**, 1352 (1959).

B. OPERATORS AND EIGENFUNCTIONS

1. Operators

a. General

The symmetric representation is generated by a special set of operators derived from the generalized angular-momentum tensor $\mathbf{\Lambda}$. This is defined in the normalized center-of-mass system by

$$ {}_m\Lambda_{ij} = {}_m\xi_i\ {}_m\pi_j - {}_m\xi_j\ {}_m\pi_i \qquad (i, j = 1, \cdots, 6) $$

or

$$ {}_m\Lambda_{ij}^{\alpha\beta} = {}_m\xi_i^{\alpha}\ {}_m\pi_j^{\beta} - {}_m\xi_i^{\beta}\ {}_m\pi_i^{\alpha} \qquad (i, j = 1, 2, 3;\ \ \alpha, \beta = 1, 2). \tag{39} $$

(In what follows, the prefix m will often be dropped where no confusion can result, and a single fixed value of m assumed.) The components of $\mathbf{\Lambda}$ obey the identity

$$ \Lambda_{ij}\Lambda_{kl} + \Lambda_{ik}\Lambda_{lj} + \Lambda_{il}\Lambda_{jk} = 0 \tag{40} $$

as well as the basic commutation rule

$$ [\Lambda_{ij}, \Lambda_{ik}] = i\hbar\Lambda_{jk}, $$

$$ [\Lambda_{ij}, \Lambda_{mk}] = 0 \qquad (i, j, m, k \ \text{ all unequal}). \tag{41} $$

All these operators commute with the kinetic energy T.

From the 3×3 tensors ${}_m\mathbf{\Lambda}^{\alpha\beta}$ one can construct the total ordinary angular momentum

$$ \mathbf{L} = {}_m\mathbf{\Lambda}^{11} + {}_m\mathbf{\Lambda}^{22}, \tag{42} $$

the symmetric togetherness tensor,

$$ \mathbf{\Sigma} = {}_m\mathbf{\Lambda}^{12} - {}_m\mathbf{\Lambda}^{21}, \tag{43} $$

and the two other antisymmetric tensors

$$ {}_m\mathbf{Y} = {}_m\mathbf{\Lambda}^{11} - {}_m\mathbf{\Lambda}^{22} $$

and

$$ {}_m\mathbf{A} = {}_m\mathbf{\Lambda}^{12} + {}_m\mathbf{\Lambda}^{21}. \tag{44} $$

Under the kinematic rotation of Eq. (10), which corresponds to an interchange of particles, $\mathbf{L}$ and $\mathbf{\Sigma}$ are invariant and ${}_m\mathbf{Y}$ and ${}_m\mathbf{A}$ transform like components of a vector:

$$ {}_k\mathbf{Y} = {}_m\mathbf{Y}\cos 2\beta_{mk} + {}_m\mathbf{A}\sin 2\beta_{mk} $$

$$ {}_k\mathbf{A} = -{}_m\mathbf{Y}\sin 2\beta_{mk} + {}_m\mathbf{A}\cos 2\beta_{mk}. \tag{45} $$

It is thus convenient to break up the 6×6 tensor in this way:

$$ {}_m\mathbf{\Lambda} = \mathbf{M} + {}_m\mathbf{N}, $$

with

$$ 2\mathbf{M} = \begin{Bmatrix} \mathbf{L} & \mathbf{\Sigma} \\ -\mathbf{\Sigma} & \mathbf{L} \end{Bmatrix} \text{ and } 2\ {}_m\mathbf{N} = \begin{Bmatrix} {}_m\mathbf{Y} & {}_m\mathbf{A} \\ {}_m\mathbf{A} & -{}_m\mathbf{Y} \end{Bmatrix}. \tag{46} $$

Most of these tensors are antisymmetric and have no trace, but they have another fundamental invariant that is essentially the trace of the square of the matrix; to conform to the usual notation for angular momenta, let us define for any tensor $\mathbf{B}$ the scalar square

$$ B^2 = \tfrac{1}{2}\sum_{i,j}(B_{ij})^2. \tag{47} $$

In addition, $\mathbf{\Sigma}$ has a nonvanishing trace which can be taken as

$$ \Sigma_t = \tfrac{1}{2}\sum_{i}(\Sigma_{ii}). \tag{48} $$

An important identity, derived from Eq. (2), is

$$ \Lambda^2 - L^2 - \Sigma^2 + \Sigma_t^2 = 0. \tag{49} $$

The symmetric representation involves a family of commuting operators derived from $\mathbf{L}$ and $\mathbf{\Sigma}$. For interactions in 3-space, five independent operators are needed besides the kinetic energy. These include Λ^2, L^2, L_z, Σ_t, and a scalar operator Σ_d. If the angular momentum $\mathbf{L}$ is written in vector form,

$$ L_i = \sum_{jk}\epsilon_{ijk}L_{jk}, \tag{50} $$

Σ_d is

$$ \Sigma_d = \mathbf{L}\cdot\mathbf{\Sigma}\cdot\mathbf{L} = \sum_{jk}(L_j\Sigma_{jk}L_k). \tag{51} $$

In the planar problem to which this paper is restricted, three operators are needed. These are Λ^2, L_z, and Σ_t. In the collinear problem the only angular operator is Λ; it behaves just like a rotation operator in the plane, with eigenvalues $\pm\lambda\hbar$ and eigenfunctions $\exp(\pm i\lambda\chi)$.

b. The Planar Problem

The following operators enter into this problem:

$$ L = \Lambda_{12} + \Lambda_{45}, \quad Y = \Lambda_{12} - \Lambda_{45}, \quad A = \Lambda_{15} - \Lambda_{24}, $$

$$ \Sigma_t = \tfrac{1}{2}(\Sigma_{11} + \Sigma_{22}) = \Lambda_{14} + \Lambda_{25}, $$

$$ \Sigma_- = \tfrac{1}{2}(\Sigma_{11} - \Sigma_{22}) = \Lambda_{14} - \Lambda_{25}, $$

$$ \Sigma_{12} = \Lambda_{15} + \Lambda_{24}. \tag{52} $$

They obey the identities

$$ \Lambda^2 = L^2 + \Sigma_-^2 + \Sigma_{12}^2 = \Sigma_t^2 + Y^2 + A^2, \tag{53} $$

[proved by expansion and use of Eq. (40)]. With

regard to the comutation rules, they fall into two families that may be written as vectors:

$$\mathbf{P} = (L, \Sigma_-, \Sigma_{12}),$$
$$\mathbf{Q} = (\Sigma_t, Y, A). \tag{54}$$

Members of one group commute with those of the other, and within a group they behave much like ordinary angular momenta:

$$[P_i, Q_i] = [P_i, Q_i] = 0,$$
$$[P_i, P_j] = 2i\hbar P_k, \qquad [Q_i, Q_j] = 2i\hbar Q_k, \tag{55}$$

if (ijk) is a cyclic permutation of (123). This is a case of a well-known property of the 4-dimensional rotation group (see, for instance, Biedenharn[6]). All these operators, of course, commute with Λ^2. By Eq. (53),

$$\Lambda^2 = \sum_i P_i^2 = \sum_i Q_i^2. \tag{56}$$

Any set (Λ^2, P_i, Q_j) is a suitable family of commuting operators for the angular part of the planar 3-body problem. All the operators of $\mathbf{P}$ are invariant under the kinematic rotation, Eq. (10), while an ordinary spatial rotation through an angle θ leaves only L invariant and changes the pair (Σ_-, Σ_{12}) like a 2-vector rotated through the angle 2θ. All the operators of $\mathbf{Q}$ are invariant under the ordinary spatial rotation, but a kinematic rotation transforms (Y, A) like a 2-vector by Eq. (45) and leaves Σ_t invariant. The most symmetric representation therefore uses the family of operators (Λ^2, L, Σ_t). However, the set $(\Lambda^2, L, {}_mY)$ is also important since it generates the representation in which two angular momenta are uncoupled:

$${}_mL_1 = {}_m\Lambda_{12} = \tfrac{1}{2}(L + {}_mY),$$
$${}_mL_2 = {}_m\Lambda_{45} = \tfrac{1}{2}(L - {}_mY). \tag{57}$$

The representation $(\Lambda^2, L, {}_mY) \doteq (\Lambda^2, {}_mL_1, {}_mL_2)$ is obviously most useful when two particles are interacting and the third, m, is far away, while the (Λ^2, L, Σ_t) representation is appropriate when all three are far apart or close together.

Just as with ordinary angular momenta, the relations (55) and (56) lead us to define the raising and lowering operators associated with any diagonal operator from $\mathbf{P}$ or $\mathbf{Q}$. Associated with P_i are

$$P_i^* = P_j \pm iP_k. \tag{58}$$

Then we find

$$[P_i^+, P_i^-] = 4\hbar P_i, \qquad [P_i^*, P_i] = \mp 2\hbar P_i^*, \tag{59}$$
$$[P_i^*, \Lambda^2] = [P_i^*, Q_i] = 0,$$

and

$$P_i^* P_i^* = \Lambda^2 - P_i^2 \pm 2\hbar P_i. \tag{60}$$

If the eigenfunctions are labeled by (Λ^2, P_i, Q_j), we can write

$$P_i\psi(\Lambda^2, p_i, q_i) = p_i\hbar\psi(\Lambda^2, p_i, q_i). \tag{61}$$

Because of Eq. (59), we have

$$P_i^*\psi(\Lambda^2, p_i, q_i) = B_i^*(\Lambda^2, p_i \pm 1, q_i)\psi(\Lambda^2, p_i \pm 2, q_i), \tag{62}$$

where B is a normalizing factor to be evaluated presently. Applying one of these operators, say P_i^+, repeatedly, we get a ladder of solutions that terminates at an upper bound, say $p_i(\max) = \lambda$, since $\Lambda^2 - P_i^2 \geq 0$ by Eq. (56). Then

$$P_i^+\psi(\Lambda^2, \lambda, q_i) = 0,$$

and

$$\Lambda^2\psi(\Lambda^2, \lambda, q_i) = (P_i^2 + 2\hbar P_i + P_i^- P_i^+)\psi(\Lambda^2, \lambda, q_i)$$
$$= \hbar^2\lambda(\lambda + 2)\psi(\Lambda^2, \lambda, q_i). \tag{63}$$

Similarly, applying P_i^- we arrive at a lowest value p_i (min) $= -\lambda'$, with $\Lambda^2 = \hbar^2\lambda'(\lambda' + 2)$. Since Λ^2 commutes with P_i and P_i^* we must have $\lambda = \lambda'$; and λ must be an integer since the steps of the ladder are $\pm 2\hbar$. The permissible values of p_i are thus even or odd along with λ.

To establish the value of the normalizing factor B it suffices to note that Eq. (60) when applied to any eigenfunction $\psi(\lambda, p_i, q_i)$ must lead to an identity. We find that B is independent of q_i, and its value is

$$B(\lambda, p_i + 1) = B_i^+(\lambda, p_i + 1, q_i)$$
$$= B_i^-(\lambda, p_i + 1, q_i)$$
$$= \hbar[\lambda(\lambda + 2) - p_i(p_i + 2)]^{1/2}$$
$$= \hbar[(\lambda - p_i)(\lambda + p_i + 2)]^{1/2}. \tag{64}$$

Symmetry shows that the same argument applies when Q_j^* is used to raise or lower q_j. Thus, p_i and q_j run independently through the even or odd integers from λ to $-\lambda$. For each value of λ we get just $\omega(\lambda) = (\lambda + 1)^2$ independent solutions. They are related to each other by

$$\psi(\lambda, p_i \pm 2, q_i) = B^{-1}(\lambda, p_i \pm 1)P_i^*\psi(\lambda, p_i, q_i)$$

and

$$\psi(\lambda, p_i, q_i \pm 2) = B^{-1}(\lambda, q_i \pm 1)Q_i^*\psi(\lambda, p_i, q_i). \tag{65}$$

For future use these can be rewritten to give the matrix elements of P_i^* and Q_i^*:

$$(\lambda, p_i \pm 2, q_i \,|P_i^*|\, \lambda, p_i, q_i) = \hbar[(\lambda \mp p_i)(\lambda \pm p_i + 2)]^{1/2}$$

$$(\lambda, p_i, q_i \pm 2 \,|Q_i^*|\, \lambda, p_i, q_i) = \hbar[(\lambda \mp q_i)(\lambda \pm q_i + 2)]^{1/2}. \quad (66)$$

In Eq. (65) the arbitrary phase factor, which might be assigned differently to each wave function, has been taken as unity in each case. This choice will prove convenient, and Eq. (65) defines the convention on relative phases that I shall regularly use.

2. The Eigenfunctions

To construct the eigenfunctions I shall copy the now standard procedure of H. P. Robertson[8] for generating the spherical harmonics.

For the (L, Σ_t) representation, the operators are best expressed in the coordinates $(\Theta, {}_k\Phi, \varphi)$. The transformation is tedious but straightforward (see Appendix), and the result is

$$L = P_1 = \frac{\hbar}{i}\frac{\partial}{\partial\varphi},$$

$$P_1^* = \Sigma_- \pm i\Sigma_{12} = \frac{\hbar}{i}e^{\pm i2\varphi} \times \left(\sec 2\Theta \frac{\partial}{\partial {}_k\Phi} + \tan 2\Theta \frac{\partial}{\partial\varphi} \pm i\frac{\partial}{\partial\Theta}\right),$$

$$\Sigma_t = Q_1 = \frac{\hbar}{i}\frac{\partial}{\partial\, {}_k\Phi},$$

$$Q_{1\pm}^* = Y \pm iA = \frac{\hbar}{i}e^{\pm i2{}_k\Phi} \times \left(\sec 2\Theta \frac{\partial}{\partial\varphi} + \tan 2\Theta \frac{\partial}{\partial\, {}_k\Phi} \pm i\frac{\partial}{\partial\Theta}\right). \quad (67)$$

In these coordinates,

$$\Lambda^2 = -\hbar^2\left[\sec 2\Theta \frac{\partial}{\partial\Theta}\left(\cos 2\Theta \frac{\partial}{\partial\Theta}\right) + \sec^2 2\Theta\left(\frac{\partial^2}{\partial\, {}_k\Phi^2} + 2\sin 2\Theta \frac{\partial^2}{\partial\, {}_k\Phi\, \partial\varphi} + \frac{\partial^2}{\partial\varphi^2}\right)\right]$$

$$= -\hbar^2 \sec 2\Theta \frac{\partial}{\partial\Theta}\left(\cos 2\Theta \frac{\partial}{\partial\Theta}\right) + \sec^2 2\Theta\left(\sum_i^2 + 2\sin 2\Theta \sum_i L + L^2\right). \quad (68)$$

[8] See E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra* (Cambridge University Press, New York, 1935), p. 50 ff.

For the (L, Y) representation, the coordinates $({}_k\chi, {}_k\varphi_+, {}_k\varphi_-)$ are most convenient. (As long as only one value of k is under consideration, that subscript can be omitted on the coordinates and on ${}_k\Phi$.)

$$L = P_1 = \frac{\hbar}{i}\frac{\partial}{\partial\varphi_+},$$

$$P_1^* = \frac{\hbar}{i}e^{\pm 2i(\varphi_+ - \pi/4)}\left(\csc 2\chi \frac{\partial}{\partial\varphi_-} - \cot 2\chi \frac{\partial}{\partial\varphi_+} \pm i\frac{\partial}{\partial\chi}\right),$$

$$Y = Q_2 = \frac{\hbar}{i}\frac{\partial}{\partial\varphi_-},$$

$$Q_2^* = \frac{\hbar}{i}e^{\pm 2i\varphi}\left(\csc 2\chi \frac{\partial}{\partial\varphi_+} - \cot 2\chi \frac{\partial}{\partial\varphi_-} \pm i\frac{\partial}{\partial\chi}\right). \quad (69)$$

Λ^2 becomes

$$\Lambda^2 = -\hbar^2\left[\csc 2\chi \frac{\partial}{\partial\chi}\left(\sin 2\chi \frac{\partial}{\partial\chi}\right) + \csc^2 2\chi\left(\frac{\partial^2}{\partial\varphi_+^2} - 2\cos 2\chi \frac{\partial^2}{\partial\varphi_+\, \partial\varphi_-} + \frac{\partial^2}{\partial\varphi_-^2}\right)\right]$$

$$= -\hbar^2 \csc 2\chi \frac{\partial}{\partial\chi}\left(\sin 2\chi \frac{\partial}{\partial\chi}\right) + \csc^2 2\chi(L^2 - 2\cos 2\chi LY + Y^2). \quad (70)$$

The construction of the eigenfunctions is simplified if the variable Θ in Eq. (67) is replaced by

$$\Theta' = \Theta + \pi/4, \quad (71)$$

so that the operators become

$$P_1^* = \frac{\hbar}{i}e^{\pm i2\varphi}\left(\csc 2\Theta' \frac{\partial}{\partial\Phi} - \cot 2\Theta' \frac{\partial}{\partial\varphi} \pm i\frac{\partial}{\partial\Theta'}\right)$$

and

$$Q_1^* = \frac{\hbar}{i}e^{\pm i2\Phi}\left(\csc 2\Theta' \frac{\partial}{\partial\varphi} - \cot 2\Theta' \frac{\partial}{\partial\Phi} \pm i\frac{\partial}{\partial\Theta'}\right). \quad (72)$$

The similarity with Eq. (31) will be reflected in the form of the eigenfunctions.

Let us now write the eigenvalues of the various operators as

$$L = \hbar p_1 = \hbar m, \qquad \Sigma_t = \hbar q_1 = \hbar s,$$
$$Y = \hbar q_2 = \hbar m_-,$$
$$L_1 = \Lambda_{12} = \hbar m_1 = \tfrac{1}{2}\hbar(m + m_-),$$
$$L_2 = \Lambda_{45} = \hbar m_2 = \tfrac{1}{2}\hbar(m - m_-). \quad (73)$$

The eigenfunctions can be written as products of simple functions if the appropriate coordinates are used in each case:

$$\psi_{\lambda,m,s}(\Theta', \Phi, \varphi) = (2\pi)^{-1}e^{im\varphi}e^{is\Phi}f_{\lambda,m,s}(\Theta'),$$

$$\psi_{\lambda,m,m_-}(\chi, \varphi_+, \varphi_-)$$
$$= (2\pi)^{-1}e^{im(\varphi_+ - \pi/4)}e^{im_-\varphi_-}f_{\lambda,m,m_-}(\chi)$$
$$= \psi_{\lambda,m_1,m_2}(\chi, \varphi_1, \varphi_2)$$
$$= (2\pi)^{-1}e^{im_1(\varphi_1-\pi/4)}e^{im_2(\varphi_2-\pi/4)}$$
$$\times f_{\lambda,(m_1+m_2),(m_1-m_2)}(\chi). \tag{74}$$

The functions $f_{\lambda,p,q}$ are connected by a recurrence relation deduced by combining Eq. (65) with Eq. (69) or (72):

$$f_{\lambda,p\pm2,q}(\theta) = [(\lambda \mp p)(\lambda \pm p + 2)]^{-1/2}$$
$$\times (q \csc 2\theta - p \cot 2\theta \pm d/d\theta)f_{\lambda,p,q}(\theta)$$
$$= \mp 4[(\lambda \mp p)(\lambda \pm p + 2)]^{-1/2}$$
$$\times \sin^{\pm\frac{1}{2}(p\pm2-q)}\theta \cos^{\pm\frac{1}{2}(p\pm2+q)}\theta$$
$$\times (d/d \cos 2\theta)[\sin^{\mp\frac{1}{2}(p-q)}\theta$$
$$\times \cos^{\mp\frac{1}{2}(p+q)}\theta f_{\lambda,p,q}(\theta)]. \tag{75}$$

The same relation applies if p and q are interchanged.

The particular function $f_{\lambda\lambda\lambda}(\theta)$ can be found from the circumstance that $P^+\psi_{\lambda\lambda\lambda} = 0$, which leads to the equation

$$(d/d \cos 2\theta)[\cos^{-\lambda}\theta f_{\lambda\lambda\lambda}(\theta)] = 0. \tag{76}$$

Its solution is

$$f_{\lambda\lambda\lambda}(\theta) = (2\lambda + 2)^{1/2}\cos^{\lambda}\theta, \tag{77}$$

where the normalization is chosen so that

$$\tfrac{1}{2}\int_0^{\pi/2} f_{\lambda\lambda\lambda}{}^2(\chi) \sin 2\chi \, d\chi$$
$$= \tfrac{1}{2}\int_{-\pi/4}^{\pi/4} f_{\lambda\lambda\lambda}{}^2(\Theta + \pi/4)\cos 2\Theta \, d\Theta = 1. \tag{78}$$

Applying the lowering operation, Eq. (75), $\frac{1}{2}(\lambda - p)$ times to $f_{\lambda\lambda\lambda}$ one gets

$$f_{\lambda,p,\lambda}(\theta) = \left[\frac{2(\lambda+1)!}{\left(\frac{\lambda-p}{2}\right)!\left(\frac{\lambda+p}{2}\right)!}\right]^{1/2}$$
$$\times \sin^{\frac{1}{2}(\lambda-p)}\theta \cos^{\frac{1}{2}(\lambda+p)}\theta, \tag{79}$$

and a similar expression for $f_{\lambda,\lambda,q}$. In the same way the second index can be lowered $\frac{1}{2}(\lambda - q)$ times, giving

$$f_{\lambda,p,q}(\theta) = \left[2(\lambda+1)\left(\frac{\lambda-q}{2}\right)!\left(\frac{\lambda+q}{2}\right)!\right.$$
$$\left.\times\left(\frac{\lambda-p}{2}\right)!\left(\frac{\lambda+p}{2}\right)!\right]^{\frac{1}{2}}\sin^{\lambda-\frac{1}{2}(p+q)}\theta\cos^{\frac{1}{2}(p+q)}\theta$$
$$\times \sum_{j=0}^{\frac{1}{2}(\lambda-p)}(-1)^j\left[j!\left(\frac{p+q}{2}+j\right)!\right.$$
$$\left.\times\left(\frac{\lambda-q}{2}-j\right)!\left(\frac{\lambda-p}{2}-j\right)!\right]^{-1}\cot^{2j}\theta. \tag{80}$$

Note that the upper limit of the summation is $\frac{1}{2}(\lambda - q)$ if $q > p$—but this is automatically taken care of by the factorials. Likewise if $p + q$ is negative the lower limit of the summation is $j = -\frac{1}{2}(p + q)$. From Eq. (80) some special cases can be quickly derived:

$$f_{\lambda,-\lambda,-\lambda} = (-1)^{\lambda}[2(\lambda+1)]^{1/2}\cos^{\lambda}\theta = (-1)^{\lambda}f_{\lambda\lambda\lambda}(\theta),$$
$$f_{\lambda,-\lambda,\lambda} = f_{\lambda,\lambda,-\lambda} = [2(\lambda+1)]^{1/2}\sin^{\lambda}\theta, \tag{81}$$
$$f_{\lambda,p,q} = f_{\lambda,q,p};$$

also,

$$f_{000} = 2^{1/2},$$
$$f_{111} = -f_{1,-1,-1} = 2\cos\theta,$$
$$f_{1,-1,1} = f_{1,1,-1} = 2\sin\theta,$$
$$f_{222} = f_{2,-2,-2} = 6^{1/2}\cos^2\theta,$$
$$f_{2,0,2} = f_{2,2,0} = 2 \times 3^{1/2}\sin\theta\cos\theta$$
$$= -f_{2,0,-2} = -f_{2,-2,0},$$
$$f_{2,2,-2} = f_{2,-2,2} = 6^{1/2}\sin^2\theta,$$
$$f_{2,0,0} = 6^{1/2}(\sin^2\theta - \cos^2\theta). \tag{82}$$

The functions $f_{\lambda,p,q}$ are simply related to the Jacobi polynomials (compare references 1 and 2). As in the case of the ordinary spherical harmonic function $\Theta_{lm}(\theta)$, they are not in general orthogonal to each other except when p and q are identical and only λ changes.

3. Coupling Coefficients

The explicit form of the eigenfunctions in the representations (Λ^2, L, Σ_i), (Λ^2, L, Y), and (Λ^2, L_1, L_2) have been found, and Eq. (74) gives the connection between the latter two, which involves merely a change in phase. This section will be devoted to finding the coupling coefficients connecting the first two representations.

The coupling coefficients $a_\lambda(m_-, s)$ are defined by the equation

$$\psi_{\lambda,m,s} = \sum_{m_-}\psi_{\lambda,m,m_-}a_\lambda(m_-, s); \tag{83}$$

the normalization of the wave functions requires that $\mathbf{a}_\lambda$ be a unitary matrix. That $\mathbf{a}_\lambda$ is actually independent of m can be shown by operating on both sides of Eq. (83) by P_1.

A connection between the rows of $\mathbf{a}_\lambda$ can now be found. Using Eq. (65) to obtain $\psi_{\lambda,m,s}$ from $\psi_{\lambda,m,\lambda}$ we get

$$\psi_{\lambda,m,s} = (2\hbar)^{-\frac{1}{2}(\lambda-s)} \times \left[\frac{\left(\frac{\lambda+s}{2}\right)!}{\lambda!\left(\frac{\lambda-s}{2}\right)!}\right]^{\frac{1}{2}} (Q_1^-)^{\frac{1}{2}(\lambda-s)}\psi_{\lambda,m,\lambda}. \quad (84)$$

Substituting on both sides from Eq. (83), multiplying by $\psi^*_{\lambda,m,m_-'}$, and integrating, we find

$$a_\lambda(m_-, s) = (2\hbar)^{-\frac{1}{2}(\lambda-s)}\left[\frac{\left(\frac{\lambda+s}{2}\right)!}{\lambda!\left(\frac{\lambda-s}{2}\right)!}\right]^{\frac{1}{2}} \times \sum_{m_-'} (\lambda, m, m_- |(Q_1^-)^{\frac{1}{2}(\lambda-s)}| \lambda, m, m_-')a_\lambda(m_-', s). \quad (85)$$

The matrix elements can be got by using the fact that

$$Q_1^- = Y - iA = Y - \tfrac{1}{2}i(Q_2^+ + Q_2^-), \quad (86)$$

and applying Eq. (28).

In just the same way one can define the inverse transformation by

$$\psi_{\lambda,m,m_-} = \sum_s \psi_{\lambda,m,s}b_\lambda(s, m_-); \quad (87)$$

but since $\mathbf{a}_\lambda$ is unitary,

$$\mathbf{b}_\lambda = \mathbf{a}_\lambda^{-1} = \mathbf{a}_\lambda^\dagger. \quad (88)$$

A connection between the columns of $\mathbf{a}_\lambda$ is deduced in the same way as Eq. (85):

$$a_\lambda(m_-, s) = (2\hbar)^{-\frac{1}{2}(\lambda-m_-)}\left[\frac{\left(\frac{\lambda+m_-}{2}\right)!}{\lambda!\left(\frac{\lambda-m_-}{2}\right)!}\right]^{\frac{1}{2}} \times \sum_{s'} (\lambda, m, s |(Q_2^-)^{\frac{1}{2}(\lambda-m-)}| \lambda, m, s')a_\lambda(m_-, s'). \quad (89)$$

Here

$$Q_2^- = A - i\Sigma_t = -i[\Sigma_t + \tfrac{1}{2}(Q_1^+ - Q_1^-)]. \quad (90)$$

To obtain the explicit form of $\mathbf{a}_\lambda$ we can use the fact that

$$Y = \tfrac{1}{2}(Q_1^+ + Q_1^-) \quad \text{and} \quad \Sigma_t = \tfrac{1}{2}i(Q_2^- - Q_2^+). \quad (91)$$

Applying the operator Σ_t to Eq. (83), multiplying by $\psi^*_{\lambda m m_-}$, and integrating, one finds

$$\hbar s a_\lambda(m_-, s) = \sum_{m_-'} (\lambda m m_- |\Sigma_t| \lambda m_-')a_\lambda(m_-', s)$$

$$= \tfrac{1}{2}i\hbar[(\lambda - m_-)(\lambda + m_- + 2)]^{1/2}a_\lambda(m_- + 2, s)$$

$$- \tfrac{1}{2}i\hbar[(\lambda + m_-)(\lambda - m_- + 2)]^{1/2}a_\lambda(m_- - 2, s). \quad (92)$$

Starting with $a_\lambda(\lambda, s)$, the successive coefficients have the recurrence relation in m_-:

$$a_\lambda(m_- - 2, s) = i2s[(\lambda + m_-)(\lambda - m_- + 2)]^{-1/2}a_\lambda(m_-, s) + \left[\frac{(\lambda - m_-)(\lambda + m_- + 2)}{(\lambda + m_-)(\lambda - m_- + 2)}\right]^{1/2} a_\lambda(m_- + 2, s). \quad (93)$$

For each value of s, these are to be normalized by

$$\sum_{|m-|\leq\lambda} |a_\lambda(m_-, s)|^2 = 1, \quad (94)$$

leaving only a phase arbitrary.

Similarly, starting from Eq. (87), we can find a recurrence relation in s:

$$a_\lambda(m_-, s - 2) = \frac{2m_-}{[(\lambda + s)(\lambda - s + 2)]^{1/2}}\, a_\lambda(m_-, s) - \left[\frac{(\lambda - s)(\lambda + s + 2)}{(\lambda + s)(\lambda - s + 2)}\right]^{1/2} a_\lambda(m_-, s + 2), \quad (95)$$

and the normalization

$$\sum_{|s|\leq\lambda} |a_\lambda(m_-, s)|^2 = 1. \quad (96)$$

We can now write

$$a_\lambda(m_-, s) = e^{i\alpha_\lambda}e^{i m_- \pi/4}b_\lambda(m_-, s), \quad (97)$$

where

$$b_\lambda(m_-, s) = b_\lambda(s, m_-) \quad (98)$$

is real and satisfies equations of the form (95) in both m_- and s. α_λ is a phase to be evaluated presently. The recurrence relation in s is satisfied if $b_\lambda(m_-, -s) = \pm b_\lambda(m_-, s)$, but that in m_- requires that

$$b_\lambda(m_-, -s) = (-1)^{\frac{1}{2}(\lambda-m_-)}b_\lambda(m_-, s). \quad (99)$$

To get numerical values for the coefficients, it is convenient to write $\mu = \frac{1}{2}(\lambda - s)$, $\nu = \frac{1}{2}(\lambda - m_-)$ and

$$b_\lambda(m_-, s) = b_\lambda(\lambda - 2\nu, \lambda - 2\mu) = \left[\binom{\lambda}{\nu}\binom{\lambda}{\mu}2^{-\lambda}\right]^{1/2} j(\lambda; \nu, \mu). \quad (100)$$

$j(\lambda; \nu, \mu)$ satisfy the relations

$$j(\lambda; \nu, \mu + 1) = \frac{\lambda - 2\nu}{\lambda - \mu}\, j(\lambda; \nu, \mu) - \frac{\mu}{\lambda - \mu}\, j(\lambda; \nu, \mu - 1)$$

and

$$\sum_{\mu=0}^{\lambda} \frac{j^2(\lambda; \nu, \mu)}{\mu!\,(\lambda - \mu)} = 2^\lambda \nu!\,(\lambda - \nu)!\,. \quad (101)$$

The following symmetry properties follow from Eq. (99):

$$j(\lambda;\mu,\nu) = j(\lambda;\nu,\mu) = (-1)^{\nu} j(\lambda;\nu,\lambda-\mu)$$
$$= (-1)^{\mu} j(\lambda;\lambda-\nu,\mu)$$
$$= (-1)^{\lambda+\mu+\nu} j(\lambda;\lambda-\nu,\lambda-\mu). \quad (102)$$

One also can easily show that

$$j(\lambda;\nu,0) = j(\lambda;0,\mu) = 1,$$

so that

$$j(\lambda;\nu,\lambda) = (\lambda;\lambda,\nu) = (-1)^{\nu}. \quad (103)$$

For each value of λ, the matrix of j's can be constructed—taking advantage of the symmetry, the whole matrix follows from a single octant. The nth row or column is characterized by having the nth differences vanish. Here are explicit expressions for the first few rows:

$$j(\lambda;\nu,0) = 1$$
$$j(\lambda;\nu,1) = 1 - \frac{2\nu}{\lambda}$$
$$j(\lambda;\nu,2) = 1 - \frac{4\nu}{\lambda-1} + \frac{4\nu^2}{\lambda(\lambda-1)}. \quad (104)$$

A tabulation for small values of λ is given in Table I.

It remains to fix the phase α_λ of Eq. (97). For this it suffices to ensure that Eq. (83) is satisfied for a single value of the coordinates.This may be taken at the point ($\xi_1 = \rho; \xi_2 = \xi_4 = \xi_5 = 0$), where

$$\chi = \varphi_+ = \varphi_- = 0$$
and $\Theta = \varphi = \Phi = 0$, or $\Theta' = \pi/4$. (105)

We need only take the case $s = l = \lambda$, for which

$$\psi_{\lambda\lambda\lambda}(0,0,\pi/4) = (2\pi)^{-1} f_{\lambda\lambda\lambda}(\pi/4)$$
$$= (2\pi)^{-1}(2\lambda+2)^{1/2} 2^{-\lambda/2}. \quad (106)$$

This is expanded in terms of the functions

$$\psi_{\lambda\lambda m_-}(0,0,0) = (2\pi)^{-1} e^{-i\lambda\pi/4} f_{\lambda\lambda m_-}(0)$$
$$= (2\pi)^{-1} e^{-i\lambda\pi/4} f_{\lambda\lambda\lambda}(0)\ \delta_{m_-,\lambda}$$
$$= (2\pi)^{-1} e^{-i\lambda\pi/4}(2\lambda+2)^{1/2}\ \delta_{m_-,\lambda}. \quad (107)$$

Introducing the expansion, Eq. (83), and the expressions (97), (101), we find

$$\alpha_\lambda = 0. \quad (108)$$

This completes the definition of the coupling coefficients.

4. Conclusion

The operators, eigenfunctions, and coupling coefficients have now been obtained for the symmetric representation of three particles in a plane and for an asymmetric representation involving a preliminary coupling between two particles and a second coupling of the third to their center of mass. The index k, identifying the third particle, is an implied label on the coordinates, operators, eigenvalues, and eigenfunctions of the asymmetric representation. In the symmetric representation the diagonal operators (Λ^2, L, Σ_t) are independent of k, as are the coordinates Θ and φ, but the coordinate ${}_k\Phi$ depends on k through the additive relation of Eq. (22),

$${}_k\Phi = \beta_{ki} + {}_i\Phi. \quad (109)$$

Thus the wave functions are not completely indeendent of k in the symmetric representation, but they depend on it only through a phase factor:

$${}_k\psi_{\lambda,m,s}(\Theta,\varphi,{}_k\Phi) = e^{is\beta_{kj}}\, {}_i\psi_{\lambda,m,s}(\Theta,\varphi,{}_i\Phi). \quad (110)$$

Using the coupling coefficients $a_\lambda(m_-, s)$ and Eq. (110) the connection between the asymmetric

TABLE I. The coefficients $j(\lambda;\mu,\nu)$.

	μ \ ν	0	1	2	3	4	5	6	7
λ = 0	0	1							
λ = 1	0	0	1						
	1	1	−1						
λ = 2	0	1	1	1					
	1	1	0	−1					
	2	1	−1	1					
λ = 3	0	1	1	1	1				
	1	1	1/3	−2/3	−1				
	2	1	−1/3	−2/3	1				
	3	1	−1	1	−1				
λ = 4	0	1	1	1	1	1			
	1	1	1/2	0	−1/2	−1			
	2	1	0	−1/3	0	1			
	3	1	−1/2	0	1/2	−1			
	4	1	−1	1	−1	1			
λ = 5	0	1	1	1	1	1	1		
	1	1	3/5	1/5	−1/5	−3/5	−1		
	2	1	1/5	−1/5	−1/5	1/5	1		
	3	1	−1/5	−1/5	1/5	1/5	−1		
	4	1	−3/5	1/5	1/5	−3/5	1		
	5	1	−1	1	−1	1	−1		
λ = 6	0	1	1	1	1	1	1	1	
	1	1	2/3	1/3	0	−1/3	−2/3	−1	
	2	1	1/3	−1/15	−1/5	−1/15	1/3	1	
	3	1	0	−1/5	0	1/5	0	−1	
	4	1	−1/3	−1/15	1/5	−1/15	−1/3	1	
	5	1	−2/3	1/3	0	−1/3	2/3	−1	
	6	1	−1	1	−1	1	−1	1	
λ = 7	0	1	1	1	1	1	1	1	1
	1	1	5/7	3/7	1/7	−1/7	−3/7	−5/7	−1
	2	1	3/7	1/21	−1/7	−1/7	1/21	1/7	1
	3	1	1/7	−1/7	−3/35	3/35	1/7	−1/7	−1
	4	1	−1/7	−1/7	3/35	3/35	−1/7	−1/7	1
	5	1	−3/7	1/21	1/7	−1/7	−1/21	3/7	−1
	6	1	−5/7	3/7	−1/7	−1/7	3/7	−5/7	1
	7	1	−1	1	−1	1	−1	1	−1

representations involving different values of k can be found. To begin with,

$$ {}_k\psi_{\lambda,m,{}_km_-}({}_k\chi,\ {}_k\varphi_+,\ {}_k\varphi_-) = \sum_s {}_k\psi_{\lambda,m,s}(\Theta,\ \varphi,\ {}_k\Phi)a^\dagger(s,\ {}_km_-), \tag{111} $$

$$ {}_i\psi_{\lambda,m,s}(\Theta,\ \varphi,\ {}_i\Phi) = \sum_{{}_im_-} {}_i\psi_{\lambda,m,{}_im_-}({}_i\chi,\ {}_i\varphi_+,\ {}_i\varphi_-)a_\lambda({}_im_-,\ s). \tag{112} $$

Hence

$$ {}_k\psi_{\lambda,m,{}_km_-}({}_k\chi,\ {}_k\varphi_+,\ {}_k\varphi_-) = \sum_{{}_im_-} {}_i\psi_{\lambda,m,{}_im_-}({}_i\chi,\ {}_i\varphi_+,\ {}_i\varphi_-)c_\lambda({}_im_-,\ {}_km_-) \tag{113} $$

where the coupling coefficients are

$$ c_\lambda({}_im_-,\ {}_km_-) = \sum_s a_\lambda({}_im_-,\ s)e^{is\beta_{ki}}a_\lambda^\dagger(s,\ {}_km_-). \tag{114} $$

As we have seen, the wave functions have the same structure in all these representations, though the arguments of the functions and the quantum numbers are variously defined. The same functional form appears in still other problems involving three independent angular variables. For instance, the complete angular wave functions are closely related to Wigner's representation coefficients for the 3-dimensional rotation group[9,10]:

$$ \psi_{\lambda,m,s}(\Theta',\ \varphi,\ \Phi) = \frac{[2(\lambda+1)]^{1/2}}{2\pi} \times \mathfrak{D}^{(\lambda/2)}_{m/2,-s/2}(2\varphi,\ \pi - 2\Theta',\ -2\Phi). \tag{115} $$

The well-known connection between these functions and the symmetric rotor wavefunctions also comes to mind, and the operational derivation of Sec. B.2 resembles the procedure used by Shaffer[11] in the latter problem. It follows that operators raising and lowering the index λ can be constructed by the method of Shaffer and Louck.[12]

Functions of three angles with a different structure appear in other problems involving rotations in a 4-dimensional space.[13] Two different types of hyperspherical harmonics are possible in 4-space because there are two structurally different ways to set up angular parameters in that space.[6] The possibilities increase with higher dimensions: there are three distinct parameterizations in 5-space, and six in 6-space.

[9] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), especially p. 167.
[10] I am indebted to Professor C. F. Curtiss for suggesting this connection.
[11] W. H. Shaffer, J. Mol. Spectroscopy **1**, 69 (1957).
[12] W. H. Shaffer and J. D. Louck, J. Mol. Spectroscopy **3**, 123 (1959).
[13] V. Fock, Z. Physik **98**, 145 (1935); A. Z. Dolginov, Soviet Phys.—JETP **3**, 589 (1956).

ACKNOWLEDGMENTS

It is a pleasure to acknowledge helpful conversations with my colleagues Dr. Joel L. Brenner and Dr. Robert C. Whitten. This work would have been impossible without the loyal support of Dr. Clinton M. Kelley and of Stanford Research Institute's Physical and Biological Sciences Divisional Research Committee, under whose auspices earlier work in this area was done.

APPENDIX
REPRESENTATIONS OF THE OPERATORS

1. The Symmetric Representation

We seek the explicit representation of the six angular-momentum operators of Eq. (52) in terms of the symmetric coordinate system $(\rho, \Theta, {}_k\Phi, \varphi)$. The operators are initially described in Cartesian coordinates $(\xi_1, \xi_2, \xi_4, \xi_5)$ as linear combinations of the

$$ \Lambda_{ij} = \frac{\hbar}{i}\left(\xi_i\frac{\partial}{\partial\xi_j} - \xi_j\frac{\partial}{\partial\xi_i}\right). \tag{A1} $$

Let us write

$$ \alpha_0 = \rho, \quad \alpha_1 = \Theta, \quad \alpha_2 = {}_k\Phi, \quad \alpha_3 = \varphi. \tag{A2} $$

Using the chain rule

$$ \frac{i}{\hbar}\Lambda_{ij} = \sum_k \lambda_{k;ij}\frac{\partial}{\partial\alpha_k}, \tag{A3} $$

where

$$ \lambda_{k;ij} = \xi_i\frac{\partial\alpha_k}{\partial\xi_j} - \xi_j\frac{\partial\alpha_k}{\partial\xi_i}. \tag{A4} $$

The definition of the coordinates involves the quantities

$$ \begin{aligned} C_1 &= 4A = 2(\xi_1\xi_5 - \xi_2\xi_4) = \rho^2\sin 2\Theta, \\ C_2 &= {}_kv = 2(\xi_1\xi_4 + \xi_2\xi_5) = \rho^2\cos 2\Theta\sin 2\,{}_k\Phi, \\ C_3 &= t = 2(\xi_1\xi_2 + \xi_4\xi_5) = \rho^2\cos 2\Theta\sin 2\varphi, \end{aligned} \tag{A5} $$

and

$$ \begin{aligned} D_1 &= \xi_1^2 + \xi_2^2 + \xi_4^2 + \xi_5^2 = \rho^2, \\ D_2 &= {}_ku = \xi_1^2 + \xi_2^2 - \xi_4^2 - \xi_5^2 = \rho^2\cos 2\Theta\cos 2\,{}_k\Phi, \\ D_3 &= s = \xi_1^2 - \xi_2^2 + \xi_4^2 - \xi_5^2 = \rho^2\cos 2\Theta\cos 2\varphi. \end{aligned} \tag{A6} $$

In addition, we shall ultimately need

$$ \begin{aligned} C_4 &= 2(\xi_1\xi_5 + \xi_2\xi_4), \\ C_5 &= 2(\xi_1\xi_4 - \xi_2\xi_5), \\ C_6 &= 2(\xi_1\xi_2 - \xi_4\xi_5), \end{aligned} $$

$$D_4 = \xi_1^2 - \xi_2^2 - \xi_4^2 + \xi_5^2. \tag{A7}$$

Differentiating $D_1 = \rho^2$ with respect to ξ_i and ξ_j, it is obvious that $\lambda_{0;ij} = 0$ identically, so the operators will only involve the angular coordinates, For $k = 1, 2, 3$ we have

$$\begin{aligned} \sin 2\alpha_1 &= C_1/D_1, \\ \tan 2\alpha_2 &= C_2/D_2, \\ \tan 2\alpha_3 &= C_3/D_3, \end{aligned} \tag{A8}$$

so that we can differentiate and write

$$\partial\alpha_k/\partial\xi_i = (A_{k;i}F_k - B_{k;i}G_k)/\rho^4 \cos^2 2\Theta, \tag{A9}$$

where

$$A_{k;i} = \tfrac{1}{2}\,\partial C_k/\partial\xi_i, \qquad B_{k;i} = \tfrac{1}{2}\,\partial D_k/\partial\xi_i, \tag{A10}$$

and

$$\begin{aligned} F_1 &= D_1 \cos 2\Theta, \qquad G_1 = C_1 \cos 2\Theta, \\ F_2 &= D_2, \qquad G_2 = C_2, \\ F_3 &= D_3, \qquad G_3 = C_3. \end{aligned} \tag{A11}$$

The matrices $A_{k;i}$ and $B_{k;i}$ are

$$\{A\} = \begin{Bmatrix} \xi_5 & -\xi_4 & -\xi_2 & \xi_1 \\ \xi_4 & \xi_5 & \xi_1 & \xi_2 \\ \xi_2 & \xi_1 & \xi_5 & \xi_4 \end{Bmatrix},$$

$$\{B\} = \begin{Bmatrix} \xi_1 & \xi_2 & \xi_4 & \xi_5 \\ \xi_1 & \xi_2 & -\xi_4 & -\xi_5 \\ \xi_1 & -\xi_2 & \xi_4 & -\xi_5 \end{Bmatrix}. \tag{A12}$$

Now we can write

$$\lambda_{k;ij} = F_k a_{k;ij} - G_k b_{k;ij}, \tag{A13}$$

where

$$\begin{aligned} a_{k;ij} &= \xi_i A_{k;j} - \xi_j A_{k;i}, \\ b_{k;ij} &= \xi_i B_{k;j} - \xi_j B_{k;i}. \end{aligned} \tag{A14}$$

The angular-momentum operators we want are of the form:

$$Q = \Lambda_{ij} \pm \Lambda_{mn} = (\hbar/i) \sum_k [F_k(a_{k;ij} \pm a_{k;mn}) - G_k(b_{k;ij} \pm b_{k;mn})]. \tag{A15}$$

Thus we have

$$\begin{aligned} L &= \Lambda_{12} + \Lambda_{45}, \quad [a_{k;12} + a_{k;45}] = [O, O, D_3], \\ &\qquad [b_{k;12} + b_{k;45}] = [O, O, -C_3], \\ \Sigma_- &= \Lambda_{14} - \Lambda_{25}, \quad [a_{k;14} - a_{k;25}] = [-C_3, D_4, C_1], \\ &\qquad [b_{k;14} - b_{k;25}] = [O, -C_5, O], \\ \Sigma_{12} &= \Lambda_{15} + \Lambda_{24}, \quad [a_{k;15} + a_{k;24}] = [D_3, C_6, O], \\ &\qquad [b_{k;15} + b_{k;24}] = [O, -C_4, -C_1], \\ \Sigma_t &= \Lambda_{14} + \Lambda_{25}, \quad [a_{k;14} + a_{k;25}] = [O, D_2, O], \\ &\qquad [b_{k;14} + b_{k;25}] = [O, -C_2, O], \\ Y &= \Lambda_{12} - \Lambda_{45}, \quad [a_{k;12} - a_{k;45}] = [-C_2, C_1, D_4], \\ &\qquad [b_{k;12} - b_{k;45}] = [O, O, -C_6], \\ A &= \Lambda_{15} - \Lambda_{24}, \quad [a_{k;15} - a_{k;24}] = [D_2, O, C_5], \\ &\qquad [a_{k;15} - a_{k;24}] = [O, -C_1, -C_4]. \end{aligned} \tag{A16}$$

To evaluate the resulting expressions we need the identities

$$\begin{aligned} &D_3F_3 + C_3G_3 = D_2F_2 + C_2G_2 = \rho^4 \cos^2 2\Theta. \\ &D_4F_2 + C_5G_2 = \rho^2 D_3, \\ &D_4F_3 + C_6G_3 = \rho^2 D_2, \\ &C_6F_2 + C_4G_2 = \rho^2 C_3, \\ &C_5F_3 + C_4G_3 = \rho^2 C_2. \end{aligned} \tag{A17}$$

As a result, we find

$$\begin{aligned} L &= \frac{\hbar}{i}\frac{\partial}{\partial\varphi}, \\ \Sigma_- &= \frac{\hbar}{i}\left(-\sin 2\varphi \frac{\partial}{\partial\Theta} + \sec\Theta \cos 2\varphi \frac{\partial}{\partial\Phi} + \tan\Theta \cos 2\varphi \frac{\partial}{\partial\varphi}\right), \\ \Sigma_{12} &= \frac{\hbar}{i}\left(\cos 2\varphi \frac{\partial}{\partial\Theta} + \sec\Theta \sin 2\varphi \frac{\partial}{\partial\Phi} + \tan\Theta \sin 2\varphi \frac{\partial}{\partial\varphi}\right), \\ \Sigma_t &= \frac{\hbar}{i}\frac{\partial}{\partial\Phi}, \\ Y &= \frac{\hbar}{i}\left(-\sin 2\Phi \frac{\partial}{\partial\Theta} + \tan 2\Theta \cos 2\Phi \frac{\partial}{\partial\Phi} + \sec 2\Theta \cos 2\Phi \frac{\partial}{\partial\varphi}\right), \\ A &= \frac{\hbar}{i}\left(\cos 2\Phi \frac{\partial}{\partial\Theta} + \tan 2\Theta \sin 2\Phi \frac{\partial}{\partial\Phi} + \sec 2\Theta \sin 2\Phi \frac{\partial}{\partial\varphi}\right). \end{aligned} \tag{A18}$$

From these it follows that

$$P_1^{\pm} = \Sigma_- \pm i\,\Sigma_{12} = \frac{\hbar}{i} e^{\pm i2\varphi}$$

$$\times \left(\pm i \frac{\partial}{\partial \Theta} + \sec 2\Theta \frac{\partial}{\partial \Phi} + \tan 2\Theta \frac{\partial}{\partial \varphi}\right),$$

$$Q_1^{\pm} = Y \pm iA = \frac{\hbar}{i} e^{\pm i2\Phi}$$

$$\times \left(\pm i \frac{\partial}{\partial \Theta} + \sec 2\Theta \frac{\partial}{\partial \varphi} + \tan 2\Theta \frac{\partial}{\partial \Phi}\right). \quad \text{(A19)}$$

2. The Asymmetric Representation

The operators are to be expressed in the angles φ_+, φ_-, χ. It is convenient to start with the angular coordinates

$$\alpha_1 = \varphi_1 = \varphi_+ + \varphi_-, \quad \alpha_2 = \varphi_2 = \varphi_+ - \varphi_-, \quad \alpha_3 = \chi, \quad \text{(A20)}$$

so that

$$\tan \alpha_1 = \xi_2/\xi_1,$$
$$\tan \alpha_2 = \xi_5/\xi_4,$$
$$\tan \alpha_3 = (\xi_4^2 + \xi_5^2)/(\xi_1^2 + \xi_2^2). \quad \text{(A21)}$$

With this definition of the α's, Eqs. (A1), (A3), and (A4) are still valid. Now take

$$g_1 = \sin \varphi_1 \cos \varphi_2 = \tfrac{1}{2}(\sin 2\varphi_+ + \sin 2\varphi_-),$$
$$g_2 = \cos \varphi_1 \sin \varphi_2 = \tfrac{1}{2}(\sin 2\varphi_+ - \sin 2\varphi_-),$$
$$g_3 = \cos \varphi_1 \cos \varphi_2 = \tfrac{1}{2}(\cos 2\varphi_+ + \cos 2\varphi_-),$$
$$g_4 = \sin \varphi_1 \sin \varphi_2 = \tfrac{1}{2}(-\cos 2\varphi_+ + \cos 2\varphi_-). \quad \text{(A22)}$$

and

$$f_1 = \tan \chi, \quad f_2 = \cot \chi. \quad \text{(A23)}$$

Then

$$\lambda_{1,12} = 1, \quad \lambda_{2,12} = \lambda_{3,12} = 0,$$
$$\lambda_{1,45} = \lambda_{3,45} = 0, \quad \lambda_{2,45} = 1,$$
$$\lambda_{1,14} = f_1 g_1, \quad \lambda_{2,14} = -f_2 g_2, \quad \lambda_{3,14} = g_3,$$
$$\lambda_{1,15} = f_1 g_4, \quad \lambda_{2,15} = f_2 g_3, \quad \lambda_{3,15} = g_2,$$
$$\lambda_{1,24} = -f_1 g_3, \quad \lambda_{2,24} = -f_2 g_4, \quad \lambda_{3,24} = g_1,$$
$$\lambda_{1,25} = -f_1 g_2, \quad \lambda_{2,25} = f_2 g_1, \quad \lambda_{3,25} = g_4. \quad \text{(A24)}$$

The coefficients appropriate to the angles φ_+, φ_- follow by writing

$$\lambda_{\pm,ij} = \tfrac{1}{2}(\lambda_{1,ij} \pm \lambda_{2,ij}). \quad \text{(A25)}$$

Taking the proper combinations of these one finds the contributions to the desired operators. These become

$$L = \frac{\hbar}{i} \frac{\partial}{\partial \varphi_+},$$

$$\Sigma_- = \frac{\hbar}{i}\left(\cos 2\varphi_+ \frac{\partial}{\partial \chi} + \csc 2\chi \sin 2\varphi_+ \frac{\partial}{\partial \varphi_-} - \cot 2\chi \sin 2\varphi_+ \frac{\partial}{\partial \varphi_+}\right),$$

$$\Sigma_{12} = \frac{\hbar}{i}\left(\sin 2\varphi_+ \frac{\partial}{\partial \chi} - \csc 2\chi \cos 2\varphi_+ \frac{\partial}{\partial \varphi_-} + \cos 2\chi \cos 2\varphi_+ \frac{\partial}{\partial \varphi_+}\right),$$

$$Y = \frac{\hbar}{i} \frac{\partial}{\partial \varphi_-},$$

$$A = \frac{\hbar}{i}\left(-\sin 2\varphi_- \frac{\partial}{\partial \chi} - \cot 2\chi \cos 2\varphi_- \frac{\partial}{\partial \varphi_-} + \csc 2\chi \cos 2\varphi_- \frac{\partial}{\partial \varphi_+}\right),$$

$$\Sigma_t = \frac{\hbar}{i}\left(\cos 2\varphi_- \frac{\partial}{\partial \chi} - \cot 2\chi \sin 2\varphi_- \frac{\partial}{\partial \varphi_-} + \csc 2\chi \sin 2\varphi_- \frac{\partial}{\partial \varphi_+}\right). \quad \text{(A26)}$$

From these it follows that

$$P_1^{\pm} = \Sigma_- \pm i\Sigma_{12} = \frac{\hbar}{i} e^{\pm i2(\varphi_+ - \pi/4)}$$

$$\times \left(\pm i \frac{\partial}{\partial \chi} + \csc 2\chi \frac{\partial}{\partial \varphi_-} - \cot 2\chi \frac{\partial}{\partial \varphi_+}\right)$$

and

$$Q_2^{\pm} = A \pm i\Sigma_t = \frac{\hbar}{i} e^{\pm i2\varphi_-}$$

$$\times \left(\pm i \frac{\partial}{\partial \chi} + \csc 2\chi \frac{\partial}{\partial \varphi_+} - \cot 2\chi \frac{\partial}{\partial \varphi_-}\right). \quad \text{(A27)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Ordering Energy Levels of Interacting Spin Systems

ELLIOTT LIEB AND DANIEL MATTIS
Thomas J. Watson Research Center, International Business Machines Corporation, Yorktown Heights, New York
(Received October 6, 1961)

The total spin S is a good quantum number in problems of interacting spins. We have shown that for rather general antiferromagnetic or ferrimagnetic Hamiltonians, which need not exhibit translational invariance, the lowest energy eigenvalue for each value of S [denoted $E(S)$] is ordered in a natural way. In antiferromagnetism, $E(S+1) > E(S)$ for $S \geq 0$. In ferrimagnetism, $E(S+1) > E(S)$ for $S \geq \mathbb{S}$, and in addition the ground state belongs to $S \leq \mathbb{S}$. $\mathbb{S}$ is defined as follows: Let the maximum spin of the A sublattice be S_A and of the B sublattice S_B; then $\mathbb{S} \equiv |S_A - S_B|$. Antiferromagnetism is treated as the special case of $\mathbb{S} = 0$. We also briefly discuss the structure of the lowest eigenfunctions in an external magnetic field.

INTRODUCTION

THE general Heisenberg Hamiltonian for interacting spins on a lattice (in any number of dimensions) is

$$H = 2 \sum J_{ij} S_i \cdot S_j. \tag{1}$$

This describes theories of ferromagnetism, ferrimagnetism, and antiferromagnetism, depending on the geometry of the lattice, the structure of the symmetric matrix J_{ij}, and the magnitude of the intrinsic spins (which may vary from site to site). In fact, it is conceivable that these factors be such that the spin system displays a mixture of the three magnetic properties. But we shall restrict the discussion to *ferrimagnetic* arrays, of which a special case is *antiferromagnetism*.

We consider only those arrays for which an A and a B sublattice can be defined. The definition of these two sublattices is circular, and perhaps not unique, for the only requirement in defining them is that there exist a constant $g^2 \geq 0$ such that for all sites $i(A)$ on one sublattice and $i(B)$ on the other,

$$J_{i(A),j(A)} \leq g^2, \qquad J_{i(B),j(B)} \leq g^2, \qquad \text{and} \quad J_{i(A),j(B)} \geq g^2. \tag{2}$$

In general, there might be several ways to decompose the lattice in such a way that (2) is obeyed, or there may be none. In the latter case, the system is not necessarily ferromagnetic, and only explicit solutions will reveal its properties. But if (2) is obeyed, we shall show that one is definitely dealing with ferrimagnetism or antiferromagnetism. Note that the number of sites in each sublattice and the magnitude of the intrinsic spin on each site is irrevalent, so that only the topology of the lattice and the structure of J_{ij} counts. Note also that for $g = 0$, and the A sublattice consisting of the nearest neighbors to the sites on (i.e., intermeshing with) the B sublattice, the requirement (2) gives a tendency for nearest neighbors to align antiparallel and next-nearest neighbors to align parallel, and therefore reduces to the usual definition of ferrimagnetism (when the spins are of unequal magnitude) and of antiferromagnetism (when all spins are equal).

The intrinsic spin of an electron is 1/2, but we may be dealing with various species of magnetized atoms or nuclei, so let the intrinsic spin angular momentum on each site be s_i. The maximum possible spin S_A on the A sublattice is therefore

$$S_A \equiv \sum_{i(A)} s_{i(A)}, \tag{3a}$$

and on the B sublattice

$$S_B \equiv \sum_{i(B)} s_{i(B)}. \tag{3b}$$

Defining

$$\mathbb{S} \equiv |S_A - S_B|, \tag{3c}$$

we shall prove that the ground state of H belongs *at most* to total spin $S = \mathbb{S}$. Moreover, if we denote by $E(S)$ the lowest energy eigenvalue belonging to total spin S, then we shall also prove

$$E(S+1) > E(S) \quad \text{for all} \quad S \geq \mathbb{S}, \quad \text{and} \quad E(S) > E(\mathbb{S}) \quad \text{for} \quad S < \mathbb{S} \quad \text{and} \quad g^2 = 0. \tag{4}$$

(Antiferromagnetism is when $\mathbb{S} = 0$, and the ground state belongs to total spin zero.) This can be regarded *either* as a theorem in ferri- or antiferromagnetism, *or* as a proof that the conditions in Eq. (2) and above eliminate the possibility of ferromagnetism (insofar as it costs energy to raise the total spin value over and above its ground-state value, and that this ground-state value is far from the maximum per-

missible value of $S_A + S_B$). It also indicates that a large class of apparently different Hamiltonians (1) have really a similar structure, as summarized in Eq. (4), and in the properties of the corresponding eigenfunctions which we shall find below.

W. Marshall was the first to show[1] that the ground state of an antiferromagnet is a singlet; Elsewhere,[2] we have commented on and strengthened his proof. In the present work, we succeed in removing the requirement of translational invariance, and also apply the method to identify the excited states. The M-subspace arguments presented here were previously found useful in the classification of the states of an electron system, and have been used to disprove the possibility of ferromagnetism in linear chains of atoms in s states.[3]

We shall now restrict the discussion to the special case $g^2 = 0$, until the end of the proof.

M SUBSPACES

With the help of the total spin operator

$$\mathbf{S} \equiv \sum \mathbf{s}_i$$

we can construct two operators which commute with each other and with H, namely, $\mathbf{S}^2$ and $\mathbf{S}_z$, which possess eigenvalues $S(S+1)$ and M, respectively. It is known from the theory of angular momentum that $S \geq |M|$. From the rotational invariance of the Hamiltonian we infer the $(2S+1)$-fold degeneracy of each energy level belonging to S, one degenerate level for each value of M in the range $-S \leq M \leq S$. It therefore follows that every energy eigenvalue has a corresponding eigenfunction (representative) in the $M = 0$ subspace of eigenfunctions; that every energy level except those belonging to $S = 0$ has a representative in the $M = 1$ subspace; similarly for all except $S = 0$ and $S = 1$ in the $M = 2$ subspace, and so forth. The theorem, Eq. (4), will be proved if we can show that the lowest energy in an M subspace belongs to $S = M$, for spin $S + 1$ also has a representative in that subspace and therefore $E(S) < E(S+1)$. If the ground state belongs to $S = S_0$ (we still have to prove that $S_0 \leq \mathcal{S}$), we need only consider the subspaces of $|M| \geq S_0$, for the ground states of the remaining subspaces will always belong to S_0.

The mechanics of the proof are this: The ground state of H in an M subspace is *not* orthogonal to the ground state of a soluble Hamiltonian in the same subspace, and the latter is *known* to belong to $S = M$ for $M \geq \mathcal{S}$; therefore, so does the former. Now let us go into more detail.

PROOF

In an M subspace, choose the basis set to consist of all distinct eigenfunctions of the $\mathbf{s}_i^z$ compatible with eigenvalue M. We denote each configuration in the set by ϕ_a, where a is an index which runs over all members of the set. Shortly, we shall specify a convenient choice of phase for each configuration. But first, perform a canonical transformation on H by letting

$$\mathbf{S}_{i(A)}^x \to -\mathbf{S}_{i(A)}^x, \qquad \mathbf{S}_{i(A)}^y \to -\mathbf{S}_{i(A)}^y, \qquad \mathbf{S}_{i(A)}^z \to +\mathbf{S}_{i(A)}^z \tag{5}$$

but leaving the spins on the B-sublattice invariant. In the *new* language, the Hamiltonian can be written as $H_0 + H_1$, where the diagonal part is

$$H_0 = 2 \sum J_{ij} \mathbf{S}_i^z \mathbf{S}_j^z, \tag{6}$$

and the nondiagonal part is

$$H_1 = -\{\sum |J_{ij}| \, S_i^+ S_j^- + \text{H.c.}\}. \tag{7}$$

We recall that g^2 of Eq. (2) is zero: the generalization for $g^2 > 0$ comes below.

In a given state ϕ_a, S_i^z has eigenvalue m_i. Choose the phase of ϕ_a in the following manner:

$$\phi_a = C(S_1^+)^{S_1+m_1}(S_2^+)^{S_2+m_2} \cdots (S_N^+)^{S_N+m_N}\chi, \tag{8}$$

where χ is the state in which $m_i = -S_i$, and C is a positive normalization constant. With this definition in mind, it is clear that if we define $K_{\beta a}$ to be

$$K_{\beta a} = \langle \phi_\beta \, | H_1 | \, \phi_a \rangle, \tag{9}$$

then

$$K_{\beta a} \leq 0, \quad \text{or equivalently,} \quad K_{\beta a} = -|K_{\beta a}|. \tag{10}$$

The ground state in the M subspace is denoted ψ, belongs to the ground-state energy E_M, and can be expanded in our complete set in terms of the amplitudes f_a,

$$\psi = \sum f_a \phi_a. \tag{11}$$

Since H_0 is diagonal, denote its eigenvalues by e_a,

$$H_0 \phi_a = e_a \phi_a, \tag{12}$$

and therefore the Schrödinger equation reads

$$-\sum_\beta |K_{\beta a}| \cdot f_\beta + e_a f_a = E_M f_a. \tag{13}$$

The variational energy of any trial function exceeds E_M, unless it is also a ground-state eigenfunction.

[1] W. Marshall, Proc. Roy. Soc. (London) **A232**, 48 (1955).
[2] E. Lieb, T. Schultz, D. Mattis, Ann. Phys. **16**, 407 (1961), particularly Appendix B.
[3] E. Lieb and D. Mattis, Phys. Rev. **125**, 164 (1962).

But

$$\psi' = \sum |f_\alpha| \phi_\alpha \tag{14}$$

is a trial function with variational energy E_M, and therefore

$$-\sum |K_{\beta\alpha}| \, |f_\beta| + e_\alpha |f_\alpha| = E_M |f_\alpha|. \tag{15}$$

Moreover,

$$e_\alpha - E_M > 0, \quad \text{for all} \quad \alpha \tag{16}$$

(otherwise, some *one* ϕ_α would be the ground state, which is in general impossible.) Therefore, taking the absolute value of $(e_\alpha - E_M)f_\alpha$ as given be Eq. (11) and combining with Eq. (15), we obtain

$$|(\sum |K_{\beta\alpha}| \, f_\beta)| = \sum |K_{\beta\alpha}| \, |f_\beta|. \tag{17}$$

This is a contradiction unless

$$f_\beta \geq 0 \quad \text{for all} \quad \beta. \tag{18}$$

In general, we have a slightly stronger result,

$$f_\beta > 0, \quad \text{for all} \quad \beta. \tag{19}$$

For, if some f_α vaaished, then Eq. (15) would read:

$$\sum K_{\beta\alpha} |f_\beta| = 0,$$

and by succeeding applications of the Hamiltonian, one could establish that *all* the amplitudes vanished, unless the Hamiltonian splits into sets of noninteracting spins in which case only the weaker result (18) holds. Therefore, in general, all amplitudes are positive and nonvanishing, and hence E_M in nondegenerate. This last statement follows from the impossibility of constructing states orthogonal to ψ without some changes of sign, and consequent violation of the ground-state property (19).

Next consider the special Hamiltonian where $J_{i(A)j(A)} = J_{i(B)j(B)} = 0$ and $J_{i(A)j(B)} = J$, a positive constant. The eigenvalues are readily calculated. The lowest energy belonging to each spin is given by $E(S)$, for $S \geq \mathcal{S}$, and the ground state belongs to $S = \mathcal{S}$.

$$E(S) = J\{S(S+1) - S_A(S_A+1) - S_B(S_B+1)\} \quad \text{for} \quad S \geq \mathcal{S}. \tag{20}$$

By the previous arguments, the ground-state eigenfunctions of this special Hamiltonian in a given M subspace satisfy Eq. (18) or (19) and are therefore not all orthogonal to the corresponding ground state of H. The special Hamiltonian has an $S = M$ ground state in each M subspace, provided $M \geq \mathcal{S}$. Therefore, so does H and this completes the proof for $g^2 = 0$.

When $g^2 > 0$, we have proved the theorem (4) for $H - g^2\mathbf{S}^2$ and it is therefore true *a fortiori* for H. However, the lowest ground state no longer necessarily belongs to $\mathcal{S}$, but belongs to $S \leq \mathcal{S}$.

MAGNETIC FIELD

A magnetic field in the z direction but of arbitrary and variable amplitude B_i modifies H_0 but not H_1, and therefore (18) or (19) are still valid for the ground state in an M subspace. The absolute ground state of the system is no longer necessarily in the $M \leq \mathcal{S}$ subspace nor is S a good quantum number in the presence of such a magnetic field.

ACKNOWLEDGMENT

It is a pleasure to thank Dr. T. D. Schultz for helpful discussion.

Proof of a Conjecture by Dyson in the Statistical Theory of Energy Levels

J. GUNSON
Department of Mathematical Physics, University of Birmingham, England
(Received February 1, 1962)

A conjectured identity relating the statistical properties of two types of ensembles occurring in a statistical theory of the distribution of energy levels in nuclei and other complex systems is proved.

IN a recent series of papers,[1] Dyson has introduced new types of statistical ensembles in an attempt to provide a mathematically tractable description of statistical properties of energy levels in nuclei and other complex systems. They possess the probability distribution functions

$$P_{N\beta}(\theta_1, \cdots, \theta_N) = C_{N\beta} \prod_{\nu=1}^{N} \prod_{\mu=1}^{\nu-1} |\exp(i\theta_\mu) - \exp(i\theta_\nu)|^\beta, \quad (1)$$

$$\beta > 0 \quad 0 \leq \theta_\mu \leq 2\pi$$

for N points $\{\exp(i\theta_\mu)\}$ distributed round the unit circle $|z| = 1$ in the complex z plane. Two important conjectures appear in the papers. The first gives the normalising constant $C_{N\beta}$ in Eq. (1) to be

$$C_{N\beta} = (2\pi)^{-N}[\Gamma(1 + \tfrac{1}{2}\beta)]^N/\Gamma(1 + \tfrac{1}{2}N\beta). \quad (2)$$

This is proved to hold for all integers N and real positive numbers β by Wilson[2] and independently by the author. The second conjecture appears in part III of reference 1 and can be stated as the following theorem.

Theorem. Let $P_{NM}(\theta_1, \cdots, \theta_N)$ be the probability distribution function describing the statistical properties of a set of N points $\{\exp(i\theta_\mu)\}$ constructed as follows: Take two independent sets each consisting of N points distributed according to the probability distribution function of Eq. (1) with $\beta = 1$ (Dyson's "orthogonal" ensemble[1]), superimpose the two sets on the unit circle, and pick out a set of N alternate points. Then

$$P_{NM}(\theta_1, \cdots, \theta_N) = P_{N2}(\theta_1, \cdots, \theta_N), \quad (3)$$

where P_{N2} is the distribution function of the "unitary" ensemble.[1]

Proof. From Eqs. (1) and (2) we obtain the probability of finding *any* point in the interval $[\alpha_1, \alpha_1 + d\alpha_1]$, *any other* point in the interval $[\alpha_2, \alpha_2 + d\alpha_2]$, etc., where $\alpha_1 < \alpha_2 < \cdots < \alpha_N < \alpha_1 + 2\pi$, to be

$$N!\, C_{N1} \prod_{\mu=1}^{N} \prod_{\nu=1}^{\mu-1} 2 \sin[(\alpha_\mu - \alpha_\nu)/2]\, d\alpha_1\, d\alpha_2 \cdots d\alpha_N. \quad (4)$$

This is correctly normalised to unity over the domain $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_N < 2\pi$. We can now express $N!\, P_{NM}(\theta_1, \cdots, \theta_N)\, d\theta_1 \cdots d\theta_N$ as a sum of partial probabilities over configurations, a particular configuration being specified by the assignment of the $2N$ angles $\alpha_1, \cdots, \alpha_N$ and $\beta_1, \cdots, \beta_N$ from the orthogonal ensembles A and B into the $2N$ intervals $I_\lambda = [\theta_\lambda, \theta_\lambda + d\theta_\lambda]$ and $G_\lambda = [\theta_\lambda + d\theta_\lambda, \theta_{\lambda+1}]$ $\lambda = 1, 2, \cdots, N$, one in each interval. A typical contribution can be written

$$(C_{N1}N!)^2 \int_{\theta_N}^{\theta_1+2\pi} d\phi_N \cdots \int_{\theta_2}^{\theta_3} d\phi_2 \int_{\theta_1}^{\theta_2} d\phi_1 \prod_{\nu=1}^{N} \prod_{\mu=1}^{\nu-1} \times |4 \sin[(\alpha_\mu - \alpha_\nu)/2] \sin[(\beta_\mu - \beta_\nu)/2]|\, d\theta_1 \cdots d\theta_N \quad (5)$$

in which, for a given integer m, $0 \leq m \leq N$, we require: m of the angles α are identified with m of the angles ϕ, $N - m$ of the angles β are identified with the remainder of the ϕ's, The remainder of the angles α and β are identified with the angles θ, in a one-to-one manner. For a given m, we then have $\binom{N}{m}^2$ essentially different possible identifications and thus the same number of different contributions to P_{NM}. Each of these terms must be positive, each being an independent contribution to a probability, so we may write Eq. (5) in the form

$$(C_{N1}N!)^2 \int_{\theta_N}^{\theta_1+2\pi} d\phi_N \cdots \int_{\theta_1}^{\theta_2} d\phi_1 \det[e^{ik_\mu\alpha_\nu}] \times \det[e^{ik_\mu\beta_\nu}]\, d\theta_1 \cdots d\theta_N \quad (6)$$

[1] F. J. Dyson, J. Math. Phys. **3**, 140 (1962); **3**, 157 (1962); **3**, 166 (1962).
[2] K. Wilson, J. Math. Phys. (to be published).

on using the identities

$$\prod_{\mu=1}^{N}\prod_{\nu=1}^{\mu-1}(2\sin[(\alpha_\mu-\alpha_\nu)/2]) = i^{N(N+1)/2}\det(e^{ik_\mu\alpha_\nu})$$

$$k_\mu = \mu - (N-1)/2;\quad \mu,\nu = 1, 2, \cdots N. \tag{7}$$

We choose the ordering in the assignment of the angles α and β in such a manner as to make the integrand positive at some point in the region of integration. It is then positive over the whole region of integration and so we are justified in dropping the modulus signs. Let us define normal ordering as that in which the angles α (and likewise the angles β) are identified successively as one of the angles θ or ϕ in increasing order of value. The integral in Eq. (6) may then be either positive or negative and we assert that its sign is given by the expression

$$(-1)^{P_\theta}(-1)^{P_\phi} \tag{8}$$

in which $P_{\theta,\phi}$ is the number of interchanges of the angles θ, ϕ between the two determinants in the integrand required to transform a particular positive term, taken as standard, into the term under consideration. To see this, consider the interchange of the integrations over the interval G of the unit circle in ensemble A with that over G' in ensemble B (Fig. 1).

This interchange produces a new term in which the normal ordering has been destroyed, but it can be restored by interchanging columns within each determinant in the manner suggested by the solid arrows in Fig. 1. The essential point about the latter is that this permutation involves in all an *odd* number of interchanges of columns in the two determinants, as the variable in every interval appears either in the A or B determinants and there are always an odd number of intervals (and hence θ's and ϕ's) between two intervals of type G. The same argument holds, with the necessary changes made, for interchanges of the θ's.

The next step is to sum the integrands of all the $\binom{N}{m}^2$ terms when written in normal form with appropriate sign factors included. To do this, we express the determinants in terms of alternating functions and use the identity

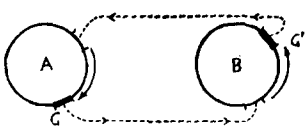

FIG. 1. Interchange of two integrations over the intervals G and G' in a contribution to $P_{NM}(\theta_1, \cdots, \theta_N)$.

$$\sum(-1)^{P_x}(-1)^{P_y}\,\Delta_N(x_1,\cdots,x_r,y_{r+1},\cdots,y_N)\times\Delta_N(y_1,\cdots,y_r,x_{r+1},\cdots,x_N) = \binom{N}{r}\Delta_N(x_1,\cdots,x_N)\,\Delta_N(y_1,\cdots,y_N), \tag{9}$$

in which

$$\Delta_N(z_1,\cdots,z_N) = \prod_{\nu=1}^{N}\prod_{\mu=1}^{\nu-1}(z_\mu - z_\nu) = \det[z_\mu^{N-\nu}]$$

and where the sum $\sum$ is taken over all partitions of the x's and y's between the arguments of the alternating functions Δ. This identity is an immediate consequence of Cauchy's result that any polynomial in N variables which is completely antisymmetric in the variables and contains powers of degree no higher than $(N-1)$ in each variable is a multiple of Δ_N. The sum of all the $\binom{N}{m}^2$ terms then becomes

$$N!\,C_{N1})^2\int_{\theta_N}^{\theta_1+2\pi}d\phi_N\cdots\int_{\theta_1}^{\theta_2}d\phi_1\binom{N}{m}\det[e^{ik_\mu\theta_\nu}]\det[e^{ik_\mu\phi_\nu}]\,d\theta_1\,d\theta_2\cdots d\theta_N. \tag{10}$$

The integrals over the ϕ's can now be evaluated to give

$$\begin{aligned} N!\,P_{Nm}(\theta_1,\cdots\theta_N) &= (N!\,C_{N1})^2\sum_{m=0}^{N}\binom{N}{m}2\times\left(\frac{2^N\Gamma(1+\tfrac{1}{2}N)}{N!}\right)^2(\det[e^{ik_\mu\theta_\nu}])^2\cdot\tfrac{1}{2}\\ &= (2\pi)^{-N}(\det[e^{ik_\mu\theta_\nu}])^2\\ &= N!\,P_{N2}(\theta_1,\theta_2,\cdots,\theta_N) \end{aligned} \tag{11}$$

as required. The factor $\frac{1}{2}$ in the second expression is introduced to express the probability that a particular alternate series of N points out of the two possible choices is taken.

The author would like to thank Professor Dyson for valuable criticism of the proof, Professor Peierls for the hospitality of the Department of Mathematical Physics of the University of Birmingham, and the Department of Scientific and Industrial Research for a Research Fellowship.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Proof that the Two-Dimensional Shape of Minimum Surface Free Energy is Convex

WILLIAM W. MULLINS
Carnegie Institute of Technology
Pittsburgh, Pennsylvania
(Received July 31, 1961)

Published proofs that the Wulff construction yields the shape of minimum surface free-energy for a crystal of fixed volume (or area in the two-dimensional case) are all based on the tacit but unproven assumption that the equilibrium shape is convex. In the present work, this assumption is justified for the two-dimensional case by establishing from a special sequence of constructions that corresponding to every nonconvex curve of line (surface) energy E, there is a convex curve enclosing the same area but having a line energy less than E. Therefore the two-dimensional equilibrium form must be convex and the existing proofs of the Wulff construction in two dimensions are vindicated; unfortunately the method of proof cannot be extended to three dimensions.

ONE of the oldest problems in the field of capillarity is that of determining the equilibrium shape of a crystal whose surface free-energy is a specified function of orientation. One seeks the particular shape of a crystal of fixed volume that will minimize the free-energy integral

$$E = \int \gamma(\mathfrak{d})\, ds, \tag{1}$$

where $\gamma(\mathfrak{d})$ is a specified continuous function of the unit surface normal $\mathfrak{d}$. The commonly accepted solution is given by Wulff's geometrical construction[1] carried out on the three-dimensional polar plot of $\gamma(\mathfrak{d})$. All published proofs that the Wulff construction yields either a relative or an absolute minimum free-energy, however, are based on the assumption that the equilibrium shape is convex. Many proofs assume at the outset that the equilibrium shape belongs to the class of convex polyhedra.[2-5] Proofs not falling in this category either are (1) restricted to two dimensions[6,7] and utilize analytic expressions which are well defined (e.g., single valued) only if the shape is convex or (2) are based on an application of Brunn-Minkowski inequality for mixed volumes[1] which has been established for convex bodies only.[8]

The purpose of the present paper is to supply the missing step in the two-dimensional case by demonstrating from first principles that the curve describing the two-dimensional equilibrium shape must be convex. The procedure is to show that corresponding to any nonconvex curve, there is a convex curve having the same enclosed area A but a lower free-energy E [given by Eq. (1) interpreted as a line integral]. This is accomplished by (1) applying a sequence of unfolding operations to the original curve each of which leaves E constant, increases A, and also makes the curve approach convexity, then (2) at some nth stage replacing the unfolded curve by a bonafide convex curve at the maximum cost of an arbitrarily small increment in E and A (for large enough n), and finally (3) uniformly shrinking this convex curve until its area equals that of the original curve; the free energy of this final convex curve is then proved to be less than that of the original curve. The result justifies the use of the convexity assumption in two-dimensional proofs, many of which are otherwise quite straightforward; unfortunately the method used to establish the proof does not seem to admit of extension to three dimensions.

DEFINITION AND SIMPLE PROPERTIES OF THE UNFOLDING OPERATION

Let the initial shape be specified by a closed, simply connected curve Γ, whose tangent is defined at all points except possibly for a finite number of cusp points. Let θ be the angle made by the tangent with a fixed line, and s be the arc length along the curve measured from some origin on the curve; $\theta(s)$ will then be a piecewise continuous function with jump discontinuities at the cusp points. We

[1] C. Herring, in *Structure and Properties of Solid Surfaces*, edited by C. S. Smith and R. Gomer (University of Chicago Press, Chicago, Illinois, 1953).
[2] H. Hilton, *Mathematical Crystallography* (Oxford University Press, New York, 1903).
[3] H. Liebmann, Z. Krist. **53**, 171 (1914).
[4] M. von Laue, Z. Krist. **105**, 124, (1943).
[5] A. Dinghas, Z. Krist. **105**, 304 (1944).
[6] L. D. Landau, *Collection of Papers in Honor of Seventieth Birthday of A. F. Ioffe* (Akademüa Nauk S.S.S.R., Moscow, U.S.S.R., 1950) p. 44.
[7] W. K. Burton, N. Cabrerra, and F. C. Frank, Phil. Trans. Roy. Soc. London **243**, 351 (1951).
[8] T. Bonnesen and W. Fenchel, *Theorie der Konvexen Körper* (Verlag Julius Springer, Berlin, Germany, 1934), p. 88.

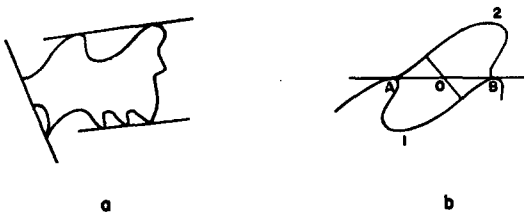

FIG. 1. Contact lines and the unfolding operation.

suppose further that $\theta(s)$ is of bounded variation[9] and denote by $\Theta(s)$ the total variation of $\theta(s)$ between the origin $s = 0$ and s. $\Theta(s)$ is a monotonically increasing function of s with jump discontinuities at the cusps; it may be thought of intuitively as a measure of the total twist of the tangent, disregarding sign, between 0 and s; i.e., $\int_0^s |d\theta/ds|\, ds = \int_0^S |d\theta|$, if $d\theta/ds$ is defined.

Define a contact line to be a line such that (1) no part of the curve lies on one of the two sides of the line and (2) the line has more than one point in common with the curve but does not entirely coincide with the curve between the two most widely separated of the common points, hereafter called the contact points. Figure 1(a) shows three contact lines of a closed curve. A contact line is said to subtend that portion of the curve Γ lying between the two contact points (and within the rest of Γ and the contact line). It is easy to show that Γ has one or more contact lines if and only if it is not convex.

We define an unfolding operation about a contact line as follows: A center point is located on the contact line midway between the two contact points [e.g., point O between A and B of Fig. 1(b)]. The subtended curve is then inverted through this center point. This is equivalent to a reflection through the contact line followed by a reflection through a line passing perpendicularly through its center. The five properties of the unfolding operation that we require are stated below. The proofs of (iii) and (iv) below are given or indicated; (i), (ii), and (v) are obvious and are stated without proof.

(i) The length L of the curve is invariant.

(ii) The area enclosed by Γ is increased, by unfolding, by twice the area included between the contact line and the subtended curve.

(iii) The value of the integral $E = \int \gamma(\theta)\, ds$ is invariant where we have rewritten Eq. (1) for the two-dimensional case and have changed the argument of γ from the unit normal to θ. The value of $\int \gamma(\theta)\, ds$ over the part of Γ not subtended by the contact line is obviously unaffected by unfolding. To prove the value over the subtended part is also unaffected, place a Cartesian frame with the x axis along the contact line and the origin at the center of inversion. Consider the original curve to be given parametrically in terms of s, i.e., $x(s)$, $y(s)$. Noting that θ is determined by $dy/dx = \dot{y}/\dot{x}$, where the dot denotes differentiation with respect to s, we have for the free-energy the expression $E = \int_{s_1}^{s_2} \gamma[\theta(\dot{y}/\dot{x})](\dot{x}^2 + \dot{y}^2)^{1/2}\, ds$, where $s_2 - s_1$ is the total length of subtended curve. It is clear that this expression is invariant under the unfolding transformation $x(s) \to -x(s)$, $y(s) \to -y(s)$ for $s_1 \leq s \leq s_2$.

(iv) The total variation of the entire curve $\Theta(L)$ is never increased by unfolding. Furthermore unfolding never results in cusps with a greater $|\Delta\theta|$ than those already present. To establish these statements we first note that the total variation over any portion of the curve not including the contact points is unaffected by unfolding. To see what happens at these points, we examine the three possible cases illustrated in Fig. 2(a), (b), and (c).

If $\theta(s)$ is continuous at both contact points, as shown in Fig. 2(a), the contact line coincides with the tangents at both points, before as well as after unfolding. Since the contact points therefore remain uncusped, they make no discrete contribution to the total variation $\Theta(L)$ which is thus unaffected by unfolding.

If there is a cusp at one of the two contact points [A of Fig. 2(b)], unfolding will generally ($\alpha \neq 0$) yield a cusp at both points (A and B). Each cusped point contributes the magnitude of its jump discontinuity $|\Delta\theta|$ to the total variation. In spite of the new cusp, Table I records what is easily seen from Fig. 2(b), namely, that the sum of $|\Delta\theta|$ for the two cusps and therefore the total variation remains unaltered by unfolding. An obvious corollary is that neither cusp after unfolding has a bigger $|\Delta\theta|$ than the original one.

Finally Fig. 2(c) shows, and Table I again records, the case in which both contact points are cusped. At A, $|\Delta\theta|$ is initially $\alpha_1 + \beta_1$ and after unfolding becomes either $\alpha_1 - \beta_2$ if $\alpha_1 > \beta_2$, or $\beta_2 - \alpha_1$ if $\beta_2 > \alpha_1$. Similarly at B, $|\Delta\theta|$ is initially $\alpha_2 + \beta_2$, and changes either to $\alpha_2 - \beta_1$ if $\alpha_2 > \beta_1$, or to $\beta_1 - \alpha_2$ if $\beta_1 > \alpha_2$. Consider point A: if $\alpha_1 > \beta_2$ then from the preceding inequalities the final value of $|\Delta\theta|$ at A cannot be greater than the initial one. On the other hand if $\beta_2 > \alpha_1$ then the final value of $|\Delta\theta|$ at A cannot be greater than the initial value at B. A similar argument applies to point B. We conclude that unfolding cannot yield a greater $|\Delta\theta|$. Also,

[9] E. C. Titchmarsh, *The Theory of Functions* (Oxford University Press, New York, 1939), p. 355, 2nd. Ed.

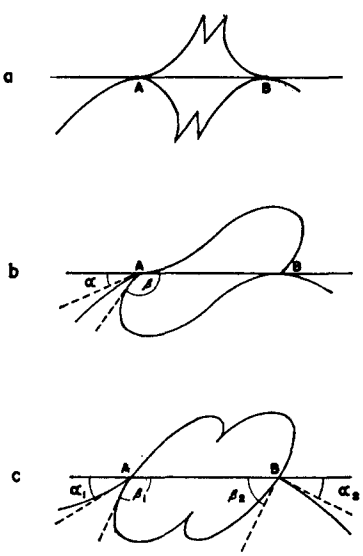

FIG. 2. The three cusp configurations occurring during unfolding.

the expression $\alpha_1 + \beta_1 + \alpha_2 + \beta_2$ giving the initial sum of $|\Delta\theta|$ for the two cusps cannot increase when some of the signs become negative as required to give the final sum. Property (iv) is thus verified for all cases.

In addition, we note the following special consequences of the unfolding operation: If the curve contains no cusps originally, it will contain none after unfolding; furthermore the total variation $\Theta(L)$ of the curve will remain constant so that the wiggles can never be removed. (Also the curvature can never be increased at all points where it is and remains defined.) If, on the other hand, cusps are originally present, their number may be increased by unfolding but neither the sum nor the largest of the corresponding $|\Delta\theta|$'s can increase.

(v) It is clear, from (iv), that the piecewise continuity and bounded variation of the function $\theta(s)$, corresponding to Γ, are preserved by unfolding. Therefore the preceding properties (i)–(iv) hold true for any unfolding operation carried out on a curve Γ_n derived from the original Γ by a sequence of n previous unfoldings.

Theorem I—Subtended distance theorem. We define the distance between a contact line and its subtended curve to be the maximum of the perpendicular distances of the points on the curve to the contact line; the points lying at this maximum distance will be called the d points of the subtended curve. We further define a d-unfolding to be an unfolding about a contact line whose subtended curve lies at a distance not exceeded by the distances between the other contact lines and their corresponding subtended curves.

Consider a curve Γ_n derived from an original curve by n, d-unfoldings. Let d_i^n be the distance between the ith contact line and its subtended curve in Γ_n, and suppose d_M^n is the largest of the d_i^n. Then we assert (Theorem I) that for a sufficiently large N either that Γ_N is convex or that for $n \geq N$, $d_M^n < \epsilon$ for arbitrarily small ϵ. To prove the theorem, suppose there are initially c_0 contact lines with corresponding distances $d_1^0, d_2^0, \cdots d_{c_0}^0$. Perform the first d-unfolding about a contact line corresponding to d_M^0 thereby obtaining Γ_1. There will now be c_1 contact lines with the corresponding distances $d_1^1\, d_2^1 \cdots d_{c_1}^1$ whose maximum will be d_M^1. Continuing at this stage, and all succeeding stages, to perform d-unfoldings we will generate the array

$$d_1^0, d_2^0, \cdots d_{c_0}^0; d_M^0$$
$$d_1^1, d_2^1, \cdots d_{c_1}^1; d_M^1$$
$$d_1^n, d_2^n, \cdots d_{c_n}^n; d_M^n .$$

If this array terminates for some $(n + 1)$st row, there are no more contact lines and the curve Γ_n must be convex. If not, there will be an infinite sequence of d_M's.

Assume, contrary to Theorem I, that there is an infinite subsequence of d_M's with a lower bound ϵ_0, i.e., $d_M^{1'}, d_M^{2'}, d_M^{3'}, \cdots d_M^{n'} \cdots > \epsilon_0$. Denote by β the angle which the outward normal to a contact line makes with a fixed direction so that each d corresponds to a β. The infinite sequence of β's, $\beta_M^{1'}, \beta_M^{2'}, \cdots$, corresponding to the infinite sequence of bounded d's, must have at least one point of accumulation β_0, since $0 \leq \beta \leq 2\pi$. We proceed to show that the width[10] of the curve along the direction of β_0 will grow indefinitely as d-unfoldings proceed, thus contradicting the constancy of the perimeter (length L), necessitating rejection of the postulated bound ϵ_0, and proving the theorem.

Since β_0 is a point of accumulation of the β se-

TABLE I. The sum of $|\Delta\theta|$ for the two Cusp configurations given in Fig. 2.[a]

		$\Delta\theta$ at A	$\Delta\theta$ at B
Fig. 2(b)	Before unfolding	$\alpha + \beta$	0
	After unfolding	α	$-\beta$
Fig. 2(c)	Before unfolding	$\alpha_1 + \beta_1$	$\alpha_2 + \beta_2$
	After unfolding	$\alpha_1 - \beta_2$	$\alpha_2 - \beta_1$

[a] Note that $\Delta\theta$ is counted as positive when clockwise as figures are traced from left to right; α and β are non-negative.

[10] We use the customary definition of width[8] as the distance between two parallel supporting lines that just sandwich the curve.

quence, there must be an infinite subsequence $d_M^{1''}$, $d_M^{2''}$, $d_M^{3''}$, $\cdots$ of the bounded sequence $d_M^{1'}$, $d_M^{2'}$, $\cdots$, whose terms correspond to contact lines lying within the range of orientations $\beta_0 - \Delta\beta$ to $\beta_0 + \Delta\beta$, where $\Delta\beta$ is arbitrarily small. As unfolding proceeds, the particular unfolding steps corresponding to the $d_M^{1''}$, $d_M^{2''}$, which we term growth steps, will be reached consecutively. Let Fig. 3 represent the curve just before a growth step, with OP along β_0, XY the (top) supporting line perpendicular to OP, and O itself a point of contact of the curve with the support line XY. Suppose the contact line corresponding to the growth step to be AB. It cannot be any lower than shown or it would pass below O and hence inside the curve; it may be higher. Because of the fixed perimeter, the d point of the subtended curve cannot possibly lie to the right of line $B'B$ whose perpendicular distance to O is $L/2$. As a result of unfolding, line AB is advanced parallel to itself to position $A'B'$ by an amount of $S > \epsilon_0$. The corresponding advance in the support line XY and hence in the diameter of the figure along β_0 would be smallest if the d point were on BB'. In this case (and since $|\Delta\chi| \leq \Delta\beta$) the advance of XY would be at least $(\epsilon_0 - L/2 \tan \Delta\beta) \cos \Delta\beta$. Clearly $\Delta\beta$ may be chosen so small (and a contact line still guaranteed) that this quantity exceeds say $\epsilon_0/2$. Since the same choice of $\Delta\beta$ will suffice for all growth steps, we see that each of the growth steps, that is, each of the unfoldings corresponding to a $d_M^{1''}$, $d_M^{2''}$, etc., increases the width of the curve along β_0 by at least $\epsilon_0/2$. But none of the other unfoldings occurring between the growth steps can decrease the width along β_0, or for that matter, along any other direction since curve Γ_n is always included in the derived, unfolded curve Γ_{n+1}. We conclude that after n steps where $n\epsilon_0/2 > L/2$ or $n > L/\epsilon_0$, the diameter along β_0 would exceed $L/2$ which is impossible. Therefore there can be no lower bound ϵ_0 of the d_M sequence for d-unfoldings, and there must be an N such that d_1^n, d_2^n, $\cdots$, $d_{c_n}^n \leq d_M^n < \epsilon$ for $n \geq N$.

Theorem II—Large angle curve theorem: The total length l of large angle (L.A.) curve, defined to be that subtended curve whose angle θ differs at all

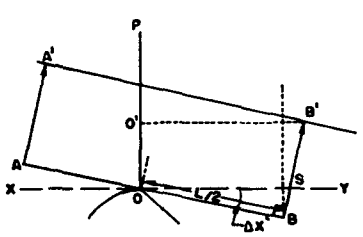

FIG. 3. Growth of width along OP as a result of unfolding.

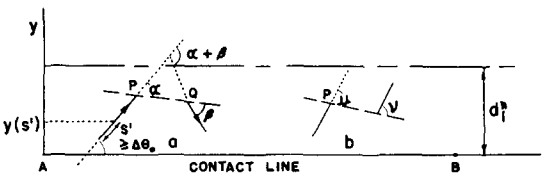

FIG. 4. The contribution to Θ of several configurations.

points[11] from that of its corresponding contact line by a magnitude lying between $\pi/2$, and $\Delta\theta_0$, where $0 \leq \Delta\theta_0 < \pi/2$, can be made arbitrarily small by a sufficiently large number n of d-unfoldings. Loosely speaking, the tangent of the subtended curve approaches arbitrarily closely that of its contact line.[12]

Suppose, on the contrary, that the length l of L.A. curve had a lower bound l_0, independent of the number n of d-unfoldings. We proceed to show that the total variation Θ could then be made arbitrarily large by taking n large enough. The contradiction with property (4) of unfoldings show that l cannot have a lower bound.

Let the c_n subtended curves lie at the maximum distances of d_1^n, d_2^n, $\cdots$, $d_{c_n}^n$ from their respective contact lines, have total lengths of L_1, L_2, $\cdots$, L_{c_n}, and contain lengths l_1, l_2, $\cdots$, l_{c_n} of large angle curve. Consider a particular subtended curve with values d_i^n, L_i, and l_i. Choose a Cartesian frame (see Fig. 4) whose x axis coincides with the contact line and whose positive y axis points toward the Γ_n curve. Suppose s to be the arc length measured along the subtended curve from A (one contact point) and $s' = \psi(s)$ to be the length of L.A. curve lying between A and s. (Throughout this discussion, the reference line for θ is taken to be the contact line.) Let $y(s')$ be the y coordinate of the point on the L.A. curve corresponding to s'; it follows that $0 \leq y(s') \leq d_i^n$, and also that $|dy/ds'| \geq \sin \Delta\theta_0$ at all points at which the derivative is defined. Denote by $Y(s) = \int_0^s |\sin \theta(t)|\, d\psi/dt\, dt$ the cumulative projected length of the L.A. curve on the y axis; it is a monotone increasing function of s.[12a]

We assert that every increase in $Y(s)$ by an amount of $d_i^n + \epsilon'$, where ϵ' is arbitrarily small, contributes at least $2\Delta\theta_0$ to Θ. To establish this consider any arbitrary initial point s_1', and suppose

[11] Points of jump discontinuities in $\theta(s)$ at which either value of θ lies in the stated range are included in the L.A. curve.

[12] The same proof as that given here establishes the theorem for subtended curve (if any) whose angle differs from that of its corresponding contact line by a magnitude lying between $\pi - \Delta\theta_0$ and $\pi/2$. Although it can probably be shown that such curve cannot be present after a sufficiently large number of d-unfoldings, the matter is of no importance to our main result.

[12a] Define $d\psi/dt = 1$ at the cusps of $\psi(t)$.

for definiteness that $y(s')$ is an increasing function at s_1' (i.e., at s_1', y either has a positive right-hand slope or a jump discontinuity to a larger value). Then if Y increases by at least $d_i^n + \epsilon' \geq [d_i^n - y(s_1')] + \epsilon'$ either (1) dy/ds' must change sign at a point of continuity of $y(s')$ (i.e., a cusp) or (2) $y(s')$ must undergo a jump discontinuity to a smaller value. If neither of these events occurred, so that dy/ds' remained positive and no downward jump discontinuity occurred, y would overshoot d_i^n which is impossible; that is,

$$d_i^n + \epsilon \leq \Delta Y = \int_{s_1}^{s_2} |\sin\theta| \frac{d\psi}{ds}\, ds = \int_{s_1'}^{s_2'} \left|\frac{dy}{ds'}\right| ds' = y(s_2') - y(s_1'),$$

or $y(s_2') \geq y(s_1') + d_i^n + \epsilon > d_i^n$. But either event, (1) or (2), increases Θ by at least $2\ \Delta\theta_0$. Thus the behavior of the L.A. portions (heavy lines) of the Γ_n curve corresponding to the two events is shown in Fig. 4 in which the arrows show the direction of increasing s'; event (1) corresponds to configuration a with P and Q at the same level (equal y projections); event (2) corresponds to configuration a or b with Q lower than P, as actually shown.[13] Now it is easy to see that the total variation Θ of θ between two points[14] of the Γ_n curve is at least as great as that between the same two points along a substituted curve consisting merely of a straight line joining the two points. But since the tangent of the L.A. curve makes an angle of at least $\Delta\theta_0$ with the x axis, it follows from elementary geometry that there is a contribution to Θ of at least $2\Delta\theta_0$ from configurations a and b. Thus the contribution from a is $\alpha + \beta$ which must be at least $2\Delta\theta_0$, whereas the contribution from b is the sum of two angles μ and ν each of which exceeds $\Delta\theta_0$. The same argument clearly applies if y is a decreasing function at s_1'. Therefore the assertion is established.

In addition to the preceding contribution to Θ arising from the crowding of L.A. curve between the contact line and the d line there are two end contributions, each with a magnitude of at least $\Delta\theta_0$ and arising between a contact point and the nearest point (via the curve) of the L.A. curve. To see this we note that the variation between two points cannot be less than the magnitude of the difference of θ at the two points (i.e., the values of θ giving the largest difference in case either point is a cusp). At the end of the L.A. curve the magnitude of θ is at least $\Delta\theta_0$. At the contact point (e.g., A of Fig. 4), θ must be set equal to 0 (and A will be the contact point of only one contact line) provided the left-hand slope of Γ_n is zero; if not, A will be the contact point for two contact lines (possibly coincidental), but the variation between the two ends of L.A. curve on either side of A cannot be overestimated by still associating, on either side of A, a value of $\theta = 0$, and calculating the end contribution from each side separately, since any angle between the two contact lines themselves can only further raise the variation so estimated. Hence, a minimum contribution to the variation of $\Delta\theta_0$ may be associated with each end of the L.A. portion of a given subtended curve.

We conclude that the contribution of the L.A. curve to the variation is at least

$$2\ \Delta\theta_0\ \{\text{int}\ [Y(L_i)/(d_i^n + \epsilon')] + 1\},$$

where int denotes the integral part of the bracket quotient. But, choosing ϵ' to be sufficiently small so that this expression is greater than $2\ \Delta\theta_0[Y(L_i)/d_i^n]$, and taking into account the relation

$$Y(L_i) = \int_0^{L_i} |\sin\theta| \frac{d\psi}{ds}\, ds \geq \sin\Delta\theta_0 \int_0^{L_i} \frac{d\psi}{ds}\, ds = \sin\Delta\theta_0 l_i,$$

we see that the minimum contribution to Θ made by the L.A. curve subtended by the ith contact line is $\sin(\Delta\theta_0) l_i 2\Delta\theta_0/d_i^n$. Adding the contribution from the L.A. curve subtended by each of the contact lines, we have

$$\Theta > 2\ \Delta\theta_0 \sin(\Delta\theta_0) \sum_{i=1}^{c_n} \frac{l_i}{d_i^n} \geq 2\ \Delta\theta_0 \sin(\Delta\theta_0) \frac{l_0}{d_M^n},$$

where $l_0 = \sum_{i=1}^{c_n} l_i$ is the postulated lower bound for the total length of high angle curve. Now for any fixed $\Delta\theta_0$, d_M^n may be made arbitrarily small and therefore Θ arbitrarily large by a sufficiently large number n of d-unfoldings. But this contradicts the existence of a fixed upper bound for Θ and proves that the total length of large angle curve cannot in fact have a finite lower bound l_0, but rather can be made arbitrarily small by selecting

[13] If point P should lie to the right of Q in either configuration, the same argument as that given shows the minimum contribution to the variation is $\pi/2 > 2\Delta\theta_0$.

[14] We use the term variation between two points here and later on to mean the total variation along the path connecting the points including any discrete contributions arising from cusps at the end points.

n to be sufficiently large; furthermore, this is true no matter how small a value is chosen for $\Delta\theta_0$.

Theorem III—Replacement theorem. The possible elevation ΔE in free energy $E = \int \gamma(\theta)\, ds$ occasioned by a replacement transformation, in which the n times d-unfolded curve Γ_n is transformed into the convex curve Γ_n' by replacing all subtended curves by their contact lines, can be made less than an arbitrarily small ϵ by choosing n large enough.

Let λ_i denote the length of the contact line i, and γ_i its energy. Then the energy change ΔE upon replacement is

$$\Delta E = \int_{\Gamma_n'} \gamma(\theta)\, ds - \int_{\Gamma_n} \gamma(\theta)\, ds$$

$$= \sum_{i=1}^{c_n} \gamma_i\lambda_i - \sum_{i=1}^{c_n} \int_i \gamma(\theta)\, ds < \sum_{i=1}^{c_n} \gamma_i\lambda_i - \sum_{i=1}^{c_n} \gamma_i^*\lambda_i + l\gamma_M$$

where the integrals $\int_i$ are taken over the curve subtended by the ith contact line in Γ_n, l is the total length of the L.A. curve in all the intervals, γ_M the largest value γ ever assumes, and γ_i^* the smallest value of $\gamma(\theta)$ in an interval of $\pm\Delta\theta_0$ about the ith contact line. The inequality comes about by replacing each of the integrals by the smaller value obtained, by (1) including only those contributions of the integral coming from nonoverlapping curve, defined to be that subtended curve whose x projection (on the contact line) does not coincide with that of any other portion of the subtended curve, and (2) by using the following inequality for each of the integrals (dropping the subscript i)

$$\int \gamma(\theta)\, ds \geq \int_{LA} \gamma(\theta)\, ds + \int_{SA} \gamma^*\, ds > \int_{LA} \gamma(\theta)\, ds + \gamma^* l_{SAX} + (\gamma^* l_{LAX} - \gamma_M l_{LA}) \geq \gamma^*\lambda - \gamma_M l_{LA},$$

where LA and SA denote large angle and small angle parts of the given subtended curve, X denotes the x projection, and the subscripted l's are the corresponding lengths. Now since γ is uniformly continuous (continuous at all points of a closed interval), $\Delta\theta_0$ may be chosen so small that for all θ $|\gamma(\theta) - \gamma(\theta + \Delta\theta)| < \epsilon/2L$ for all $|\Delta\theta| < \Delta\theta_0$. But the arguments of γ_i and γ_i^* differ by less than $\Delta\theta_0$; therefore $|\gamma_i - \gamma_i^*| < \epsilon/2L$ for all i. For this particular $\Delta\theta_0$; Theorem III guarantees that n may be chosen so large that $l < \epsilon/2\gamma_M$. Therefore, substituting into the previous expression for ΔE we obtain for sufficiently large n, $\Delta E < \epsilon/2 + \epsilon/2 < \epsilon$ for arbitrarily small ϵ.

Theorem IV—The convex equilibrium form theorem. Corresponding to any nonconvex curve with a certain value of $E = \int \gamma(\theta)\, ds$, there is always a convex curve Γ'' of the same area but a lower value of E. Therefore, the equilibrium form must be convex.

We proceed to give a construction that will convert a nonconvex curve Γ to a convex curve Γ'' with the same area but a lower energy E. If the curve is not convex there will be at least one contact line. Unfolding about this line will increase the area from an initial value A, to $A + \Delta A$, where ΔA is some finite quantity. This is then followed by the number (if any) of d-unfoldings required to permit a replacement transformation to a curve Γ', guaranteed by Theorem III, that gives $\Delta E < E(1 - k_1)/k_1$, where k_1 is a fraction given by $k_1^2 = A/(A + \Delta A)$. Since these unfoldings (if any) leave E fixed and increase A, we have for the convex curve Γ', the relations

$$E' = E + \Delta E < E/k_1$$

$$A' \geq A + \Delta A.$$

The curve Γ' is now subjected to a uniform contraction to a curve Γ'' such that all radius vectors from some origin are multiplied by a constant $k < 1$; since $r'' = kr'$ and $ds'' = kds'$, we have

$$\frac{A''}{A'} = \frac{\int r''^2\, d\theta}{\int r'^2\, d\theta} = k^2$$

$$\frac{E''}{E'} = \frac{\int \gamma\, ds''}{\int \gamma\, ds'} = k$$

The constant k is chosen so that the final area is equal to the original area, i.e.,

$$A = A'' = k^2 A'$$

Combining this equation with the preceding ones, we see that k^2 has the fractional upper bound

$$k^2 \leq A/(A + \Delta A) = k_1^2$$

The corresponding change in energy from the original value is

$$\Delta E_T = E'' - E' + E' - E = E'(k - 1) + \Delta E$$

$$\leq E'(k_1 - 1) + \Delta E < (E/k_1)(k_1 - 1) + E(1 - k_1)/k_1 = 0$$

Therefore ΔE_T is negative and Theorem IV is established.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Traces of Products of Angular Momentum Matrices. II. Spherical Basis

E. AMBLER, J. C. EISENSTEIN, AND J. F. SCHOOLEY*
National Bureau of Standards, Washington, D. C.
(Received March 2, 1962)

Closed formulas are given for $\mathrm{Tr}\, J_\alpha{}^p J_\beta{}^q J_\gamma{}^r \cdots$ where $\alpha, \beta, \gamma \cdots$ are equal to $+1$, 0, or -1, and p, q, r are non-negative integers for which $p + q + r + \cdots \leq 9$. The procedures used in evaluating the traces are described. In order to facilitate numerical evaluation from existing tables, simple relationships are given between traces in the spherical basis and traces in the Cartesian basis. The use of these tables for evaluating expectation values of certain operators is discussed in relation to other methods based on the coupling and recoupling theory of angular momentum.

IN a previous paper[1] we gave formulas and tables for evaluating traces of products of the Cartesian components of angular momentum operators. As an application we cited the calculation of thermodynamic properties of paramagnetic salts from the spin-Hamiltonian $\mathfrak{IC}$. One needs to evaluate quantities such as $\langle O\mathfrak{IC}^n\rangle \equiv (\mathrm{Tr}\, O\mathfrak{IC}^n)/(\mathrm{Tr}\, 1)$. In this article we give similar formulas for the angular momentum components in the spherical basis. This additional compilation seemed to be useful for certain problems in paramagnetic resonance and relaxation. For example, in calculating the moments of an absorption line where the width is due to dipolar and exchange interactions, it is necessary to write these parts of the Hamiltonian in terms of the raising and lowering operators.[2] In this way the Hamiltonian can be divided into parts corresponding to zero, one, and two spin flips, and the terms which contribute to the moments of the different order Zeeman lines can be separated. It did not seem obvious, moreover, how to obtain the traces in one representation from those in the other. Before going on to discuss our method of evaluation of the traces we should like to discuss briefly two alternative approaches to the evaluation of quantities such as $\langle O\mathfrak{IC}^n\rangle$.

The first of these has been given by Rose,[3] who obtained formulas for quantities defined as

$$Z_n \equiv \mathrm{Tr}\,(T^{k_1}_{q_1} T^{k_2}_{q_2} \cdots T^{k_n}_{q_n})$$

in terms of 3-j and 6-j symbols. The T^k_q are components of tensor operators. Although the formulas for Z_n form a discernible pattern, they are not particularly simple to evaluate when the product contains more than a few components, say for $n > 3$. The method does have special advantages, however, when the spin-Hamiltonian represents an S-state ion. Consider, for example, the case of the Gd^{3+} ion in $GdCl_3$,[4] where $S = 7/2$ and $\mathfrak{IC}$ contains powers of $\mathbf{S}$ up to the sixth. A straightforward expansion of $\langle O\mathfrak{IC}^n\rangle$ in terms of the components of $\mathbf{S}$ leads rapidly to products with a large number of factors. We note, however, that the spin-Hamiltonian may be written in terms of tensor operators,

$$\mathfrak{IC} = G \sum_q (-)^q T^1_q(\mathbf{H}) T^1_{-q}(\mathbf{S}) + A_{20} T^2_0(\mathbf{S}) + A_{40} T^4_0(\mathbf{S}) + A_{60} T^6_0(\mathbf{S}) + A_{66}\{T^6_6(\mathbf{S}) + T^6_{-6}(\mathbf{S})\}$$

so that powers of $\mathfrak{IC}$ fall naturally into the form discussed by Rose. This procedure of writing the spin-Hamiltonian in terms of tensor operators can be carried out for any value of S, of course, but there is no great advantage in doing so for the majority of cases of interest where S is small, so that, in general, the use of tables of the kind we have prepared is more economical.

Another method[5] that has been suggested for handling the problem is to write the spin-Hamiltonian in terms of the scalar products of vectors, thus,

$$\mathfrak{IC} = g_\parallel \beta H_z S_z + g_\perp \beta (H_x S_x + H_y S_y) + D[S_z^2 - \tfrac{1}{3}S(S+1)] + A S_z I_z + B(S_x I_x + S_y I_y)$$

becomes

$$\mathfrak{IC} = (g_\parallel - g_\perp)\beta(\mathbf{H}\cdot\mathbf{e})(\mathbf{S}\cdot\mathbf{e}) + g_\perp \beta(\mathbf{H}\cdot\mathbf{S}) + D[(\mathbf{S}\cdot\mathbf{e})^2 - \tfrac{1}{3}S(S+1)] + (A - B)(\mathbf{S}\cdot\mathbf{e})(\mathbf{I}\cdot\mathbf{e}) + B\mathbf{S}\cdot\mathbf{I},$$

where $\mathbf{e}$ is a unit vector along the z axis defined by

* National Academy of Sciences, National Research Council, Resident Research Associate. This note should also have appeared in reference 1.

[1] E. Ambler, J. C. Eisenstein, and J. F. Schooley, J. Math. Phys. **3**, 118 (1962).

[2] J. H. Van Vleck, Phys. Rev. **74**, 1168 (1948); A. Wright, *ibid.* **76**, 1826 (1949); A. G. Anderson, *ibid.* **115**, 863 (1959); Hung Cheng, *ibid.* **124**, 1359 (1961).

[3] M. E. Rose, J. Math. Phys. **3**, 409 (1962).

[4] C. A. Hutchison, Jr., B. R. Judd, and D. F. D. Pope, Proc. Phys. Soc. (London) **B70**, 514 (1957).

[5] See, e.g., U. Fano, Revs. Modern Phys. **29**, 74 (1957).

the crystal. In a power series expansion, terms containing a number of scalar products will appear, and the aim is to use recoupling procedures[6] to pick out tensors of rank zero, since the trace is nonvanishing only for the zero rank tensors. Thus, for example, we proceed with $(\mathbf{S}\cdot\mathbf{e})^2$ as follows:

$$(\mathbf{S}\cdot\mathbf{e})^2 = (\mathbf{S}\cdot\mathbf{e})(\mathbf{S}\cdot\mathbf{e}) = \sum_{k=0}^{2} [\mathbf{S}\times\mathbf{S}]^{(k)}\cdot[\mathbf{e}\times\mathbf{e}]^{(k)}$$

$$\begin{aligned}\mathrm{Tr}\,(\mathbf{S}\cdot\mathbf{e})^2 &= \mathrm{Tr}\,\{[\mathbf{S}\times\mathbf{S}]^{(0)}[\mathbf{e}\times\mathbf{e}]^{(0)}\}\\ &= (2S+1)^{\frac{1}{2}}\langle S\,||[\mathbf{S}\times\mathbf{S}]^{(0)}||\,S\rangle/\sqrt{3}\\ &= \tfrac{1}{3}S(S+1)(2S+1).\end{aligned}$$

This method is not simple to apply when the number of factors in the product is large, so a tabulation of the traces of products of angular momentum operators should be useful.

We shall take the operators in the spherical basis to be $J_\pm = J_x \pm iJ_y$ and J_z. Their products, like the Cartesian products, can all be derived from $\mathrm{Tr}\,J_z^{2n}$ where n is a positive integer. In the systematic derivation one starts with the nonzero traces J_z^2 and J_+J_-, goes on to $J_+J_zJ_-$ and then to the traces which contain four angular momentum operators. All possible traces are evaluated for a given number of operators. The number of operators is then increased by one and the process repeated.

It is necessary to use the commutation relations

$$J_\pm J_z - J_zJ_\pm = \mp J_\pm$$

and

$$J_+J_- - J_-J_+ = 2J_z.$$

Note also that J_+J_- and J_-J_+ are functions only of J_z.

$$J_\pm J_\mp = J^2 - J_z^2 \pm J_z.$$

When the angular momentum operators are expressed in a Cartesian basis the number of traces which it is necessary to evaluate can be drastically reduced by using symmetry arguments. When the operators are expressed in the spherical basis symmetry arguments can again be used to reduce the labor of calculation. The reduction in the number of traces is not so great, however, because the number of symmetry operations is smaller. Thus, in the simple case of the second-order traces it is necessary to evaluate only J_z^2 for the Cartesian basis but for the spherical basis both J_z^2 and J_+J_- are needed.

The rules we have used to simplify the calculations are the following:

(1) The trace is zero unless the number of raising operators is the same as the number of lowering operators.

(2) If J_+ is substituted for J_- and J_- for J_+ the trace is unchanged when J_z occurs an even number of times, and changes only in sign when J_z occurs an odd number of times. A rotation through 180° about a line in the xy plane which bisects the angle between the x and y axes in the first quadrant sends x into y, y into x, and z into $-z$. Hence J_+ goes into iJ_-, J_- into $-iJ_+$ and J_z into $-J_z$. Since J_+ and J_- occur the same number of times in all nonvanishing traces rule (2) clearly holds.

(3) If the order of the operators is reversed the trace changes sign if J_z occurs an odd number of times and is unchanged if J_z occurs an even number of times. Consider some arbitrary product such as $J_+J_-J_+J_z\cdots J_-$. We have

$$\mathrm{Tr}\,(J_+J_-J_+J_z\cdots J_-) = \mathrm{Tr}\,(J_+J_-J_+J_z\cdots J_-)^*$$

since all the traces we obtain here are real.

$$\mathrm{Tr}\,(J_+J_-J_+J_z\cdots J_-)^* = \mathrm{Tr}\,(J_+J_-J_+J_z\cdots J_-)^\dagger$$

since only diagonal elements are involved in the trace.

$$\mathrm{Tr}\,(J_+J_-J_+J_z\cdots J_-)^\dagger = \mathrm{Tr}\,(J_-^\dagger\cdots J_z^\dagger J_+^\dagger J_-^\dagger J_+^\dagger)$$

by the well known rule for hermitian conjugate of a product of operators.

$$\mathrm{Tr}\,(J_-^\dagger\cdots J_z^\dagger J_+^\dagger J_-^\dagger J_+^\dagger) = \mathrm{Tr}\,(J_+\cdots J_zJ_-J_+J_-)$$

since $J_-^\dagger = J_+$, $J_+^\dagger = J_-$, and $J_z^\dagger = J_z$.

$$\mathrm{Tr}\,(J_+\cdots J_zJ_-J_+J_-) = \pm\mathrm{Tr}\,(J_-\cdots J_zJ_+J_-J_+)$$

by rule (2). Consequently, rule (3) holds.

It follows immediately that the symmetrical products such as $J_+J_-J_zJ_-J_+$ have zero trace.

A useful application of the tables which follow is in the evaluation of Rose's quantity Z_n when the rank of the statistical operators occurring therein is not too large. For example, we have

$$\begin{aligned}&\mathrm{Tr}\,T_2^2T_1^2T_{-1}^2T_{-2}^2\\ &\quad = -(9/4)(a_2)^4J_+^2(J_zJ_+ + J_+J_z)(J_zJ_- + J_-J_z)J_-^2,\end{aligned}$$

where T_q^k and a_k are defined in Table IV of reference 1. One easily finds, by using results given in Table I,

[6] U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic Press Inc., New York, 1959).

that

$$\text{Tr } T_2^2 T_1^2 T_{-1}^2 T_{-2}^2 = \frac{-5(4J^2+4J+21)(J-1)(J+2)}{7J(J+1)(2J+1)(2J-1)(2J+3)}.$$

This example illustrates one advantage of using our table; namely, that the traces are obtained as quotients of simple polynomials in J.

Table I gives analytical expressions for traces of products of angular momentum operators in the spherical basis. Since it is frequently tedious to evaluate these expressions for particular values of J, Table I also gives the traces in the Cartesian basis which yield the same analytical expressions. The traces in the spherical basis can then be readily evaluated by referring to Table III of reference 1.

Finally, further to facilitate the calculations, the values of frequently occurring numerical factors are given in Table II. In the table the following abbreviations are used.

$$A = J(J+1)(2J+1),$$
$$B = (2J-1)(2J+3),$$
$$C = (J-1)(J+2),$$
$$D = (2J-3)(2J+5),$$
$$E = (J-2)(J+3).$$

These quantities are related to the normalization constants given in Table IV of reference 1 in the following way:

$$a_1 = \sqrt{3}\, A^{-\frac{1}{2}},$$
$$a_2 = \sqrt{5}\, (AB)^{-\frac{1}{2}},$$
$$a_3 = \sqrt{7}\, (ABC)^{-\frac{1}{2}},$$
$$a_4 = \tfrac{3}{2}(ABCD)^{-\frac{1}{2}}.$$

TABLE I. Traces of products of angular momentum operators. $A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_z^2 = \tfrac{1}{3}A$$

$$\text{Tr } J_+J_- = \tfrac{2}{3}A = 2\,\text{Tr } J_z^2$$

$$\text{Tr } J_+J_zJ_- = -\tfrac{1}{3}A = 2i\,\text{Tr } J_xJ_yJ_z$$

$$\text{Tr } J_z^4 = (1/15)A(3J^2+3J-1)$$

$$\text{Tr } J_+J_z^2J_- = (1/15)A(2J^2+2J+1) = 2\,\text{Tr } J_x^2J_y^2$$

$$\text{Tr } J_+J_zJ_-J_z = (2/15)AC = 2\,\text{Tr } J_xJ_yJ_xJ_y$$

$$\text{Tr } J_+^2J_-^2 = (2/15)AB = 2\,\text{Tr } (J_x^2J_y^2 + J_xJ_yJ_xJ_y)$$

$$\text{Tr } J_+J_-J_+J_- = (4/15)A(2J^2+2J+1) = 8\,\text{Tr } J_x^2J_y^2$$

$$\text{Tr } J_+J_z^3J_- = -(1/15)A(3J^2+3J-1) = 2i\,\text{Tr } J_x^3J_yJ_z$$

$$\text{Tr } J_+J_z^2J_-J_z = -(1/15)AC = 2i\,\text{Tr } J_x^2J_yJ_xJ_z$$

$$\text{Tr } J_+^2J_zJ_-^2 = -(2/15)AB = 4i\,\text{Tr } (J_x^3J_yJ_z + J_x^2J_yJ_xJ_z)$$

$$\text{Tr } J_+J_zJ_-J_+J_- = -(2/15)A(2J^2+2J+1) = 4i\,\text{Tr } (J_x^3J_yJ_z - J_x^2J_yJ_xJ_z)$$

$$\text{Tr } J_z^6 = (1/21)A(3J^4+6J^3-3J+1)$$

$$\text{Tr } J_+J_z^4J_- = (1/105)A(6J^4+12J^3+14J^2+8J-5) = 2\,\text{Tr } J_x^4J_y^2$$

$$\text{Tr } J_+J_z^3J_-J_z = (1/105)AC(6J^2+6J-1) = 2\,\text{Tr } J_z^3J_yJ_xJ_y$$

$$\text{Tr } J_+J_z^2J_-J_z^2 = (2/105)AC(3J^2+3J-4) = 2\,\text{Tr } J_x^2J_yJ_z^2J_y$$

$$\text{Tr } J_+^2J_z^2J_-^2 = (2/105)AB(J^2+J+5) = 4\,\text{Tr } (J_x^4J_y^2 - J_x^2J_y^2J_z^2)$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_- = (2/105)ABC = 2\,\text{Tr } (J_x^2J_yJ_z^2J_y + J_x^2J_yJ_zJ_yJ_z)$$

TABLE I. Continued

$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+^2 J_z J_-^2 J_z = (2/105)AB(J^2 + J - 9)$$
$$= 4 \text{ Tr } (J_x^2 J_y^2 J_z^2 + J_z J_y J_z J_x J_y J_z)$$
$$\text{Tr } J_+^2 J_- J_z^2 J_- = (2/105)ABC$$
$$\text{Tr } J_+ J_z J_+ J_- J_z J_- = (2/105)ABC$$
$$\text{Tr } J_+ J_z^2 J_- J_+ J_- = (2/105)A(4J^4 + 8J^3 + 14J^2 + 10J - 1)$$
$$= 4 \text{ Tr } (J_x^2 J_y^2 J_z^2 + J_x^2 J_y J_z^2 J_y)$$
$$\text{Tr } J_+ J_z J_- J_+ J_z J_- = (2/105)A(4J^4 + 8J^3 + 14J^2 + 10J - 1)$$
$$\text{Tr } J_+ J_z J_- J_+ J_- J_z = (8/105)AC(J^2 + J + 1) = 8 \text{ Tr } J_x^2 J_y J_z J_y J_z$$
$$\text{Tr } J_+^3 J_-^3 = (4/35)ABC$$
$$\text{Tr } J_+^2 J_- J_+ J_-^2 = (4/105)AB(3J^2 + 3J + 1) = 8 \text{ Tr } (J_x^4 J_y^2 + J_x^3 J_y J_z J_y)$$
$$\text{Tr } J_+ J_- J_+ J_- J_+ J_- = (4/105)A(12J^4 + 24J^3 + 21J^2 + 9J + 4)$$
$$= 8 \text{ Tr } (J_x^6 - 3J_x^3 J_y J_z J_y)$$
$$\text{Tr } J_+ J_z^5 J_- = -(1/21)A(3J^4 + 6J^3 - 3J + 1) = 2i \text{ Tr } J_z^5 J_y J_z$$
$$\text{Tr } J_+ J_z^4 J_- J_z = -(1/105)AC(9J^2 + 9J - 5) = 2i \text{ Tr } J_x^4 J_y J_x J_z$$
$$\text{Tr } J_+ J_z^3 J_- J_z^2 = -(1/105)AC(3J^2 + 3J - 4) = 2i \text{ Tr } J_x^3 J_y J_x^2 J_z$$
$$\text{Tr } J_+^2 J_z^3 J_-^2 = -(2/105)AB(3J^2 + 3J + 1)$$
$$= 4i \text{ Tr } (J_x^5 J_y J_z - J_x^3 J_y J_x^2 J_z)$$
$$\text{Tr } J_+^2 J_z^2 J_- J_z J_- = -(4/105)ABC = 8i \text{ Tr } (J_z^2 J_y^2 J_x J_y J_z - J_z^2 J_y J_z J_y J_z J_y)$$
$$\text{Tr } J_+^2 J_z^2 J_-^2 J_z = -(2/105)AB(J^2 + J - 9)$$
$$= 4i \text{ Tr } (J_z^2 J_y^2 J_x J_y J_z + J_x^2 J_y^2 J_x J_z J_y)$$
$$\text{Tr } J_+^2 J_z J_- J_z^2 J_- = -(2/105)ABC$$
$$\text{Tr } J_+ J_z J_+ J_z J_- J_z J_- = -(2/105)ABC$$
$$\text{Tr } J_+ J_z^3 J_- J_+ J_- = -(2/105)A(6J^4 + 12J^3 + 14J^2 + 8J - 5)$$
$$= 4i \text{ Tr } (J_x^5 J_y J_z - J_x^4 J_y J_x J_z)$$
$$\text{Tr } J_+ J_z^2 J_- J_+ J_z J_- = -(2/105)A(6J^4 + 12J^3 + 14J^2 + 8J - 5)$$
$$\text{Tr } J_+ J_z^2 J_- J_+ J_- J_z = -(4/105)AC(J^2 + J + 1)$$
$$= 4i \text{ Tr } (-J_x^2 J_y J_x J_y J_z J_y + J_z J_y J_x J_y J_z J_y J_z)$$
$$\text{Tr } J_+ J_z J_- J_+ J_z J_- J_z = -(4/105)AC(J^2 + J + 1)$$
$$\text{Tr } J_+^3 J_z J_-^3 = -(6/35)ABC$$
$$\text{Tr } J_+^3 J_- J_z J_-^2 = -(2/35)ABC$$
$$\text{Tr } J_+^2 J_z J_- J_+ J_-^2 = -(2/105)AB(5J^2 + 5J + 4)$$
$$= 8i \text{ Tr } (J_x^4 J_y J_x J_z - J_x^2 J_y^2 J_x J_z J_y)$$
$$\text{Tr } J_+^2 J_z J_-^2 J_+ J_- = -(2/15)AB(J^2 + J)$$
$$= 8i \text{ Tr } (J_x^5 J_y J_z - J_x^2 J_y J_x J_y J_z J_y)$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+^2J_-J_+J_-J_zJ_- = (2/105)ABC$$

$$\text{Tr } J_+^2J_-J_zJ_+J_-^2 = (2/105)ABC$$

$$\text{Tr } J_+J_zJ_-J_+J_-J_+J_- = -(2/105)A(12J^4+24J^3+21J^2+9J+4)$$
$$= 8i\,\text{Tr}\,(J_x^5J_yJ_z - 3J_x^3J_y^2J_zJ_y)$$

$$\text{Tr } J_z^8 = (1/45)A(5J^6+15J^5+5J^4-15J^3-J^2+9J-3)$$

$$\text{Tr } J_+J_z^6J_- = (1/315)A(10J^6+30J^5+55J^4+60J^3-23J^2-48J+21) = 2\,\text{Tr}\,J_x^6J_y^2$$

$$\text{Tr } J_+J_z^5J_-J_z = (1/315)AC(10J^4+20J^3+10J^2-3)$$
$$= 2\,\text{Tr}\,J_x^5J_yJ_zJ_y$$

$$\text{Tr } J_+J_z^4J_-J_z^2 = (1/315)AC(10J^4+20J^3-17J^2-27J+12)$$
$$= 2\,\text{Tr}\,J_x^4J_yJ_x^2J_y$$

$$\text{Tr } J_+J_z^3J_-J_z^3 = (2/315)AC(5J^4+10J^3-13J^2-18J+12)$$
$$= 2\,\text{Tr}\,J_x^3J_yJ_x^3J_y$$

$$\text{Tr } J_+^2J_z^4J_-^2 = (2/315)AB(J^4+2J^3+14J^2+13J-9)$$
$$= 4\,\text{Tr}\,(J_x^6J_y^2 - J_x^2J_yJ_xJ_y^2J_xJ_y)$$

$$\text{Tr } J_+^2J_z^3J_-J_zJ_- = (2/315)ABC(J^2+J+6)$$
$$= 4\,\text{Tr}\,(J_x^4J_yJ_zJ_yJ_z + J_x^3J_yJ_zJ_xJ_zJ_y)$$

$$\text{Tr } J_+^2J_z^3J_-^2J_z = (2/315)AB(J^4+2J^3-4J^2-5J-15)$$
$$= 4\,\text{Tr}\,(J_x^4J_yJ_x^2J_y - J_x^2J_y^2J_x^2J_y^2)$$

$$\text{Tr } J_+^2J_z^2J_-J_z^2J_- = (2/315)ABC(J^2+J)$$
$$= 4\,\text{Tr}\,(J_x^4J_yJ_x^2J_y - J_xJ_yJ_xJ_yJ_xJ_yJ_xJ_y)$$

$$\text{Tr } J_+^2J_z^2J_-J_zJ_-J_z = (2/315)ABCE$$
$$= 4\,\text{Tr}\,(J_x^4J_yJ_x^2J_y - J_x^3J_y^2J_xJ_y^2)$$

$$\text{Tr } J_+^2J_z^2J_-^2J_z^2 = (2/315)AB(J^4+2J^3-10J^2-11J+39)$$
$$= 4\,\text{Tr}\,(J_x^2J_yJ_zJ_x^2J_yJ_z + J_x^2J_yJ_zJ_x^2J_zJ_y)$$

$$\text{Tr } J_+^2J_zJ_-J_z^3J_- = (2/315)ABC(J^2+J-3)$$
$$= 4\,\text{Tr}\,(J_x^2J_yJ_xJ_zJ_yJ_xJ_z + J_xJ_yJ_xJ_yJ_xJ_zJ_xJ_z)$$

$$\text{Tr } J_+^2J_zJ_-J_z^2J_-J_z = (2/315)ABCE$$

$$\text{Tr } J_+^2J_-J_z^4J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_zJ_+J_z^2J_-J_zJ_- = (2/315)ABC(J^2+J)$$

$$\text{Tr } J_+J_zJ_+J_zJ_-J_z^2J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_zJ_+J_zJ_-J_zJ_-J_z = (2/315)ABCE$$

$$\text{Tr } J_+J_zJ_+J_-J_z^3J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_z^2J_+J_-J_z^2J_- = (2/315)ABC(J^2+J-3)$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\begin{aligned}
\operatorname{Tr} J_+J_z^4J_-J_+J_- &= (2/315)A(2J^2+2J-1)(2J^4+4J^3+14J^2+12J+3)\\
&= 4\operatorname{Tr}(J_x^4J_y^4 - J_x^4J_yJ_zJ_yJ_z)\\
\operatorname{Tr} J_+J_z^3J_-J_+J_zJ_- &= (2/315)A(2J^2+2J-1)(2J^4+4J^3+14J^2+12J+3)\\
\operatorname{Tr} J_+J_z^3J_-J_+J_-J_z &= (4/315)AC(2J^4+4J^3+8J^2+6J-3)\\
&= 4\operatorname{Tr}(J_x^2J_y^2J_xJ_zJ_xJ_z + J_x^2J_yJ_xJ_zJ_xJ_zJ_y)\\
\operatorname{Tr} J_+J_z^2J_-J_+J_z^2J_- &= (2/315)A(2J^2+2J-1)(2J^4+4J^3+14J^2+12J+3)\\
\operatorname{Tr} J_+J_z^2J_-J_+J_zJ_-J_z &= (4/315)AC(2J^4+4J^3+8J^2+6J-3)\\
\operatorname{Tr} J_+J_z^2J_-J_+J_-J_z^2 &= (4/315)AC(2J^4+4J^3+5J^2+3J-6)\\
&= 8\operatorname{Tr} J_x^2J_y^2J_zJ_x^2J_z\\
\operatorname{Tr} J_+J_z^2J_-J_zJ_+J_-J_z &= (4/315)AC(2J^4+4J^3+5J^2+3J-6)\\
\operatorname{Tr} J_+J_zJ_-J_zJ_+J_zJ_-J_z &= (4/315)AC(2J^4+4J^3+5J^2+3J-6)\\
\operatorname{Tr} J_+^3J_z^2J_-^3 &= (2/315)ABC(2J^2+2J+33)\\
&= 8\operatorname{Tr}(5J_x^2J_y^2J_xJ_yJ_xJ_y + 5J_x^4J_y^2J_z^2 - 9J_xJ_yJ_xJ_yJ_xJ_zJ_xJ_z - 9J_xJ_yJ_xJ_zJ_xJ_yJ_xJ_z)\\
\operatorname{Tr} J_+^3J_zJ_-J_zJ_-^2 &= (4/315)ABC(J^2+J+3)\\
&= 8\operatorname{Tr}(J_x^2J_yJ_zJ_y^2J_xJ_y - J_x^4J_y^2J_z^2)\\
\operatorname{Tr} J_+^3J_zJ_-^2J_zJ_- &= (2/315)ABC(2J^2+2J-21)\\
&= 8\operatorname{Tr}(J_x^2J_yJ_xJ_zJ_xJ_yJ_z + J_x^2J_yJ_xJ_zJ_yJ_xJ_z)\\
\operatorname{Tr} J_+^3J_zJ_-^3J_z &= (4/315)ABC(J^2+J-24)\\
&= -8\operatorname{Tr}(2J_x^4J_yJ_z^2J_y - J_x^2J_y^2J_x^2J_y^2 - J_x^3J_y^2J_xJ_z^2)\\
\operatorname{Tr} J_+^3J_-J_z^2J_-^2 &= (2/315)ABC(2J^2+2J-3)\\
&= 8\operatorname{Tr}(J_x^3J_y^2J_zJ_xJ_z + J_x^2J_y^2J_zJ_x^2J_z)\\
\operatorname{Tr} J_+^3J_-J_zJ_-J_zJ_- &= (4/315)ABCE\\
\operatorname{Tr} J_+^2J_zJ_+J_-J_zJ_-^2 &= (2/315)ABC(2J^2+2J-3)\\
\operatorname{Tr} J_+^2J_zJ_+J_-^2J_zJ_- &= (4/315)ABCE\\
\operatorname{Tr} J_+^2J_z^2J_-J_+J_-^2 &= (2/315)AB(2J^2+2J+3)(J^2+J+4)\\
&= 8\operatorname{Tr}(J_x^3J_y^2J_xJ_z^2 + J_x^2J_y^2J_xJ_zJ_xJ_z)\\
\operatorname{Tr} J_+^2J_z^2J_-^2J_+J_- &= (2/315)AB(2J^2+2J-1)(J^2+J+12)\\
&= 8\operatorname{Tr}(J_x^4J_yJ_z^2J_y + J_x^2J_yJ_xJ_zJ_xJ_zJ_y)\\
\operatorname{Tr} J_+^2J_zJ_-J_+J_zJ_-^2 &= (2/315)AB(2J^2+2J+3)(J^2+J+4)\\
\operatorname{Tr} J_+^2J_zJ_-J_+J_-J_zJ_- &= (4/315)ABC(J^2+J)\\
\operatorname{Tr} J_+^2J_zJ_-J_+J_-^2J_z &= (2/315)AB(2J^4+4J^3-17J^2-19J-12)\\
&= 8\operatorname{Tr}(J_x^3J_yJ_xJ_zJ_yJ_z + J_x^3J_yJ_zJ_xJ_yJ_z)\\
\operatorname{Tr} J_+^2J_zJ_-J_zJ_+J_-^2 &= (4/315)ABC(J^2+J)\\
\operatorname{Tr} J_+^2J_zJ_-J_zJ_-J_+J_- &= (4/315)ABC(J^2+J+3)
\end{aligned}$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+^2 J_z J_-^2 J_+ J_z J_- = (4/315)ABC(J^2+J+3)$$

$$\text{Tr } J_+^2 J_- J_+ J_z J_- J_z J_- = (4/315)ABC(J^2+J)$$

$$\text{Tr } J_+^2 J_- J_+ J_- J_z^2 J_- = (2/315)ABC(2J^2+2J+3)$$
$$= 8\,\text{Tr}\,(J_x^4 J_y J_z J_y J_z + J_x^3 J_y J_x J_z J_y J_z)$$

$$\text{Tr } J_+^2 J_- J_z J_+ J_- J_z J_- = (2/315)ABC(2J^2+2J+3)$$

$$\text{Tr } J_+^2 J_- J_z^2 J_+ J_-^2 = (2/315)ABC(2J^2+2J+3)$$

$$\text{Tr } J_+ J_z J_+ J_- J_+ J_- J_z J_- = (2/315)ABC(2J^2+2J+3)$$

$$\text{Tr } J_+ J_z^2 J_- J_+ J_- J_+ J_- = (2/315)A(8J^6+24J^5+50J^4+60J^3+53J^2+27J-12)$$
$$= 8\,\text{Tr}\,(3J_x^3 J_y J_x J_z J_y J_z - J_x^3 J_y J_z J_x J_y J_z)$$

$$\text{Tr } J_+ J_z J_- J_+ J_z J_- J_+ J_- = (2/315)A(8J^6+24J^5+50J^4+60J^3+53J^2+27J-12)$$

$$\text{Tr } J_+ J_z J_- J_+ J_- J_+ J_- J_z = (4/315)AC(4J^4+8J^3+7J^2+3J+6)$$
$$= 8\,\text{Tr}\,(3J_x^2 J_y J_x J_z J_x J_z J_y - J_x^3 J_y J_x J_z J_y J_z)$$

$$\text{Tr } J_+ J_z J_- J_+ J_- J_z J_+ J_- = (4/315)AC(4J^4+8J^3+7J^2+3J+6)$$

$$\text{Tr } J_+^4 J_-^4 = (8/315)ABCD$$
$$= 32\,\text{Tr}\,(J_x^2 J_y^2 J_x J_y J_x J_y + J_x^4 J_y^2 J_z^2)$$

$$\text{Tr } J_+^3 J_- J_+ J_-^3 = (4/315)ABC(8J^2+8J-3)$$
$$= 16\,\text{Tr}\,(J_x^4 J_y J_x^2 J_y + J_x^3 J_y^3 J_x J_y)$$

$$\text{Tr } J_+^3 J_-^2 J_+ J_-^2 = (8/315)ABC(4J^2+4J+3)$$
$$= 16\,\text{Tr}\,(J_x^5 J_y J_x J_y + J_x^3 J_y^2 J_x J_y^2)$$

$$\text{Tr } J_+^2 J_- J_+ J_- J_+ J_-^2 = (4/315)AB(J^2+J)(8J^2+8J+5)$$
$$= 16\,\text{Tr}\,(J_x^6 J_y^2 + J_x^2 J_y^2 J_x J_y J_x J_y)$$

$$\text{Tr } J_+^2 J_- J_+ J_-^2 J_+ J_- = (8/315)AB(4J^4+8J^3+5J^2+J+3)$$
$$= 16\,\text{Tr}\,(J_x^4 J_y J_z^2 J_y + J_x^3 J_y^3 J_x J_y + J_x^2 J_y^2 J_x J_y J_x J_y + J_x^4 J_y^2 J_z^2)$$

$$\text{Tr } J_+^2 J_-^2 J_+^2 J_-^2 = (8/315)AB(4J^4+8J^3+5J^2+J+3)$$

$$\text{Tr } J_+ J_- J_+ J_- J_+ J_- J_+ J_- = (16/315)A(J^2+J+1)(8J^4+16J^3+8J^2+3)$$
$$= 32\,\text{Tr}\,(J_x^3 J_y^2 J_x J_z^2 + 3J_x J_y J_x J_y J_z J_x J_z)$$

$$\text{Tr } J_+ J_z^7 J_- = -(1/45)A(5J^6+15J^5+5J^4-15J^3-J^2+9J-3)$$
$$= 2i\,\text{Tr}\,J_x^7 J_y J_z$$

$$\text{Tr } J_+ J_z^6 J_- J_z = -(1/315)AC(25J^4+50J^3-20J^2-45J+21)$$
$$= 2i\,\text{Tr}\,J_x^6 J_y J_x J_z$$

$$\text{Tr } J_+ J_z^5 J_- J_z^2 = -(1/105)AC(5J^4+10J^3-10J^2-15J+8)$$
$$= 2i\,\text{Tr}\,J_x^5 J_y J_x^2 J_z$$

$$\text{Tr } J_+ J_z^4 J_- J_z^3 = -(1/315)AC(5J^4+10J^3-13J^2-18J+12)$$
$$= 2i\,\text{Tr}\,J_x^4 J_y J_x^3 J_z$$

TABLE I. Continued

$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\begin{aligned}
\mathrm{Tr}\, J_+^2 J_z^5 J_-^2 &= -(2/315)AB(5J^4 + 10J^3 + 10J^2 + 5J - 9) \\
&= 4i\,\mathrm{Tr}\,(J_x^5 J_y^3 J_z + J_x^5 J_y^2 J_z J_y) \\
\mathrm{Tr}\, J_+^2 J_z^4 J_- J_z J_- &= -(8/315)ABC(J^2 + J) \\
&= 4i\,\mathrm{Tr}\,(J_x^4 J_y^3 J_x J_z + J_x^4 J_y^2 J_z J_x J_y) \\
\mathrm{Tr}\, J_+^2 J_z^4 J_-^2 J_z &= -(2/105)AB(J^4 + 2J^3 - 6J^2 - 7J + 3) \\
&= 4i\,\mathrm{Tr}\,(J_x^4 J_y^2 J_x J_y J_z + J_x^4 J_y^2 J_x J_z J_y) \\
\mathrm{Tr}\, J_+^2 J_z^3 J_- J_z^2 J_- &= -(2/105)ABC^2 \\
&= 4i\,\mathrm{Tr}\,(J_x^3 J_y^2 J_z J_x^2 J_y - J_x^2 J_y J_x^2 J_y J_x J_z J_y) \\
\mathrm{Tr}\, J_+^2 J_z^3 J_- J_z J_- J_z &= -(4/315)ABCE \\
&= 8i\,\mathrm{Tr}\,(J_x^5 J_y^2 J_z J_y + J_x^3 J_y J_x J_y J_x J_z J_y) \\
\mathrm{Tr}\, J_+^2 J_z^3 J_-^2 J_z^2 &= -(2/315)AB(J^4 + 2J^3 - 10J^2 - 11J + 39) \\
&= 4i\,\mathrm{Tr}\,(J_x^3 J_y^2 J_x^2 J_y J_z + J_x^3 J_y^2 J_x^2 J_z J_y) \\
\mathrm{Tr}\, J_+^2 J_z^2 J_- J_z^3 J_- &= -(4/315)ABC(J^2 + J - 3) \\
&= 8i\,\mathrm{Tr}\,(J_x^3 J_y^2 J_z J_x^2 J_y - J_x^4 J_y J_x^3 J_z) \\
\mathrm{Tr}\, J_+^2 J_z^2 J_- J_z^2 J_- J_z &= -(2/315)ABCE \\
\mathrm{Tr}\, J_+^2 J_z J_- J_z^4 J_- &= -(2/315)ABC(J^2 + J - 3) \\
\mathrm{Tr}\, J_+ J_z J_+ J_z^3 J_- J_z J_- &= -(2/105)ABC^2 \\
\mathrm{Tr}\, J_+ J_z J_+ J_z^2 J_- J_z^2 J_- &= -(4/315)ABC(J^2 + J - 3) \\
\mathrm{Tr}\, J_+ J_z J_+ J_z^2 J_- J_z J_- J_z &= -(2/315)ABCE \\
\mathrm{Tr}\, J_+ J_z J_+ J_z J_- J_z^3 J_- &= -(2/315)ABC(J^2 + J - 3) \\
\mathrm{Tr}\, J_+ J_z^2 J_+ J_z J_- J_z^2 J_- &= -(2/315)ABC(J^2 + J - 3) \\
\mathrm{Tr}\, J_+ J_z^5 J_- J_+ J_- &= -(2/315)A(10J^6 + 30J^5 + 55J^4 + 60J^3 - 23J^2 - 48J + 21) \\
&= 4i\,\mathrm{Tr}\,(J_x^5 J_y^3 J_z - J_x^5 J_y^2 J_z J_y) \\
\mathrm{Tr}\, J_+ J_z^4 J_- J_+ J_z J_- &= -(2/315)A(10J^6 + 30J^5 + 55J^4 + 60J^3 - 23J^2 - 48J + 21) \\
\mathrm{Tr}\, J_+ J_z^4 J_- J_+ J_- J_z &= -(2/105)AC(2J^4 + 4J^3 + 7J^2 + 5J - 4) \\
&= -4i\,\mathrm{Tr}\,(J_x^4 J_y^2 J_z J_x J_y + J_x^4 J_z J_y^2 J_x J_y) \\
\mathrm{Tr}\, J_+ J_z^3 J_- J_+ J_z^2 J_- &= -(2/315)A(10J^6 + 30J^3 + 55J^4 + 60J^3 - 23J^2 - 48J + 21) \\
\mathrm{Tr}\, J_+ J_z^3 J_- J_+ J_z J_- J_z &= -(2/105)AC(2J^4 + 4J^3 + 7J^2 + 5J - 4) \\
\mathrm{Tr}\, J_+ J_z^3 J_- J_+ J_- J_z^2 &= -(2/315)AC(2J^4 + 4J^3 + 5J^2 + 3J - 6) \\
&= 4i\,\mathrm{Tr}\,(J_x^3 J_y J_z^2 J_y^2 J_z - J_x^3 J_y^2 J_z J_x^2 J_y) \\
\mathrm{Tr}\, J_+ J_z^3 J_- J_z J_+ J_- J_z &= -(2/315)AC(2J^4 + 4J^3 + 5J^2 + 3J - 6) \\
\mathrm{Tr}\, J_+ J_z^2 J_- J_+ J_z^2 J_- J_z &= -(2/105)AC(2J^4 + 4J^3 + 7J^2 + 5J - 4) \\
\mathrm{Tr}\, J_+ J_z^2 J_- J_+ J_z J_- J_z^2 &= -(2/315)AC(2J^4 + 4J^3 + 5J^2 + 3J - 6)
\end{aligned}$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+J_z^2J_-J_zJ_+J_zJ_-J_z = -(2/315)AC(2J^4+4J^3+5J^2+3J-6)$$

$$\text{Tr } J_+^3J_z^3J_-^3 = -(2/35)ABC(J^2+J+3)$$
$$= 72i\,\text{Tr}\,(J_x^3J_yJ_x^2J_yJ_zJ_y - J_xJ_yJ_xJ_zJ_yJ_xJ_zJ_yJ_z)$$

$$\text{Tr } J_+^3J_z^2J_-J_zJ_-^2 = -(2/315)ABC(7J^2+7J-6)$$
$$= 8i\,\text{Tr}\,(J_x^4J_y^2J_xJ_yJ_z + J_x^2J_yJ_zJ_xJ_yJ_zJ_yJ_z)$$

$$\text{Tr } J_+^3J_z^2J_-^2J_zJ_- = -(2/315)ABC(5J^2+5J-39)$$
$$= 8i\,\text{Tr}\,(3J_x^5J_y^2J_zJ_y + 3J_x^3J_yJ_xJ_yJ_xJ_zJ_y - J_x^3J_yJ_z^2J_yJ_zJ_y + J_xJ_yJ_xJ_zJ_yJ_xJ_zJ_yJ_z)$$

$$\text{Tr } J_+^3J_z^2J_-^3J_z = -(2/105)ABC(J^2+J-24)$$
$$= 24i\,\text{Tr}\,(J_x^3J_y^2J_xJ_zJ_xJ_y + J_x^3J_yJ_z^2J_yJ_zJ_y)$$

$$\text{Tr } J_+^3J_zJ_-J_z^2J_-^2 = -(2/315)ABC(5J^2+5J-12)$$
$$= 8i\,\text{Tr}\,(J_x^3J_yJ_xJ_y^2J_xJ_z - J_xJ_yJ_xJ_zJ_xJ_zJ_yJ_zJ_y)$$

$$\text{Tr } J_+^3J_zJ_-J_zJ_-J_zJ_- = -(2/105)ABCE$$

$$\text{Tr } J_+^3J_zJ_-J_zJ_-^2J_z = -(2/315)ABC(J^2+J-24)$$

$$\text{Tr } J_+^3J_zJ_-^2J_z^2J_- = -(2/315)ABC(J^2+J+3)$$

$$\text{Tr } J_+^3J_-J_z^3J_-^2 = -(2/105)ABC(J^2+J-3)$$

$$\text{Tr } J_+^3J_-J_z^2J_-J_zJ_- = -(2/315)ABCE$$

$$\text{Tr } J_+^2J_zJ_+J_zJ_-J_zJ_-^2 = -(2/315)ABC(5J^2+5J-12)$$

$$\text{Tr } J_+^2J_zJ_+J_zJ_-^2J_zJ_- = -(2/105)ABCE$$

$$\text{Tr } J_+^2J_zJ_+J_-J_z^2J_-^2 = -(2/105)ABC(J^2+J-3)$$

$$\text{Tr } J_+^2J_zJ_+J_-J_zJ_-J_zJ_- = -(2/315)ABCE$$

$$\text{Tr } J_+^2J_zJ_+J_-J_zJ_-^2J_z = (2/315)ABC(J^2+J+3)$$

$$\text{Tr } J_+^2J_zJ_+J^2-J_z^2J_- = (2/315)ABCE$$

$$\text{Tr } J_+^2J_z^3J_-J_+J_-^2 = -(2/315)AB(J^2+J)(5J^2+5J+11)$$
$$= 8i\,\text{Tr}\,(J_z^5J_y^3J_z - J_x^3J_y^2J_z^2J_yJ_z)$$

$$\text{Tr } J_+^2J_z^3J_-^2J_+J_- = -(2/315)AB(7J^4+14J^3+26J^2+19J-24)$$
$$= 8i\,\text{Tr}\,(J_x^5J_y^3J_z - J_xJ_yJ_xJ_zJ_xJ_yJ_zJ_yJ_z)$$

$$\text{Tr } J_+^2J_z^2J_-J_+J_zJ_-^2 = -(2/315)AB(J^2+J)(5J^2+5J+11)$$

$$\text{Tr } J_+^2J_z^2J_-J_+J_-J_zJ_- = -(2/105)ABC(J^2+J+2)$$
$$= 8i\,\text{Tr}\,(J_x^3J_yJ_z^2J_y^2J_z - J_x^2J_y^2J_z^2J_xJ_yJ_z)$$

$$\text{Tr } J_+^2J_z^2J_-J_+J_-^2J_z = -(2/315)AB(J^4+2J^3-10J^2-11J-24)$$
$$= 8i\,\text{Tr}\,(-J_x^2J_yJ_x^2J_yJ_xJ_zJ_y + J_x^2J_yJ_z^2J_yJ_xJ_zJ_y)$$

$$\text{Tr } J_+^2J_z^2J_-J_zJ_+J_-^2 = -(2/105)ABC(J^2+J+2)$$

$$\text{Tr } J_+^2J_z^2J_-J_zJ_-J_+J_- = -(2/315)ABC(5J^2+5J+6)$$
$$= 8i\,\text{Tr}\,(J_x^3J_yJ_xJ_yJ_xJ_yJ_z + J_x^2J_yJ_z^2J_xJ_yJ_zJ_y)$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+^2J_z^2J_-^2J_+J_zJ_- = -(2/315)ABC(5J^2+5J+6)$$

$$\text{Tr } J_+^2J_z^2J_-^2J_+J_-J_z = -(2/105)AB(J^2+J)(J^2+J-9)$$
$$= 8i\,\text{Tr}\,(J_x^4J_y^2J_xJ_zJ_y + J_x^3J_y^2J_z^2J_yJ_z)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_zJ_-J_zJ_- = -(2/105)ABC(J^2+J+2)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_zJ_-^2J_z = -(2/315)AB(J^4+2J^3-10J^2-11J-24)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_-J_z^2J_- = -(2/315)ABC(J^2+J+6)$$
$$= 8i\,\text{Tr}\,(-J_x^3J_y^3J_x^2J_z + J_x^3J_y^2J_z^2J_yJ_z)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_-J_zJ_-J_z = (2/315)ABCE$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_+J_zJ_-^2 = -(2/105)ABC(J^2+J+2)$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_+J_-J_zJ_- = -(2/315)ABC(J^2+J+6)$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_+J_-^2J_z = (2/315)ABCE$$

$$\text{Tr } J_+^2J_zJ_-J_z^2J_+J_-^2 = -(2/315)ABC(J^2+J+6)$$

$$\text{Tr } J_+^2J_zJ_-J_z^2J_-J_+J_- = -(2/105)ABC(J^2+J)$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_-J_+J_zJ_- = -(2/105)ABC(J^2+J)$$

$$\text{Tr } J_+^2J_zJ_-J_zJ_-J_zJ_+J_- = -(2/315)ABCE$$

$$\text{Tr } J_+^2J_zJ_-^2J_+J_z^2J_- = -(2/105)ABC(J^2+J)$$

$$\text{Tr } J_+^2J_-J_+J_z^2J_-J_zJ_- = -(2/105)ABC(J^2+J+2)$$

$$\text{Tr } J_+^2J_-J_+J_zJ_-J_z^2J_- = -(2/315)ABC(J^2+J+6)$$

$$\text{Tr } J_+^2J_-J_+J_-J_z^3J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+^2J_-J_zJ_+J_zJ_-J_zJ_- = -(2/315)ABC(J^2+J+6)$$

$$\text{Tr } J_+^2J_-J_zJ_+J_-J_z^2J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+^2J_-J_z^2J_+J_-J_zJ_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+^2J_-J_z^3J_+J_-^2 = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_zJ_+J_zJ_-J_+J_-J_zJ_- = -(2/315)ABC(J^2+J+6)$$

$$\text{Tr } J_+J_zJ_+J_zJ_-J_zJ_-J_+J_- = -(2/105)ABC(J^2+J)$$

$$\text{Tr } J_+J_zJ_+J_-J_+J_-J_z^2J_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_zJ_+J_-J_zJ_+J_-J_zJ_- = (2/315)ABC(J^2+J-3)$$

$$\text{Tr } J_+J_z^3J_-J_+J_-J_+J_- = -(2/105)A(4J^6+12J^3+19J^4+18J^3+16J^2+9J-8)$$
$$= 8i\,\text{Tr}\,(J_x^3J_y^3J_z^3 - 3J_x^3J_yJ_x^2J_yJ_zJ_y)$$

$$\text{Tr } J_+J_z^2J_-J_+J_zJ_-J_+J_- = -(2/105)A(4J^6+12J^3+19J^4+18J^3+16J^2+9J-8)$$

$$\text{Tr } J_+J_z^2J_-J_+J_-J_+J_-J_z = -(2/315)AC(4J^4+8J^3+7J^2+3J+6)$$
$$= 8i\,\text{Tr}\,(3J_x^3J_yJ_x^2J_yJ_zJ_y - J_zJ_yJ_zJ_zJ_zJ_yJ_zJ_yJ_z)$$

$$\text{Tr } J_+J_z^2J_-J_+J_-J_zJ_+J_- = -(2/315)AC(4J^4+8J^3+7J^2+3J+6)$$

$$\text{Tr } J_+J_zJ_-J_+J_zJ_-J_+J_zJ_- = -(2/105)A(4J^6+12J^5+19J^4+18J^3+16J^2+9J-8)$$

TABLE I. Continued
$A = J(J+1)(2J+1)$, $B = (2J-1)(2J+3)$, $C = (J-1)(J+2)$, $D = (2J-3)(2J+5)$ and $E = (J-2)(J+3)$.

$$\text{Tr } J_+J_zJ_-J_+J_zJ_-J_+J_-J_z = -(2/315)AC(4J^4 + 8J^3 + 7J^2 + 3J + 6)$$

$$\text{Tr } J_+J_zJ_-J_+J_-J_+J_zJ_-J_z = -(2/315)AC(4J^4 + 8J^3 + 7J^2 + 3J + 6)$$

$$\text{Tr } J_+^4J_zJ_-^4 = -(16/315)ABCD$$

$$= 64i \text{ Tr } (J_x^4J_yJ_xJ_yJ_zJ_y + J_x^2J_y^2J_xJ_zJ_x^2J_y)$$

$$\text{Tr } J_+^4J_-J_zJ_-^3 = -(8/315)ABCD$$

$$\text{Tr } J_+^3J_zJ_-J_+J_-^3 = -(4/315)ABC(10J^2 + 10J + 3)$$

$$= 18i \text{ Tr } (J_x^4J_y^3J_xJ_z + J_x^4J_y^2J_zJ_xJ_y) + 16i \text{ Tr } (J_x^3J_yJ_x^2J_yJ_zJ_y - J_xJ_yJ_xJ_zJ_yJ_zJ_yJ_z)$$

$$\text{Tr } J_+^3J_zJ_-^2J_+J_-^2 = -(4/105)ABC(4J^2 + 4J + 3)$$

$$= 48i \text{ Tr } (J_x^5J_y^2J_zJ_y + J_x^3J_y^3J_x^2J_z)$$

$$\text{Tr } J_+^3J_zJ_-^3J_+J_- = -(8/315)ABC(7J^2 + 7J - 6)$$

$$= -16i \text{ Tr } (J_x^6J_yJ_xJ_z + J_x^3J_y^3J_x^2J_z)$$

$$\text{Tr } J_+^3J_-J_+J_-J_zJ_-^2 = -(8/315)ABC(J^2 + J + 3)$$

$$\text{Tr } J_+^3J_-J_+J_-^2J_zJ_- = (4/105)ABC(2J^2 + 2J - 3)$$

$$= -16i \text{ Tr } (J_x^4J_y^2J_xJ_yJ_z + J_x^2J_yJ_xJ_yJ_xJ_zJ_xJ_y)$$

$$\text{Tr } J_+^3J_-J_zJ_+J_-^3 = -(8/315)ABC(J^2 + J + 3)$$

$$\text{Tr } J_+^3J_-J_zJ_-J_+J_-^2 = -(4/315)ABC(4J^2 + 4J + 3)$$

$$\text{Tr } J_+^3J_-^2J_+J_zJ_-^2 = -(4/315)ABC(4J^2 + 4J + 3)$$

$$\text{Tr } J_+^2J_zJ_+J_-^2J_+J_-^2 = -(4/315)ABC(4J^2 + 4J + 3)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_-J_+J_-^2 = -(8/105)AB(J^4 + 2J^3 + 2J^2 + J + 1)$$

$$= 16i \text{ Tr } (J_x^3J_y^3J_z^3 + J_x^3J_yJ_zJ_yJ_zJ_yJ_z)$$

$$\text{Tr } J_+^2J_zJ_-J_+J_-^2J_+J_- = -(8/315)AB(4J^4 + 8J^3 + 5J^2 + J + 3)$$

$$= 16i \text{ Tr } (J_x^5J_y^3J_z + J_x^3J_y^2J_zJ_yJ_xJ_z + 2J_xJ_yJ_xJ_zJ_yJ_xJ_zJ_yJ_z - 2J_x^3J_yJ_x^2J_yJ_zJ_y)$$

$$\text{Tr } J_+^2J_zJ_-^2J_+^2J_-^2 = -(8/315)AB(4J^4 + 8J^3 + 5J^2 + J + 3)$$

$$\text{Tr } J_+^2J_zJ_-^2J_+J_-J_+J_- = -(8/315)AB(5J^4 + 10J^3 + 7J^2 + 2J - 3)$$

$$= 16i \text{ Tr } (J_x^5J_y^3J_z + J_x^3J_y^2J_zJ_yJ_xJ_z)$$

$$\text{Tr } J_+^2J_-J_+J_zJ_-J_+J_-^2 = -(8/105)AB(J^4 + 2J^3 + 2J^2 + J + 1)$$

$$\text{Tr } J_+^2J_-J_+J_-J_+J_-J_zJ_- = (4/315)ABC(2J^2 + 2J + 3)$$

$$= -16i \text{ Tr } (J_x^5J_y^2J_zJ_y - J_x^3J_yJ_x^2J_yJ_zJ_y)$$

$$\text{Tr } J_+^2J_-J_+J_-J_zJ_+J_-^2 = (4/315)ABC(2J^2 + 2J + 3)$$

$$\text{Tr } J_+^2J_-J_+J_-^2J_+J_zJ_- = 0$$

$$\text{Tr } J_+J_zJ_-J_+J_-J_+J_-J_+J_- = -(8/315)A(J^2 + J + 1)(8J^4 + 16J^3 + 8J^2 + 3)$$

$$= 8i \text{ Tr } (2J_x^5J_y^3J_z - 7J_x^2J_yJ_zJ_xJ_yJ_zJ_yJ_z - 7J_x^2J_yJ_zJ_xJ_zJ_yJ_zJ_y)$$

TABLE II. Numerical values of A, AB, AC, ABC, $ABCD$, and $ABCE$ for $J = \frac{1}{2}$ to 10.

J	A	AB	AC	ABC	$ABCD$	$ABCE$
1/2	3/2	0	$-3\cdot5/2^3$	0	0	0
1	$2\cdot3$	$2\cdot3\cdot5$	0	0	0	0
3/2	$3\cdot5$	$2^2\cdot3^2\cdot5$	$3\cdot5\cdot7/2^2$	$3^2\cdot5\cdot7$	0	$-3^4\cdot5\cdot7/2^2$
2	$2\cdot3\cdot5$	$2\cdot3^2\cdot5\cdot7$	$2^3\cdot3\cdot5$	$2^3\cdot3^2\cdot5\cdot7$	$2^3\cdot3^4\cdot5\cdot7$	0
5/2	$3\cdot5\cdot7/2$	$2^4\cdot3\cdot5\cdot7$	$3^4\cdot5\cdot7/2^3$	$2^2\cdot3^4\cdot5\cdot7$	$2^4\cdot3^4\cdot5^2\cdot7$	$3^4\cdot5\cdot7\cdot11$
3	$2^2\cdot3\cdot7$	$2^2\cdot3^3\cdot5\cdot7$	$2^3\cdot3\cdot5\cdot7$	$2^3\cdot3^3\cdot5^2\cdot7$	$2^3\cdot3^4\cdot5^2\cdot7\cdot11$	$2^4\cdot3^4\cdot5^2\cdot7$
7/2	$2\cdot3^2\cdot7$	$2^3\cdot3^3\cdot5\cdot7$	$3^2\cdot5\cdot7\cdot11/2$	$2\cdot3^3\cdot5^2\cdot7\cdot11$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11$	$3^4\cdot5^2\cdot7\cdot11\cdot13/2$
4	$2^2\cdot3^2\cdot5$	$2^2\cdot3^2\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5$	$2^3\cdot3^4\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5\cdot7^2\cdot11$
9/2	$3^2\cdot5\cdot11/2$	$2^4\cdot3^3\cdot5\cdot11$	$3^2\cdot5\cdot7\cdot11\cdot13/2^3$	$2^2\cdot3^3\cdot5\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5\cdot7^2\cdot11\cdot13$	$3^4\cdot5^3\cdot7\cdot11\cdot13$
5	$2\cdot3\cdot5\cdot11$	$2\cdot3^3\cdot5\cdot11\cdot13$	$2^3\cdot3\cdot5\cdot7\cdot11$	$2^3\cdot3^3\cdot5\cdot7\cdot11\cdot13$	$2^3\cdot3^4\cdot5^2\cdot7^2\cdot11\cdot13$	$2^6\cdot3^4\cdot5\cdot7\cdot11\cdot13$
11/2	$3\cdot11\cdot13$	$2^2\cdot3\cdot5\cdot7\cdot11\cdot13$	$3^4\cdot5\cdot11\cdot13/2^2$	$3^4\cdot5^2\cdot7\cdot11\cdot13$	$2^7\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$	$3^4\cdot5^2\cdot7^2\cdot11\cdot13\cdot17/2^2$
6	$2\cdot3\cdot7\cdot13$	$2\cdot3^2\cdot5\cdot7\cdot11\cdot13$	$2^4\cdot3\cdot5\cdot7\cdot13$	$2^4\cdot3^2\cdot5^2\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5^2\cdot7\cdot11\cdot13\cdot17$	$2^6\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$
13/2	$3\cdot5\cdot7\cdot13/2$	$2^5\cdot3^2\cdot5\cdot7\cdot13$	$3\cdot5\cdot7\cdot11\cdot13\cdot17/2^3$	$2^3\cdot3^2\cdot5\cdot7\cdot11\cdot13\cdot17$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11\cdot13\cdot17$	$2\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19$
7	$2^3\cdot3\cdot5\cdot7$	$2^3\cdot3\cdot5\cdot7\cdot13\cdot17$	$2^4\cdot3^4\cdot5\cdot7$	$2^4\cdot3^4\cdot5\cdot7\cdot13\cdot17$	$2^4\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19$	$2^5\cdot3^4\cdot5^3\cdot7\cdot13\cdot17$
15/2	$2^2\cdot3\cdot5\cdot17$	$2^4\cdot3^3\cdot5\cdot7\cdot17$	$3\cdot5\cdot13\cdot17\cdot19$	$2^2\cdot3^3\cdot5\cdot7\cdot13\cdot17\cdot19$	$2^6\cdot3^4\cdot5^2\cdot7\cdot13\cdot17\cdot19$	$3^4\cdot5\cdot7^2\cdot11\cdot13\cdot17\cdot19$
8	$2^3\cdot3^2\cdot17$	$2^3\cdot3^3\cdot5\cdot7\cdot19$	$2^4\cdot3^2\cdot5\cdot7\cdot17$	$2^4\cdot3^3\cdot5^2\cdot7\cdot17\cdot19$	$2^4\cdot3^4\cdot5^2\cdot7^2\cdot13\cdot17\cdot19$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11\cdot17\cdot19$
17/2	$3^2\cdot17\cdot19/2$	$2^5\cdot3^2\cdot5\cdot17\cdot19$	$3^4\cdot5\cdot7\cdot17\cdot19/2^3$	$2^3\cdot3^4\cdot5^2\cdot7\cdot17\cdot19$	$2^5\cdot3^4\cdot5^2\cdot7^2\cdot11\cdot17\cdot19$	$2\cdot3^4\cdot5^2\cdot7\cdot13\cdot17\cdot19\cdot23$
9	$2\cdot3^2\cdot5\cdot19$	$2\cdot3^3\cdot5\cdot7\cdot17\cdot19$	$2^4\cdot3^2\cdot5\cdot11\cdot19$	$2^4\cdot3^3\cdot5\cdot7\cdot11\cdot17\cdot19$	$2^4\cdot3^4\cdot5^2\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^6\cdot3^4\cdot5\cdot7^2\cdot11\cdot17\cdot19$
19/2	$3\cdot5\cdot7\cdot19$	$2^2\cdot3^3\cdot5\cdot7\cdot11\cdot19$	$3\cdot5\cdot7\cdot17\cdot19\cdot23/2^2$	$3^3\cdot5\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^7\cdot3^4\cdot5\cdot7\cdot11\cdot17\cdot19\cdot23$	$3^4\cdot5^4\cdot7\cdot11\cdot17\cdot19\cdot23/2^2$
10	$2\cdot3\cdot5\cdot7\cdot11$	$2\cdot3\cdot5\cdot7\cdot11\cdot19\cdot23$	$2^3\cdot3^4\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5\cdot7\cdot11\cdot19\cdot23$	$2^3\cdot3^4\cdot5^3\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^6\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot19\cdot23$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

The Baker-Hausdorff Formula and a Problem in Crystal Physics*

G. H. WEISS
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
AND
A. A. MARADUDIN
Westinghouse Research Laboratories, Beulah Road, Churchill Borough, Pittsburgh, Pennsylvania
(Received February 15, 1962)

A derivation is given of the Baker-Hausdorff formula for $z = \ln e^x e^y$ where x and y are noncommuting operators. This result is then used to obtain an expression for z in the case that $X = P^2 + Q^2$ and $y = \alpha P + \beta Q$, where P and Q are two operators whose commutator $[P,Q] = c$ is a c number. With the aid of this result a very simple derivation of the expression for the intensity of x rays scattered by the thermal vibrations of a crystal is presented.

I. INTRODUCTION

ONE of the major differences between classical and quantum physics is the fundamental role assigned to noncommuting operators in quantum physics. The algebra of noncommuting operators has been studied by mathematicians in many contexts over the years and many results have been derived which are of potential value in quantum mechanical and statistical mechanical calculations. It is our purpose in the present paper to summarize some of the theory as it has so far been developed, and to derive by its use a result for particular operators which is then used to calculate the quantum mechanical intensity factor for the scattering of x rays by crystal lattices. The theory that will be presented is abstracted mainly from the excellent review paper by Magnus,[1] a following paper by Finklestein,[2] and the original and quite readable paper by Hausdorff.[3]

* This research was supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, Air Research and Development Command under Contract #AF 18(600)1315.

[1] W. Magnus, Communs. Pure and Appl. Math. 7, 649 (1954).
[2] D. Finklestein, Communs. Pure and Appl. Math. 8, 245 (1955).
[3] F. Hausdorff, Ber. Verhandl. sachs. Akad. Wiss. Leipzig, Math-naturwiss. 58, 19 (1906).

TABLE II. Numerical values of A, AB, AC, ABC, $ABCD$, and $ABCE$ for $J = \frac{1}{2}$ to 10.

J	A	AB	AC	ABC	$ABCD$	$ABCE$
1/2	3/2	0	$-3\cdot5/2^3$	0	0	0
1	$2\cdot3$	$2\cdot3\cdot5$	0	0	0	0
3/2	$3\cdot5$	$2^2\cdot3^2\cdot5$	$3\cdot5\cdot7/2^2$	$3^2\cdot5\cdot7$	0	$-3^4\cdot5\cdot7/2^2$
2	$2\cdot3\cdot5$	$2\cdot3^2\cdot5\cdot7$	$2^3\cdot3\cdot5$	$2^3\cdot3^2\cdot5\cdot7$	$2^3\cdot3^4\cdot5\cdot7$	0
5/2	$3\cdot5\cdot7/2$	$2^4\cdot3\cdot5\cdot7$	$3^4\cdot5\cdot7/2^3$	$2^2\cdot3^4\cdot5\cdot7$	$2^4\cdot3^4\cdot5^2\cdot7$	$3^4\cdot5\cdot7\cdot11$
3	$2^2\cdot3\cdot7$	$2^2\cdot3^3\cdot5\cdot7$	$2^3\cdot3\cdot5\cdot7$	$2^3\cdot3^3\cdot5^2\cdot7$	$2^3\cdot3^4\cdot5^2\cdot7\cdot11$	$2^4\cdot3^4\cdot5^2\cdot7$
7/2	$2\cdot3^2\cdot7$	$2^3\cdot3^3\cdot5\cdot7$	$3^2\cdot5\cdot7\cdot11/2$	$2\cdot3^3\cdot5^2\cdot7\cdot11$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11$	$3^4\cdot5^2\cdot7\cdot11\cdot13/2$
4	$2^2\cdot3^2\cdot5$	$2^2\cdot3^2\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5$	$2^3\cdot3^4\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5\cdot7^2\cdot11$
9/2	$3^2\cdot5\cdot11/2$	$2^4\cdot3^3\cdot5\cdot11$	$3^2\cdot5\cdot7\cdot11\cdot13/2^3$	$2^2\cdot3^3\cdot5\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5\cdot7^2\cdot11\cdot13$	$3^4\cdot5^3\cdot7\cdot11\cdot13$
5	$2\cdot3\cdot5\cdot11$	$2\cdot3^3\cdot5\cdot11\cdot13$	$2^3\cdot3\cdot5\cdot7\cdot11$	$2^3\cdot3^3\cdot5\cdot7\cdot11\cdot13$	$2^3\cdot3^4\cdot5^2\cdot7^2\cdot11\cdot13$	$2^6\cdot3^4\cdot5\cdot7\cdot11\cdot13$
11/2	$3\cdot11\cdot13$	$2^2\cdot3\cdot5\cdot7\cdot11\cdot13$	$3^4\cdot5\cdot11\cdot13/2^2$	$3^4\cdot5^2\cdot7\cdot11\cdot13$	$2^7\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$	$3^4\cdot5^2\cdot7^2\cdot11\cdot13\cdot17/2^2$
6	$2\cdot3\cdot7\cdot13$	$2\cdot3^2\cdot5\cdot7\cdot11\cdot13$	$2^4\cdot3\cdot5\cdot7\cdot13$	$2^4\cdot3^2\cdot5^2\cdot7\cdot11\cdot13$	$2^4\cdot3^4\cdot5^2\cdot7\cdot11\cdot13\cdot17$	$2^6\cdot3^4\cdot5^2\cdot7\cdot11\cdot13$
13/2	$3\cdot5\cdot7\cdot13/2$	$2^5\cdot3^2\cdot5\cdot7\cdot13$	$3\cdot5\cdot7\cdot11\cdot13\cdot17/2^3$	$2^3\cdot3^2\cdot5\cdot7\cdot11\cdot13\cdot17$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11\cdot13\cdot17$	$2\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19$
7	$2^3\cdot3\cdot5\cdot7$	$2^3\cdot3\cdot5\cdot7\cdot13\cdot17$	$2^4\cdot3^4\cdot5\cdot7$	$2^4\cdot3^4\cdot5\cdot7\cdot13\cdot17$	$2^4\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19$	$2^5\cdot3^4\cdot5^3\cdot7\cdot13\cdot17$
15/2	$2^2\cdot3\cdot5\cdot17$	$2^4\cdot3^3\cdot5\cdot7\cdot17$	$3\cdot5\cdot13\cdot17\cdot19$	$2^2\cdot3^3\cdot5\cdot7\cdot13\cdot17\cdot19$	$2^6\cdot3^4\cdot5^2\cdot7\cdot13\cdot17\cdot19$	$3^4\cdot5\cdot7^2\cdot11\cdot13\cdot17\cdot19$
8	$2^3\cdot3^2\cdot17$	$2^3\cdot3^3\cdot5\cdot7\cdot19$	$2^4\cdot3^2\cdot5\cdot7\cdot17$	$2^4\cdot3^3\cdot5^2\cdot7\cdot17\cdot19$	$2^4\cdot3^4\cdot5^2\cdot7^2\cdot13\cdot17\cdot19$	$2^5\cdot3^4\cdot5^2\cdot7\cdot11\cdot17\cdot19$
17/2	$3^2\cdot17\cdot19/2$	$2^5\cdot3^2\cdot5\cdot17\cdot19$	$3^4\cdot5\cdot7\cdot17\cdot19/2^3$	$2^3\cdot3^4\cdot5^2\cdot7\cdot17\cdot19$	$2^5\cdot3^4\cdot5^2\cdot7^2\cdot11\cdot17\cdot19$	$2\cdot3^4\cdot5^2\cdot7\cdot13\cdot17\cdot19\cdot23$
9	$2\cdot3^2\cdot5\cdot19$	$2\cdot3^3\cdot5\cdot7\cdot17\cdot19$	$2^4\cdot3^2\cdot5\cdot11\cdot19$	$2^4\cdot3^3\cdot5\cdot7\cdot11\cdot17\cdot19$	$2^4\cdot3^4\cdot5^2\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^6\cdot3^4\cdot5\cdot7^2\cdot11\cdot17\cdot19$
19/2	$3\cdot5\cdot7\cdot19$	$2^2\cdot3^3\cdot5\cdot7\cdot11\cdot19$	$3\cdot5\cdot7\cdot17\cdot19\cdot23/2^2$	$3^3\cdot5\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^7\cdot3^4\cdot5\cdot7\cdot11\cdot17\cdot19\cdot23$	$3^4\cdot5^4\cdot7\cdot11\cdot17\cdot19\cdot23/2^2$
10	$2\cdot3\cdot5\cdot7\cdot11$	$2\cdot3\cdot5\cdot7\cdot11\cdot19\cdot23$	$2^3\cdot3^4\cdot5\cdot7\cdot11$	$2^3\cdot3^4\cdot5\cdot7\cdot11\cdot19\cdot23$	$2^3\cdot3^4\cdot5^3\cdot7\cdot11\cdot17\cdot19\cdot23$	$2^6\cdot3^4\cdot5\cdot7\cdot11\cdot13\cdot19\cdot23$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

The Baker-Hausdorff Formula and a Problem in Crystal Physics*

G. H. WEISS
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
AND
A. A. MARADUDIN
Westinghouse Research Laboratories, Beulah Road, Churchill Borough, Pittsburgh, Pennsylvania
(Received February 15, 1962)

A derivation is given of the Baker-Hausdorff formula for $z = \ln e^x e^y$ where x and y are noncommuting operators. This result is then used to obtain an expression for z in the case that $X = P^2 + Q^2$ and $y = \alpha P + \beta Q$, where P and Q are two operators whose commutator $[P,Q] = c$ is a c number. With the aid of this result a very simple derivation of the expression for the intensity of x rays scattered by the thermal vibrations of a crystal is presented.

I. INTRODUCTION

ONE of the major differences between classical and quantum physics is the fundamental role assigned to noncommuting operators in quantum physics. The algebra of noncommuting operators has been studied by mathematicians in many contexts over the years and many results have been derived which are of potential value in quantum mechanical and statistical mechanical calculations. It is our purpose in the present paper to summarize some of the theory as it has so far been developed, and to derive by its use a result for particular operators which is then used to calculate the quantum mechanical intensity factor for the scattering of x rays by crystal lattices. The theory that will be presented is abstracted mainly from the excellent review paper by Magnus,[1] a following paper by Finklestein,[2] and the original and quite readable paper by Hausdorff.[3]

* This research was supported in part by the U. S. Air Force through the Air Force Office of Scientific Research, Air Research and Development Command under Contract #AF 18(600)1315.

[1] W. Magnus, Communs. Pure and Appl. Math. 7, 649 (1954).
[2] D. Finklestein, Communs. Pure and Appl. Math. 8, 245 (1955).
[3] F. Hausdorff, Ber. Verhandl. sachs. Akad. Wiss. Leipzig, Math-naturwiss. 58, 19 (1906).

A problem which is often met in practice is that of expressing the product of two exponential operators in an equivalent form. When the operators x and y commute with their commutator $[x, y] = xy - yx$ then the following result is well known

$$e^x e^y = e^{x+y+(1/2)[x,y]}. \tag{1.1}$$

The generalization of this formula to the case when x and y do not commute with their commutator is less familiar, but has been extensively discussed in the mathematical literature in reference to the theory of groups[4] and the theory of differential equations.[5] The resulting formula, an identity of the form

$$e^z e^y = e^z, \tag{1.2}$$

is known as the Baker-Hausdorff formula. The problem of finding an explicit expression for z was first attacked by Campbell[4] and was soon thereafter followed by the investigations of Baker[5] and of Hausdorff. The last named author found an expression for z in terms of repeated commutators of x and y. This derivation will be repeated below. More recently, Goldberg[6] has given an expression for the coefficient in the expansion of

$$C = \ln (e^A e^B) = \sum_{n_1} \cdots \sum_{n_k} \cdots C(n_1, n_2, \cdots) \times A^{n_1} B^{n_2} A^{n_3} \cdots . \tag{1.3}$$

However, the expression is quite unwieldy, and is less useful than an expression in terms of commutators. Knowing the $C(n_1, n_2, \cdots)$, one can find the coefficients of the expansion in terms of commutators by methods due to Dynkin[7] and to Specht.[8] In addition to these expansions, there is still another of the form

$$e^{x+y} = e^x e^y e^{z_1} e^{z_2} e^{z_3} \cdots , \tag{1.4}$$

where z_j is a polynomial of jth degree in x and y and can be computed by a linear recurrence method. This expansion is due to Zassenhaus, and a derivation of the recurrence procedure has been given by Magnus.[1]

We will sketch Hausdorff's calculation of $z = \ln (e^x e^y)$ in Sec. III since it is a systematic procedure easily adapted to specific calculations, and since it is found in a rather inaccessible journal. In Sec. IV we use this result to obtain

$$\ln [\exp (x^2 + y^2) \exp (\beta y)]$$

and in Sec. V we apply our result to a problem arising in the theory of the scattering of x rays by lattice vibrations.

II. DEFINITIONS AND BASIC RESULTS

We begin by introducing a definition and deriving several preliminary results which will be useful in Sec. III. Our starting point is the Taylor series definition

$$F(x + u) = F(x) + u(\partial/\partial x)F(x) + \tfrac{1}{2}[u(\partial/\partial x)]^2 F(x) + \cdots , \tag{2.1}$$

where u is a variable which doesn't necessarily commute with x and where the operator $u\ \partial/\partial x$ is defined in the following manner: Let $F(x)$ consist of a polynomial in x of the form $Ax^m Bx^n \cdots$. Then, if we operate with $u\ \partial/\partial x$ on this polynomial, it replaces the x's one at a time by u. Thus, we would have

$$u(\partial/\partial x)(Ax^2Bx) = AxuBx + AuxBx + Ax^2Bu, \tag{2.2}$$

where A and B are independent of x. We remark that if $F = F_0 + F_1 + \cdots$ is a sum of polynomials of the order 0, 1, 2, $\cdots$ in x, then

$$x(\partial/\partial x)F(x) = F_1 + 2F_2 + 3F_3 + \cdots . \tag{2.3}$$

The two-dimensional Taylor's series also follows from Eq. (2.1) by noting that $u\ \partial/\partial x$ and $v\ \partial/\partial y$ commute. We then find

$$F(x + u, y + v) = F(x_1 y) + \left(u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}\right)F(x_1 y) + \tfrac{1}{2}(u\ \partial/\partial x + v\ \partial/\partial y)^2 F(x_1 y) + \cdots . \tag{2.4}$$

In what follows, we will use the notation

$$\{u\} = u,\ \{u, x\} = [u, x],\ \{u, x^n\} = [\{u, x^{n-1}\}, x]. \tag{2.5}$$

This notation can be extended to

$$\{u, p(x)\} = \sum_{n=0}^{\infty} p_n \{u, x^n\}, \tag{2.6}$$

where

$$p(x) = \sum p_n x^n \tag{2.7}$$

with constant coefficients p_n.

It follows as a trivial consequence of Eq. (2.1) that

$$u(\partial/\partial x)e^x = u + (1/2!)(ux + xu) + (1/3!)(ux^2 + xux + x^2u). + \cdots . \tag{2.8}$$

[4] J. E. Campbell, Proc. London Math. Soc. **29**, 14 (1898).
[5] H. F. Baker, Proc. London Math. Soc. **34**, 347 (1902); **35**, 333 (1903); **2**, 293 (1904); **3**, 24 (1904).
[6] K. Goldberg, Duke Math. J. **23**, 13 (1956).
[7] E. B. Dynkin, Doklady Akad. Nauk S.S.S.R. **57**, 323 (1947).
[8] W. Specht, Math. Z. **51**, 367 (1948).

If we now set

$$u = [w, x] \tag{2.9}$$

in Eq. (2.8) and expand the resulting expressions we find

$$[w, x]\,(\partial/\partial x)e^x = [w, x] + (1/2!)[w, x^2] + (1/3!)[w, x^3] + \cdots = [w, e^x]. \tag{2.10}$$

Another formula which will be of some use is

$$e^{-x}we^x = w + \{w, x\} + (1/2!)\{w, x^2\} + \cdots, \tag{2.11}$$

which is derived by expansion of the terms involved, and a simple induction. From this it follows by multiplication from the left by e^x that

$$[w, e^x] = e^x(\{w, x\} + (1/2!)\{w, x^2\} + (1/3!)\{w, x^3\} + \cdots) \tag{2.12}$$

and also, from Eqs. (2.10) and (2.12), that

$$u(\partial/\partial x)e^x = [w, e^x] = e^x[u + (1/2!)\{u, x\} + (1/3!)\{u, x^2\} + \cdots], \tag{2.13}$$

which can be written as

$$u(\partial/\partial x)e^x = e^x\phi(u, x), \tag{2.14}$$

where

$$\phi(u, x) = u + (1/2!)\{u, x\} + (1/3!)\{u, x^2\} + \cdots \tag{2.15}$$

$$= \{u, (e^x - 1)/x\}. \tag{2.16}$$

In a similar fashion, one can show that

$$u(\partial/\partial x)e^x = \psi(u, x)e^x, \tag{2.17}$$

where

$$\psi(u, x) = \phi(u, -x). \tag{2.18}$$

We will also require the reversion of the series

$$P = \phi(u, x) = u + (1/2!)\{u, x\} + (1/3!)\{u, x^2\} + \cdots, \tag{2.19}$$

i.e., we wish to express u in terms of p and x. We can do this purely by scalar multiplication, addition, and subtraction. Starting from Eq. (2.19) we can successively write

$$\begin{aligned} \{p, x\} &= \{u, x\} + (1/2!)\{u, x^2\} + (1/3!)\{u, x^3\} + \cdots, \\ \{p, x^2\} &= \{u, x^2\} + (1/2!)\{u, x^3\} + \cdots, \\ \{p, x^3\} &= \{u, x^3\} + \cdots, \end{aligned} \tag{2.20}$$

so that we can successively eliminate the terms $\{u, x^2\}$, $\{u, x^3\}$, $\cdots$. Now, if we consider the purely scalar equation

$$p = u + (1/2!)ux + (1/3!)ux^2 + \cdots = u[(e^x - 1)/x], \tag{2.21}$$

we see that the same manipulations will serve to solve for u in terms of p and x. Hence, if we can solve Eq. (2.21) for x in terms of p and u, we can solve Eq. (2.20) for their operators' counterparts. But Eq. (2.21) can be written

$$u = \frac{px}{e^x - 1} = P\sum_{n=0}^{\infty}\frac{B_n}{n!}x^n, \tag{2.22}$$

where the B_n are the Bernouilli numbers

$$B_1 = -1/2, \quad B_2 = 1/6, \quad B_4 = -1/30, \quad B_6 = 1/42, \cdots, \quad B_{2n+1} = 0 \quad n > 0.$$

Thus, if

$$u = \psi(p, x) \tag{2.23}$$

is the solution to $p = \phi(u, x)$ then

$$\psi(p, x) = \sum_{n=0}^{\infty}\frac{B_n}{n!}\{p, x^n\} = \left\{P, \frac{x}{e^x - 1}\right\}. \tag{2.24}$$

Similarly, if

$$u = \omega(q, x) \tag{2.25}$$

is the solution to $q = \psi(u, x)$, then

$$\omega(q, x) = \psi(q, -x) = \{q, x/(1 - e^{-x})\}. \tag{2.26}$$

III. THE BAKER-HAUSDORFF FORMULA

With the foregoing results in hand we now turn to the calculation of $z = \ln(e^xe^y)$. If x is changed to $x + \alpha u$ and y is changed to $y - \alpha v$ where u and v are arbitrary in such a way that $z(x, y) = z(x + \alpha u, y - \alpha v)$ then a Taylor's series expansion yields the identity

$$u(\partial/\partial x)z = v(\partial/\partial y)z. \tag{3.1}$$

In the present case if

$$e^{x+\alpha u}e^{y-\alpha v} = e^x[1 + \alpha\phi(u, x) + \cdots] \times [1 - \alpha\psi(v, y) + \cdots]e^y = e^xe^y, \tag{3.2}$$

we must have

$$\phi(u, x) = \psi(v, y). \tag{3.3}$$

Since u is arbitrary we may choose it equal to x, from which we find

$$x(\partial/\partial x)z = v(\partial/\partial y)z \tag{3.4}$$

and

$$\phi(u, x) = x. \tag{3.5}$$

As a consequence we find, from Eq. (3.3),

$$\psi(v, y) = x \quad \text{and} \quad v = \omega(x, y). \tag{3.6}$$

We now expand z as a series

$$z = z_0 + z_1 + z_2 + \cdots , \tag{3.7}$$

where z_k is a polynomial containing k factors of x. Since v contains x only to the first power [see Eq. (2.27)] we can rewrite Eq. (3.4) as

$$z_1 + 2z_2 + 3z_3 + \cdots = v(\partial/\partial y) \times (z_0 + z_1 + z_2 + \cdots) \tag{3.8}$$

using Eq. (2.3). Hence we have the identities

$$z_1 = v(\partial/\partial y)z_0, \qquad 2z_2 = v(\partial/\partial y)z_1, \qquad 3z_3 = v(\partial/\partial y)z_2, \cdots , \tag{3.9}$$

where $z_0 = y$, and $z_1 = v = \omega(x, y)$. If we define

$$\begin{aligned} v &= \omega(x, y) \\ v(\partial/\partial y)\omega(x, y) &= \omega_1(x, y) \\ v(\partial/\partial y)\omega_1(x, y) &= \omega_2(x, y) \\ &\vdots \qquad\qquad \vdots \end{aligned} \tag{3.10}$$

then

$$z = y + \omega(x, y) + (1/2!)\omega_1(x, y) + (1/3!)\omega_2(x, y) + \cdots . \tag{3.11}$$

In the same way, using the substitution $v = y$, we find

$$y(\partial/\partial y)z = u(\partial/\partial x)z \quad \text{with} \quad u = \psi(y, x) \tag{3.12}$$

and the result analogous (and equivalent) to Eq. (3.11) is

$$z = x + \psi(y, x) + (1/2!)\psi_1(y, x) + (1/3!)\psi_2(y, x) + \cdots , \tag{3.13}$$

where

$$\begin{aligned} u &= \psi(y, x) \\ u(\partial/\partial x)\psi(y, x) &= \psi_1(y, x) \\ u(\partial/\partial x)\psi_1(y, x) &= \psi_2(y, x), \\ &\vdots \qquad\qquad \vdots \end{aligned} \tag{3.14}$$

The expansion of z to fifth order is

$$\begin{aligned} z = x + y &+ (1/2)\{x, y\} \\ &+ (1/12)\{x, y^2\} + (1/12)\{y, x^2\} \\ &+ (1/24)\{y, x^2, y\} - (1/720)\{x, y^4\} \\ &+ (1/360)\{x, y^3, x\} \\ &+ (1/360)\{y, x^3, y\} \\ &- (1/120)\{x, y^2, x, y\} \\ &- (1/120)\{y, x^2, y, x\} + \cdots , \end{aligned} \tag{3.15}$$

where, for example,

$$\{y, x^2, y\} = \{\{y, x^2\}, y\}. \tag{3.16}$$

An important extension of the Baker-Hausdorff formula results when one considers the solution of the matrix differential equation

$$d\mathbf{Y}(t)/dt = \mathbf{A}(t)\mathbf{y}(t) \tag{3.17}$$

with the initial condition $\mathbf{y}(0) = \mathbf{I}$. We follow closely the work of Magnus in this section. In analogy with with the one-dimensional case, one would like a solution in the form

$$\mathbf{Y}(t) = \exp \boldsymbol{\Omega}(t) \tag{3.18}$$

and the problem then becomes one of finding an expression for $\boldsymbol{\Omega}(t)$. It follows from Eq. (2.14) that

$$\frac{d\mathbf{y}}{dt} = \left(\frac{d\boldsymbol{\Omega}}{dt}\frac{\partial}{\partial\boldsymbol{\Omega}}\right)\exp \boldsymbol{\Omega} = \left\{\frac{d\boldsymbol{\Omega}}{dt}, \frac{1 - \exp(-\boldsymbol{\Omega})}{\boldsymbol{\Omega}}\right\}\exp \boldsymbol{\Omega}. \tag{3.19}$$

Noting Eq. (3.18), and substituting this result into the original differential equation, we find

$$\mathbf{A}(t) = \left\{\frac{d\boldsymbol{\Omega}}{dt}, \frac{1 - \exp(-\boldsymbol{\Omega})}{\boldsymbol{\Omega}}\right\}. \tag{3.20}$$

If we now use a reversion scheme similar to that used in the derivation of Eq. (2.22) we find

$$\frac{d\boldsymbol{\Omega}}{dt} = \left\{\mathbf{A}, \frac{\boldsymbol{\Omega}}{1 - \exp(-\boldsymbol{\Omega})}\right\} = \sum_{n=0}^{\infty} \beta_n\{\mathbf{A}, \boldsymbol{\Omega}^n\}, \tag{3.21}$$

where

$$\beta_3 = \beta_5 = \cdots = 0$$

and

$$\beta_{2n} = (-1)^{n-1}B_{2m}/(2m)!,$$

where the B_{2m} are the Bernouilli numbers. Finally, this equation is solved by iteration, by setting

$$\begin{aligned} \boldsymbol{\Omega}_0 &= 0, \\ \boldsymbol{\Omega}_1(t) &= \int_0^t \mathbf{A}(\tau)\, d\tau \\ \boldsymbol{\Omega}_n(t) &= \int_0^t \Big(\mathbf{A}(\tau) + \frac{1}{2}[\mathbf{A}_1\boldsymbol{\Omega}_{n-1}] \\ &\quad + \frac{1}{12}[[\mathbf{A}, \boldsymbol{\Omega}_{n-1}], \boldsymbol{\Omega}_{n-1}] + \cdots\Big)\, d\tau \end{aligned} \tag{3.22}$$

and putting $\boldsymbol{\Omega}(t) = \lim_{n\to\infty} \boldsymbol{\Omega}_n(t)$. This leads to a result

$$\begin{aligned} \boldsymbol{\Omega}(t) &= \int_0^t \mathbf{A}(\tau)\, d\tau + \frac{1}{2}\int_0^t \left[\mathbf{A}(\tau), \int_0^\tau \mathbf{A}(\sigma)\, d\sigma\right] d\tau \\ &+ \frac{1}{4}\int_0^t \left[\mathbf{A}(\tau), \int_0^\tau \left[\mathbf{A}(\sigma), \int_0^\sigma \mathbf{A}(\rho)\, d\rho\right] d\sigma\right] d\tau \\ &+ \frac{1}{12}\int_0^t \left[\left[\mathbf{A}(\tau), \int_0^\tau \mathbf{A}(\sigma)\, d\sigma\right], \int_0^\tau \mathbf{A}(\sigma)\, d\sigma\right] d\tau \\ &+ \cdots , \end{aligned} \tag{3.23}$$

which is the continuous analog to the Baker-Hausdorff formula.

IV. EVALUATION OF ln [exp ($P^2 + Q^2$) exp ($\alpha P + \beta Q$)]

In the present section we derive an expression for

$$z = \ln [\exp (P^2 + Q^2) \exp (\alpha P + \beta Q)], \tag{4.1}$$

where P and Q satisfy

$$[P, Q] = c. \tag{4.2}$$

With c a scalar constant in Eq. (4.1) we can put $\alpha = 0$ without loss of generality since, if it is not equal to zero, we can define new operators

$$\begin{aligned} p &= (\alpha^2 + \beta^2)^{-1/2}(\beta P - \alpha Q), \\ q &= (\alpha^2 + \beta^2)^{-1/2}(\alpha P + \beta Q), \end{aligned} \tag{4.3}$$

which have the property

$$\begin{aligned} P^2 + Q^2 &= p^2 + q^2, \\ \alpha P + \beta Q &= (\alpha^2 + \beta^2)^{1/2} q, \\ [P, Q] &= [p, q]. \end{aligned} \tag{4.4}$$

We will calculate the successive terms in Eq. (3.13). In preparation for what follows, we note the formulas

$$\begin{aligned} \{y, x^{2n}\} &= (-1)^n (2c)^{2n} \beta Q, \\ \{y, x^{2n+1}\} &= (-1)^{n+1} (2c)^{2n+1} \beta P. \end{aligned} \tag{4.5}$$

where we have made the identifications

$$x = P^2 + Q^2, \qquad y = \beta Q. \tag{4.6}$$

Equations (4.5) are easily proved by induction.

The first term to be calculated is

$$\begin{aligned} \psi(y, x) &= \sum_{n=0}^{\infty} \frac{B_n}{n!} \{y, x^n\}, \\ &= \beta c P - \beta c \cot cQ, \end{aligned} \tag{4.7}$$

where we have substituted the results of Eq. (4.5) into the first line, and used the identity

$$x \cot x = \sum_{n=0}^{\infty} (-1)^n B_{2n} \frac{(2x)^{2n}}{(2n)!}. \tag{4.8}$$

The second term in the expansion of Eq. (3.13) is

$$\psi_1(y, x) = \psi(\partial/\partial x)\psi(y, x). \tag{4.9}$$

We may calculate $\psi_1(y, x)$ by the following device. Consider the equation

$$\psi(y, x + \lambda\psi) = \{y, (x + \lambda\psi)/(e^{x+\lambda\psi} - 1)\}, \tag{4.10}$$

where λ is a small ordering parameter (small in the sense that the resulting series is convergent). According to the Taylor series expansion of this function we have

$$\begin{aligned} \psi(y, x + \lambda\psi) &= \psi(y, x) \\ &\quad + \lambda\psi(y, x)(\partial/\partial x)\psi(y, x) + \cdots . \end{aligned} \tag{4.11}$$

It will be observed that the term linear in λ is just $\psi_1(y, x)$. Hence, we need only calculate the term linear in λ in the expansion of $\psi(y, x + \lambda\psi)$,

$$\psi(y, x + \lambda\psi) = \sum_{n=0}^{\infty} \frac{B_n}{n!} \{y, (x + \lambda\psi)^n\}. \tag{4.12}$$

Let us set

$$\begin{aligned} T_n &= \{y, x^n\}, \\ S_n &= \{y, (x + \lambda\psi)^n\}. \end{aligned} \tag{4.13}$$

Then we have

$$\begin{aligned} S_n &= [S_{n-1}, P^2 + Q^2 + \lambda\beta c(P + Q \cot c)] \\ &= T_n + \lambda\beta c[T_{n-1}, P + Q \cot c], \end{aligned} \tag{4.14}$$

where we have made use of the fact that $[y, \psi]$ is a c number so that only the first power of λ remains in the expression for S_n, i.e., S_n is of the form $S_n = U_n + \lambda a_n$ where a_n is a scalar. We are only interested in the term linear in λ, i.e.,

$$U_n = \beta c[T_{n-1}, P + \cot cQ]. \tag{4.15}$$

For U_n we therefore find

$$\begin{aligned} U_{2n} &= (-1)^n \beta 2^{2n-1} c^{2n+1} \cot c, \\ U_{2n+1} &= (-1)^{n+1} \beta 2^{2n} c^{2n+2}. \end{aligned} \tag{4.16}$$

When the U_n are substituted into the expression

$$\psi_1 = \sum_{n=0}^{\infty} \frac{B_n}{n!} U_n \tag{4.17}$$

and the resulting summations carried out, we find

$$\psi_1(y, x) = \tfrac{1}{2}\beta^2 c^2 + \tfrac{1}{2}\beta^2 c \cot c(c \cot c - 1). \tag{4.18}$$

Since $\psi_1(y, x)$ is a scalar quantity rather than an operator, $\psi_2, \psi_3 \cdots$ all vanish and

$$z = x + \psi(y, x) + \tfrac{1}{2}\psi_1(y, x) \tag{4.19}$$

exactly.

We have therefore obtained the result that

$$\begin{aligned} z &= (P + \tfrac{1}{2}\beta c)^2 \\ &\quad + (Q + \tfrac{1}{2}\beta c \cot c)^2 - \tfrac{1}{4}\beta^2 c \cot c. \end{aligned} \tag{4.20}$$

For the case $\alpha \neq 0$ we can use the transformation of Eq. (4.3) to establish the result

$$\begin{aligned} z &= P^2 + Q^2 + (\beta + \alpha \cot c)cP \\ &\quad + (\beta \cot c - \alpha)cQ + \tfrac{1}{4}c^2(\alpha^2 + \beta^2) \\ &\quad + \tfrac{1}{4}(\alpha^2 + \beta^2)c \cot c(c \cot c - 1) \\ &= \{P + \tfrac{1}{2}c(\beta + \alpha \cot c)\}^2 \\ &\quad + \{Q + \tfrac{1}{2}c(\beta \cot c - \alpha)\}^2 \\ &\quad - \tfrac{1}{4}(\alpha^2 + \beta^2)c \cot c. \end{aligned} \tag{4.21}$$

V. X-RAY INTENSITY FORMULA

The intensity of x rays scattered by the thermal vibrations of a monatomic Bravais lattice is, in

electron units,[9]

$$I = |f_0|^2 \sum_{l,l'} \exp\{i\mathbf{k}\cdot[\mathbf{x}(l) - \mathbf{x}(l')]\} \times \exp\{i\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]\}. \quad (5.1)$$

In this expression f_0 is the atomic scattering factor, $\mathbf{x}(l)$ is the position vector of the equilibrium l'th atom in the crystal, $\mathbf{u}(l)$ is the displacement of the l'th atom from its equilibrium position, and the sums extend over all the atoms in the crystal. The vector $\mathbf{k}$, called the scattering vector, is given by

$$\mathbf{k} = (2\pi/\lambda)(\mathbf{s} - \mathbf{s}_0), \quad (5.2)$$

where $\mathbf{S}_0$ and $\mathbf{S}$ are unit vectors in the directions of the normals to the wavefronts to the incident and scattered waves, respectively, while λ is the wavelength of the x rays.

The atomic displacements $\mathbf{u}(l)$ are time-dependent so that Eq. (5.1) gives the instantaneous intensity. What is observed, however, is the average of the intensity over a time long compared with the period of the atomic vibrations. It is easier and more usual to replace the time average by an ensemble average over a canonical distribution given by

$$\rho = e^{-\beta H}/\mathrm{Tr}\,(e^{-\beta H}), \quad (5.3)$$

where H is the crystal Hamiltonian. The ensemble average of the expectation value of an operator O is given by

$$\langle O\rangle = \mathrm{Tr}\,(e^{-\beta H}O)/\mathrm{Tr}\,(e^{-\beta H}). \quad (5.4)$$

We are therefore required to evaluate the thermal average

$$\langle \exp i\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]\rangle.$$

The crystal Hamiltonian in the harmonic approximation is given by

$$H = \sum_{l,\alpha} \frac{p_\alpha^2(l)}{2M} + \tfrac{1}{2}\sum_{l,\alpha}\sum_{l',\beta} \Phi_{\alpha\beta}(l, l')u_\alpha(l)u_\beta(l'), \quad (5.5)$$

where $p_\alpha(l) = M\dot{u}_\alpha(l)$, M is the atomic mass, $\Phi_{\alpha\beta}(l, l')$ is a second derivative of the crystal's potential energy, and α, β label the Cartesian axes.

Equation (5.5) can be put into a simpler form if we make the normal coordinate transformation

$$u_\alpha(l) = (NM)^{-\frac{1}{2}} \sum_{\mathbf{k},j} e_\alpha(\mathbf{k}\mid j)\{q(\mathbf{k}\mid j)\cos 2\pi\mathbf{k}\cdot\mathbf{x}(l) - \dot{q}(\mathbf{k}\mid j)\sin 2\pi\mathbf{k}\cdot\mathbf{x}(l)\},$$

$$\dot{u}_\alpha(l) = (NM)^{-\frac{1}{2}} \sum_{\mathbf{k},j} e_\alpha(\mathbf{k}\mid j)\{\dot{q}(\mathbf{k}\mid j)\cos 2\pi\mathbf{k}\cdot\mathbf{x}(l) + q(\mathbf{k}\mid j)\sin 2\pi\mathbf{k}\cdot\mathbf{x}(l)\}. \quad (5.6)$$

Here, $\mathbf{e}(\mathbf{k}_j)$ is the polarization vector of the lattice wave described by the wave vector $\mathbf{k}$ and polarization index j. The allowed values of $\mathbf{k}$ are uniformly and densely distributed throughout a unit cell of the reciprocal lattice, and j takes the values 1, 2, 3. $\omega_j(\mathbf{k})$ is the frequency of the mode $\mathbf{k}_j$.

With the aid of the transformation (5.6) the expression for H becomes

$$H = \tfrac{1}{2}\sum\{\dot{q}^2(\mathbf{k}\mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k}\mid j)\} \quad (5.7)$$

and $\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]$ is transformed into

$$\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')] = (NM)^{-\frac{1}{2}} \sum_{\mathbf{k},j} (\mathbf{k}\cdot\mathbf{e}(\mathbf{k}\mid j) \times \{q(\mathbf{k}\mid j)[\cos 2\pi\mathbf{k}\cdot\mathbf{x}(l) - \cos 2\pi\mathbf{k}\cdot\mathbf{x}(l')] - [\dot{q}(\mathbf{k}\mid j)/\omega_j(\mathbf{k})][\sin 2\pi\mathbf{k}\cdot\mathbf{x}(l) - \sin 2\pi\mathbf{k}\cdot\mathbf{x}(l')]$$

$$= \sum_{\mathbf{k},j} \{\alpha(\mathbf{k}\mid j)\dot{q}(\mathbf{k}\mid j) + \gamma(\mathbf{k}\mid j)q(\mathbf{k}\mid j)\}. \quad (5.8)$$

This equation defines the coefficients $\alpha(\mathbf{k}|j)$ and $\gamma(\mathbf{k}|j)$. It should be kept in mind that they are functions of l and l'. The thermal average $\langle \exp i\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]\rangle$ becomes

$$\langle \exp i\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]\rangle = \frac{\mathrm{Tr}\,(\exp[-\frac{1}{2}\beta\sum_{\mathbf{k},j}\dot{q}^2(\mathbf{k}\mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k}\mid j)]\exp[i\sum_{\mathbf{k},j}\{\alpha(\mathbf{k}\mid j)\dot{q}(\mathbf{k}\mid j) + \gamma(\mathbf{k}\mid j)q(\mathbf{k}\mid j)\}])}{\mathrm{Tr}\,(\exp[-\frac{1}{2}\beta\sum_{\mathbf{k},j}\{\dot{q}^2(\mathbf{k}\mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k}\mid j)\}])}. \quad (5.9)$$

The variables $q(\mathbf{k}|j)$ and $\dot{q}(\mathbf{k}|j)$ must be regarded as quantum-mechanical operators and obey the commutation relations

$$[q(\mathbf{k}\mid j), \dot{q}(\mathbf{k}'\mid j')] = i\hbar\,\Delta(\mathbf{k} - \mathbf{k}')\,\delta_{jj'}, \quad (5.10a)$$

where $\Delta(\mathbf{k})$ is unity if $\mathbf{k}$ is zero or a translation vector of the reciprocal lattice and vanish otherwise. All other pairs of operators commute

$$[q(\mathbf{k}\mid j), q(\mathbf{k}'\mid j')] = [\dot{q}(\mathbf{k}\mid j), \dot{q}(\mathbf{k}'\mid j')] = 0. \quad (5.10b)$$

These results mean that we can factor the thermal average (5.9) into a product of thermal averages, one for each mode. Thus we have the relation

[9] R. W. James, *The Optical Principles of the Diffraction of X-Rays: The Crystalline State* (G. Bell and Sands, London, England, 1958).

$$\langle \exp i\mathbf{k}\cdot[\mathbf{u}(l) - \mathbf{u}(l')]\rangle$$

$$= \prod_{\mathbf{k},j} \frac{\operatorname{Tr} \exp\{-\frac{1}{2}\beta[\dot{q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k} \mid j)]\} \exp\{i[\alpha(\mathbf{k} \mid j)\dot{q}(\mathbf{k} \mid j) + \gamma(\mathbf{k} \mid j)q(\mathbf{k} \mid j)]\}}{\operatorname{Tr} \exp\{-\frac{1}{2}\beta[\dot{q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k} \mid j)]\}}. \tag{5.11}$$

The application of the results of the preceding section to this problem are obvious. If we make the following identifications

$$\begin{aligned} P &= i(\tfrac{1}{2}\beta)^{\frac{1}{2}}\dot{q}(\mathbf{k} \mid j), \qquad \alpha = (2/\beta)^{\frac{1}{2}}\alpha(\mathbf{k} \mid j), \\ Q &= i(\tfrac{1}{2}\beta)^{\frac{1}{2}}\omega_j(\mathbf{k})q(\mathbf{k} \mid j), \\ C &= i\frac{\beta\hbar\omega_j(\mathbf{k})}{2}, \qquad \beta = \left(\frac{2}{\beta}\right)^{\frac{1}{2}}\frac{\gamma(\mathbf{k} \mid j)}{\omega_j(\mathbf{k})}, \end{aligned} \tag{5.12}$$

we find that

$$\exp[-\tfrac{1}{2}\beta[\dot{q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k} \mid j)]] \exp[i[\alpha(\mathbf{k} \mid j)\dot{q}(\mathbf{k} \mid j) + \gamma(\mathbf{k} \mid j)q(\mathbf{k} \mid j)]]$$

$$= \exp\{-\tfrac{1}{2}\beta[\dot{Q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})Q^2(\mathbf{k} \mid j)] - \tfrac{1}{4}\hbar(\alpha^2(\mathbf{k} \mid j)\omega_j^2(\mathbf{k}) + \gamma^2(\mathbf{k} \mid j))\}\frac{\coth \frac{1}{2}\beta\hbar\omega_j(\mathbf{k})}{\omega_j(\mathbf{k})}, \tag{5.13}$$

where

$$\begin{aligned} \dot{Q}(\mathbf{k} \mid j) &= \dot{q}(\mathbf{k} \mid j) + \frac{\hbar\omega_j(\mathbf{k})}{2}\left(\frac{\gamma(\mathbf{k} \mid j)}{\omega_j(\mathbf{k})} - i\alpha(\mathbf{k} \mid j)\coth \tfrac{1}{2}\beta\hbar\omega_j(\mathbf{k})\right), \\ Q(\mathbf{k} \mid j) &= q(\mathbf{k} \mid j) - \frac{\hbar}{2}\left(\alpha(\mathbf{k} \mid j) + i\frac{\gamma(\mathbf{k} \mid j)}{\omega_j(\mathbf{k})}\coth \tfrac{1}{2}\beta\hbar\omega_j(\mathbf{k})\right). \end{aligned} \tag{5.14}$$

Thus, Eq. (5.11) becomes

$$\langle \exp i\mathbf{k}\cdot(\mathbf{u}(l) - \mathbf{u}(l'))\rangle = \prod_{\mathbf{k},j} \exp\{-\tfrac{1}{4}\hbar(\alpha^2(\mathbf{k} \mid j)\omega_j^2(\mathbf{k}) + \gamma^2(\mathbf{k} \mid j))\}\frac{\coth \frac{1}{2}\beta\hbar\omega_j(\mathbf{k})}{\omega_j(\mathbf{k})}$$

$$\times \frac{\operatorname{Tr} \exp\{-\frac{1}{2}\beta[\dot{Q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})Q^2(\mathbf{k} \mid j)\}}{\operatorname{Tr} \exp\{-\frac{1}{2}\beta[\dot{q}^2(\mathbf{k} \mid j) + \omega_j^2(\mathbf{k})q^2(\mathbf{k} \mid j)]\}}. \tag{5.15}$$

Since the transformation from the q variables to the Q variables is a canonical transformation, the two traces in Eq. (5.15) cancel each other and we are left with

$$\langle \exp i\mathbf{k}\cdot(\mathbf{u}(l) - \mathbf{u}(l'))\rangle = \exp\left\{-\frac{\hbar}{4}\sum_{\mathbf{k},j}\frac{(\alpha^2(\mathbf{k} \mid j)\omega_j^2(\mathbf{k}) + \gamma^2(\mathbf{k} \mid j))}{\omega_j(\mathbf{k})}\coth \tfrac{1}{2}\beta\hbar\omega_j(\mathbf{k})\right\}. \tag{5.16}$$

If we substitute into this result the explicit expressions for $\alpha(\mathbf{k}|j)$ and $\gamma(\mathbf{k}|j)$ from Eq. (5.8) we obtain finally that

$$\langle \exp i\mathbf{k}\cdot(\mathbf{u}(l) - \mathbf{u}(l'))\rangle = \exp\left\{-\frac{\hbar}{2NM}\sum_{\mathbf{k},j}\frac{(\mathbf{k}\cdot\mathbf{e}(\mathbf{k} \mid j))^2}{\omega_j(\mathbf{k})}\coth \tfrac{1}{2}\beta\hbar\omega_j(\mathbf{k}) \times (1 - \cos 2\pi\mathbf{k}\cdot(\mathbf{x}(l) - \mathbf{x}(l'))\right\}. \tag{5.17}$$

This result has been obtained by several authors[10-12] using different mathematical techniques.

[10] H. Ott, Ann. Physik **23**, 169 (1935).
[11] M. Born, Reports on Progress in Physics, **IX**, 294 (1943).
[12] A. A. Maradudin, G. H. Weiss, and E. W. Montroll, *Theory of Lattice Dynamics in the Harmonic Approximation* (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Exact Statistical Mechanics of a One-Dimensional System with Coulomb Forces. II. The Method of Functional Integration

S. F. EDWARDS
Department of Theoretical Physics, Manchester University, Manchester, England
AND
A. LENARD*
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received January 11, 1962)

The statistical mechanics of a one-dimensional system of charged sheets is studied in the formalism of the grand canonical ensemble. It is shown that the grand partition function may be expressed as a Wiener integral, i. e., as an average of a certain functional of Brownian motion paths. This functional integral is then expressed in terms of the fundamental solution of a partial differential equation of diffusion type. This depends on a theorem of Kac whose proof is also given. The generality of this method is discussed. When all charges are integral multiples of a common unit the problem is reduced to the determination of the largest characteristic value of an ordinary differential operator with periodic coefficients. An invariance property of the thermodynamic potential is shown to imply charge neutrality in the infinite system limit. A theorem is proven which, in certain cases, excludes the possibility of a thermodynamic phase transition. The method is generalized to yield exact expressions for the n-particle reduced density functions. Some properties of the two-particle functions are discussed.

1. INTRODUCTION

In two recent papers the thermodynamics of a one-dimensional "plasma" has been determined rigorously.[1,2] This system consists of a great number of uniformly charged plane sheets freely able to move and cross each other in the direction normal to the planes. It was shown in I that the thermodynamic functions are expressible in terms of a single function which may be defined by the Sturm-Liouville characteristic value problem of the Mathieu functions. An analysis of this result showed that the system behaves as a true plasma when its mean kinetic energy is much larger than its mean potential energy; in the opposite case it resembles a collection of infrequently colliding neutral molecules. The calculation was rendered possible by the observation that the partition function of the constant-pressure ensemble is a sum of simple algebraic expressions, and that the problem of summation over very many terms could be discussed systematically by using generating functions.

In the present paper we propose to discuss this problem with the aid of a more powerful method. It is a variant of a method used by one of us[3] for the calculation of the thermodynamic functions of a real (i.e., three-dimensional) plasma. Its essence is the expression of the grand partition function as the average of a certain functional over its argument function. An averaging over functions (or "functional integration"; we shall use the terms interchangeably) may be defined in a variety of ways. In probability theory a set of functions over which a method of averaging with non-negative weight is prescribed is termed a stochastic (or random) process. It turns out that for the one-dimensional plasma the appropriate random process is of a familiar and simple kind. It is the "Wiener process" which is the mathematical idealization of the Brownian motion of a particle.[4,5]

There are several motivations for this work. One of them is that in three dimensions the presence of the singularity of the interparticle force at small distances makes a theory based on a pure Coulomb force mathematically meaningless, and if this is remedied by introducing a repulsive hard core for the particles the resulting extra complication obscures the essentials of the functional integration method. There is no need for such a complication in the one-dimensional model since the interparticle potential remains regular with vanishing distance. Secondly, the present method offers greater insight

* Supported by the U. S. Atomic Energy Commission.
[1] A. Lenard, J. Math. Phys. 2, 682 (1961). This paper will be referred to as I.
[2] S. Prager, *Advances in Chemical Physics*, edited by I. Prigogine (Interscience Publishers, Inc., New York, 1961), Vol. IV.
[3] S. F. Edwards, Phil. Mag. 4, 1171 (1959). The author regrets that several calculational errors were inadvertently allowed to occur in this article.
[4] M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Inc., New York, 1959), Chap. IV.
[5] J. M. Gel'fand and A. M. Yaglom, J. Math. Phys. 1, 48 (1960), where other references are to be found.

into some mathematical features of the one-dimensional model. In particular, the characteristic value problem of the Mathieu functions enters here in a quite natural way. This stands in contrast to the treatment in I where it appears as a trick whose origin remains somewhat of a mystery. The reduction to the characteristic value problem is accomplished by the use of a theorem due to Kac[6] which allows the expression of the average of a functional of a certain type in terms of the fundamental solution of a parabolic partial differential equation. This reduction has no analog in the three-dimensional case for reasons which will be pointed out below.

With the present method we also obtain some results that go far beyond those contained in I. For one, we are not restricted to the case of a two-component gas containing particles of equal and opposite charge, although this remains the simplest case in which concrete results can be obtained. More importantly, we obtain rigorous expressions for the reduced density functions involving an arbitrary number of particles, and, in particular, a fast converging expansion for the simplest of these, the two-particle density function.

Our plan is as follows. We first set up the machinery of statistical mechanics in the grand canonical ensemble formalism and explain the reason for the appropriate choice of the potential energy function (Sec. 2). Next, we give a brief resume of the Wiener process and reproduce a proof of a theorem due to Kac which is of decisive importance (Sec. 3). We then show how the grand partition function may be regarded as a certain functional integral (or average) over the Wiener process, and reduce its determination to an initial value problem for a partial differential equation of the diffusion type. We discuss the role that is played in this result by two properties of the Wiener process, its Markoffian and its Gaussian nature (Sec. 4). If all particles possess charges which are integral multiples of a common unit, the problem may be further reduced to a characteristic value problem of an ordinary differential operator with periodic coefficients. Two general questions are discussed, the question of charge neutrality and the question of the continuity of the thermodynamic quantities (Sec. 5). The dominant behavior in the "plasma limit" is derived by a simple heuristic approximation scheme on the appropriate partial differential equation (Sec. 6). Finally, we explain how the reduced density functions are determined and discuss, in particular, the properties of the two-particle function (Sec. 7).

[6] M. Kac, *Proceedings of the Second Berkeley Symposium on Probability and Statistics* (University of California Press, Berkeley, California, 1951), p. 189. See also reference 4.

2. THE GRAND PARTITION FUNCTION

The most convenient method from the present point of view is that based on the grand canonical ensemble. We suppose our system to consist of a great number of parallel charged sheets located in the region

$$0 \leq x \leq L. \tag{1}$$

These sheets (which in the following will be referred to as particles) carry surface charge densities which fall into a finite number of groups. We denote by N' the number of particles carrying charge σ', N'' carrying charge σ'', etc., and let $N = N' + N'' + \cdots$ stand for the total number. For the moment we make no assumptions regarding the charges σ', σ'', $\cdots$ which may be different or equal, positive, negative, or even zero, and need not be integral multiples of a common unit. A more restrictive hypothesis will be imposed later.

In the grand canonical ensemble the probability density that N arbitrary elements dx_1, dx_2, $\cdots$, dx_N within the region (1) are occupied by particles of charges σ_1, σ_2, $\cdots$, σ_N, respectively, and that no other particles are in the region is

$$\frac{1}{\Omega} z'^{N'} z''^{N''} \cdots \times \exp\left\{-\frac{V_{N'N''\cdots}(x_1, x_2, \cdots, x_N)}{\theta}\right\}. \tag{2}$$

Here θ is the temperature; z', z'', $\cdots$ are positive parameters of the dimension of a number density; $V_{N'N''\cdots}$ is the total potential energy as a function of the coordinates; and $\Omega = \Omega(L, \theta, z', z'', \cdots)$ is the grand partition function. The latter is defined by

$$\Omega = \sum_{N'=0}^{\infty} \frac{z'^{N'}}{N'!} \sum_{N''=0}^{\infty} \frac{z''^{N''}}{N''!} \cdots \times \int_0^L dx_N \cdots \int_0^L dx_1 \exp\left\{-\frac{V_{N'N''\cdots}}{\theta}\right\}. \tag{3}$$

As is well known, the asymptotic behavior of Ω in the infinite system limit $L \to \infty$ specifies the thermodynamic properties of the system. In particular, the pressure is

$$P = \theta \lim_{L\to\infty} \frac{\ln \Omega}{L}. \tag{4}$$

Once $P = P(\theta, z', z'', \cdots)$ is determined, the densi-

ties of the several components may be derived by differentiation

$$n' = (z'/\theta)\,\partial P/\partial z',$$
$$n'' = (z''/\theta)\,\partial P/\partial z'', \tag{5}$$
$$\text{etc.}$$

The internal energy per unit volume is

$$u = \tfrac{1}{2}z'\,\partial P/\partial z' + \tfrac{1}{2}z''\,\partial P/\partial z'' + \cdots + \theta\,\partial P/\partial\theta - P. \tag{6}$$

We now turn to the consideration of the potential energy. If two particles, placed at x_1 and x_2, have charges σ_1 and σ_2, respectively, the potential energy of their mutual interaction is $-2\pi\sigma_1\sigma_2\,|x_1 - x_2|$. It would be incorrect, however, to take for the total potential energy merely the sum

$$-2\pi \sum_{1\le k<l\le N}\sum \sigma_k\sigma_l\,|x_k - x_l|. \tag{7}$$

We must imagine that the system is in contact with an infinite reservoir (in the region $x \le 0$, say) which exchanges particles with it giving rise to the statistical fluctuations in particle numbers. Take the system *plus reservior* electrically neutral as a whole and consider the state of affairs when some particular number of particles, N' of charge σ', N'' of charge σ'', etc., are in the system. Then there is a total charge

$$\sigma_{\text{res}} = -(N'\sigma' + N''\sigma'' + \cdots) \tag{8}$$

in the reservior. This creates a constant electric field of magnitude $2\pi\sigma_{\text{res}}$ in the region $x \ge 0$. The particles of the system are subject to this electric field in addition to the forces they exert on each other. Thus, there is an additional contribution

$$-2\pi\sigma_{\text{res}} \sum_{1\le k\le N} \sigma_k x_k \tag{9}$$

to the potential energy. *We define* $V_{N'N'\ldots}$ *to be the sum of the functions* (7) *and* (9). The contribution (9) comes physically from the interaction of the particles inside the system with those inside the reservoir. That it must be properly attributed to the system as an additional potential energy and cannot be neglected as a "boundary effect" is due, of course, to the fact that the interparticle force does not tend to zero with increasing distance.

The expression for the potential energy may be conveniently simplified by the device made already use of in I. The grand partition function may be decomposed into a sum of integrals in each of which the integration variables are restricted by the inequalities

$$0 \le x_1 \le x_2 \le \cdots \le x_N \le L \tag{10}$$

and each being associated with a definite succession of charges (a "configuration")

$$C\colon \sigma_1, \sigma_2, \cdots, \sigma_N. \tag{11}$$

In addition to the sums over the numbers $N', N'', \cdots$ we now have also a sum over all sequences C restricted only by the requirement that N' of the σ_k have value σ', N'' the value σ'', etc. Let us consider now a definite configuration and introduce the quantities

$$E_0 = 4\pi\sigma_{\text{res}},$$
$$E_k = 4\pi\left(\sigma_{\text{res}} + \sum_{l=1}^{k}\sigma_l\right), \tag{12}$$
$$(k = 1, 2, \cdots, N).$$

E_k is just the electric field in the space between the kth and $(k+1)$st particles. A simple calculation shows that the potential energy, as we have defined it, is equal to

$$V_{N',N'',\cdots} = \sum_{k=1}^{N} \frac{E_{k-1}^2}{8\pi}(x_k - x_{k-1}) \tag{13}$$

($x_0 = 0$, by definition), i.e., *the total electrostatic field energy within the boundaries of the system.*

Making use of this we have the grand partition function in the form

$$\Omega = \sum_{N'=0}^{\infty} z'^{N'} \sum_{N''=0}^{\infty} z''^{N''} \cdots \sum_{C} \int_0^L dx_N \cdots \times \int_0^{x_2} dx_1 \exp\left\{-\frac{1}{8\pi\theta}\sum_{k=1}^{N} E_{k-1}^2(x_k - x_{k-1})\right\}. \tag{14}$$

The symbol $\sum_C$ denotes the summation over the

$$\frac{N!}{N'!\,N''!\cdots}$$

configurations.

3. THE WIENER PROCESS AND THE THEOREM OF KAC

We shall now collect the mathematical apparatus needed for exhibiting the grand partition function as a functional integral. This material is not new and it is included here only in order to enable a reader not thoroughly familiar with it to follow the argument without interruption. If more detail and mathematical rigor than can be given here is desired, reference may be made to the review of Gel'fand and Yaglom,[5] or to the book of Kac.[4]

We consider the mathematical description of the Brownian motion of a particle. Let $\phi(x)$ stand for one of its coordinates as a function of time.[7] This function is specified only in a statistical sense, namely, by the following two postulates:

(a) The conditional probability that $\phi(x_1)$ is in an element $d\phi_1$ around ϕ_1 given that $\phi(x_0) = \phi_0$ is $P(\phi_1 - \phi_0, x_1 - x_0)\, d\phi_1$ for any $x_1 > x_0$. The function P is

$$P(\phi, x) = (4\pi Dx)^{-\frac{1}{2}} e^{-(\phi^2/4Dx)}, \tag{15}$$

D being a positive constant.

(b) Displacements in nonoverlapping time intervals are statistically independent quantities.

The two postulates are compatible on account of the identity

$$P(\phi_2 - \phi_0, x_2 - x_0) = \int_{-\infty}^{\infty} d\phi_1 P(\phi_2 - \phi_1, x_2 - x_1) \times P(\phi_1 - \phi_0, x_1 - x_0), \quad (x_0 < x_1 < x_2), \tag{16}$$

whose meaning is that the displacement $\phi(x_2) - \phi(x_0)$ can be regarded as the sum of two statistically independent displacements $\phi(x_1) - \phi(x_0)$ and $\phi(x_2) - \phi(x_1)$. If $x_0, x_1, \cdots x_n$ in any increasing sequence of time values, the probability that $\phi(x_k)$ is in $d\phi_k (k = 1, 2, \cdots, n)$ given that $\phi(x_0) = x_0$ is the product

$$\prod_{k=1}^{n} P(\phi_k - \phi_{k-1}, x_k - x_{k-1})\, d\phi_k. \tag{17}$$

Any quantity which is an expression involving $\phi(x)$ is a random variable whose average value may be calculated by the above assignment of probabilities. It was Wiener[8] who first recognized that this probabilistic model of Brownian motion amounts to assigning a measure in the space of continuous functions $\phi(x)$, the "possible paths of the Brownian particle," and that statistical averaging corresponds to a theory of integration over this space.[4,5]

We shall be interested exclusively in averages of the form

$$\left\langle \exp \int_{x_0}^{x} dx' F(\phi(x'), x') \mid \phi(x_0) = \phi_0 \right\rangle, \tag{18}$$

where $F(\phi, x)$ is some given function. We shall always use this bracket notation to indicate averaging over $\phi(x)$ by the method just explained. The condition following the bar is to be satisfied by all Brownian paths over which the averaging is carried out. When misunderstanding is not likely we shall omit writing out this condition explicitly.

A method of calculating (18) will now be given.

Expand the exponential in a power series, interchange the averaging with the time integration, and make use of the symmetry of the integrand in the integration variables. The following series expansion results:

$$1 + \sum_{n=1}^{\infty} \int_{x_0}^{x} dx_n \int_{x_0}^{x_n} dx_{n-1} \cdots \int_{x_0}^{x_2} dx_1 \times \langle F(\phi_n, x_n) F(\phi_{n-1}, x_{n-1}) \cdots F(\phi_1, x_1) \rangle, \tag{19}$$

where we have set $\phi(x_k) = \phi_k$ as an abbreviation. The successive terms of this series will now be related to each other by an integral-recursion formula. We write the first term

$$1 = \int_{-\infty}^{\infty} d\phi Q_0(\phi, x; \phi_0, x_0) \tag{20}$$

with

$$Q_0 = P(\phi - \phi_0, x - x_0). \tag{21}$$

The next term ($n = 1$) may be written as

$$\int_{x_0}^{x} dx_1 \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\phi_1 P(\phi - \phi_1, x - x_1) \times F(\phi_1, x_1) P(\phi_1 - \phi_0, x_1 - x_0) = \int_{-\infty}^{\infty} d\phi Q_1(\phi, x; \phi_0, x_0). \tag{22}$$

The first factor P may appear superfluous inasmuch as the integration over ϕ can be trivially carried out. However, it is convenient to include it because this puts in evidence that

$$Q_1 = \int_{x_0}^{x} dx_1 \int_{-\infty}^{\infty} d\phi_1 P(\phi - \phi_1, x - x_1) \times F(\phi_1, x_1) Q_0(\phi_1, x_1; \phi_0, x_0). \tag{23}$$

The next term ($n = 2$) is

$$\int_{x_0}^{x} dx_2 \int_{x_0}^{x_2} dx_1 \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\phi_2 \int_{-\infty}^{\infty} d\phi_1 \times P(\phi - \phi_2, x - x_2) F(\phi_2, x_2) P(\phi_2 - \phi_1, x_2 - x_1) \times F(\phi_1, x_1) P(\phi_1 - \phi_0, x_1 - x_0) = \int_{-\infty}^{\infty} d\phi Q_2(\phi, x; \phi_0, x_0), \tag{24}$$

where

$$Q_2 = \int_{x_0}^{x} dx_2 \int_{-\infty}^{\infty} d\phi_2 P(\phi - \phi_2, x - x_2) \times F(\phi_2, x_2) Q_1(\phi_2, x_2; \phi_0, x_0). \tag{25}$$

[7] The reason for this unconventional notation will become clear below.

[8] N. Wiener, J. Math. and Phys. 2, 131 (1923).

The structure of this formula is the same as (23), and this can be continued indefinitely. The general term of the sum (19) is

$$\int_{-\infty}^{\infty} d\phi Q_n(\phi, x; \phi_0, x_0), \tag{26}$$

and the successive integrands are related to each other by

$$Q_n(\phi, x; \phi_0, x_0) = \int_{x_0}^{x} dx' \int_{-\infty}^{\infty} d\phi'$$

$$\times P(\phi - \phi', x - x')F(\phi', x')Q_{n-1}(\phi', x'; \phi_0, x_0). \tag{27}$$

Let

$$Q = \sum_{n=0}^{\infty} Q_n. \tag{28}$$

Then we have shown that *the average value* (18) *can* be expressed as

$$\int_{-\infty}^{\infty} d\phi Q(\phi, x; \phi_0, x_0), \tag{29}$$

the function Q being the solution of the integral equation

$$Q(\phi, x; \phi_0, x_0) = P(\phi - \phi_0, x - x_0)$$

$$+ \int_{x_0}^{x} dx' \int_{-\infty}^{\infty} d\phi' P(\phi - \phi', x - x')$$

$$\times F(\phi', x')Q(\phi', x'; \phi_0, x_0). \tag{30}$$

This theorem and this derivation of it are due to Kac.[9,4] Other derivations have been given by Rosenblatt,[10] by Blanc-Lapierre and Fortet,[11] and by Darling and Siegert.[12]

The integral equation (30), in turn, is equivalent to a partial differential equation with an appropriate boundary condition. The function P, given explicitly in Eq. (15), satisfies the equation

$$\{\partial/\partial x - D\,\partial^2/\partial\phi^2\}P(\phi, x) = 0 \tag{31}$$

and the initial condition

$$P(\phi, 0) = \delta(\phi). \tag{32}$$

Therefore, applying the operator $\{\partial/\partial x - D\partial^2/\partial\phi^2\}$ on the right-hand side of (30) only one term survives; thus,

$$\{\partial/\partial x - D\,\partial^2/\partial\phi^2 - F(\phi, x)\}Q(\phi, x; \phi_0, x_0) = 0. \tag{33}$$

Equation (30) also shows that

$$Q(\phi, x_0; \phi_0, x_0) = \delta(\phi - \phi_0). \tag{34}$$

This means that Q is the fundamental solution of the partial differential equation (33).

The theorem of Kac can be interpreted physically in a variety of ways, but our concern with it is due only to the interesting fact that the grand partition function of the system under investigation can be identified as a functional average of the precise form (18). The next section is devoted to the demonstration of this fact and a discussion of its significance.

4. THE GRAND PARTITION FUNCTION AS A FUNCTIONAL AVERAGE

We intend to show that

$$\Omega = \left\langle \exp \int_0^L dx' F(\phi(x')) \mid \phi(0) = 0 \right\rangle, \tag{35}$$

where the function F is

$$F(\phi) = z'e^{i\sigma'\phi} + z''e^{i\sigma''\phi} + \cdots, \tag{36}$$

and the diffusion constant of the Wiener process [cf. Eq. (15)] is

$$D = 2\pi/\theta. \tag{37}$$

To show this we begin by expanding the exponential inside the bracket into a multiple power series of z', z'', etc., and then we interchange averaging with integrating over x. The series

$$\sum_{N'=0}^{\infty} \frac{z'^{N'}}{N'!} \sum_{N''=0}^{\infty} \frac{z''^{N''}}{N''!} \cdots \int_0^L dx_1 \cdots \int_0^L dx_N$$

$$\times \left\langle \exp\left\{ i \sum_{k=1}^{N} \sigma_k \phi_k \right\} \right\rangle \tag{38}$$

results, where again $\phi_k = \phi(x_k)$ is used as an abbreviation. The quantities $\sigma_1, \sigma_2 \cdots, \sigma_N$ are defined as follows: The first N' of them are equal to σ', the next N'' to σ'', etc., and $N = N' + N'' + \cdots$. The integration is over a domain symmetric in the N variables, so that the integrand, although not symmetric, may be replaced by its symmetric part

$$\frac{N'!\,N''!\cdots}{N!} \sum_{C} \left\langle \exp\left\{ i \sum_{k=1}^{N} \sigma_k \phi_k \right\} \right\rangle. \tag{39}$$

Here C is a symbol for an arbitrary succession of quantities $\sigma_1, \sigma_2 \cdots, \sigma_N$ with the property that *some* N' of them equal σ', *some* N'' equal σ'', etc., in other words it has the same significance as in Sec. 2. The summation goes over all such sequences, whose total number is just the inverse of the factor in front. Once (39) is substituted as the integrand of (38) it is permissible to integrate over the restricted domain (10) at the price of multiplying by

[9] Reference 6, Sec. 3.
[10] M. Rosenblatt, Trans. Am. Math. Soc. 77, 120 (1951).
[11] A. Blanc-Lapierre and R. Fortet, *Theorie des functions aletoires* (Masson & Cie, Paris, 1953), Chap. VII.
[12] D. A. Darling and A. J. F. Siegert, Proc. Natl. Acad. Sci. U. S. 42, 525 (1956).

a factor $N!$. Thus, (35) becomes

$$\sum_{N'=0}^{\infty} z'^{N'} \sum_{N''=0}^{\infty} z''^{N''} \cdots \sum_{C} \int_0^L dx_N \cdots \times \int_0^{x_2} dx_1 \left\langle \exp\left\{ i \sum_{k=1}^{N} \sigma_k \phi_k \right\} \right\rangle. \quad (40)$$

This is identical to Ω, in the form given by Eq. (14), if

$$\left\langle \exp\left\{ i \sum_{k=1}^{N} \sigma_k \phi_k \right\} \right\rangle = \exp\left\{ -\frac{1}{8\pi\theta} \sum_{k=1}^{N} E_{k-1}^2 (x_k - x_{k-1}) \right\}. \quad (41)$$

This identity is established as follows. We recall the definition of the quantities E_k given by Eqs. (12) and (8) which can also be written

$$(1/4\pi)(E_k - E_{k-1}) = \sigma_k \quad (k = 1, 2, \cdots, N), \qquad E_N = 0. \quad (42)$$

Now multiply (42) by ϕ_k and sum over k from 1 to N. In view of $E_N = 0$ and $\phi_0 = 0$ this sum is

$$-\frac{1}{4\pi} \sum_{k=1}^{N} E_{k-1}(\phi_k - \phi_{k-1}) = \sum_{k=1}^{N} \sigma_k \phi_k. \quad (43)$$

Thus, the quantity to be averaged on the left-hand side of (41) is a product of N *statistically independent* quantities

$$\prod_{k=1}^{N} \exp\left\{ -\frac{i}{4\pi} E_{k-1}(\phi_k - \phi_{k-1}) \right\},$$

so that the product sign may be taken outside the bracket

$$\prod_{k=1}^{N} \left\langle \exp\left\{ -\frac{i}{4\pi} E_{k-1}(\phi_k - \phi_{k-1}) \right\} \right\rangle = \prod_{k=1}^{N} \int_{-\infty}^{\infty} d\phi P(\phi, x_k - x_{k-1}) \exp\left\{ -\frac{i}{4\pi} E_{k-1}\phi \right\} = \prod_{k=1}^{N} \exp\left\{ -\frac{D E_{k-1}^2 (x_k - x_{k-1})}{16\pi^2} \right\}. \quad (44)$$

This is identical to the right-hand side of (41) provided D is set equal to $2\pi/\theta$, as claimed.

The expression for Ω thus obtained is of the general form discussed in the previous section, therefore it may be expressed in yet another form, namely as an integral over the fundamental solution of a certain partial differential equation. Application of the theorem of Kac to Eq (35) establishes the following

Theorem 1. *The grand partition function is expressible as*

$$\Omega = \int_{-\infty}^{\infty} d\phi Q(\phi, L; 0, 0), \quad (45)$$

where $Q = Q(\phi, x; \phi_0, x_0)$ *is the fundamental solution of the partial differential equation*

$$\{\partial/\partial x - D\,\partial^2/\partial\phi^2 - F(\phi)\}Q = 0, \quad (46)$$

that is to say the solution satisfying the initial condition

$$Q(\phi, x_0; \phi_0, x_0) = \delta(\phi - \phi_0). \quad (47)$$

$F(\phi)$ *is given by* (36) *and the constant* D *by* (37). The function Q depends only on the difference $x - x_0$ because F does not depend explicitly on x. Thus, we shall more appropriately write $Q = Q(\phi, \phi_0, x)$ and for abbreviation often omit the argument ϕ_0 when it is understood to have the value zero.

Theorem 1 is the central result of our paper. Let us make some comments on its significance.

It is clear, first of all, that our result has its origin in two distinct calculations, the proof of the theorem of Kac and the representation of the grand partition function as a functional average. They represent two ways of "evaluating" the functional average and it is in the connection between the two end products that our interest lies. The details of the calculations are of a quite familiar kind, and it is apparent that the intermediate notion of functional average could be eliminated altogether. The really essential point is that in the iteration solution (Neumann series) of the integral equation (30) each term can be explicitly evaluated by a repeated application of the elementary formula

$$\int_{-\infty}^{\infty} d\phi \frac{1}{(4a\pi)^{\frac{1}{2}}} \exp\left\{ -\frac{\phi^2}{4a} + ib\phi \right\} = \exp\{-ab^2\}, \quad (48)$$

$$(a > 0)$$

and the terms turn out to be identical to the terms in the power series expansion (14) of the grand partition function.

Nevertheless it would be wrong to conceal the importance of the functional average concept with its probabilistic interpretation, because it sheds light on the question: To what extent is the success of our analysis due to the particular features of our model? The two aspects of the calculation are characterized by different assumptions. In the proof of the theorem of Kac—at least in its integral equation formulation—no use is made of the special Gaussian form (15) of the transition probability, but only that the joint probability for succeeding steps is the product (17). This is the assumption that the random process $\phi(x)$ is Markoffian. On the other hand, the explicit identification of the grand

partition function with the functional average is based on the Gaussian property, and on that alone.

In order to bring out this point we shall make a digression from the principal topic of this paper and consider the averaging over general Gaussian random processes in some detail. A general Gaussian random process $\phi(x)$ is defined by the postulate that for any finite number of points $x_1, x_2, \cdots, x_N$ the joint probability density for $\phi(x_k)$ in $d\phi_k$ is of the form

$$\frac{(\det B)^{\frac{1}{2}}}{(2\pi)^{N/2}} \exp\left\{-\tfrac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} B_{kl}\phi_k\phi_l\right\}, \qquad (49)$$

where the elements of the matrix B are functions of the x_k alone. The matrix B must be positive definite because otherwise the normalization integral would not converge, but this is not the only condition to which it is subject. Let $\alpha_k (k = 1, 2, \cdots, N)$ be arbitrary real numbers. Then

$$\left\langle \exp\left\{i\sum_{k=1}^{N}\alpha_k\phi(x_k)\right\}\right\rangle = \exp\left\{-\tfrac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} C_{kl}\alpha_k\alpha_l\right\}, \qquad (C = B^{-1}) \quad (50)$$

because the left-hand side is just the N-dimensional Fourier transform of the function (49) which can be explicitly evaluated by the obvious generalization of the formula (48). If (50) is differentiated with respect to α_k and α_l (not excluding $k = l$) and then all α are set zero, one obtains

$$\langle\phi(x_k)\phi(x_l)\rangle = C_{kl} = C(x_k, x_l). \qquad (51)$$

Thus, B is necessarily the inverse of a matrix C whose elements are given by the second equation of (51) with $C(x_1, x_2)$ a single function (the "covariance function") of two variables. This function then completely characterizes the statistical nature of $\phi(x)$; however, it too is not arbitrary, but subject to the condition that for *any choice* of the points $x_k (k = 1, 2, \cdots, N$; N arbitrary) the matrix C_{kl} must be positive definite.

All this generalizes in an obvious manner to the case when the independent variable x ranges over a three-dimensional space, say, and also to the case when a certain finite number $\phi_1(x), \phi_2(x), \cdots \phi_k(x)$ of processes are involved. In the latter case the covariance function is replaced by the covariance matrix

$$\langle\phi_\alpha(x_1)\phi_\beta(x_2)\rangle = C_{\alpha\beta}(x_1, x_2). \qquad (52)$$

The formula (50) is the basis for the application to statistical mechanics. If all the α's are put equal to a common constant we get, integrating over the x's,

$$\left\langle\left[\int_0^L dx e^{i\alpha\phi(x)}\right]^N\right\rangle = \int_0^L dx_1 \cdots \int_0^L dx_N \times \exp\left\{-\frac{\alpha^2}{2}\sum_{k=1}^{N}\sum_{l=1}^{N} C(x_k, x_l)\right\}. \qquad (53)$$

If it were not for the terms $k = l$, this would be the partition function of a canonical ensemble for a system consisting of N particles between which there are two-body forces with a potential $C(x_1, x_2)$.[13] When several Gaussian functions $\phi_\alpha(x)$ are involved then the corresponding formula refers to the statistical mechanics of a system which contains different species of particles and (52) then just gives the interaction potential between the α and β species.

This is the point of contact of the present work with the paper of Edwards.[3] There the averaging over function space was done in a formal way; however, the method prescribed[14] shows that averaging over a Gaussian random process is involved. It is clear that no property of the interparticle potential needs to be used other than that it is the covariance function (or matrix, in the case of more than one species) of a Gaussian random process. This shows that the representation of the partition function as a functional average has great generality, although it should be said as a precaution that the positive definiteness condition on the covariance function is certainly not a natural condition if that function is to be interpreted as a potential.[15] We also see now that the really special feature of our model is the Markoffian nature of the random process associated with it in the sense just explained. This is a severe restriction. There is no possibility for application to three-dimensional systems, because the concept of a Markoffian process involves the idea of a succession in "time" and this is meaningless when there is more than one independent variable. More importantly, even a one-dimensional Gaussian process is not Markoffian in general. An interesting process, both Gaussian and Markoffian, is the Ornstein-Uhlenbeck process

[13] This observation lies at the root of a recent paper on the statistical mechanics of a one-dimensional system, M. Kac, Phys. Fluids 2, 8 (1959).

[14] Reference 3, Eq. (1.7).

[15] If the positive definiteness condition is not respected then one deals with "integrals" of the type (48) with negative parameter a. This may not be fatal as long as the integral is only a symbolic expression which stands for the right-hand side. However, it is quite easy this way to get into contradictions because formal manipulations of such "integrals" may be illegitimate on account of their lack of convergence. There is a legitimate way of overcoming this difficulty due to A. J. F. Siegert, but we do not want to go into this subject here. One of us (A. L.) is indebted to Professor Siegert for letting him see his recent work before publication.

defined by the covariance function[16]

$$\langle\phi(x_1)\phi(x_2)\rangle = C_1 e^{-C_2|x_1-x_2|}, \quad (C_1, C_2 > 0). \tag{54}$$

With an interparticle potential of this form it is then possible to investigate the statistical mechanics in much the same detail as for our model.[17] For the general case the representation of the partition function as a functional integral is a device which may or may not be a convenient aid in approximate calculations, but progress of the kind that we are reporting here on the one-dimensional plasma cannot be expected in general.

It remains to point out the specific connection between our work and the preceding considerations. Let $\phi(x)$ be the Wiener process as defined in Sec. 3 and normalized by the condition $\phi(0) = 0$, and let us associate with a particle species (having charge σ', say) the process which is just a numerical multiple of it

$$\phi'(x) = \sigma'\phi(x). \tag{55}$$

The processes $\phi'(x), \phi''(x), \cdots$ are Gaussian and their covariance matrix is given by the elements

$$C'''(x_1, x_2) = \langle\phi'(x_1)\phi''(x_2)\rangle = \sigma'\sigma''\langle\phi(x_1)\phi(x_2)\rangle = 2D\sigma'\sigma'' \min(x_1, x_2), \tag{56}$$

as one may verify easily by carrying out the relevant integrals. On the other hand, the total potential energy, i.e., the sum of (7) and (9), is

$$\begin{aligned} V &= -2\pi \sum_{k=2}^{N} \sum_{l=1}^{k-1} \sigma_k\sigma_l |x_k - x_l| + 2\pi \sum_{k=1}^{N} \sigma_k \sum_{l=1}^{N} \sigma_l x_l \\ &= 2\pi \sum_{k=1}^{N} \sum_{l=1}^{N} \sigma_k\sigma_l \left\{ -\frac{|x_k - x_l|}{2} + \frac{x_k + x_l}{2} \right\} \\ &= 2\pi \sum_{k=1}^{N} \sum_{l=1}^{N} \sigma_k\sigma_l \min(x_k, x_l). \end{aligned} \tag{57}$$

Thus, in view of (37)

$$\frac{V}{\theta} = \frac{1}{2} \sum_{k=1}^{N} \sum_{l=1}^{N} C_{kl}(x_k, x_l), \tag{58}$$

the subscripts being interpreted in the obvious sense as referring to the charges of the particles which label the rows and columns of the matrix C. Now the identification of the right-hand side of Eq. (35) with that of Eq. (3) is a simple matter of using the formula (53). This derivation of (35) is even simpler than the one presented at the beginning of this section. We have presented the latter because is presupposes less familiarity with the ideas of a Gaussian random process.[18]

One final remark. The Wiener process is not "stationary," i.e., explicitly independent of time, since the condition $\phi(0) = 0$ puts the Brownian particle at the origin at a given time thus distinguishing this instant from all others. This implies that the covariance function is not a function of the difference $x_1 - x_2$ alone [cf. Eq. (56)]. The *general* identification of the covariance with the interparticle potential demands that the process be stationary because the interparticle potential is a function of the difference of the two position variables. But the one-dimensional plasma is an exception. Here the total potential energy is not just the sum of pair potentials [cf. discussion in Sec. 2]. This explains how it comes about that the Gaussian random process appropriate to this problem is the nonstationary Wiener process.

5. THE CHARACTERISTIC VALUE PROBLEM

We now assume that all charges are rational multiples of each other, so that with a proper choice of the charge unit σ', σ'', etc., will be integers. For convenience we also choose the unit of length in such a way that $D = 1$.

The coefficient function $F(\phi)$ in Eq. (46) is now periodic with period 2π. This makes it possible to reduce the problem posed by Theorem 1 to the characteristic value problem of an ordinary differential operator on a finite interval of the independent varible. Let

$$\hat{Q}(\phi, x) = \sum_{n=-\infty}^{\infty} Q(\phi + 2\pi n, x). \tag{59}$$

This function is the "periodic fundamental solution" of the partial differential equation (46), i.e., for $x = 0$ it reduces to

$$\hat{Q}(\phi, 0) = \sum_{n=-\infty}^{\infty} \delta(\phi + 2\pi n). \tag{60}$$

It is clear that

$$\Omega = \int_{-\pi}^{\pi} d\phi \hat{Q}(\phi, L). \tag{61}$$

Since F does not depend on x, we may use the method of separation of variables. This leads to the charac-

[16] M. C. Wang, and G. E. Uhlenbeck, Revs. Modern Phys. 17, 323 (1945).

[17] M. Kac, "Probability in Classical Physics," *Proceedings of the Seventh Symposium in Applied Mathematics of the American Mathematical Society* (McGraw-Hill Book Company, Inc., New York, 1957). It appears to be the first published suggestion of using stochastic averaging as an aid in calculating partition functions.

[18] An excellent general introduction to this subject is contained in Wang and Uhlenbeck, reference 16.

teristic value problem

$$\{d^2/d\phi^2 + F(\phi)\}y(\phi) = \gamma y(\phi) \tag{62}$$

with the boundary condition

$$y(\phi + 2\pi) = y(\phi). \tag{63}$$

We shall label the characteristic values and functions with a subscript m ($m = 0, 1, 2, \cdots$) if the need arises. The functions are assumed to be normalized, orthogonal

$$\int_{-\pi}^{\pi} d\phi y_m(\phi) y_{m'}(\phi) = \delta_{mm'}, \tag{64}$$

and to form a complete set for functions with period 2π. The expansion of $\hat{Q}$ in terms of these functions is

$$\hat{Q}(\phi, x) = \sum_{m=0}^{\infty} e^{\gamma_m x} y_m(0) y_m(\phi). \tag{65}$$

Hence, the grand partition function appears exhibited in the following form

$$\Omega = \sum_{m=0}^{\infty} A_m e^{\gamma_m L}, \tag{66}$$

where the coefficients are

$$A_m = y_m(0) \int_{-\pi}^{\pi} d\phi y_m(\phi). \tag{67}$$

The γ_m and the y_m depend parametrically on z', z'', etc., which enter into the definition of $F(\phi)$.

Let us make some general remarks on the characteristic value problem. It is a generalization of the type known in the mathematical literature as Hill's problem.[19] The only difference is that our function $F(\phi)$ is complex in general while in the Hill problem the corresponding function is real. This has the consequence that our problem, while being self adjoint in the real sense, is not necessarily *Hermitian*. Thus, it cannot be excluded *a priori* that characteristic values and functions be complex, but it seems likely that the expansion theorem still holds.[20] The function F has the symmetry

$$F^*(-\phi) = F(\phi), \tag{68}$$

which does imply that together with any characteristic function $y(\phi)$ another characteristic function is $y^*(-\phi)$, the corresponding values of γ being complex conjugates. But from this it only follows that $\hat{Q}$ has the same symmetry (68) and that the terms of the expansion (66) are either real or occur in complex conjugate pairs.

Let us consider now the asymptotic behavior of Ω as $L \to \infty$. It is clear that those terms of (66) survive which have the largest real parts for the γ's. Thus,

$$\Omega(L) \sim \sum_{m}{}' A_m e^{\gamma_m L}, \qquad (L \to \infty), \tag{69}$$

where the dash indicates that the summation is extended only over those terms for which Re γ_m is the maximum.[21] As just remarked, the terms of this sum are either real or occur in complex conjugate pairs, but one at least must be real otherwise the sum could not be positive for all L which would preclude the asymptotic equality with the positive function $\Omega(L)$. Let us call this real value γ_0. By taking the logarithm and comparing with (4) we have obtained

Theorem 2. *The thermodynamic pressure is given by*

$$P = 2\pi\gamma_0, \tag{70}$$

where γ_0 is the largest real characteristic value of the problem (62) *with F being given by* (36).[22]

Regarding the magnitude of γ_0 we have the inequalities

$$0 < \gamma_0 \leq (z' + z'' + \cdots), \tag{71}$$

the equality being possible only in the trivial case of no interaction (all σ's zero). This follows from the corresponding inequalities

$$1 < \Omega(L) \leq \exp[(z' + z'' + \cdots)L] \tag{72}$$

for the grand partition function, the second of which depends on the non-negative nature of the potential energy [see Eq. (13)].

We shall now discuss two general questions concerning the thermodynamic properties of the model. The first of these is the question of charge neutrality. It is intuitively evident that in the infinite system limit the total charge density should be zero

$$\sigma' n' + \sigma'' n'' + \cdots = 0. \tag{73}$$

This is not in contradiction with the fact that for any fixed value of L the mean particle numbers can be freely adjusted by varying the parameters z', z'' etc. The point is that as L becomes larger it becomes

[19] E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis* (Cambridge University Press, New York, 1927), Chap. XIX.

[20] See E. Hilb, Math. Ann. **71**, 76 (1912) where a closely related problem is studied.

[21] If is not excluded, of course, that this sum contains a single term; indeed, this should be expected to be the case in general. However, for special values of the z's the possibility of a degeneracy cannot be excluded.

[22] This (in a more restricted case) is the content of Theorem 4 of I.

"more and more difficult" to do this, and as $L \to \infty$ for any fixed z values, (73) becomes established as a limiting relation. We shall now prove this.

Hitherto ϕ was always a real variable, but once the characteristic values γ for Eq. (62) are determined the corresponding functions $y(\phi)$ can be analytically continued into the complex ϕ plane. This follows from a general theorem on differential equations with analytic coefficients.[23] Indeed, our equation has no finite singular points so that its solutions are entire functions, i.e., regular in the whole open complex plane. Together with any characteristic function $y(\phi)$ with characteristic value γ, consider the new function $Y(\phi) = y(\phi - i\kappa)$ with κ an arbitrary real number. Evidently

$$\{d^2/d\phi^2 + F(\phi - i\kappa)\}\, Y(\phi) = \gamma Y(\phi), \qquad (74)$$

and furthermore Y has the period 2π because $Y(\phi + 2\pi) - Y(\phi)$ is an entire function which vanishes on the line Im $\phi = i\kappa$ and hence vanishes everywhere. Now $F(\phi - i\kappa)$, regarded as a function of the real variable ϕ, arises out of $F(\phi)$ by the substitutions

$$\begin{aligned} z' &\to z'e^{\sigma'\kappa}, \\ z'' &\to z''e^{\sigma''\kappa}, \\ &\text{etc.} \end{aligned} \qquad (75)$$

Thus, we have constructed a characteristic function with the same value of γ of the problem (62) transformed by the substitution (75). This means that we have proven the following.

Theorem 3. *The characteristic values γ of the problem* (62), *regarded as functions of the parameters z', z'', etc., are invariant under the group of substitutions* (75). This means that the quantities

$$\gamma(z'e^{\sigma'\kappa}, z''e^{\sigma''\kappa}, \cdots) \qquad (76)$$

are in effect independent of κ for any values of the z's. Differentiating with respect to κ and setting it equal to zero it follows that

$$\{\sigma'z'\,\partial/\partial z' + \sigma''z''\,\partial/\partial z'' + \cdots\}\gamma(z', z'', \cdots) = 0. \qquad (77)$$

In particular, this is true for $\gamma = \gamma_0$ and now the stated conclusion follows from Eqs. (5) and Theorem 2.

The charge neutrality condition implies that if all nonvanishing[24] σ values are of the same sign the corresponding densities vanish. This is quite obvious physically: If only charges of one sign are admitted through the "membrane" separating the system from the reservoir [cf. discussion of Sec. 2], a "sheath" builds up which prevents particles to enter beyond the sheath region into the system. It is less obvious what mathematical property of the characteristic value problem is responsible for this radically different behavior of the case of all σ's having one sign. To clear this up it is convenient to reformulate the problem in terms of the complex variable $u = e^{i\phi}$. We then have

$$\{-u^2\, d^2/du^2 - u\, d/du + z'u^{\sigma'} + z''u^{\sigma''} + \cdots\}y(u) = \gamma y(u), \qquad (78)$$

and the periodicity in ϕ is replaced by the demand that $y(u)$ be *single valued* in u. Now, the differential equation (78) has two singular points, $u = 0$ and $u = \infty$. They are of the irregular type when σ's with both signs occur; $u = 0$ is of the regular type, $u = \infty$ irregular, when all σ's are positive; and the other way round when all σ's are negative.[25] To be specific, let all of them be positive. Then two independent solutions of the equation exist[25] of the form

$$y(u) = u^{\sqrt{-\gamma}}\, f(u), \qquad (79)$$

where $f(u)$ is regular in the neighborhood of the origin. Evidently this is single valued only when

$$\begin{aligned} \gamma &= -m^2, \\ (m &= 0, 1, 2, \cdots). \end{aligned} \qquad (80)$$

The remarkable fact is that the characteristic values are *independent* of the z's. In particular $\gamma_0 = 0$ which, as we have seen, implies the vanishing of the densities as well as the pressure. This analysis shows that certain striking features of the characteristic value problem, formulated at first in the real domain, are uncovered easiest by going into the complex plane. We have here the example of a problem where the "perturbation" does not affect the spectrum of characteristic values.

The second general question concerns the continuity of thermodynamic functions. We have seen how the latter are derivable from γ_0 given as a function of z', z'', etc. γ_0 has been defined as the largest real characteristic value for any given values of the parameters, but in some respect this is not a natural labeling. Characteristic values may be labeled alternatively by means of their association with the functions $y(\phi)$ which change smoothly as

[23] Reference 19, p. 194.

[24] Vanishing charge numbers mean added ideal gas components without interaction. This is easily seen to be compatible with Theorem 2.

[25] Reference 19, p. 197.

the z parameters are varied from zero up. In such a labeling one of the γ's tends to zero with decreasing z's which is thus γ_0 for sufficiently small z values. But it is in principle possible that a lower value may with increasing z's, actually "overtake" γ_0. It is evident that if this happens "the largest characteristic value" suffers a discontinuity in at least one of its z derivatives which leads to a discontinuity of some of the densities [cf. Eqs. (5)]. As the pressure is continuous, we get in the pressure-density diagram the horizontal portion typical of a phase transition.

Does this happen? We do not have a definite answer to this question. For a one-dimensional system of particles interacting with forces of finite range the answer is negative,[26] but our system is not of this type and a special investigation is needed. We have a partial result in this direction, and in order to formulate and prove it we need to broaden somewhat the characteristic value problem (62), namely, by replacing the condition of periodicity by

$$y(\phi + 2\pi) = \pm y(\phi). \tag{81}$$

Characteristic values where the plus sign applies will be called of type I, where the minus sign applies, of type II.

Theorem 4. *Let $\mathfrak{R}$ be a connected region in the space of z', z'', $\cdots$ which includes the point $z' = z'' = \cdots = 0$ and which is such that for all of its points all characteristic values (of both types) are real. Then within $\mathfrak{R}$ the largest characteristic value is not crossed by another one.*

When all z's vanish $F(\phi) = 0$ and the problem is trivial. The characteristic values of type I are given be (80) and those of type II by

$$\gamma = -(m + \tfrac{1}{2})^2, \qquad (m = 0, 1, 2, \cdots). \tag{82}$$

All of them are doubly degenerate except the single value $\gamma_0 = 0$ of type I. Moreover, those of type I alternate on the real line with those of type II. It is clear that when the z's are varied in $\mathfrak{R}$ the largest γ cannot be crossed by another one without this ordering being violated and if that is so there is a point in $\mathfrak{R}$ at which one of the characteristic values of type II coincides with one of type I (which may or may not be the largest one). But this is impossible because the two corresponding characteristic functions would have to satisfy

$$(d/d\phi)(y_{\mathrm{I}}\, dy_{\mathrm{II}}/d\phi - y_{\mathrm{II}}\, dy_{\mathrm{I}}/d\phi) = 0, \tag{83}$$

but the quantity inside the bracket is not a constant. It changes sign upon increasing the argument by 2π. This proves the theorem.

As a minimum the region $\mathfrak{R}$ includes those points where $z_1 = z_{-1}$, $z_2 = z_{-2}$, etc. (the subscripts are the values or the corresponding σ's). Then the problem (62) is real, thus also Hermitian, so that all γ's are real. Within this region then no phase transition can occur. It must be remembered that due to the invariance under the transformation (75) actually more points belong to $\mathfrak{R}$: all of those for which

$$z_1/z_{-1} = (z_2/z_{-2})^{\frac{1}{2}} = \cdots . \tag{84}$$

By a transformation of the form (75) the common value of these ratios may be brought to unity. In the case studied in I, namely a two-component system with charges equal in magnitude and opposite in sign, this condition is fulfilled trivially. We have excluded the possibility of a phase transition in those cases when a "detailed" balancing of charge obtains, i.e., a balancing for any pair of components carrying a charge of the same magnitude and opposite sign. Of course, the region $\mathfrak{R}$ may be larger, and it would be interesting to know how large, and also (if it is smaller than the totality of all z values) whether the possibility of a phase transition can be excluded on grounds independent of Theorem 4. It goes without saying that if a phase transition is possible,[27] a proof would be of great interest, as well as an analysis of its physical nature.

6. THE PLASMA LIMIT

In this section we restrict our considerations to the simplest case studied in I. Thus, we have only two kinds of particles with $\sigma = \pm 1$, and the corresponding values of z may be assumed equal without loss of generality. The problem is then to solve for the function $Q(\phi, x)$ which is the fundamental solution of the equation

$$\{\partial/\partial x - \partial^2/\partial\phi^2 - 2z\cos\phi\}Q(\phi, x) = 0. \tag{85}$$

The problem depends parametrically on the quantity z,[28] and we propose to investigate it in the limit of large z. As pointed out in I,[29] this limit corresponds to the true "plasma state," i.e., a state when the charges move independently of one another in the first approximation, and the kinetic energy is large compared to the potential energy.

[26] L. Van Hove, Physica **16**, 137 (1950).

[27] We do not think this is likely.

[28] $z/D\sigma^2 = \theta z/2\pi\sigma^2$ in ordinary units.

[29] Section 8.

In I the treatment was based on the asymptotic analysis of the characteristic value problem (the Mathieu equation). Here we want to show that a more direct, even if somewhat heuristic, treatment of Eq. (85) gives easily the dominant behavior in the limit of large z.

It is convenient to split off a factor and write

$$Q(\phi, x) = e^{2zx}Q_1(\phi, x), \tag{86}$$

so that Q_1 is the fundamental solution of

$$\{\partial/\partial x - \partial^2/\partial\phi^2 + 4z\sin^2\tfrac{1}{2}\phi\}Q_1(\phi, x) = 0. \tag{87}$$

Now that the coefficient function is non-negative, the equation can be interpreted as a diffusion problem with absorption. The "absorption coefficient" $4z\sin^2\frac{1}{2}\phi$ is large everywhere except in the immediate neighborhoods of the points $\phi = 2\pi n$ (n integer). Hence, Q_1 can be expected to be very small everywhere except in these neighborhoods. But again, all of these except $\phi = 0$ can be ignored because initially Q_1 has a δ singularity at $\phi = 0$, and the other points with no absorption are separated from this by regions of very strong absorption. Thus, it seems reasonable that a good approximation to Q can be obtained by replacing (87) by the equation

$$\{\partial/\partial x - \partial^2/\partial\phi^2 + z\phi^2\}Q_1(\phi, x) = 0, \tag{88}$$

which represents correctly the behavior of the absorption coefficient near $\phi = 0$. But this problem has the exact analytic solution

$$Q_1(\phi, x) = \left(\frac{z^{1/2}}{2\pi\sinh(2z^{1/2}x)}\right)^{1/2}\exp\left\{-\frac{z^{1/2}\phi^2}{2\tanh(2z^{1/2}x)}\right\} \tag{89}$$

as may be verified by substitution. In addition, the integration over ϕ can be carried out explicitly and one obtains

$$\Omega = e^{2zL}/(\cosh 2z^{\frac{1}{2}}L)^{\frac{1}{2}}. \tag{90}$$

This is only an approximation, of course, the true Ω being larger because the "absorpton coefficient" has been increased in the transition from Eq. (87) to (88). Remembering the definition of γ_0 we get in the same approximation

$$\gamma_0 = 2z - z^{\frac{1}{2}}. \tag{91}$$

This is the beginning of the asymptotic development of γ_0 in inverse powers of $z^{1/2}$ quoted in I.[29] The first term leads to the ideal gas law, the next to the "Debye-Hückel correction."[30]

It does not seem easy to improve on the formula (91) with the present method. The root of the difficulty is that we are not looking for an approximation to $Q(\phi, x)$, nor even to $\Omega = \Omega(L)$, but rather to the ultimate exponential rate of growth of Ω as $L \to \infty$. That this should not be easy is not surprising because the asymptotic analysis of the periodic solution of the Mathieu equation is involved and that is quite an elaborate affair.[31] Nevertheless, it is gratifying that at least the dominant behavior can be understood with the heuristic discussion of this section. If Eq. (88) is attacked by the method of separation of variables the characteristic value problem for the Hermite functions results.[32] That the analysis in the limit of large z leads to this problem has already been observed by Prager.[2] An approach to the Mathieu problem based on the Hermite functions as a first approximation is given by Morse and Feschbach.[33]

7. THE REDUCED DENSITY FUNCTIONS

We come now to a subject which demonstrates forcefully the superiority of the functional integration method over the direct treatment of I. We shall construct explicit and exact expressions for the reduced density functions.

These functions form a hierarchy of ever increasing complexity and are defined as follows. $f_{\sigma_1}(x_1)\,dx_1$ is the probability that the element dx_1 is occupied by a particle of charge σ_1; $f_{\sigma_1\sigma_2}(x_1x_2)\,dx_1\,dx_2$ is the joint probability that the two elements dx_1 and dx_2 are occupied by particles of charge σ_1 and σ_2, respectively; and so on. In contrast to the probability densities (2) the function f refer to occupation probabilities *regardless of the presence of other particles* in the system volume. There is a connection between the f's and the probability densities (2) which can be written down in the most concise form by making use of generating functionals introduced by Bogoliubov[34] and recently discussed by Green.[35] Let us introduce arbitrary functions $\zeta_{\sigma'}(x)$, $\zeta_{\sigma''}(x)$, etc., one associated

[30] For a discussion of the validity of the Debye-Hückel treatment see N. Balazs, Phys. Fluids **4**, 1259 (1961).

[31] E. L. Ince, Proc. Roy. Soc. Edinburgh **46**, 316 (1926). Ince wrote "If anyone had the courage to push the development a stage or two further he would greatly enhance the value of an important expansion. But any reader who attempts to verify [my] results given above will realize that the work involved would be tremendous."

[32] The equation differs only in a factor i in front of the first term from the Schrödinger equation for a harmonic oscillator.

[33] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1416.

[34] N. N. Bogoliubov, *Problems of a Dynamical Theory in Statistical Physics* (in Russian). Also available in English translation by E. K. Gora, *Studies in Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962), Vol. I.

[35] M. S. Green, "Some Applications of the Generating Functional of the Molecular Distribution Functions" in Vol. III of *Lectures in Theoretical Physics* held at the University of Colorado in 1960 (Interscience Publishers, Inc., New York, 1961).

with each species of particles, and form the quantity

$$\Phi = \frac{1}{\Omega} \sum_{N'=0}^{\infty} \sum_{N''=0}^{\infty} \cdots \frac{1}{N'!\,N''! \cdots} \int_0^L dx_1 \cdots$$

$$\times \int_0^L dx_N \prod_{k=1}^{N} [z_k + \zeta_{\sigma_k}(x_k)] \exp\left\{-\frac{V_{N'N'' \cdots}}{\theta}\right\}. \quad (92)$$

It is to be regarded as a functional of the ζ's. We have then the following formulas for the reduced density functions

$$f_{\sigma_1}(x_1) = z_1 \, \delta\Phi/\delta\zeta_{\sigma_1}(x_1),$$

$$f_{\sigma_1\sigma_2}(x_1x_2) = z_1z_2 \frac{\delta^2\Phi}{\delta\zeta_{\sigma_1}(x_1)\,\delta\zeta_{\sigma_2}(x_2)}, \quad (93)$$

and so on. It is understood that after the functional differentiation all ζ's are set identically equal zero.[34,35]

We now make the important observation that Φ (apart from the factor Ω^{-1}) has a structure very closely resembling that of the partition function itself. In fact, the difference is only that with each integration variable x_k we now have a factor $z_k + \zeta_{\sigma_k}(x_k)$ instead of just z_k as before. Therefore the analysis of Sec. 4 which led us to the formula (35) can be duplicated with these obvious and simple changes. The analog of Eq. (35) for Φ is then as follows

$$\Phi = \frac{1}{\Omega}\left\langle \exp \int_0^L dx F(\phi(x), x) \mid \phi(0) = 0 \right\rangle, \quad (94)$$

where the function F is now

$$F(\phi, x) = [z' + \zeta_{\sigma'}(x)]e^{i\sigma'\phi} + [z'' + \zeta_{\sigma''}(x)]e^{i\sigma''\phi} + \cdots. \quad (95)$$

The functional dependence of Φ on $\zeta_{\sigma'}$, $\zeta_{\sigma''}$, etc., comes through the functional dependence of F. The formula (94) together with (95) represents a *closed, exact expression* for the Bogoliubov functional of the reduced density functions in the one-dimensional plasma.

The theorem of Kac [cf. Sec. 3] could now be applied thus expressing Φ in terms of the fundamental solution of a partial differential equation. The method of separation of variables is inapplicable, however, because the coefficient function (35) is explicitly dependent on x through the functions $\zeta_{\sigma'}$, $\zeta_{\sigma''}$, etc. Fortunately this causes no difficulty because we are really not interested in Φ but rather in its functional derivatives with respect to the ζ's. For these a different method of evaluation suggests itself.

Since the integral over F occurs as the argument of the exponential function, every functional differentiation with respect to

$$\zeta_{\sigma_1}(x_1), \quad \zeta_{\sigma_2}(x_2), \quad \text{etc.},$$

brings down factors

$$e^{i\sigma_1\phi(x_1)}, \quad e^{i\sigma_2\phi(x_2)}, \quad \text{etc.} \quad (96)$$

Thus, we have

$$f_{\sigma_1\sigma_2 \cdots}(x_1x_2 \cdots) = \frac{z_1z_2 \cdots}{\Omega}\left\langle \exp\left\{i\sigma_1(x_1) + i\sigma_2(x_2) + \cdots \int_0^L dx F(\phi(x))\right\} \,\middle|\, \phi(0) = 0 \right\rangle, \quad (97)$$

where F is now just (36) since the ζ's are to be put zero after the functional differentiation with respect to them. It is clearly possible to arrange the notation so that the successive arguments of f form an increasing sequence, provided we leave the corresponding charges arbitrary. Thus, it will be understood that

$$0 < x_1 < x_2 < \cdots < L. \quad (98)$$

We now break up the integral in the exponent into a sum

$$\int_0^L dx = \int_0^{x_1} dx + \int_{x_1}^{x_2} dx + \cdots. \quad (99)$$

The Wiener average is over all continuous functions such that $\phi(0) = 0$. We carry out this averaging by making, initially, a number of additional restrictions $\phi(x_1)$ in $d\phi_1$, $\phi(k_2)$ in $d\phi_2$, etc., and afterward integrating over all ϕ_1, ϕ_2, etc. In the restricted Wiener integrals the factors (96) are just constants, and according to (99) we just have product of factors of the form

$$\left\langle \exp\left\{\int_{x_{k-1}}^{x_k} dx F(\phi(x))\right\} \,\middle|\, \phi(x_{k-1}) = \phi_{k-1}, \phi(x_k) = \phi_k \right\rangle. \quad (100)$$

We have seen in Sec. 3 that this is just $Q(\phi_k, \phi_{k-1}, x_k - x_{k-1})$, the function Q being defined in Theorem 1. Thus, we obtain the following results

$$f_{\sigma_1}(x_1) = \frac{z_1}{\Omega} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\phi_1 \times Q(\phi, \phi_1, L - x_1)e^{i\sigma_1\phi_1}Q(\phi_1, 0, x_1), \quad (101)$$

$$f_{\sigma_1\sigma_2}(x_1x_2) = \frac{z_1z_2}{\Omega} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\phi_2 \int_{-\infty}^{\infty} d\phi_1 \times Q(\phi, \phi_2, L - x_2)e^{i\sigma_2\phi_2}Q(\phi_2, \phi_1, x_2 - x_1) \times e^{i\sigma_1\phi_1}Q(\phi_1, 0, x_1), \quad (102)$$

and so on. When the σ's are integers (which will be assumed in the following) the function Q may be replaced in these formulas by

$$\hat{Q}(\phi, \phi', x) = \sum_{n=-\infty}^{\infty} Q(\phi + 2\pi n, \phi', x) \qquad (103)$$

and the integration limits changed to $-\pi$ and π. $\hat{Q}$ is periodic in *both* variables ϕ and ϕ', and its expansion in terms of the characteristic functions of the problem (62) is

$$\hat{Q} = \sum_{m=0}^{\infty} e^{\gamma_m x} y_m(\phi) y_m(\phi'). \qquad (104)$$

The results we have obtained are exact but they do not quite correspond to the customary definitions of the density functions. The reason is that our functions depend not only on all position variables, but even on L, the size of the system. In the customary definition the limiting form of these functions is meant, namely, the limit

$$\begin{gathered} x_1 \to \infty, \\ L - x_n \to \infty, \\ x_2 - x_1, \quad x_3 - x_2, \cdots, x_n - x_{n-1} \quad \text{fixed.} \end{gathered} \qquad (105)$$

These limiting functions truly express the bulk properties of the model inasmuch as "surface effects" have been removed. It is quite easy to carry out the limit (105) provided one notes that as $x \to \infty$

$$Q(\phi, \phi', x) \sim e^{\gamma_0 x} y_0(\phi) y_0(\phi') \qquad (106)$$

and of course the corresponding limiting form of Ω.[36] The results are

$$f_{\sigma_1} = z_1 \int_{-\pi}^{\pi} d\phi_1 y_0(\phi_1) e^{i\sigma_1\phi_1} y_0(\phi_1), \qquad (107)$$

$$f_{\sigma_1\sigma_2} = z_1 z_2 e^{-\gamma_0(x_2 - x_1)} \int_{-\pi}^{\pi} d\phi_2 \int_{-\pi}^{\pi} d\phi_1 y_0(\phi_2) e^{i\sigma_2\phi_2} \times \hat{Q}(\phi_2, \phi_1, x_2 - x_1) e^{i\sigma_1\phi_1} y_0(\phi_1), \qquad (108)$$

$$f_{\sigma_1\sigma_2\sigma_3} = z_1 z_2 z_3 e^{-\gamma_0(x_3 - x_1)} \int_{-\pi}^{\pi} d\phi_3 \int_{-\pi}^{\pi} d\phi_2 \int_{-\pi}^{\pi} d\phi_1 \times y_0(\phi_3) e^{i\sigma_3\phi_3} \hat{Q}(\phi_3, \phi_2, x_3 - x_2) \times e^{i\sigma_2\phi_2} \hat{Q}(\phi_2, \phi_1, x_2 - x_1) e^{i\sigma_1\phi_1} y_0(\phi_1), \qquad (109)$$

and so on. These functions are manifestly translation invariant. Furthermore, they satisfy the "product condition." This means that if a group of arguments is far removed from another group than the f corresponding to the whole group is the product of the two f's corresponding to the two separated groups. This is shown by making use of the asymptotic relation (106) for that factor $\hat{Q}$ whose x argument is the separation between the two groups.

Let us consider the simplest of these functions. $f_{\sigma'}\, dx$ is the probability that an element dx is occupied by a particle of charge σ', therefore, $f_{\sigma'}$ should also be the mean number density for this species. This, we have seen, is

$$n' = z' \, \partial\gamma_0/\partial z'. \qquad (110)$$

Let us show that the right-hand sides of Eqs. (107) and (110) are identical. We have

$$d^2 y_0/d\phi^2 + F y_0 = \gamma_0 y_0. \qquad (111)$$

Differentiate this with respect to the parameter z'.

$$\frac{d^2}{d\phi^2}\left(\frac{\partial y_0}{\partial z'}\right) + \frac{\partial F}{\partial z'} y_0 + F \frac{\partial y_0}{\partial z'} = \frac{\partial \gamma_0}{\partial z'} y_0 + \gamma \frac{\partial y_0}{\partial z'} \qquad (112)$$

Now multiply (111) by $\partial y_0/\partial z'$, (112) by y_0 and integrate them between $-\pi$ and π. The difference of the two equations so resulting is just

$$\int_{-\pi}^{\pi} \frac{\partial F}{\partial z'} y_0^2 \, d\phi = \frac{\partial \gamma_0}{\partial z'} \int_{-\pi}^{\pi} y_0^2 \, d\phi \qquad (113)$$

which, in view of the normalization of y_0, is the identity desired.

Let us consider now the two particle density functions (108). By means of the expansion (104) we obtain[37]

$$f_{\sigma_1,\sigma_2}(x) = \sum_{m=0}^{\infty} B_{m,\sigma_1} B_{m,\sigma_2} e^{-(\gamma_0 - \gamma_m)x}. \qquad (114)$$

The coefficients are

$$B_{m,\sigma} = z_\sigma \int_{-\pi}^{\pi} d\phi y_0(\phi) y_m(\phi) e^{i\sigma\phi}. \qquad (115)$$

Clearly f_{σ_1,σ_2} is symmetric in the two subscripts which run over all particle species.

Certain general properties of these functions will now be shown under the hypothesis of "detailed charge neutrality." This means that the species occur in pairs of equal and opposite charge and the corresponding values of z are equal

$$z_\sigma = z_{-\sigma}. \qquad (116)$$

This implies that the corresponding densities are also equal

$$n_\sigma = n_{-\sigma}. \qquad (117)$$

In more generality, all reduced density functions are invariant under a simultaneous change of sign of all their charge subscripts.

[36] We ignore here the possibility of degeneracy for the largest characteristic value.

[37] The variable x is here always taken positive. This is no restriction on generality.

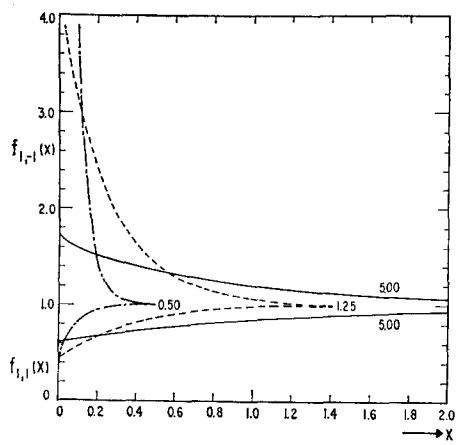

FIG. 1. Two particle density functions.

Theorem 5. *Suppose there is detailed charge neutrality. Let ξ_σ be arbitrary real numbers, one associated with each particle species. Then the function*

$$\sum_{\sigma_1} \sum_{\sigma_2} \xi_{\sigma_1}\xi_{\sigma_2} f_{\sigma_1,-\sigma_2}(x) \tag{118}$$

is positive and monotone decreasing. To prove this, note that in the case of detailed charge neutrality

$$F(\phi) = 2z_1 \cos\phi + 2z_2 \cos 2\phi + \cdots, \tag{119}$$

so that the basic characteristic value problem (62) is real. Hence, in this case, all γ_m are real as well as all $y_m(\phi)$. But then

$$B_{m,-\sigma} = B^*_{m,\sigma}, \tag{120}$$

and therefore the function (117) is

$$\sum_{m=0}^{\infty} |\sum_{\sigma} \xi_\sigma B_{m,\sigma}|^2 e^{-(\gamma_0-\gamma_m)x}. \tag{121}$$

The conclusion of the theorem is hereby exhibited.

The physical significance is best appreciated by a choice of suitable values for the ξ_σ. It is easy to see, for instance, that the functions $f_{\sigma,-\sigma}(x)$ and $f_{\sigma,\sigma}(x) + f_{\sigma,-\sigma}(x)$ are monotone decreasing. Also $f_{\sigma,-\sigma}(x) - f_{\sigma,\sigma}(x)$ are positive and monotone decreasing. The latter fact means that *it is more likely for two particles, a given distance apart, to have equal and opposite charges than the same charge, and the difference decreases with distance.* This is quite reasonable. It is a manifestion of the attraction (repulsion) between charges of opposite (same) sign.

There is a more subtle effect connected with the tendency for the two particles in dx_1 and dx_2 to have the same *magnitude* of their charge. To show this, we consider any even function of the charge σ, such as $|\sigma|$ for instance, and calculate its covariance at the two points x_1 and $x_2 = x_1 + x$. Let ξ_σ be this function, and let us denote by $\langle\cdots\rangle_{\text{av}}$ averaging over charges. The covariance is

$$\begin{aligned}\langle(\xi_{\sigma_1} - \langle\xi_{\sigma_1}\rangle_{\text{av}})(\xi_{\sigma_2} - \langle\xi_{\sigma_2}\rangle_{\text{av}})\rangle_{\text{av}} \\ &= \langle\xi_{\sigma_1}\xi_{\sigma_2}\rangle_{\text{av}} - \langle\xi_{\sigma_1}\rangle_{\text{av}}\langle\xi_{\sigma_2}\rangle_{\text{av}} \\ &= \sum_{\sigma_1}\sum_{\sigma_2} \xi_{\sigma_1}\xi_{\sigma_2}[f_{\sigma_1,\sigma_2}(x) - f_{\sigma_1}f_{\sigma_2}].\end{aligned} \tag{122}$$

Since $\xi_\sigma = \xi_{-\sigma}$ we can ignore the minus sign in (118), and now the theorem implies that the expression (122) is positive, since it obviously tends to zero as $x \to \infty$. Thus *ξ_σ at the point x_1 and ξ_σ at the point x_2 are positively correlated random variables.*

Certain rather plausible properties of the two particle density functions are not implied by our theorem. In particular, no assertion is made about the functions $f_{\sigma,\sigma}(x)$ although it seems likely on physical grounds that they are monotone increasing functions of x. To check this, we have performed numerical computations in the simplest case of a two-component plasma with $\sigma = \pm 1$ and $z_1 = z_{-1}$. In this case the characteristic value problem is just the Mathieu equation

$$d^2y(\phi)/d\phi^2 + 2z\cos\phi y(\phi) = \gamma y(\phi), \tag{123}$$

and detailed tables are available on the periodic solutions as well as the corresponding values of γ.[38] The Mathieu functions are given in terms of the coefficients of their Fourier series which converge rapidly. Therefore the integrals $B_{m,\sigma}$ of Eq. (115) can be also expressed as fast converging series with tabulated numerical coefficients. The decrease with m is also fairly rapid, although this depends on the value of z,[28] and becomes worse with increasing z. We present the results on the accompanying figure for three representative values of z. For convenience of physical interpretation the unit of distance is the mean interparticle distance n^{-1}. The decreasing functions are $f_{1,-1}(x)$, the increasing ones $f_{1,1}(x)$. We note that for larger z the approach to the asymptotic value is slower and the deviation from it smaller at small distances. For very small values of z the approach to the asymptotic value takes place in a distance small compared to the mean interparticle distance. This is a manifestation of the "condensation" into pairs of oppositely charged particles in this limit. The distance between members of a pair becomes small compared to the distance between pairs. This is a confirmation of the picture of the system obtained in I on thermodynamic grounds.[39]

[38] National Bureau of Standards, *Tables Relating to Mathieu Functions* (Columbia University Press, New York, 1951).
[39] Section 9.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Dynamic Behavior of a Set of Weakly Coupled Ising Spins*

PAUL H. E. MEIJER, TOMOYASU TANAKA,† AND JEREMIAH BARRY
Department of Physics, Catholic University of America, Washington, D. C.
(Received March 16, 1961)

The master equation of a set of independent equivalent spins contains only one undetermined constant, the rate constant. If one assumes the local field to be altered by the field produced by the two neighboring spins, one can formulate a set of equations for the average of one, two, three, etc., spins. On assuming an Ising interaction between the spins, weak compared to the coupling with the heat bath, we can terminate the hierarchy and solve the problem of a linear chain with periodic boundary conditions by Fourier-transformation. The resulting secular equation determines two sets of relaxation times and two sets of eigensolutions. An explicit solution for the spin averages is given for the initial condition describing a localized excitation. Similarities with, and differences between, this and the random walk problem is pointed out.

1. INTRODUCTION

THE master equation for a set of Ising spins can be obtained in a simple way provided the coupling between the spins is small compared to the coupling between the spins and the temperature bath. The problem is explicitly solvable for a one-dimensional chain as long as the spin can take only two values. In this case the conditions of detailed balancing determines the rate matrix elements except for a common proportionality factor which determines the time scale. Extension to more than one dimension cannot be done without further assumptions, hence the present calculation serves as an exploratory model only. It is not possible to describe a cooperative phenomenon this way. This will be done in a subsequent paper.

The calculations are based on a model suggested by Glauber.[1] His calculations are in the strong coupling limit, ours in the weak coupling limit in the sense that the external field is considered predominant.

2. UNCOUPLED N-SPIN SYSTEM

The probability distribution of an N-particle system of spins that has two components only is given by a function $P(\sigma_1, \sigma_2 \cdots \sigma_N; t)$ which has 2^N possible values for its spin arguments: $\sigma_i = \pm\frac{1}{2}$ $(i = 1, \cdots N)$. Hence, the time behavior of such a function can be described by a matrix

$$P(\{\sigma_i\}; t + \tau) = \sum \Omega(\{\sigma_i\} \mid \{\sigma_i'\}; \tau)P(\{\sigma_i'\}; t). \quad (1)$$

In case the spins are independent, the matrix is the direct product of all individual 2×2 matrices

$$\Omega = \prod_{i=1}^{N} \omega_i = \omega_1 \times \omega_2 \cdots \times \omega_N. \quad (2)$$

These matrices describe evolution in time of the probability of the individual spins.

$$P(\sigma_i; t + \tau) = \sum_{\sigma_i' = \pm 1/2} \omega(\sigma_i \mid \sigma_i'; \tau)P(\sigma_i'; t). \quad (3)$$

Equation (3) can be obtained by reducing $P(\{\sigma_k\})$, i.e., summing over all spin variables σ_k excluding σ_i, and by using the special form of Ω given by (2).

In order to preserve the normalization of the probability at all times, we have to require that each sum over a row is equal to one.

$$\sum_{\{\sigma_i\}} \Omega(\{\sigma_i\} \mid \{\sigma_i'\}; \tau) = 1. \quad (4)$$

In particular, in Eq. (3),

$$\omega_{++} + \omega_{+-} = 1; \qquad \omega_{-+} + \omega_{--} = 1. \quad (5)$$

In order to obtain the master equation or Pauli equation of the system, we subtract from each side $P(\sigma; t)$, divide by τ, and take the limit for $\tau \to 0$.

$$\frac{\partial P(\sigma)}{\partial t} = \sum_{\sigma'} W(\sigma \mid \sigma')P(\sigma') - \sum_{\sigma'} W(\sigma' \mid \sigma)P(\sigma). \quad (6)$$

(To simplify notation let σ without an index refer to the entire set of σ_i's.) The transition coefficients W are

$$W(\sigma \mid \sigma') = \lim_{\tau \to 0} \tau^{-1}\Omega(\sigma \mid \sigma'; \tau); \qquad \sigma \neq \sigma', \quad (7)$$

and the condition (4) is replaced by

$$\sum_{\sigma}' W(\sigma \mid \sigma') = 0. \quad (8)$$

In the same way we can introduce the master equation for a single spin i:

* Supported by a U. S. Air Force Contract.
† On leave of absence from Kyusyu University, Hukuoka, Japan.
[1] R. J. Glauber, Bull. Am. Phys. Soc. 5, 296 (1960).

$$\frac{\partial P(\sigma_i)}{\partial t} = \sum_{\sigma_i'} w(\sigma_i \mid \sigma_i')P(\sigma_i') - \sum_{\sigma_i'} w(\sigma_i' \mid \sigma_i)P(\sigma_i). \quad (9)$$

The four elements in the matrix w have to fulfill condition (8) and also detailed balancing; hence, they are determined apart from a proportionality factor[2]

$$w_{+-} = -w_{++} = \nu \exp(-\mu H/kT), \quad (10a)$$

$$w_{-+} = -w_{--} = \nu \exp(\mu H/kT), \quad (10b)$$

$$\text{and} \quad \bar{\bar{\omega}} = \tau\bar{\bar{w}} - 1, \quad (11)$$

where the rate constant ν is determined by the interaction with the heat bath.

We can calculate the matrix $W(\sigma \mid \sigma')$ for a system of N independent spins from (11), (10), (2), and (7):

$$W(\sigma \mid \sigma') = \lim_{\tau \to 0} \tau^{-1}\left[\prod_{i=1}^{N} (\tau\bar{\bar{w}}_i - 1) + 1\right], \quad (12)$$

where the product sign refers to a direct product.

Only the linear term in τ survives in the limit. Hence, the only nondiagonal elements of the matrix that are nonzero are those where only one σ, say σ_i, is different in the final state $\sigma_i \neq \sigma_i'$, while all other σ's are the same, $\sigma_k = \sigma_k'$ $(k \neq i)$. This means that only one spin flips at a time, in accordance with the picture of independent spins.

Another property of the special choice for W made in Eq. (2) or (12) is the reducibility of the master equation. If we introduce the reduced probability function

$$\sum_{\sigma_1}\sum_{\sigma_2}\cdots\sum_{\sigma_{i-1}}\sum_{\sigma_{i+1}}\cdots\sum_{\sigma_N} P(\sigma) = P(\sigma_i) \quad (13)$$

in Eq. (6), we can "project" the transition probability W into a 2×2 matrix if W is given by (12). We find the projected part of W to be w. This is not surprising since the direct product rule is the composition of single particles into a system of independent particles and the "projection" is the inverse process, starting from N-independent particles and decomposing it into a set of single-particle equations.

The description above is mainly given to establish a convenient starting point for a system of dependent particles. Before going into this we will write down the explicit formulas for a one-spin system. We introduce $\beta = NH/kT$

$$\sum_{\sigma_k} \sigma_k \dot{P}(\sigma_k) = \langle \dot{\sigma}_k \rangle = \nu[e^{\beta}P(-) - e^{-\beta}P(+)]$$

$$= -\nu\{2\langle\sigma_k\rangle \cosh\beta - \sinh\beta\}. \quad (14)$$

[2] Note that ω_{-+} corresponds to a transition from spin $-\frac{1}{2}$ to spin $+\frac{1}{2}$.

The solution of this differential equation is

$$\langle\sigma(t)\rangle = \sigma_\infty + (\langle\sigma(0)\rangle - \sigma_\infty)e^{-t/\tau}, \quad (15)$$

with

$$\sigma_\infty = \tfrac{1}{2}\tanh\beta \quad \text{and} \quad \tau^{-1} = 2\nu\cosh\beta. \quad (16)$$

The appearance of the temperature in the nonequilibrium equation is due to the requirement that the master equation must fulfill detailed balancing, and the result that for $t \to \infty$ the solution goes over into the equilibrium solution is implied. The relaxation time is always finite; this model does not contain any cooperative transitions which may give rise to infinite relaxation times.

3. WEAK COUPLING

In this case the field acting upon the ith spin will be

$$H_i = H_i^{(0)} + (\sigma_{i-1} + \sigma_{i+1})M, \quad (17)$$

where H_i^0 is the external field, until now called H, and M the coupling constant. If we evaluate the expression for $\langle\dot{\sigma}_i\rangle$ in a way similar to the previous section, we find that it depends on $\langle\sigma_i\sigma_{i+1}\rangle$ and $\langle\sigma_i\sigma_{i-1}\rangle$. The time derivatives of these averages depends on the three-spin correlation function, and so on. Since we cannot solve such a hierarchy of equations we assume the coupling to be weak and make a cutoff after a finite set of equations. For instance, the time derivative of the two-spin correlation contains terms in two-spin correlations and also, terms in three-spin correlations. The latter are the only ones that contain the coupling constant M. Hence, if we take $M \ll H$ we can express the two-spin correlation function in a closed differential equation.

Using this assumption in (10a) and (10b), we have

$$w_{+-}^{(i)} = w_{+-}^{(i)0}\{1 + (\sigma_{i-1} + \sigma_{i+1})\mu M/kT\}, \quad (18a)$$

$$w_{-+}^{(i)} = w_{-+}^{(i)0}\{1 + (\sigma_{i-1} + \sigma_{i+1})m\}, \quad (18b)$$

where m stands for $\mu M/kT$. The result is that Eq. (14) is modified, using

$$\langle\dot{\sigma}_k\rangle = -\nu\{2\langle\sigma_k\rangle\cosh\beta - \sinh\beta\} - \nu m\{2\langle\sigma_k\sigma_{k+1}\rangle\cosh\beta - \langle\sigma_{k+1}\rangle\sinh\beta + 2\langle\sigma_k\sigma_{k-1}\rangle\cosh\beta - \langle\sigma_{k-1}\rangle\sinh\beta\}. \quad (19)$$

On introducing the equilibrium average, this equation simplifies to

$$\langle\dot{\sigma}_k\rangle = -\tau^{-1}\{\langle\sigma_k\rangle - \langle\sigma_\infty\rangle\} - \tau^{-1}m\{\langle\sigma_k\sigma_{k+1}\rangle + \langle\sigma_k\sigma_{k-1}\rangle - \langle\sigma_{k+1}\rangle\langle\sigma_\infty\rangle - \langle\sigma_{k-1}\rangle\langle\sigma_\infty\rangle\}, \quad (19a)$$

where τ is defined in (16). If we introduce the relative deviations from equilibrium

$$X_k = \{\langle\sigma_k\rangle - \sigma_\infty\}/\sigma_\infty, \tag{20a}$$

$$Y_k = \{\langle\sigma_k\sigma_{k+1}\rangle - \sigma_\infty^2\}/\sigma_\infty^2, \tag{20b}$$

we have

$$\tau\dot{X}_k = -X_k - \xi(Y_{k-1} + Y_k - X_{k+1} - X_{k-1}), \tag{21}$$

with $\xi = m\sigma_\infty$. This equation can be solved in conjunction with a similar equation for Y. The latter is obtained under the condition that $m \ll \beta$. We find

$$\langle\sigma_k\sigma_{k+1}\rangle = -\nu\{2\langle\sigma_k\sigma_{k+1}\rangle \cosh\beta + 2\langle\sigma_k\sigma_{k+1}\rangle \times \cosh\beta - \langle\sigma_{k+1}\rangle \sinh\beta - \langle\sigma_k\rangle \sinh\beta\}, \tag{22}$$

or by using (20a) and (20b)

$$\tau\dot{Y}_k = -2Y_k + X_k + X_{k+1}. \tag{23}$$

Equations (21) and (23) form the starting point for the solution of the problem. In Sec. 4, we discuss the behavior in time of a particular initial condition.

In order to clarify the assumptions made in the previous derivation we write the master equation for the one-spin distribution function in terms of the three-spin distribution function. The latter refers to the spin under consideration, and its two-nearest neighbors:

$$\frac{dP(\sigma_k)}{dt} = \sum_{\sigma_{k-1}\sigma_{k-1}'\sigma_k'\sigma_{k+1}'\sigma_{k+1}} \Omega(\sigma_{k-1}\sigma_k\sigma_{k+1} \mid \sigma_{k-1}'\sigma_k'\sigma_{k+1}') \times P(\sigma_{k-1}'\sigma_k'\sigma_{k+1}'). \tag{24}$$

We assume, first, that the transition probability depends on the value of the neighboring spin in a parametric way (i.e., the relation is diagonal), and second, that the dependence is linear. As a result of these assumptions, the three-spin distribution function can be reduced to a two-spin distribution function.

$$\begin{aligned}\frac{dP(\sigma_k)}{dt} &= \sum_{\sigma_{k-1}\sigma_k'\sigma_{k+1}} w(\sigma_k\sigma_k') \times [1 + m(\sigma_{k-1} + \sigma_{k+1})]P(\sigma_{k-1}\sigma_k'\sigma_{k+1}) \\ &= \sum_{\sigma_k'} w(\sigma_k\sigma_k')P(\sigma_k') \\ &\quad + m\sum_{\sigma_{k-1}\sigma_k'} w(\sigma_k\sigma_k')\sigma_{k-1}P(\sigma_{k-1}\sigma_k') \\ &\quad + m\sum_{\sigma_k'\sigma_{k+1}} w(\sigma_k\sigma_k')\sigma_{k+1}P(\sigma_k'\sigma_{k+1}).\end{aligned} \tag{25}$$

The time derivative of the two-spin distribution function is a function of the four-spin distribution function. We assume again: (1) that it depends parametrically on the neighbors, and (2) that it depends linearly on the neighboring spins. However, the terms in higher order in the coupling constant will be neglected. We find the following expression:

$$\frac{dP(\sigma_k\sigma_{k+1})}{dt} = \sum_{\sigma_k'} w(\sigma_k\sigma_k')P(\sigma_k'\sigma_{k+1}) + \sum_{\sigma_{k+1}'} w(\sigma_{k+1}\sigma_{k+1}')P(\sigma_k\sigma_{k+1}') \tag{26}$$

and a similar expression for $P(\sigma_k, \sigma_{k-1})$. The one-spin distribution function is characterized by one quantity $\langle\sigma_k\rangle$; and the two-spin distribution function $P(\sigma_k\sigma_{k+1})$ by

$$4\langle\sigma_k\sigma_{k+1}\rangle = P(++) - P(+-) - P(-+) + P(++), \tag{27a}$$

$$2\langle\sigma_k\rangle = P(++) - P(-+) + P(+-) - P(--), \tag{27b}$$

$$2\langle\sigma_{k+1}\rangle = P(++) + P(-+) - P(+-) - P(--), \tag{27c}$$

$$1 = P(++) + P(-+) + P(-+) + P(--). \tag{27d}$$

Hence, Eqs. (25) and (26) can be rewritten in terms of these averages. The calculation gives again Eqs. (19) and (19a).

Equations (21) and (23) form a set of coupled equations in X_k and Y_k. They can be solved conveniently with a Fourier transform since this automatically satisfies the boundary condition that spins o and spin N are equivalent. Hence, we introduce

$$\begin{pmatrix}X_k\\Y_k\end{pmatrix} = \frac{1}{\sqrt{N}}\sum\begin{pmatrix}a_l\\b_l\end{pmatrix}e^{2\pi ilk/N}; \times \begin{pmatrix}a_l(t)\\b_l(t)\end{pmatrix} = \begin{pmatrix}a_l\\b_l\end{pmatrix}e^{-\lambda_l t}, \tag{28}$$

which gives the following two equations for the amplitudes

$$\tau\dot{a}_l = -a_l - \xi b_l[1 + \exp(-2\pi il/N)] + 2\xi a_l \cos(2\pi l/N), \tag{29a}$$

$$\tau\dot{b}_l = -2b_l + a_l[1 + \exp(2\pi il/N)]. \tag{29b}$$

The solution of the secular determinant in the approximation of small coupling constant ($\xi \ll 1$) gives the following two eigenvalues:

$$\tau\lambda_l^{(+)} = 2[1 - \xi - \xi\cos(2\pi l/N)], \tag{30a}$$

$$\tau\lambda_l^{(-)} = 1 + \xi. \tag{30b}$$

The eigenvectors are

$$a_l^{(+)} = 2\xi \exp(-\pi il/N) \cos(\pi l/N) C_l^{+}, \quad (31a)$$

$$b_l^{(+)} = (1 - 2\xi) C_l^{(+)}, \quad (31b)$$

$$a_l^{(-)} = -\exp(-\pi il/N) C_l^{(-)}, \quad (31c)$$

$$b_l^{(-)} = -2 \cos(\pi l/N) C_l^{(-)}. \quad (31d)$$

The time dependent solution is

$$a_l(t) = a_l^{(+)} \exp(-\lambda_l^{(+)} t) + a_l^{(-)} \exp(-\lambda^{(-)} t), \quad (32a)$$

$$b_l(t) = b_l^{(+)} \exp(-\lambda_l^{(+)} t) + b_l^{(-)} \exp(-\lambda_l^{(-)} t), \quad (32b)$$

and the constants $C_l^{(*)}$ are determined by the initial conditions.

4. COMPARISON OF THE SPIN DEVIATION WITH A RANDOM WALK PROBLEM

The probability distribution of the spin deviation as a function of position and time was obtained in the preceding section. We now find the solution that results from a delta function initial distribution in space. In doing so we can compare the result with the probability distribution function of the random walk. The functions have in common that the sharp peak for $t = 0$ will have a tendency to spread out for $t > 0$. However, the random-walk function, which obeys the differential equation that is used in the heat conduction problem, has the property that the total probability is always constant. In the case of the spin deviation this will not be so. If a particular spin has a deviation from the equilibrium distribution it will disappear monotonically in time, and if a certain spin in a weakly coupled system has a deviation it will tend to "spread" to its neighbors and at the same time will tend to disappear. Hence, we expect that the area under the probability curve will decrease in time. Since the exact solution of this problem becomes rather cumbersome because of the inverse Fourier transformation, we will again assume the quantity ξ to be small. In this case the answer can be evaluated in terms of Bessel functions of imaginary argument. We find, indeed, that the spin deviation at the "hot spot" will decrease monotonically. The neighbors, which were assumed to be in equilibrium initially, display probability-deviation functions that will increase initially and decrease for large values of the time.

Due to the cyclical boundary condition it is irrelevant which spin we select to be out of equilibrium. The initial condition is: All x_k and y_k are zero except the following three,

$$x_0(0) = X; \qquad y_0(0) = y_{-1}(0) = \tfrac{1}{2}Y. \quad (33)$$

The resulting Fourier amplitudes are, if we substitute (33) into (28),

$$a_l = N^{-\frac{1}{2}} X; \qquad b_l = \tfrac{1}{2} N^{-\frac{1}{2}} Y[1 + \exp(2\pi il/N)].$$

By introducing the corresponding values for C^* with (31) and by introducing this in (32) one obtains

$$\begin{aligned} a_l(t) = {} & \{1 + 4\xi + 2\xi \cos(2\pi l/N)\} \\ & \times [2\xi N^{-\frac{1}{2}}(X - 2Y) \cos^2(\pi l/N) \\ & \times \exp(-\lambda_l^{+} t) - \{2\xi N^{-\frac{1}{2}} Y \cos^2(\pi l/N) \\ & - N^{-\frac{1}{2}} X(1 - 2\xi)\} \exp(-\lambda_l^{-} t)], \end{aligned} \quad (34a)$$

$$\begin{aligned} b_l(t) = {} & \{1 + 4\xi + 2\xi \cos(2\pi l/N\} \\ & \times [N^{-\frac{1}{2}} \cos(\pi l/N) \exp(\pi il/N)(Y - 2X) \\ & \times \exp(-\lambda_l^{+} t) - 2N^{-\frac{1}{2}} \cos(\pi l/N) \\ & \times \exp(\pi il/N)\{2\xi \cos^2(\pi l/N) Y \\ & - (1 - 2\xi)X\} \exp(-\lambda_l^{-} t). \end{aligned} \quad (34b)$$

The contributions of the initial condition from X and Y are additive.

The inverse transformation can be performed if we replace the sum over l by an integral over φ, by taking the limit $N \to \infty$. The result is

$$\begin{aligned} x_0(t) = {} & \{X + 2\xi(X - Y)\} \exp\{-(1 + 2\xi)t/\tau\} \\ & - 2\xi(X - Y)(I_0 + I_1) \exp\{-2(1 - \xi)t/\tau\}, \\ x_1(t) = {} & \xi(X - Y)[\exp\{-(1 + 2\xi)t/\tau\} \\ & - (I_0 + 2I_1 + I_2) \exp\{-2(1 + \xi)t/\tau\}], \\ y_0(t) = {} & (Y - X)\{(1 + 3\xi)I_0 \\ & + (1 + 4\xi)I_1 + \xi I_2\} \exp\{-2(1 - \xi)t/\tau\} \\ & + [X - 3\xi(Y - X)] \exp\{-(1 + 2\xi)t/\tau\}, \\ y_1(t) = {} & (Y - X)\{\xi(I_0 + I_3) \\ & + (1 + 3\xi)(I_1 + I_2)\} \exp\{-2(1 - \xi)t/\tau\} \\ & - (Y - X)\xi \exp\{-(1 + 2\xi)t/\tau\}, \\ y_k(t) = {} & (Y - X)[\xi(I_{k-1} + I_{k+2}) \\ & + (1 + 3\xi)(I_k + I_{k+1})] \exp\{-2(1 - \xi)t/\tau\}, \end{aligned} \quad (35)$$

where the argument of the imaginary Bessel function[3] $I_n(z) = i^{-n} J_n(iz)$ is $Z = 2\xi t/\tau$. Although these functions increase exponentially for large argument, their product with the exponentials will always decrease.

[3] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1945), p. 224.

The functions I_n have the following characteristics. All functions $I_n(0)$ vanish except for $n = 0$ and $I_0(0)$ equals unity. Consequently, we see that all $x_n(0)$ are zero except x_0. The argument at which the function first exceeds a given value is larger for higher n, in other words, the farther away from the origin, the later the maximum spin deviation will be observed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Thermodynamic Functions of a Fermi System of Dilute Hard Spheres at Temperatures Near Absolute Zero

CHEN-PING YANG
The Ohio State University, Columbus, Ohio
(Received September 1, 1961)

The thermodynamic functions of a Fermi system of hard spheres at temperatures near absolute zero are calculated through the evaluation of the coefficient appearing before the second-order term a^2, in the fugacity series for the system where a is the hard sphere diameter.

1. INTRODUCTION

PROGRESS has been made in recent years in the calculation of the properties of the dilute hard-sphere gas in the ground state.[1] The present paper extends these computations to the thermodynamic properties of a dilute hard-sphere Fermi system at temperatures near absolute zero.

2. ASYMPTOTIC BEHAVIOR OF THE FUGACITY SERIES

The fugacity series for a Fermi system of hard spheres with spin J to the order a^2 has been given in reference 1:

$$\begin{aligned}\lambda^3(p/kT) &= \lambda^3 \sum_1^\infty b_l^4 z^l \\ &= -(2J+1)g_{3/2}(-z) \\ &\quad - 2J(2J+1)[g_{3/2}(-z)]^2(a/\lambda) \\ &\quad - 8J^2(2J+1)g_{1/2}(-z)[g_{3/2}(-z)]^2(a/\lambda)^2 \\ &\quad + 8J(2J+1)F(-z)(a/\lambda)^2 + O(a^3/\lambda^3), \end{aligned} \tag{1}$$

where

$$g_n(z) = \sum_{l=1}^\infty l^{-n} z^n, \tag{2}$$

and

$$F(z) = \sum_{r,s,t=1}^\infty (rst)^{-1/2}(r+s)^{-1}(r+t)^{-1} z^{r+s+t}. \tag{3}$$

Since the fugacity z is equal to $\exp(\mu/kT)$ for $\mu > 0$, as $T \to 0$, $z \to +\infty$. Using the asymptotic expansions of $g_{1/2}$, $g_{3/2}$, $g_{5/2}$, and F for $z \to +\infty$, one is able to obtain the asymptotic limit of p to the order T^2, the coefficients of which are series in powers of $a\mu^{1/2}$. The series as $z \to +\infty$ of the g's are known, they are

$$\begin{aligned} -g_{1/2}(-z) &= 2\pi^{-1/2}(\ln z)^{1/2} \\ &\quad - (12)^{-1}\pi^{3/2}(\ln z)^{-3/2} + O([\ln z]^{-7/2}), \\ -g_{3/2}(-z) &= 4(9\pi)^{-1/2}(\ln z)^{3/2} \\ &\quad + 6^{-1}\pi^{3/2}(\ln z)^{-1/2} + O([\ln z]^{-5/2}), \\ -g_{5/2}(-z) &= 8(15\pi^{1/2})^{-1}(\ln z)^{5/2} \\ &\quad + 3^{-1}\pi^{3/2}(\ln z)^{1/2} + O(1). \end{aligned} \tag{4}$$

The series of F was computed to the lowest order in reference 1. We shall now extend the calculation to the next order. The result is

$$\begin{aligned} -F(-z) &= (16/105)(11 - 2\ln 2)\pi^{-3/2}(\ln z)^{7/2} \\ &\quad - \tfrac{1}{3}\pi^{1/2}(4\ln 2 - 2)(\ln z)^{3/2} + O([\ln z]^{5/4}) \end{aligned} \tag{5}$$

To obtain this we follow a procedure used in reference 1, as shown in the Appendix.

[1] See T. D. Lee and C. N. Yang, Phys. Rev. 117, 12 (1960) and the papers referred to therein. We follow the notations used there: mass of particle $= \frac{1}{2}$, $\hbar = 1$, a = hard-sphere diameter, N = number of particles, Ω = volume of box, $\rho = N/\Omega$, J = spin of particles, P_F = maximum Fermi momentum for free particles $= [6\pi^2\rho/(2J+1)]^{1/2}$ $\beta = 1/kT$, $\lambda = (4\pi\beta)^{1/2}$, μ = fugacity.

The functions I_n have the following characteristics. All functions $I_n(0)$ vanish except for $n = 0$ and $I_0(0)$ equals unity. Consequently, we see that all $x_n(0)$ are zero except x_0. The argument at which the function first exceeds a given value is larger for higher n, in other words, the farther away from the origin, the later the maximum spin deviation will be observed.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Thermodynamic Functions of a Fermi System of Dilute Hard Spheres at Temperatures Near Absolute Zero

CHEN-PING YANG
The Ohio State University, Columbus, Ohio
(Received September 1, 1961)

The thermodynamic functions of a Fermi system of hard spheres at temperatures near absolute zero are calculated through the evaluation of the coefficient appearing before the second-order term a^2, in the fugacity series for the system where a is the hard sphere diameter.

1. INTRODUCTION

PROGRESS has been made in recent years in the calculation of the properties of the dilute hard-sphere gas in the ground state.[1] The present paper extends these computations to the thermodynamic properties of a dilute hard-sphere Fermi system at temperatures near absolute zero.

2. ASYMPTOTIC BEHAVIOR OF THE FUGACITY SERIES

The fugacity series for a Fermi system of hard spheres with spin J to the order a^2 has been given in reference 1:

$$\begin{aligned}\lambda^3(p/kT) &= \lambda^3 \sum_1^\infty b_l^4 z^l \\ &= -(2J+1)g_{3/2}(-z) \\ &\quad - 2J(2J+1)[g_{3/2}(-z)]^2(a/\lambda) \\ &\quad - 8J^2(2J+1)g_{1/2}(-z)[g_{3/2}(-z)]^2(a/\lambda)^2 \\ &\quad + 8J(2J+1)F(-z)(a/\lambda)^2 + O(a^3/\lambda^3), \end{aligned} \tag{1}$$

where

$$g_n(z) = \sum_{l=1}^\infty l^{-n} z^n, \tag{2}$$

and

$$F(z) = \sum_{r,s,t=1}^\infty (rst)^{-1/2}(r+s)^{-1}(r+t)^{-1} z^{r+s+t}. \tag{3}$$

Since the fugacity z is equal to $\exp(\mu/kT)$ for $\mu > 0$, as $T \to 0$, $z \to +\infty$. Using the asymptotic expansions of $g_{1/2}$, $g_{3/2}$, $g_{5/2}$, and F for $z \to +\infty$, one is able to obtain the asymptotic limit of p to the order T^2, the coefficients of which are series in powers of $a\mu^{1/2}$. The series as $z \to +\infty$ of the g's are known, they are

$$\begin{aligned} -g_{1/2}(-z) &= 2\pi^{-1/2}(\ln z)^{1/2} \\ &\quad - (12)^{-1}\pi^{3/2}(\ln z)^{-3/2} + O([\ln z]^{-7/2}), \\ -g_{3/2}(-z) &= 4(9\pi)^{-1/2}(\ln z)^{3/2} \\ &\quad + 6^{-1}\pi^{3/2}(\ln z)^{-1/2} + O([\ln z]^{-5/2}), \\ -g_{5/2}(-z) &= 8(15\pi^{1/2})^{-1}(\ln z)^{5/2} \\ &\quad + 3^{-1}\pi^{3/2}(\ln z)^{1/2} + O(1). \end{aligned} \tag{4}$$

The series of F was computed to the lowest order in reference 1. We shall now extend the calculation to the next order. The result is

$$\begin{aligned} -F(-z) &= (16/105)(11 - 2\ln 2)\pi^{-3/2}(\ln z)^{7/2} \\ &\quad - \tfrac{1}{3}\pi^{1/2}(4\ln 2 - 2)(\ln z)^{3/2} + O([\ln z]^{5/4}) \end{aligned} \tag{5}$$

To obtain this we follow a procedure used in reference 1, as shown in the Appendix.

[1] See T. D. Lee and C. N. Yang, Phys. Rev. 117, 12 (1960) and the papers referred to therein. We follow the notations used there: mass of particle $= \frac{1}{2}$, $\hbar = 1$, a = hard-sphere diameter, N = number of particles, Ω = volume of box, $\rho = N/\Omega$, J = spin of particles, P_F = maximum Fermi momentum for free particles $= [6\pi^2\rho/(2J+1)]^{1/2}$ $\beta = 1/kT$, $\lambda = (4\pi\beta)^{1/2}$, μ = fugacity.

3. THERMODYNAMIC FUNCTIONS NEAR ZERO TEMPERATURE

Using (1), (4), and (5), one can write down the pressure p as a function of μ and T:

$$p = \left\{\frac{2J+1}{15\pi^2}\mu^{5/2} - \frac{2J(2J+1)}{9\pi^3}a\mu^3 + \frac{4J(2J+1)}{\pi^4} \times \left[\frac{2J}{9} - \frac{11-2\ln 2}{105}\right]a^2\mu^{7/2} + \cdots\right\} + \{kT\}^2\left\{\frac{2J+1}{24}\mu^{1/2} - \frac{J(2J+1)}{18\pi}a\mu + \frac{J(2J+1)}{3\pi^2}\left[\frac{5J}{9} + \ln 2 - \tfrac{1}{2}\right]a^2\mu^{3/2} + \cdots\right\} + \cdots . \quad (6)$$

The other thermodynamic functions can be obtained from (6) by differentiation, using

$$dp = \rho\, d\mu + (S/\Omega)\, dT. \quad (7)$$

where S = entropy of the system. In particular one can obtain the entropy S in terms of the energy E and the variables N and Ω. The result is

$$\frac{S}{N} = \left(\frac{24}{2J+1}\right)^{1/2}\frac{k\pi^2}{2}(\Delta\epsilon P_F^{-5})^{1/2}\left[1 - \frac{2}{3}\frac{J}{\pi}(P_F a) + \left(\frac{22}{3}J - \frac{97}{21} + \frac{26\ln 2}{5}\right)\frac{J}{\pi^2}(P_F a)^2 + \frac{24\,\Delta\epsilon}{(2J+1)P_F^5}\left\{-\frac{\pi^2}{8} + \frac{5}{8}J\pi P_F a + \left(-\frac{455}{54}J + \frac{3079}{840} - \frac{52}{15}\ln 2\right)J(P_F a)^2\right\} + \cdots\right], \quad (8)$$

where P_F is defined in footnote 1 and

$$\Delta\epsilon = (E - E_{\text{ground}})/\Omega,$$

$$\frac{1}{\Omega}E_{\text{ground}} = \frac{2J+1}{10\pi^2}P_F^5\left\{1 + \frac{20J}{9\pi}aP_F + \left(\frac{88-16\ln 2}{21}\right)\frac{J}{\pi^2}(P_F a)^2 + \cdots\right\}. \quad (9)$$

All thermodynamic functions can also be obtained from (8) using

$$d\left(\frac{S}{N}\right) = \frac{1}{T}d\left(\frac{E}{N}\right) + \frac{p}{T}d\left(\frac{\Omega}{N}\right). \quad (10)$$

Now the density of energy levels is related to the entropy by

$$\text{density of energy levels} = e^{S/k}. \quad (11)$$

Thus (8) and (11) show that to the first order in a, the density of levels is smaller for a dilute hard sphere Fermi gas than for a free Fermi gas at the same density, but the next order correction goes in the opposite direction.

ACKNOWLEDGMENTS

The author wishes to thank Professor C. N. Yang for his suggestion of this problem, and Professor T. H. Berlin for his interest and for many discussions during the progress of this work.

APPENDIX. REPRESENTING F IN TERMS OF AN INTEGRAL

$$-F(-z) = \frac{8}{\pi^{3/2}}\int_0^\infty \frac{z}{z + \exp(X^2+u+v)} \times \frac{z}{z+\exp(Y^2+u)}\frac{z}{z+\exp(Z^2+v)} \times dX\,dY\,dZ\,du\,dv. \quad (A1)$$

Let

$$G(u) = \int_0^\infty \frac{z}{z+\exp(Y^2+u)}dY = \tfrac{1}{2}\int_u^\infty \frac{ze^{-\alpha}}{1+ze^{-\alpha}}\frac{d\alpha}{(\alpha-u)^{1/2}}, \quad (A2)$$

with $\alpha = u + y^2$. Then

$$-F(-z) = \frac{8}{\pi^{3/2}}\iint_0^\infty du\,dv G(u)G(v)G(u+v). \quad (A3)$$

Observing that $G(u)$ is small for $u > \ln z$ and large z, we split $G(u)$ into two parts

$$G(u) = G_0(u) + H(u - \ln z), \quad (A4)$$

where (See Fig. 1)

$$G_0(u) = \begin{cases}\tfrac{1}{2}\int_u^{\ln z}\frac{d\alpha}{(\alpha-u)^{1/2}} = (\ln z - u)^{1/2} & u \le \ln z \\ 0 & u > \ln z\end{cases} \quad (A5)$$

and (See Fig. 2)

$$H(x) = H(u - \ln z) = \begin{cases}\frac{1}{2}\int_x^\infty \frac{d\Delta}{1+e^\Delta}\frac{1}{(\Delta-x)^{1/2}} & x \ge 0 \\ \frac{1}{2}\int_0^{-x}\frac{d\Delta}{1+e^\Delta}\left(\frac{1}{(\Delta-x)^{1/2}} - \frac{1}{(-x-\Delta)^{1/2}}\right) + \int_{-x}^\infty \frac{1/2\,d\Delta}{(1+e^\Delta)}\frac{1}{(\Delta-x)^{1/2}} & x < 0\end{cases} \quad (A6)$$

The behavior of $H(x)$ as $x \to \pm\infty$ is

$$H(x) \to \begin{cases} \frac{1}{2} e^{-x} \int_0^\infty e^{-\Delta'} d\Delta' \frac{1}{\sqrt{\Delta'}} \\ \quad = \frac{\sqrt{\pi}}{2} e^{-x} \qquad x \to +\infty \\ \left(-\frac{1}{2} \int_0^\infty \frac{\Delta\, d\Delta}{1 + e^{\Delta}}\right) \frac{1}{(-x)^{3/2}} \\ \quad = -\frac{\pi^2}{24} \frac{1}{(-x)^{3/2}} \quad x \to -\infty . \end{cases} \tag{A7}$$

The function $-F(-z)$ can now be written as a sum of eight integrals, each having as an integrand a product of three G_0's or two G_0's and one H, or two H's and one G_0, or three H's. The integral containing three G_0's is called I_{000} which gives rise to the lowest order term and is given in reference 1. The sum of the integrals containing more than one H in their integrands is, except for a constant factor $8/\pi^{3/2}$

$$\begin{aligned} &I_{110} + I_{011} + I_{101} + I_{111} \\ &\quad = \iint_0^\infty du\, dv[H(u - \ln z)H(v - \ln z)G_0(u + v) \\ &\quad + H(u - \ln z)G_0(v)H(u + v - \ln z) \\ &\quad + G_0(u)H(v - \ln z)H(u + v - \ln z) \\ &\quad + H(u - \ln z)H(v - \ln z)H(u + v - \ln z)]. \end{aligned}$$

Since H is a function which is bounded and absolutely integrable from $-\infty$ to $+\infty$ and G_0 is less than $(\ln z)^{1/2}$ the sum of the above four integrals is of an order not more than that of $(\ln z)^{1/2}$, i.e.,

$$I_{110} + I_{011} + I_{101} + I_{111} = O([\ln z]^{1/2}). \tag{A8}$$

The sum of the remaining three integrals is

$$A = I_{001} + I_{100} + I_{010}$$

where

$$I_{001} = \iint_0^{\ln z} du\, dv(\ln z - u)^{1/2} \times (\ln z - v)^{1/2} H(u + v - \ln z).$$

With the substitution $u = u' + v'$ and $v = u' - v'$,

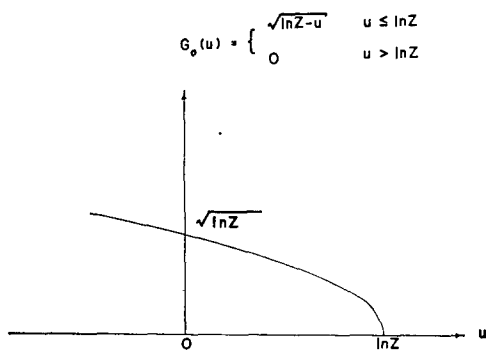

FIG. 1. Schematic graph of $G_0(u)$.

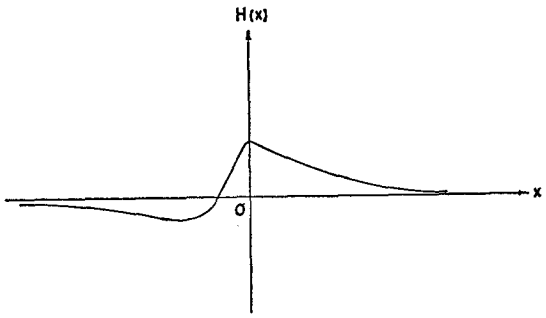

FIG. 2. Schematic graph of $H(x)$.

the v' integration becomes elementary. After performing this integration, we obtain,

$$\begin{aligned} I_{001} &= \int_0^{\ln z} dx \Big\{ H(-x) \Big[\left(\frac{\ln z + x}{2}\right)^2 \sin^{-1} \frac{\ln z - x}{\ln z + x} \\ &\quad + \frac{\ln z - x}{2} (x \ln z)^{1/2} \\ &\quad + H(x) \Big[\pi/2 \left(\frac{\ln z - x}{2}\right)^2 \Big] \Big\} \\ &= \int_0^{\ln z} \frac{\pi}{8} (\ln z)^2 [H(x) + H(-x)]\, dx \\ &\quad + \int_0^{\ln z} dx H(-x) \Big[\left(\frac{\ln z + x}{2}\right)^2 \sin^{-1} \frac{\ln z - x}{\ln z + x} \\ &\quad - \frac{\pi}{8} (\ln z)^2 + \frac{(\ln z)^{3/2}}{2} \sqrt{x} - \frac{(\ln z)^{1/2}}{2} \sqrt{x^3} \Big] \\ &\quad + \int_0^{\ln z} dx H(x) \Big[\frac{\pi}{8} (-2x \ln z + x^2) \Big]. \end{aligned}$$

We write the three integrals on the right as J_1, J_2, and J_3. Now, for any n positive,

$$\int_0^\infty |H(x)|\, x^n\, dx$$

is convergent. Hence,

$$\begin{aligned} |J_3| &< \ln z \int_0^\infty |H(x)| \frac{\pi}{4} x\, dx \\ &\quad + \frac{\pi}{8} \int_0^\infty |H(x)|\, x^2\, dx = O(\ln z) \end{aligned} \tag{A9}$$

$$J_1 = \frac{\pi}{8} (\ln z)^2 \Big[\int_0^\infty - \int_{\ln z}^\infty \Big] [H(x) + H(-x)]\, dx.$$

The $\int_\vartheta^\infty$ vanishes, as can be easily checked by substituting (A6) into the integrand and interchanging the order of integration. Furthermore, $H(x)$ behaves like e^{-x} for large x. Therefore for large z,

$$J_1 = -\frac{\pi}{8} (\ln z)^2 \int_{\ln z}^\infty H(-x)\, dx + O\left(\frac{[\ln z]^2}{z}\right)$$

Using the asymptotic formula (A7), one obtains

$$J_1 = (\pi^3/96)(\ln z)^{3/2} + O([\ln z]^{1/2}).$$

To evaluate J_2, we first put $x = y\,[\ln z]$ and obtain

$$J_2 = (\ln z)^3 \int_0^1 dy H(-y \ln z)\left[\left(\frac{1+y}{2}\right)^2 \sin^{-1}\left(\frac{1-y}{1+y}\right) - \frac{\pi}{8} + \frac{y^{1/2}}{2} - \frac{1}{2}y^{3/2}\right]. \quad \text{(A10)}$$

Put $\epsilon = (\ln z)^{-7/8}$ and evaluate $\int_0^1$ as $\int_0^\epsilon + \int_\epsilon^1$ For $\int_0^\epsilon$, the integrand is in absolute value less than $k\epsilon$ where k is a numerical constant. This is because H is bounded and the square bracket varies linearly as y plus higher order terms. Thus

$$\left|\int_0^\epsilon\right| < (\ln z)^3 \tfrac{1}{2}k\epsilon^2 = O([\ln z]^{5/4}).$$

For $\int_\epsilon^\infty$, as $z \to +\infty$ we replace $H(-y \ln z)$ by its asymptotic expression, since $\epsilon \ln z = (\ln z)^{1/8} \to +\infty$. Therefore,

$$\begin{aligned} J_2 &= (\ln z)^{3/2} \int_\epsilon^1 dy\left(-\frac{\pi^2}{24}\right)y^{-3/2}\left[\left(\frac{1+y}{2}\right)^2 \sin^{-1}\left(\frac{1-y}{1+y}\right) - \frac{\pi}{8} + \tfrac{1}{2}y^{1/2} - \tfrac{1}{2}y^{3/2}\right] \\ &\quad \times [1 + O([y \ln z]^{-1})] + o([\ln z]^{5/4}) \\ &= (\ln z)^{3/2}\left(-\frac{\pi^2}{24}\right)\left[\frac{4}{3}\ln 2 - \frac{4}{3} + \frac{\pi}{4}\right] \\ &\quad + o([\ln z]^{5/4}). \end{aligned} \quad \text{(A11)}$$

Thus,

$$\begin{aligned} I_{001} &= J_1 + J_2 + J_3 \\ &= (\ln z)^3(\pi^2/18)(1 - \ln z) + 0([\ln z]^{5/4}). \end{aligned} \quad \text{(A12)}$$

Now,

$$\begin{aligned} I_{100} + I_{010} &= 2\int_0^{\ln z} (\ln z - u)^{1/2}\,du \\ &\quad \times \int_0^{\ln z - u} dv(\ln z - u - v)^{1/2} H(v - \ln z) \\ &= \int_0^{\ln z} dx H(-x)\left[-\left(\frac{\ln z - x}{2}\right)^2 \cosh^{-1}\left(\frac{\ln z + x}{\ln z - x}\right) \right. \\ &\quad \left. + \tfrac{1}{2}(\ln z)^{3/2}x^{1/2} + \tfrac{1}{2}(\ln z)^{1/2}x^{3/2}\right]. \end{aligned}$$

Making the transformation $x = y \ln z$, we can evaluate the resulting integral in the same manner as the evaluation of (A10), and obtain

$$\begin{aligned} I_{100} + I_{010} &= (\ln z)^{3/2}\int_0^1 dy\left(-\frac{\pi^2}{24}\right)y^{-3/2} \\ &\quad \times \left[-\left(\frac{1-y}{2}\right)^2 \cosh^{-1}\left(\frac{1+y}{1-y}\right) + \tfrac{1}{2}y^{1/2} + \tfrac{1}{2}y^{3/2}\right] \\ &\quad + o([\ln z)^{5/4}) \\ &= (\ln z)^{3/2}(-\pi^2/18)(2\ln 2 - \tfrac{1}{2}) + o([\ln z]^{5/4}). \end{aligned} \quad \text{(A13)}$$

Combining (A8), (A12), and (A13), we obtain (5).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Structure of a Bloch Wall

M. W. MULLER AND S. DAWSON*
Varian Associates,
Palo Alto, California
(Received February 1, 1962)

The standard form of a domain wall in a ferromagnetic medium is the result of a one-dimensional variational principle and is valid only for an unbounded medium. In a bounded ferromagnetic body two- or three-dimensional dipolar fields arise wherever a Bloch wall meets a surface of the body, and the form of the wall is distorted by these fields. If the fields and the corresponding distortions are small, they can be calculated approximately. We carry out the calculation for 180° walls in a ferromagnetic film with uniaxial anisotropy, and with the easy axis of magnetization either parallel or perpendicular to the surfaces of the film. Numerical results are not obtained but formulas are presented ready for evaluation.

I. INTRODUCTION

At temperatures well below the Curie temperature a macroscopically demagnetized ferromagnetic body consists of uniformly magnetized domains separated by narrow walls in which the magnetization turns rapidly but quasi-continuously[1] between the directions of the adjacent domains.

The domain structure is especially simple in a

* Present address: Harvard University, Cambridge, Massachusetts.

[1] F. Bloch, Z. Physik **76**, 513 (1932).

To evaluate J_2, we first put $x = y\,[\ln z]$ and obtain

$$J_2 = (\ln z)^3 \int_0^1 dy H(-y \ln z)\left[\left(\frac{1+y}{2}\right)^2 \sin^{-1}\left(\frac{1-y}{1+y}\right) - \frac{\pi}{8} + \frac{y^{1/2}}{2} - \frac{1}{2}y^{3/2}\right]. \quad \text{(A10)}$$

Put $\epsilon = (\ln z)^{-7/8}$ and evaluate $\int_0^1$ as $\int_0^\epsilon + \int_\epsilon^1$ For $\int_0^\epsilon$, the integrand is in absolute value less than $k\epsilon$ where k is a numerical constant. This is because H is bounded and the square bracket varies linearly as y plus higher order terms. Thus

$$\left|\int_0^\epsilon\right| < (\ln z)^3 \tfrac{1}{2}k\epsilon^2 = O([\ln z]^{5/4}).$$

For $\int_\epsilon^\infty$, as $z \to +\infty$ we replace $H(-y \ln z)$ by its asymptotic expression, since $\epsilon \ln z = (\ln z)^{1/8} \to +\infty$. Therefore,

$$\begin{aligned} J_2 &= (\ln z)^{3/2} \int_\epsilon^1 dy\left(-\frac{\pi^2}{24}\right)y^{-3/2}\left[\left(\frac{1+y}{2}\right)^2 \sin^{-1}\left(\frac{1-y}{1+y}\right) - \frac{\pi}{8} + \tfrac{1}{2}y^{1/2} - \tfrac{1}{2}y^{3/2}\right] \\ &\quad \times [1 + O([y \ln z]^{-1})] + o([\ln z]^{5/4}) \\ &= (\ln z)^{3/2}\left(-\frac{\pi^2}{24}\right)\left[\frac{4}{3}\ln 2 - \frac{4}{3} + \frac{\pi}{4}\right] \\ &\quad + o([\ln z]^{5/4}). \end{aligned} \quad \text{(A11)}$$

Thus,

$$\begin{aligned} I_{001} &= J_1 + J_2 + J_3 \\ &= (\ln z)^3(\pi^2/18)(1 - \ln z) + 0([\ln z]^{5/4}). \end{aligned} \quad \text{(A12)}$$

Now,

$$\begin{aligned} I_{100} + I_{010} &= 2\int_0^{\ln z} (\ln z - u)^{1/2}\,du \\ &\quad \times \int_0^{\ln z - u} dv(\ln z - u - v)^{1/2} H(v - \ln z) \\ &= \int_0^{\ln z} dx H(-x)\left[-\left(\frac{\ln z - x}{2}\right)^2 \cosh^{-1}\left(\frac{\ln z + x}{\ln z - x}\right) \right. \\ &\quad \left. + \tfrac{1}{2}(\ln z)^{3/2}x^{1/2} + \tfrac{1}{2}(\ln z)^{1/2}x^{3/2}\right]. \end{aligned}$$

Making the transformation $x = y \ln z$, we can evaluate the resulting integral in the same manner as the evaluation of (A10), and obtain

$$\begin{aligned} I_{100} + I_{010} &= (\ln z)^{3/2}\int_0^1 dy\left(-\frac{\pi^2}{24}\right)y^{-3/2} \\ &\quad \times \left[-\left(\frac{1-y}{2}\right)^2 \cosh^{-1}\left(\frac{1+y}{1-y}\right) + \tfrac{1}{2}y^{1/2} + \tfrac{1}{2}y^{3/2}\right] \\ &\quad + o([\ln z)^{5/4}) \\ &= (\ln z)^{3/2}(-\pi^2/18)(2\ln 2 - \tfrac{1}{2}) + o([\ln z]^{5/4}). \end{aligned} \quad \text{(A13)}$$

Combining (A8), (A12), and (A13), we obtain (5).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Structure of a Bloch Wall

M. W. MULLER AND S. DAWSON*
Varian Associates,
Palo Alto, California
(Received February 1, 1962)

The standard form of a domain wall in a ferromagnetic medium is the result of a one-dimensional variational principle and is valid only for an unbounded medium. In a bounded ferromagnetic body two- or three-dimensional dipolar fields arise wherever a Bloch wall meets a surface of the body, and the form of the wall is distorted by these fields. If the fields and the corresponding distortions are small, they can be calculated approximately. We carry out the calculation for 180° walls in a ferromagnetic film with uniaxial anisotropy, and with the easy axis of magnetization either parallel or perpendicular to the surfaces of the film. Numerical results are not obtained but formulas are presented ready for evaluation.

I. INTRODUCTION

At temperatures well below the Curie temperature a macroscopically demagnetized ferromagnetic body consists of uniformly magnetized domains separated by narrow walls in which the magnetization turns rapidly but quasi-continuously[1] between the directions of the adjacent domains.

The domain structure is especially simple in a

* Present address: Harvard University, Cambridge, Massachusetts.

[1] F. Bloch, Z. Physik **76**, 513 (1932).

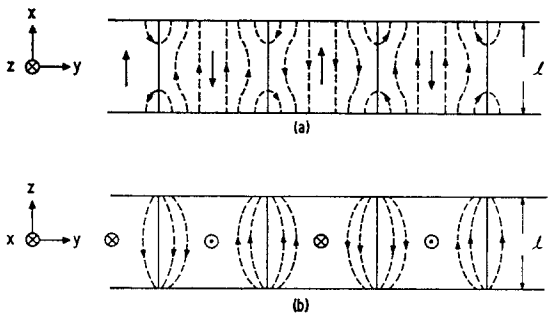

FIG. 1. Ideal domain structure in ferromagnetic films with uniaxial anisotropy, easy direction along x. Full arrows: direction of magnetization; dashed lines: dipolar field.

parallel-plane monocrystalline slab or film with uniaxial anisotropy, and with the faces cut either parallel or perpendicular to the easy axis of magnetization. Schematic pictures of the domain structure for these configurations are shown in Figs. 1(a) and 1(b). The structures are one dimensional, that is to say, uniform along z in Fig. 1(a) and uniform along x in Fig. 1(b).

The form of the magnetization in the domain walls was given by Landau and Lifshitz[2] for an unbounded medium. This result is valid in the geometry we are considering at distances from the surfaces which are large compared with a wall thickness. Near the surfaces the discontinuity of the normal component of the magnetization gives rise to a dipolar magnetic field which can distort the form of the wall. This dipolar field is shown dashed in Fig. 1.

The distortion of the wall may be large and may even change the qualitative aspect of the domain structure. Thus, in the configuration of Fig. 1(a), if the saturation magnetization $4\pi M$ is large compared with the anisotropy field $2K_1/M$, the dipolar field gives rise to domains of closure. Similarly, in the configuration of Fig. 1(b), if the thickness $2l$ of the film is small compared with the domain-wall thickness, the walls assume the configuration denoted as Néel walls.[3] In the present paper we do not consider such large deviations from the one-dimensional geometry. The calculation is confined to distortions of the walls under conditions when they remain small.

II. THEORY

A. Physics

The equilibrium distribution of the magnetization satisfies the equation

$$\mathbf{M}\times\mathbf{H}_{\text{eff}} = 0 \tag{1}$$

The effective field $\mathbf{H}_{\text{eff}}$ consists of contributions due to exchange, anisotropy, and dipolar forces (we neglect the effect of the magnetostrictive energy). Since we are assuming that the temperature is well below the Curie temperature, the magnitude of M is constant, and Eq. (1) describes the behavior of the direction cosines of $\mathbf{M}$. The component equations for uniaxial anisotropy and for the geometry of Fig. 1 are[4,5]

$$\begin{aligned} &c(\beta\nabla^2\gamma - \gamma\nabla^2\beta) - M(\beta\Phi_x - \gamma\Phi_y) = 0\\ &c(\gamma\nabla^2\alpha - \alpha\nabla^2\gamma) + 2K\alpha\gamma - M(\gamma\Phi_x - \alpha\Phi_z) = 0\\ &c(\alpha\nabla^2\beta - \beta\nabla^2\alpha) - 2K\alpha\beta - M(\beta\Phi_z - \gamma\Phi_y) = 0, \end{aligned} \tag{2}$$

where α, β, γ are the direction cosines of $\mathbf{M}$, K is the anisotropy constant, c is the exchange constant, and Φ is the magnetostatic potential.

If we focus our attention on a single domain wall (the interaction between adjacent domain walls is very small) then in zeroth approximation $\Phi = \beta = 0$, $\nabla^2 = d^2/dy^2$ and

$$c\left(\gamma_0\frac{d^2\alpha_0}{dy^2} - \alpha_0\frac{d^2\gamma_0}{dy^2}\right) + 2K\alpha_0\gamma_0 = 0. \tag{3}$$

Letting $\alpha_0 = \sin\theta_0$, $\gamma_0 = \cos\theta_0$ and using the boundary condition $\theta_0 = \pm\pi/2$ for $y \to \pm\infty$ we obtain the well-known one-dimensional solution[2]

$$\alpha_0 = \tanh(y/\delta), \qquad \gamma_0 = \operatorname{sech}(y/\delta), \tag{4}$$

where $\delta = (c/2K)^{1/2}$ is the nomial wall thickness.

In computing the next-higher approximation we will determine the magnetostatic potential Φ from the surface "charges." These are

$$\begin{aligned} \sigma &= \pm 4\pi M\sin\theta_0\\ &= \pm 4\pi M\tanh(y/\delta) \quad \text{at} \quad x = \pm l \end{aligned} \tag{5a}$$

$$\begin{aligned} \sigma &= \pm 4\pi M\cos\theta_0\\ &= \pm 4\pi M\operatorname{sech}(y/\delta) \quad \text{at} \quad z = \pm l. \end{aligned} \tag{5b}$$

[Here and throughout this paper equations labeled (a) and (b) refer to the configurations of Figs. 1(a) and 1(b), respectively.]

We now write $\theta_1 = \theta_0 + \theta$ and linearize Eqs. (2) by dropping terms of second order in θ, β, and Φ. After some manipulation this yields

$$\begin{aligned} &\{\nabla^2 + (1/\delta^2)[2\operatorname{sech}^2(y/\delta) - 1]\}\theta\\ &\quad = (M/c)[\operatorname{sech}(y/\delta)\Phi_x - \tanh(y/\delta)\Phi_z]\\ &\{\nabla^2 + (1/\delta^2)[2\operatorname{sech}^2(y/\delta) - 1]\}\beta = (M/c)\Phi_y. \end{aligned} \tag{6}$$

[2] L. Landau and E. Lifshitz, Physik. Z. Sowjetunion **8**, 153 (1935).
[3] L. Néel, Compt. rend. **241**, 533 (1955).
[4] W. F. Brown, Jr., J. Appl. Phys. **30**, 625 (1959).
[5] M. W. Muller, Phys. Rev. **122**, 1485, (1961).

Each of the two problems we are considering is two dimensional, that is to say,

$$\partial_z = 0 \tag{7a}$$

or

$$\partial_x = 0. \tag{7b}$$

The boundary conditions are

$$\theta = \beta = 0 \quad \text{as} \quad y \to \pm\infty$$

and[4]

$$\theta_x = \beta_x = 0 \quad \text{at} \quad x = \pm l \tag{8a}$$

$$\theta_z = \beta_z = 0 \quad \text{at} \quad z = \pm l. \tag{8b}$$

The approximation we are using is valid only if the calculated corrections are small. In general, this will turn out to be true if (a) $K/M^2 \gg 1$ or (b) $1/\delta \gg 1$. In principle, the procedure could be iterated to compute second and higher approximations; in this paper we do not go beyond the first approximation.

B. Mathematical Treatment

We have to solve Laplace's equation in two dimensions for a plane-parallel region with the surface charges of Eqs. (5). The potential ϕ that is obtained then fixes the inhomogeneous terms of Eqs. (6). This procedure can be carried out explicitly by working with the Fourier-transformed equations, and performing the inversion back to coordinate space as a final step in the calculation.

For convenience we introduce dimensionless coordinates $\xi = x/\delta$, $\eta = y/\delta$, $\zeta = z/\delta$, and the dimensionless film thickness $2\lambda = 2l/\delta$.

The transform of the potential Φ

$$\phi(\xi, \kappa) = \int_{-\infty}^{\infty} e^{-i\kappa\eta}\Phi(\xi, \eta)\, d\eta \tag{9}$$

obeys the ordinary differential equation

$$d^2\phi/d\xi^2 - \kappa^2\phi = 0 \tag{10a}$$

$$d^2\phi/d\zeta^2 - \kappa^2\phi = 0, \tag{10b}$$

with the two-point boundary condition[6]

$$\left(\frac{d\phi}{d\xi}\right)_{\xi=\pm\lambda} = -4\pi M \int_{-\infty}^{\infty} e^{-i\kappa\eta} \tanh \eta\, d\eta$$

$$= +4\pi^2 iM \operatorname{csch}(\pi\kappa/2), \tag{11a}$$

$$\left(\frac{d\phi}{d\zeta}\right)_{\xi=\pm\lambda} = -4\pi M \int_{-\infty}^{\infty} e^{-i\kappa\eta} \operatorname{sech} d\eta$$

$$= -4\pi^2 M \operatorname{sech}(\pi\kappa/2). \tag{11b}$$

[6] *Tables of Integral Transforms*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol I, pp. 88, 30.

The solution is

$$\phi = \begin{cases} \dfrac{4\pi^2 iM}{\kappa} \dfrac{\operatorname{csch}(\pi\kappa/2)}{\cosh \lambda\kappa} \sinh \xi\kappa & \text{(12a)} \\ -\dfrac{4\pi^2 M}{\kappa} \dfrac{\operatorname{sech}(\pi\kappa/2)}{\cosh \lambda\kappa} \sinh \zeta\kappa & \text{(12b)} \end{cases}$$

and

$$\Phi = \begin{cases} 2\pi iM \displaystyle\int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{1}{\kappa} \frac{\operatorname{csch}(\pi\kappa/2)}{\cosh \lambda\kappa} \sinh \xi\kappa\, d\kappa & \text{(13a)} \\ -2\pi M \displaystyle\int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{1}{\kappa} \frac{\operatorname{sech}(\pi\kappa/2)}{\cosh \lambda\kappa} \sinh \zeta\kappa\, d\kappa & \text{(13b)} \end{cases}$$

$$\Phi_\xi = 2\pi iM \int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{\cosh \xi\kappa}{\cosh \lambda\kappa \sinh(\pi\kappa/2)}\, d\kappa \tag{14a}$$

$$\Phi_\zeta = -2\pi M \int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{\cosh \zeta\kappa}{\cosh \lambda\kappa \cosh(\pi\kappa/2)}\, d\kappa \tag{14b}$$

$$\Phi_\eta = \begin{cases} -2\pi M \displaystyle\int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{\sinh \xi\kappa}{\cosh \lambda\kappa \sinh(\pi\kappa/2)}\, d\kappa & \text{(15a)} \\ -2\pi iM \displaystyle\int_{-\infty}^{\infty} e^{i\kappa\eta} \frac{\sinh \zeta\kappa}{\cosh \lambda\kappa \cosh(\pi\kappa/2)}\, d\kappa. & \text{(15b)} \end{cases}$$

These are the sources for Eqs. (6).

It should be noted here that of these four fields, Φ_η in Eq. (15a) diverges logarithmically as $\lambda \to \infty$, $\xi \to \lambda$. This trouble arises from the idealization we have introduced in considering a wall between infinite domains. It does mean that we must avoid passing to this limit for configuration (a).

The Green's function for Eq. (6) satisfies

$$[\nabla^2 + (2 \operatorname{sech}^2 \eta - 1)]G = \begin{cases} \delta(\xi - \xi_s, \eta - \eta_s) & \text{(16a)} \\ \delta(\zeta - \zeta_s, \eta - \eta_s), & \text{(16b)} \end{cases}$$

with boundary conditions

$$\begin{aligned} G_\xi(\pm\lambda, \eta) &= 0 \\ G(\xi, \pm\infty) &= 0 \end{aligned} \tag{17a}$$

$$\begin{aligned} G_\zeta(\pm\lambda, \eta) &= 0 \\ G(\zeta, \pm\infty) &= 0. \end{aligned} \tag{17b}$$

The variables are separated by

$$G(\xi, \xi_s; \eta, \eta_s) = \sum_{m=0}^{\infty} \left\{ g_m^c(\eta, \eta_s) \cos\frac{m\pi\xi}{\lambda} \cos\frac{m\pi\xi_s}{\lambda} + g_m^s(\eta, \eta_s) \sin\frac{n\pi\xi}{\lambda} \sin\frac{n\pi\xi_s}{\lambda} \right\} \tag{18a}$$

[using the abbreviations $n \equiv \frac{1}{2}(2m + 1)$], which satisfies the first boundary condition. Substitution in Eq. (16) and integration over ξ and ξ_s yields

$$\left(\frac{d^2}{d\eta^2} + 2\,\mathrm{sech}\,\eta - \alpha_m^2\right) g_m^c = \frac{\delta(\eta - \eta_s)}{(1 + \delta_0^m)\lambda}, \tag{19}$$

where $\alpha_m^2 = 1 + (m\pi/\lambda)^2$ and δ_0^m is the Kronecker delta; and a similar equation for g_m^s.

The complete primitive of this ordinary differential equation is[7]

$$g_m^c = A_m e^{\alpha_m \eta}(-\alpha_m + \tanh\eta) + B_m e^{-\alpha_m\eta}(\alpha_m + \tanh\eta). \tag{20}$$

To satisfy the second boundary condition we must have

$$B = 0, \quad \eta < \eta_s; \qquad A = 0, \quad \eta > \eta_s.$$

The other two constants are obtained from

$$\begin{aligned} g_m^c(\eta_s)_- - g_m^c(\eta_s)_+ &= 0 \\ g_m^{c\prime}(\eta_s)_- - g_m^{c\prime}(\eta_s)_+ &= 1/(1 + \delta_0^m)\lambda, \end{aligned} \tag{21}$$

giving

$$\begin{aligned} g_0^c &= 0 \\ g_m^c &= (\lambda/2\pi^2 m^2 \alpha_m) e^{\alpha_m(\eta - \eta_s)} \\ &\quad \times (-\alpha_m + \tanh\eta)(\alpha_m + \tanh\eta_s) \\ &\quad \text{if} \quad \eta < \eta_s \end{aligned} \tag{22}$$

and the same expressions with η and η_s permuted if $\eta > \eta_s$.

Since these expressions will occur frequently, we introduce the abbreviation

$$h_m(\eta) = e^{\alpha_m\eta}(-\alpha_m + \tanh\eta)$$

so that

$$g_m^c = -\frac{\lambda}{2\pi^2 m^2 \alpha_m} \begin{cases} h_m(\eta)h_m(-\eta_s) & \eta < \eta_s \\ h_m(-\eta)h_m(\eta_s) & \eta > \eta_s. \end{cases} \tag{23}$$

Thus the solution of the inhomogeneous problem is

$$\begin{pmatrix}\theta\\ \beta\end{pmatrix} = \frac{M}{c}\int_{-\lambda}^{\lambda} d\xi_s \int_{-\infty}^{\infty} d\eta_s G(\xi, \xi_s; \eta, \eta_s) \times \begin{bmatrix} \mathrm{sech}\,\eta_s\, \partial\Phi(\xi_s, \eta_s)/\partial\eta_s \\ \partial\Phi(\xi_s, \eta_s)/\partial\eta_s \end{bmatrix} \tag{24a}$$

$$\begin{pmatrix}\theta\\ \beta\end{pmatrix} = \frac{M}{c}\int_{-\lambda}^{\lambda} d\zeta_s \int_{-\infty}^{\infty} d\eta_s G(\zeta, \zeta_s; \eta, \eta_s) \times \begin{bmatrix} -\tanh\eta_s\, \partial\Phi(\zeta_s, \eta_s)/\partial\eta_s \\ \partial\Phi(\zeta_s, \eta_s)/\partial\eta_s \end{bmatrix}. \tag{24b}$$

Because of the symmetry of the fields, the parts of the Green's function that make a nonvanishing contribution are the g_m^c for θ, the g_m^s for β.

[7] J. M. Winter, Phys. Rev. **124**, 452, (1961).

Performing the integrations over ξ_s and ζ_s yields[8]

$$\theta = -\frac{M^2}{\pi c}\int_{-\infty}^{\infty} d\kappa \sum_{m=1}^{\infty}\left[\frac{1}{i\kappa + m\pi/\lambda} + \frac{1}{i\kappa - m\pi/\lambda}\right] \times \frac{\lambda}{m^2\alpha_m}\cos\frac{m\pi\xi}{\lambda}\frac{\tanh\kappa\lambda}{\sinh(\pi\kappa/2)} I_1(\eta) \tag{25a}$$

$$\beta = +\frac{M^2}{\pi c}\int_{-\infty}^{\infty} d\kappa \sum_{m=0}^{\infty}\left[\frac{1}{i\kappa + n\pi/\lambda} - \frac{1}{i\kappa - n\pi/\lambda}\right] \times \frac{\lambda}{n^2\alpha_n}\cos\frac{n\pi\xi}{\lambda}\frac{\tanh\kappa\lambda}{\sinh(\pi\kappa/2)} I_2(\eta)$$

$$\theta = \frac{iM^2}{\pi c}\int_{-\infty}^{\infty} d\kappa \sum_{m=1}^{\infty}\left[\frac{1}{i\kappa + m\pi/\lambda} + \frac{1}{i\kappa - m\pi/\lambda}\right] \times \frac{\lambda}{m^2\alpha_m}\cos\frac{m\pi\zeta}{\lambda}\frac{\tanh\kappa\lambda}{\cosh(\pi\kappa/2)} I_3(\eta) \tag{25b}$$

$$\beta = -\frac{iM^2}{\pi c}\int_{-\infty}^{\infty} d\kappa \sum_{m=0}^{\infty}\left[\frac{1}{i\kappa + n\pi/\lambda} - \frac{1}{i\kappa - n\pi/\lambda}\right] \times \frac{\lambda}{n^2\alpha_n}\sin\frac{n\pi\zeta}{\lambda}\frac{\tanh\kappa\lambda}{\cosh(\pi\kappa/2)} I_2(\eta),$$

where

$$I_1(\eta) = -\left[h_m(-\eta)\int_{-\infty}^{\eta} e^{i\kappa\eta_s} h_m(\eta_s)\,\mathrm{sech}\,\eta_s\, d\eta_s + h_m(\eta)\int_{\eta}^{\infty} e^{i\kappa\eta_s} h_m(-\eta_s)\,\mathrm{sech}\,\eta_s\, d\eta_s\right]$$

$$I_2(\eta) = -\left[h_n(-\eta)\int_{-\infty}^{\eta} e^{i\kappa\eta_s} h_n(\eta_s)\, d\eta_s + h_n(\eta)\int_{\eta}^{\infty} e^{i\kappa\eta_s} h_n(-\eta_s)\, d\eta_s\right]$$

$$I_3(\eta) = -\left[h_m(-\eta)\int_{-\infty}^{\eta} e^{i\kappa\eta_s} h_m(\eta_s)\tanh\eta_s\, d\eta_s + h_m(\eta)\int_{\eta}^{\infty} e^{i\kappa\eta_s} h_m(-\eta_s)\tanh\eta_s\, d\eta_s\right].$$

The integrations over η_s are carried out in Appendix I. The results are

$$I_1(\eta) = 2e^{i\kappa\eta}\left\{-\alpha_m\,\mathrm{sech}\,\eta + i\kappa e^{-\eta}\sum_{j=0}^{\infty}(-e^{-2\eta})^j \times \left[\frac{\alpha_m + \tanh\eta}{i\kappa + \alpha_m - 1 - 2j} + \frac{-\alpha_m + \tanh\eta}{-i\kappa + \alpha_m + 1 + 2j}\right]\right\}$$

$$I_2(\eta) = e^{i\kappa\eta}\left\{-(\alpha_n + 1)\frac{\alpha_n + \tanh\eta}{i\kappa + \alpha_n} + (\alpha_n - 1)\frac{-\alpha_n + \tanh\eta}{-i\kappa + \alpha_n} + 2\sum_{j=0}^{\infty}(-e^{-2\eta})^j \times \left[\frac{\alpha_n + \tanh\eta}{i\kappa + \alpha_n - 2j} + \frac{-\alpha_n + \tanh\eta}{-i\kappa + \alpha_n + 2j}\right]\right\}$$

[8] We use $\alpha_n^2 = 1 + (n\pi/\lambda)^2 = 1 + (2m+1)^2\pi^2/4\lambda^2$, etc.

$$I_3(\eta) = e^{i\kappa\eta}\Big\{-2\alpha_m \tanh\eta + (1 - i\kappa)\Big[\frac{\alpha_m + \tanh\eta}{i\kappa + \alpha_m} + \frac{(-\alpha_m + \tanh\eta)}{-i\kappa + \alpha_m}\Big] + 2i\kappa \sum_{j=0}^{\infty} (-e^{-2\eta})^j \times \Big[\frac{\alpha_m + \tanh\eta}{i\kappa + \alpha_m - 2j} + \frac{-\alpha_m + \tanh\eta}{-i\kappa + \alpha_m + 2j}\Big]\Big\}. \quad (26)$$

The summations which appear in Eqs. (26) are related to the hypergeometric function.[9]

It is now a simple but tedious task to perform the Fourier inversion indicated in Eqs. (25). For this purpose we note that the integrands in these equations vanish exponentially in the upper half-plane if $|\kappa| \to \infty$ and have no singularities on the real axis. Thus each integral over κ is just the sum of the residues in the upper half-plane. The only special caution that need be observed is to note that the expression for β contains sets of second-order poles at $\kappa = i\,(2m + 1)/2\lambda$.

We will write down only one of the resulting expressions as an example. The integrand for θ, Eq. (25a), has poles in the upper half-plane at

$$\kappa = i2k \qquad k = 1, 2, \cdots \infty$$

$$\kappa = i\frac{2l+1}{2\lambda}\pi \qquad l = 0, 1, \cdots \infty$$

$$\kappa = i(\alpha_m - 1 - 2j) \qquad j = 0, 1, \cdots j_{\max},$$

where $j_{\max}$ is the largest integer $< \frac{1}{2}(\alpha_m - 1)$.

Evaluating all the residues we obtain

$$\theta = -\frac{2M^2\lambda}{c\pi} \sum_{m=1}^{\infty} (a_m + b_m + c_m) \cos\frac{m\pi\xi}{\lambda}, \quad (27)$$

where

$$a_m = \sum_{k=1}^{\infty} 16(-1)^k \exp[-2k\eta] \frac{\tan 2k\lambda}{(m\pi/\lambda)^2 - 4k^2} \times \Big[\alpha_m \operatorname{sech}\eta + 2ke^{-\eta} \sum_{j=0}^{\infty} (-1)^j \exp[-2j\eta] \times \Big(\frac{\alpha_m \tanh\eta}{\alpha_m - 1 - 2(j+k)} + \frac{-\alpha_m + \tanh\eta}{\alpha_m + 1 + 2(j+k)}\Big)\Big]$$

$$b_m = \sum_{l=0}^{\infty} 2\exp[-(2l+1)\pi\eta/2\lambda] \csc[(2l+1)\pi^2/4\lambda] \times \frac{(2l+1)\pi}{m^2 - \frac{1}{4}(2l+1)^2\pi^2}\Big[-\alpha_m \operatorname{sech}\eta - \frac{2l+1}{2\lambda}\pi e^{-\eta} \times \sum_{j=0}^{\infty} (-1)^j \exp[-2j\eta]\Big[\frac{\alpha_m + \tanh\eta}{\alpha_m - 1 - 2\Big(j + \frac{2l+1}{2\lambda}\pi\Big)} + \frac{-\alpha_m + \tanh\eta}{\alpha_m + 1 + 2\Big(j + \frac{2l+1}{2\lambda}\pi\Big)}\Big]\Big]$$

$$c_m = -2\pi \sum_{j=0}^{j_{\max}} (-1)^j \frac{(\alpha_m - 1 - 2j)^2}{(m\pi/\lambda)^2 - (\alpha_m - 1 - 2j)^2} \times \tan[(\alpha_m - 1 - 2j)\lambda] \csc[(\alpha_m - 1 - 2j)\pi/2].$$

These expression and the corresponding formulas for the other three functions of Eqs. (25) are too lengthy for evaluation without the aid of a computer. Since we have no immediate requirement for numerical results we have not carried out any computer evaluation. Such an evaluation would be relatively simple. The analytical results have been recorded here for this purpose, and because the analytical procedure may be useful for other problems in this area.

APPENDIX I.

EVALUATION OF THE INTEGRALS I_1, I_2, I_3 FOR EQUATIONS (25)

To evaluate

$$I_1 = e^{-\alpha_m\eta}(\alpha_m + \tanh\eta) \int_{-\infty}^{\eta} e^{(i\kappa+\alpha_m)\eta_s} \times [-\alpha_m + \tanh\eta_s]\, d\eta_s + e^{\alpha_m\eta}(-\alpha_m + \tanh\eta) \int_{\eta}^{\infty} e^{(i\kappa-\alpha_m)\eta_s} \times [\alpha_m + \tanh\eta_s]\, d\eta_s$$

$$= e^{-\alpha_m\eta}(\alpha_m + \tanh\eta) \int_{-\infty}^{\eta} e^{(i\kappa-\alpha_m)\eta_s} \times \Big[-\alpha_m \operatorname{sech}\eta_s - \frac{d}{d\eta_s}\operatorname{sech}\eta_s\Big] d\eta_s + e^{\alpha_m\eta}(-\alpha_m + \tanh\eta) \int_{\eta}^{\infty} e^{(i\kappa-\alpha_m)\eta_s} \times \Big[\alpha_m \operatorname{sech}\eta_s - \frac{d}{d\eta_s}\operatorname{sech}\eta_s\Big] d\eta_s$$

$$= -2\alpha_m e^{i\kappa\eta} \operatorname{sech}\eta + i\kappa\Big[e^{-\alpha_m\eta}(\alpha_m + \tanh\eta) \times \int_{-\infty}^{\eta} e^{(i\kappa-\alpha_m)\eta_s} \operatorname{sech}\eta_s\, d\eta_s + e^{\alpha_m\eta}(-\alpha_m + \tanh\eta) \times \int_{\eta}^{\infty} e^{(i\kappa-\alpha_m)\eta_s} \operatorname{sech}\eta_s\, d\eta_s\Big].$$

[9] *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, pp. 27–30, especially Eq. (1.11.10).

Use the substitutions $\eta_s = -t + \eta$ in the first integral and $\eta_s = t + \eta$ in the second integral. Then the expression in square brackets becomes[10]

$$
\begin{aligned}
&2\Big[e^{(i\kappa-1)\eta}(\alpha_m + \tanh\eta)\int_0^\infty \frac{\exp[-(i\kappa+\alpha_m-1)t]}{\exp[2(t-\eta)]+1}\,dt \\
&\quad + e^{(i\kappa-1)\eta}(-\alpha_m + \tanh\eta) \\
&\quad \times \int_0^\infty \frac{\exp[(i\kappa-\alpha_m-1)t]}{\exp[-2(t+\eta)]+1}\Big]\,dt \\
&= 2e^{i\kappa\eta}\Big[e^{-\eta}(\alpha_m+\tanh\eta)\sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{(i\kappa+\alpha_m-1)-2j} \\
&\quad + e^{-\eta}(-\alpha_m+\tanh\eta) \\
&\quad \times \sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{-i\kappa+\alpha_m+1+2j}\Big] \quad \eta>0,
\end{aligned}
$$

similarly the integral needed for β

$$
\begin{aligned}
I_2 &= e^{-\alpha_n\eta}(\alpha_n+\tanh\eta) \\
&\quad \times \int_{-\infty}^{\eta} e^{(i\kappa+\alpha_n)\eta_s}(-\alpha_n+\tanh\eta_s)\,d\eta_s \\
&\quad + e^{+\alpha_n\eta}(\alpha_n+\tanh\eta) \\
&\quad \times \int_{\eta}^{\infty} e^{(i\kappa+\alpha_n)\eta_s}(\alpha_n+\tanh\eta_s)\,d\eta_s \\
&= e^{i\kappa\eta}\Big\{-(\alpha_n+1)\frac{\alpha_n+\tanh\eta}{i\kappa+\alpha_n} \\
&\quad - (\alpha_n-1)\frac{-\alpha_n+\tanh\eta}{i\kappa-\alpha_n} \\
&\quad + 2\Big[(\alpha_n+\tanh\eta)\sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{i\kappa+\alpha_n-2j} \\
&\quad + (-\alpha_n+\tanh\eta)\sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{-i\kappa+\alpha_n+2j}\Big]\Big\}
\end{aligned}
$$

and the integral needed for θ, case (b)

$$
\begin{aligned}
I_3 &= e^{-\alpha_m\eta}(\alpha_m+\tanh\eta) \\
&\quad \times \int_{-\infty}^{\eta} e^{(i\kappa+\alpha_m)\eta_s}(-\alpha_m+\tanh\eta_s)\tanh\eta_s\,d\eta_s \\
&\quad + e^{\alpha_m\eta}(-\alpha_m+\tanh\eta) \\
&\quad \times \int_{\eta}^{\infty} e^{(i\kappa-\alpha_m)\eta_s}(\alpha_m+\tanh\eta_s)\tanh\eta_s\,d\eta_s \\
&= e^{i\kappa\eta}\Big\{-2\alpha_m\tanh\eta + (1-i\kappa) \\
&\quad \times \Big[\frac{\alpha_m+\tanh\eta}{i\kappa+\alpha_m} - \frac{(-\alpha_m+\tanh\eta)}{i\kappa-\alpha_m}\Big] \\
&\quad + 2i\kappa\Big[(\alpha_m+\tanh\eta)\sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{i\kappa+\alpha_m-2j} \\
&\quad + (-\alpha_m+\tanh\eta)\sum_{j=0}^{\infty}\frac{(-e^{-2\eta})^j}{i\kappa+\alpha_m+2j}\Big]\Big\}.
\end{aligned}
$$

[10] D. Bierens de Haan, *Nouvelles tables d'intégrales definies* (Stechert-Haffner, Inc., New York, 1939), p. 54.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Asymptotic Theory of Hamiltonian and other Systems with all Solutions Nearly Periodic*†

MARTIN KRUSKAL
Plasma Physics Laboratory, Princeton University, Princeton, New Jersey
(Received November 27, 1961)

Consider a system of N ordinary first-order differential equations in N dependent variables, and let the independent variable s not appear explicitly. Let the system depend on a small parameter ϵ and possess a formal infinite power series expansion in ϵ, and suppose that the limiting system for $\epsilon = 0$ exists and has only periodic solutions. Then a formal solution can be constructed involving infinite power series in ϵ and satisfying the equations over large domains of s (of order $1/\epsilon$). The true solutions of the system exist over such domains and are asymptotically represented as $\epsilon \to 0$ by the formal solutions. The construction is based on the standard type of formal series solution of a "reduced" system of $N - 1$ equations in $N - 1$ dependent variables and with the new independent variable $\sigma = \epsilon s$; the omitted variable is essentially an angle variable ϕ describing the phase around the simple, closed curves. If the original system is Hamiltonian, then one can define the usual action integral $J = \int \mathbf{p} \cdot d\mathbf{q}$ to all orders; the integral is taken around the phase ring. It is proved that J is an integral of the system and that the Poisson bracket of ϕ with J is unity, both to all orders. The usefulness of this particular integral is that it is computable locally. The reduced system, after elimination of another dependent variable by means of the constancy of J, can itself be put in Hamiltonian form; if its solutions are nearly periodic, the whole procedure can be reapplied. The present theory encompasses previous proofs of adiabatic invariance to all orders for particular systems such as the harmonic oscillator, the nonlinear oscillator, the charged particle gyrating tightly in a given electromagnetic field, and the longitudinal back-and-forth motion of such a particle trapped between two "magnetic mirrors" in a weak electric field. There are many other applications.

A. INTRODUCTION

1. Historical Perspective

THE phenomenon of adiabatic invariance (constancy of an action integral under slow change of external parameters) has long been known in classical mechanics, receiving prominence because of its role in the so-called quantum conditions and of their importance during the transition from classical to quantum mechanics. Although the invariance was recognized to be only approximate and was only demonstrated to the lowest significant order,[1] the question of whether the constancy might be valid to higher order seems not to have been considered at that time or in that connection. For this there were perhaps two reasons: the first that the question probably appeared to be of only academic interest, and the second that not all adiabatic invariants are constant beyond lowest order.

The question apparently first became of significance for applications in connection with the magnetic moment of gyration of a charged particle gyrating tightly in a strong magnetic field, which was shown to be an adiabatic invariant by Alfvén.[2] On the one hand, the theory of virtually every prospective device for the production of useful energy from controlled thermonuclear fusion[3] has leaned very heavily on the constancy of this magnetic moment, and in those cases for which more or less steady operation was envisioned (stellarator, mirror machine, etc.) it was seen that the requirement that particles remain confined for periods of time encompassing many millions of gyrations could generally be met only if the magnetic moment were in fact constant to a much higher approximation. In connection with the stellarator several other related approximate results were proved to hold to all orders,[4,5] which made it seem both fitting and somewhat more likely that the constancy of the magnetic moment would follow suit. On the other hand, astrophysical theories (e.g. the Fermi mechanism[6] for cosmic-ray production) led to precisely the same question, in response to which

* This work was accomplished largely under the auspices of the U.S. Atomic Energy Commission. Some of it was accomplished while the author was a guest at the Max-Planck-Institut für Physik und Astrophysik (Munich) as a National Science Foundation Senior Postdoctoral Fellow, to both of which organizations he is very grateful. It is a pleasure to acknowledge the helpful discussions and suggestions of a number of scientists, especially E. Gerjuoy, J. Moser, and the author's colleagues at the Princeton Plasma Physics Laboratory, mainly I. Bernstein, E. Frieman, R. Kulsrud, A. Lenard, C. Oberman, and L. Spitzer, Jr.

† Presented at the International Atomic Energy Agency's Conference on Plasma Physics and Controlled Nuclear Fusion Research, September 4–9, 1961, Salzburg, Austria.

[1] M. Born and V. Fock, Z. Physik, **51**, 165 (1928).

[2] H. Alfvén, *Cosmical Electrodynamics* (Clarendon Press, Oxford, England, 1950).

[3] A. Bishop, *Project Sherwood* (Addison-Wesley, Reading, Massachussetts, 1958).

[4] L. Spitzer, Phys. Fluids, **1**, 253 (1958).

[5] M. Kruskal, U.S. Atomic Energy Commision Report NYO-998 (PM-S-5), 1952.
M. Kruskal, U.S. Atomic Energy Commission Report NYO-996 (Appendix to PM-S-3), 1951.

[6] E. Fermi, Astrophys. J. **119**, 1 (1954).

Hellwig[7] proved the constancy of the magnetic moment to the next order beyond the lowest. Kulsrud[8] considered the simpler problem of a harmonic oscillator with slowly changing coefficient of elasticity and proved that its adiabatic invariant (ratio of energy to frequency) was constant to all orders, seemingly the first result of the kind. In quick succession after this breakthrough Kruskal[9] proved the analogous result for the gyrating particle and Lenard[10] for the anharmonic oscillator. Berkowitz and Gardner[11] stiffened the results with unwonted (but wanted!) mathematical rigor by proving that the formal expansions employed to describe the motion of the gyrating particle really were correct asymptotic series for the true trajectory.

Another adiabatic invariant whose degree of constancy has become of practical importance recently is the so-called longitudinal adiabatic invariant of the gyrating particle. This obtains under more restrictive conditions than the magnetic moment invariant does, namely, if the tightly gyrating particle moves back and forth along a magnetic line of force and periodically returns nearly to its initial state. It seems to have been first employed (to lowest order) by Rosenbluth.[12] It became of interest again (and to higher order) in connection with the Van Allen belt.[13] Gardner[14] has settled the matter by proving the invariance to all orders.

2. Significance and Arrangement of Present Paper

In the present paper, a unification and simplification of all these preceding treatments of adiabatic invariance to higher order is achieved. The significant common element has turned out to be that each was concerned with a Hamiltonian system whose solutions are all nearly periodic. Lest this appear too like a truism, it may be pointed out that there are other types of adiabatic invariant associated with Hamiltonian systems whose solutions to lowest order are ergodic over surfaces of constant energy in phase space. (Incidentally, Lenard[10] has shown that descriptions and results valid to all orders can be obtained even for a system all of whose solutions are nearly only "multiply periodic" so long as it is linear.)

These two characteristics, of being Hamiltonian and of having all solutions nearly periodic, may largely be treated independently of each other, although their amalgamation (coined "almost-mechanics" by E. Gerjuoy) leads to interesting consequences (part E). Since the theory of Hamiltonian systems is already extremely well developed, everything we need is at hand, except, perhaps, for the result proved in Appendix 2. We therefore devote the major portion of the paper to discussing systems of differential equations with all solutions nearly periodic, which is of great independent interest in any case. We first derive the appropriate formal series solutions (Secs. B. 5–9). This treatment is in essence very similar to the method of Kryloff and Bogoliuboff,[15] but seems to be a generalization inasmuch as the latter method is confined to quasi-linear second-order systems.

Following the determination of the formal series we prove (Secs. B. 10–11) that the given system has exact solutions for an appropriately large range of the independent variable and that the formal series solutions represent the exact solutions asymptotically. The method seems simpler and more natural than that given by Berkowitz and Gardner[11] for the special case of the gyrating particle. An interesting feature is the "bootstrap" argument (Sec. B. 11) by which the two items are linked together and proved simultaneously, thereby avoiding a certain duplication to be found in their treatment.

Part C starts with the theorem of phase independence, the importance and utility of which for this paper can hardly be overestimated: it is fair to say that parts C, D, and E consist of almost nothing but its systematic exploitation. The full power of this ridiculously simple theorem can probably only be appreciated by those who, like the author, have spent tremendous amounts of effort and time in attempting (often unsuccessfully) to obtain various results to all orders.

3. Notation and Conventions

We consistently employ vector and dyadic notation in dealing with arrays of quantities and often indulge in the geometric terminology which accompanies it so naturally. Vectors are boldfaced throughout, and polyadics are represented by script letters. The number of components (array elements)

[7] G. Hellwig, Z. Naturforsch. **10a**, 508 (1955).
[8] R. Kulsrud, Phys. Rev. **106**, 205 (1957).
[9] M. Kruskal, Rendiconti del Terzo Congresso Internazionale sui Fenomeni D'Ionizzazione nei Gas tenuto a Venezia, Societá Italiana di Fisica, Milan (1957).
[10] A. Lenard, Ann. Phys. **6**, 261 (1959).
[11] J. Berkowitz and C. Gardner, Comm. Pure Appl. Math., Vol. XII, 501-512 (1959).
[12] M. Rosenbluth, U.S. Atomic Energy Commision Rept. LA-2030 (1956).
[13] J. Van Allen, J. Geophys. Research **64**, 1683 (1959).
[14] C. Gardner, Phys. Rev. **115**, 791 (1959).
[15] N. Kryloff and N. Bogoliuboff, *Introduction to Non-Linear Mechanics*, (Annals of Mathematics Studies, No. 11, Princeton University Press, Princeton, 1947).

of every vector (etc.) will of course be announced (if it is definite) when that vector is first introduced, but that number will not be indicated, thereafter, by any special notational feature. The unit dyadic in any number of dimensions will be denoted by $\mathcal{I}$. The inner ("dot") product of two vectors is defined only if they have the same number of components, and the double inner product is to be understood in the sense that $\mathbf{ab} : \mathbf{cd} = \mathbf{a}\cdot\mathbf{d}\,\mathbf{b}\cdot\mathbf{c}$.

Differentiation, whether ordinary or partial, will be consistently denoted by attaching as a subscript the variable with respect to which the derivative is taken. Moreover, the use of subscripts is reserved exclusively for this purpose. To maintain the correct vector ordering, it is convenient, in a few places, to put the subscript before the symbol to which it is attached; thus $_{\mathbf{x}}\mathbf{f}$ denotes the gradient of the vector field $\mathbf{f}$ and the transpose of the dyadic $\mathbf{f}_{\mathbf{x}}$.

When one set of variables (coordinates) is transformed into another, which will be symbolized by a double-headed arrow $\leftrightarrow$, the new variables will be denoted by symbols entirely distinct from the old ones. Therefore (in contrast, for instance, with the conventional notation in thermodynamics), there can arise no possible ambiguity as to what is being held fixed during a partial differentiation.

We will have mostly to do with formal infinite series in increasing integral powers of an expansion parameter ϵ, and we generally omit any indication of ϵ in the notation. If $\mathbf{f}$ denotes such a series, $\mathbf{f}^{(n)}$ denotes the coefficient of ϵ^n, and $\mathbf{f}^{[n]}$ denotes the sum of all the terms of $\mathbf{f}$ up to and including $\epsilon^n\mathbf{f}^{(n)}$. As usual, $O(\epsilon^n)$ denotes any quantity which, as $\epsilon \to 0$, approaches zero at least as fast as ϵ^n.

A variable will be called *angle-like* if it is determined up to and only up to an additive integer, so that it is like an angle variable measured in units of a complete revolution. (As here, defined terms will regularly be italicized when they are introduced.)

4. Local Dependence and Useful Integrals

For a full exposition of the significance of the results to come, it is necessary to discuss here the concept of local dependence. A function $\eta(\mathbf{x})$ on a space of points $\mathbf{x}$ which is a functional of (i.e. depends on) another such function $\xi(\mathbf{x})$ depends *locally* on $\xi(\mathbf{x})$ if the value of η at any given point $\mathbf{x}$ is independent of the values of ξ outside of arbitrarily small neighborhoods of that point. (Here ξ and η can stand indifferently also for sets of independently many functions.) In practice this usually means that the value of η at $\mathbf{x}$ is a function of $\mathbf{x}$ and of the values at $\mathbf{x}$ of ξ and its derivatives (and often in fact only a finite number of them). We shall use this definition only "metamathematically" and so may modify it or use it loosely on occasion.

To see why local dependence is significant, let us observe that one of the later results is the derivation of an approximate integral or "constant of motion" J (Sec. E. 2) for certain systems of differential equations. How can this be important, in view of the fact that such systems commonly have not merely one but a complete set of exact integrals? The answer is that J is a "useful" integral. There are two ways in which an integral of a system may fail to be useful. The first is that it may fail to be a so-called isolating integral (see Wintner[16]), the second that it may not be locally computable.

To illustrate the first possibility (without really going into the subject or even giving the definition of isolating integral) by means of a really trivial example, let us take the system which consists of two independent subsystems

$$\alpha_s = a, \qquad \beta_s = b, \tag{A1}$$

where α and β are angle-like variables, s is the independent variable, and a and b are given constants. The subsystems have each one integral, namely,

$$\alpha - as, \qquad \beta - bs, \tag{A2}$$

respectively. If we now ask for integrals which are independent of s, we see that the subsystems have none, but the combined system has the s-independent integral

$$b\alpha - a\beta. \tag{A3}$$

This integral, however, is multivalued. If a and b are commensurable (a/b rational) it is an isolating integral and is useful: Knowing the value of the integral (from initial conditions, for example) and, say, of β, we can say something about the value of α; in fact, writing a/b in standard form as m/n (m and n relatively prime integers and n positive), we can tell that α must have one of n (essentially distinct) values all of whose pairwise differences are multiples of $1/n$. On the other hand, if a and b are incommensurable, the integral is not isolating, and is clearly not useful in the same way.

The second possibility is of more direct concern. For illustration consider the motion of a particle in a given static three-dimensional potential field. To guarantee the existence of a complete set of isolating integrals let us agree to accept time-dependent integrals. (In various special cases this

[16] A. Wintner, *Analytical Foundations of Celestial Mechanics* (Princeton University Press, Princeton, 1941) Sec. 128.

can be guaranteed even for time-independent integrals, for instance, if the potential has a translational invariance, so that the corresponding spatial coordinate can be ignored, grows linearly in time, and can serve in effect as the time.) In such a case, it is sometimes said that the energy is the only integral. Strictly speaking this is nonsense, since we know that there is a complete set of integrals (for instance, the values of the dynamical variables taken on at $t = 0$, viewed as functions of the free dynamical variables). But there is no question about the exceptional status of the energy integral: In countless physical arguments it plays a unique role denied to the other integrals. The reason for this is that the energy depends only locally on the potential and there is no other integral (functionally independent of the energy) with that property. The point is that, although in every particular case (i.e., for every particular potential function) the other integrals exist, and can be perfectly well defined mathematically, the problem of finding them is equivalent to the problem of integrating the equations of motion of the particle. Only the energy integral is useful because it alone is known *a priori*, without integrating the system completely, and because it alone remains unaffected (in its functional form) in one region by a change in the potential elsewhere along the orbit of the particle.

5. Autonomous Systems

Throughout this paper we shall be dealing with systems of ordinary differential equations. Primarily for notational convenience we do not permit the functions expressing the derivatives to depend on the independent variable, but this is no essential restriction, for there is no objection to such a dependence *de facto* (so long as appearances are preserved *de nomine*) by way of the existence of a dependent variable which always equals the independent variable, i.e., has derivative unity. Denoting the dependent variables collectively by $\mathbf{x}$ and the independent variable by s, we may write a general system in the form

$$\mathbf{x}_s = \mathbf{f}(\mathbf{x}). \tag{A4}$$

This may be interpreted as a first-order differential equation in the space of points $\mathbf{x}$. We call it an *autonomous* system for short and as a faint reminder that the rate of change of $\mathbf{x}$ depends only on $\mathbf{x}$ itself. When $\mathbf{f}(\mathbf{x})$ is a formal series in powers of ϵ, we shall say that (A4) is in *standard form* if $\mathbf{f}$ starts with a term of zeroth order which nowhere (as a function of $\mathbf{x}$) vanishes (in all components simultaneously).

We assume that $\mathbf{f}$ is as smooth as needed, so all the standard theory of systems of ordinary differential equations applies. A most fundamental teaching of that theory is that for every specified "initial condition," i.e., specification of $\mathbf{x}$ at (say) $s = 0$, there exists a solution $\mathbf{x}(s)$ of (A4) which may be continued indefinitely so long as $\mathbf{x}$ stays within the domain of definition of $\mathbf{f}$. Another fundamental teaching is that such a solution is unique: Any two functions satisfying the same autonomous system and equal for $s = 0$ are equal for all values of s (in any interval which contains zero and over which both functions are defined); this is in fact a special case of the fundamental approximation theorem for ordinary differential equations (Appendix 1).

Let $\mathbf{X}(\mathbf{x}, s)$ denote the unique solution of (A4) which passes through $\mathbf{x}$ at $s = 0$, i.e., let $\mathbf{X}$ be defined by the conditions

$$\mathbf{X}_s(\mathbf{x}, s) = \mathbf{f}(\mathbf{X}(\mathbf{x}, s)), \tag{A5}$$

$$\mathbf{X}(\mathbf{x}, 0) = \mathbf{x}. \tag{A6}$$

An important property of the solutions of autonomous systems is that the one-parameter family of mappings of the form $\mathbf{x} \to \mathbf{X}(\mathbf{x}, s)$ (between parts of $\mathbf{x}$ space) constitute a group, and furthermore that the group operation on $\mathbf{X}(\mathbf{x}, s')$ and $\mathbf{X}(\mathbf{x}, s)$ is $\mathbf{X}(\mathbf{x}, s' + s)$, so that the group is isomorphic to the additive group of real numbers s. [Indeed, (A6) already states that $s = 0$ represents the identity.] To show this one must establish that

$$\mathbf{X}(\mathbf{X}(\mathbf{x}, s'), s) = \mathbf{X}(\mathbf{x}, s + s'). \tag{A7}$$

This follows from the uniqueness theorem, since the two sides are equal for $s = 0$ by (A6), and satisfy the same autonomous system because the left-hand side is obtained from $\mathbf{X}(\mathbf{x}, s)$ merely by substituting $\mathbf{X}(\mathbf{x}, s')$ for $\mathbf{x}$ and the right-hand side by substituting $s + s'$ for s, neither of which substitutions invalidates (A5).

A further property we use in the next section is now easily obtained. Using (A5) with $s = 0$ and (A6) we have

$$\mathbf{X}_s(\mathbf{x}, 0) = \mathbf{f}(\mathbf{x}). \tag{A8}$$

Differentiating (A7) with respect to s', setting $s' = 0$, and using (A6) and (A8) gives

$$\mathbf{X}_{\mathbf{x}}(\mathbf{x}, s) \cdot \mathbf{f}(\mathbf{x}) = \mathbf{X}_s(\mathbf{x}, s). \tag{A9}$$

6. Recurrent Systems

If all solutions of (A4) are periodic (not necessarily with equal periods) we call (A4) a *recurrent*

system. We also call the vector field $\mathbf{f}$ in $\mathbf{x}$ space *recurrent* under the same circumstances, namely, when its integral curves (curves which at each of their points are tangent to the field) are closed, forming topological circles; we call them *loops*. For each $\mathbf{x}$ there is then a least positive value of s for which the initial value $\mathbf{x}$ recurs; we denote this period by $S(\mathbf{x})$ and have

$$\mathbf{X}(\mathbf{x}, S(\mathbf{x})) = \mathbf{x}, \tag{A10}$$

$$\mathbf{X}(\mathbf{x}, s) \neq \mathbf{x} \quad \text{for} \quad 0 < s < S(\mathbf{x}). \tag{A11}$$

We next show by a formal argument the obvious fact that the period $S(\mathbf{x})$ is constant along a solution of (A4). We first take the s derivative of (A10) as $\mathbf{x}$ varies in accordance with (A4), or in other words, we take the $\mathbf{x}$ derivative, on the right, say, and then dot on the right with $\mathbf{f}(\mathbf{x})$; using (A9) with $s = S(\mathbf{x})$ the result may be written

$$\mathbf{X}_s(\mathbf{x}, S(\mathbf{x})) + \mathbf{X}_s(\mathbf{x}, S(\mathbf{x}))S_{\mathbf{x}}(\mathbf{x})\cdot\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}). \tag{A12}$$

But setting $s = S(\mathbf{x})$ in (A5) we have

$$\mathbf{X}_s(\mathbf{x}, S(\mathbf{x})) = \mathbf{f}(\mathbf{x}) \tag{A13}$$

in view of (A10). Therefore, the extreme terms in (A12) cancel, and, furthermore, since $\mathbf{f} \neq 0$, we have

$$S_{\mathbf{x}}(\mathbf{x})\cdot\mathbf{f}(\mathbf{x}) = 0, \tag{A14}$$

which is the desired result; to be entirely explicit,

$$[S(\mathbf{X}(\mathbf{x}, s))]_s = S_{\mathbf{x}}(\mathbf{X}(\mathbf{x}, s))\cdot\mathbf{X}_s(\mathbf{x}, s)$$
$$= S_{\mathbf{x}}(\mathbf{X}(\mathbf{x}, s))\cdot\mathbf{f}(\mathbf{X}(\mathbf{x}, s)) = 0 \tag{A15}$$

in view of (A5) and of (A14) with $\mathbf{X}(\mathbf{x}, s)$ for $\mathbf{x}$, so that the period evaluated at all points of the solution $\mathbf{X}(\mathbf{x}, s)$ is the same (independent of s). This result may be written

$$S(\mathbf{X}(\mathbf{x}, s)) = S(\mathbf{x}). \tag{A16}$$

Another hardly surprising result, a generalization of (A10), is

$$\mathbf{X}(\mathbf{x}, s + S(\mathbf{x})) = \mathbf{X}(\mathbf{x}, s). \tag{A17}$$

This can be deduced from the uniqueness theorem, or by taking (A7) with $s' = S(\mathbf{x})$ and using (A10). It is now obvious that $\mathbf{X}(\mathbf{x}, s)$ has the same value for two different values of s if and only if they differ by an integral multiple of $S(\mathbf{x})$; the "if" follows by induction from (A17), and the "only if" part since otherwise a contradiction to (A11) could be similarly obtained.

7. Splittable Systems

An important property of some systems of differential equations is the possession of an autonomous subsystem. Suppose that, of the set of N dependent variables denoted collectively by $\mathbf{x}$, there are some M (where $1 \leq M \leq N - 1$) whose s derivatives as given by (A4) are independent of the remaining $N - M$ variables. The M variables then satisfy an autonomous subsystem of their own and we call the original system *split*. More generally, we call a system *splittable* if by a local change of dependent variables it can be transformed into a split system, i.e., if there are local definitions of M new variables, functions of $\mathbf{x}$ whose derivatives, computed using (A4), are expressible as functions of the M new variables themselves alone. These variables then satisfy a *new autonomous* system of fewer variables.

By a local change of variables we mean, a change whose defining formulas depend locally on $\mathbf{f}(\mathbf{x})$ (see Sec. 4). The requirement of localness is essential to the definition; without it, for instance, every system with a complete set of isolating integrals would be not merely splittable but completely splittable into N independent autonomous subsystems of one dependent variable each.

The problem of solving a system of order N may, if the system is splittable, be split up into the problems of first solving a system of order M and then one of order $N - M$.

B. NEARLY RECURRENT SYSTEMS

1. Formulation and Description

Consider the system of first-order ordinary differential equations

$$\mathbf{x}_s = \mathbf{F}(\mathbf{x}, \epsilon), \tag{B1}$$

where $\mathbf{x}$ and $\mathbf{F}$ are vectors of N components. We suppose that $\mathbf{F}$ is defined in a suitable domain of its arguments, with properties to be described shortly, and that it and all its derivatives of all orders exist and are continuous. Let the domain be the product of a domain in $\mathbf{x}$ space with a closed interval of values for ϵ among which occurs zero. Let (B1) be a recurrent system for $\epsilon = 0$, so that $\mathbf{F}(\mathbf{x}, 0)$ is a nowhere vanishing vector field with all integral curves closed. [When we speak of loops in connection with (B1), we shall always mean these, even if it should happen, most atypically, that (B1) is also recurrent for $\epsilon \neq 0$.] We require finally that the domain of points in $\mathbf{x}$ space for which $\mathbf{F}(\mathbf{x}, \epsilon)$ is defined be closed, bounded, and N dimensional and be made up of loops.

As ϵ approaches zero any solution of (B1) gets closer and closer to a loop, so long as the range of

values of s considered remains bounded (independently of ϵ). We may say that the solutions of system (B1) are all *nearly periodic*[17] for small ϵ. The distance in $\mathbf{x}$ space (in any reasonable sense) by which a solution misses its initial point after one *gyration*, namely, after approximately following one closed loop around, is of order ϵ; in general, it is to be expected that these small deviations will accumulate, so that after a large number of gyrations (a large change in s) of order $1/\epsilon$, the solution will have *drifted* a finite amount and will be gyrating in a different region of $\mathbf{x}$ space from where it started. The curve in $\mathbf{x}$ space traversed by a solution in its entirety may be suggestively thought of as the distorted image of a helix of small pitch, as shown in Fig. 1.

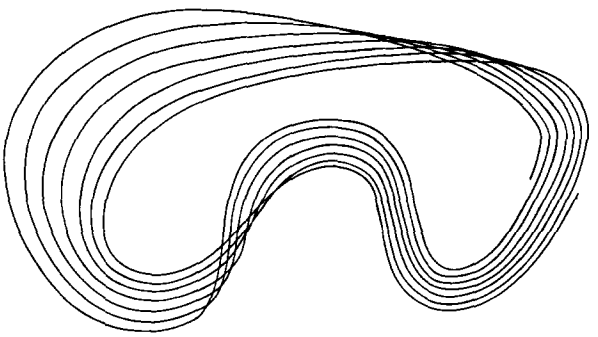

FIG. 1. Curve traversed by a solution of (B1). The curve cannot cross itself; the apparent crossings are due to projection.

2. Inadequacy of Obvious Series Solution

Our first task is to find some kind of formal solution of (B1), in terms of infinite power series in ϵ, which is asymptotically valid uniformly over large ranges of s and shows explicitly its nearly periodic nature. The obvious approach at first sight seems to be to represent $\mathbf{x}$ directly by a series,

$$\mathbf{x} \approx \mathbf{u}^{(0)}(s) + \epsilon\,\mathbf{u}^{(1)}(s) + \epsilon^2\,\mathbf{u}^{(2)}(s) + \cdots, \tag{B2}$$

and to determine the $\mathbf{u}^{(n)}$ by substituting this series into (B1), expanding the function on the right-hand side in a formal double Taylor series around its zeroth-order arguments $(\mathbf{u}^{(0)}, 0)$, and collecting and separately equating like powers of ϵ. This leads to the equations

$$\mathbf{u}_s^{(0)} = \mathbf{F}(\mathbf{u}^{(0)}, 0),$$
$$\mathbf{u}_s^{(1)} = \mathbf{F}_x(\mathbf{u}^{(0)}, 0)\cdot\mathbf{u}^{(1)} + \mathbf{F}_\epsilon(\mathbf{u}^{(0)}, 0), \cdots, \tag{B3}$$

which do indeed determine the $\mathbf{u}^{(n)}$ if some initial values are given, say at $s = 0$. [And, in fact, if $\mathbf{F}$ is analytic then the series in (B2) converges to an analytic solution of (B1) for sufficiently small ϵ.] Of course, $\mathbf{u}^{(0)}$ runs periodically around and around a loop. But (B2) obviously can represent a solution asymptotically only so long as the solution stays within $O(\epsilon)$ of $\mathbf{u}^{(0)}$, which in general means only for bounded values of s: For larger s the solution will have drifted farther from its initial loop. Therefore (B2) is not useful for our purpose.

3. Standardization

Before proceeding any further let us replace (B1) by

$$\mathbf{x}_s = \mathbf{f}(\mathbf{x}), \tag{B4}$$

where $\mathbf{f}(\mathbf{x})$ denotes the Taylor expansion

$$\mathbf{f}(\mathbf{x}) = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!}\left[\frac{d^n\mathbf{F}}{d\epsilon^n}\right]_{\epsilon=0} \tag{B5}$$

of $\mathbf{F}$ around $\epsilon = 0$, and is therefore an infinitely differentiable formal infinite series of non-negative powers of ϵ. We have two reasons for doing this: First, since we are seeking a formal solution (which later is proved to represent the true solution asymptotically, but which does not in general converge to it, nor indeed at all; see the next section) we must switch over to formal series at some point anyway, and it is simplest to do so at the start. Second, the theory about to be developed can be applied not only to true systems of differential equations, but also to merely formal systems (and in particular to the "reduced" system resulting from the first application of the theory; see Secs. C.3 and E.6).

However, a warning is in order. Certainly (B4) is the formal analog of (B1); and indeed, the formal solutions of (B4) which we shall obtain, when taken up to terms of order n, do provide approximations up to that order of the true solutions of (B1) [as do also the true solutions of the system $\mathbf{x}_s = \mathbf{f}^{[n]}(\mathbf{x})$], as we see from the basic approximation lemma of Appendix 1. But this is only for ranges of s bounded independently of ϵ; due to the exponential in the estimate (F9), no matter how large we take n, we could not expect these formal approximations to remain at all close to the true solutions for ranges of s of order $1/\epsilon$ (or of order $1/\epsilon^\alpha$, $\alpha > 0$; but we could, for instance, for ranges of order $\log \epsilon$). Or rather, we could not expect so except for the near periodicity assumed; one of our conclusions (Sec.

[17] This concept is not to be confounded with that of almost periodic functions, H. Bohr, *Fastperiodische Funktionen, Ergebnisse der Mathematik und Ihrer Grenzgebiete* (Julius Springer, Berlin, 1932), B.1, N.5, which we are not concerned with here.

B.10) is that the approximations are in fact good for ranges of order $1/\epsilon$. (But not quite so good: two orders in ϵ get lost in the shuffle.)

4. An Example

An extremely simple but doubly illuminating example is provided by the harmonic oscillator with slowly varying rest point. As our basic equation we take

$$\epsilon^2 u_{tt} + u = r(t), \tag{B6}$$

where u is the displacement of the oscillator and $r(t)$ the rest point, a given function of its argument (and allowably a series in powers of ϵ). If r were constant, every solution of (B6) would oscillate sinusoidally around r with period $2\pi\epsilon$; as it is, r is nearly constant during a period, so it should be possible to put (B6) into the form this paper deals with.

To render the period finite we introduce $s \equiv t/\epsilon$, so (B6) becomes

$$u_{ss} + u = r(\epsilon s), \tag{B7}$$

a version of the equation we might have started with; here the right-hand side varies slowly because of the factor ϵ in the argument. To render (B7) as a formally first-order system we introduce $v \equiv u_s$. To eliminate the dependence (of r) on the independent variable s we bring back t, but now as a new dependent variable. (It wouldn't do to introduce a new dependent variable equal to s itself, because s does not recur, even to lowest order.) Altogether, we end up with a system in standard form (B4) with

$$\begin{aligned} \mathbf{x} &\equiv (u, v, t), \\ \mathbf{f} &\equiv (v, r(t) - u, \epsilon). \end{aligned} \tag{B8}$$

This system is nearly recurrent because to zeroth order, t remains constant and u and v oscillate periodically.

Having used this example to illustrate how a given system may be transformed as required by the theory, we now use it to illustrate why the formal series we are led to are, in general, only asymptotic, and not convergent as one might have hoped (when $\mathbf{f}$ is convergent, i.e., analytic). For this we need only the original form (B6). Since the corresponding homogeneous equation has the general solution

$$u = \alpha \cos(t/\epsilon) + \beta \sin(t/\epsilon), \tag{B9}$$

we may confine ourselves to a particular solution of (B6). One particular solution, for that matter, may be singled out as "oscillation free" (a concept which may be rendered precise by interpreting it to mean the vanishing of J; see Sec. E.2). This is a unique characterization to all orders, though ill defined for finite ϵ. The oscillations referred to are naturally those of small period $2\pi\epsilon$, and are excluded by requiring that t differentiation not affect (not lower) the order in ϵ, in obvious contrast to its effect on the general solution (B9). Thus $\epsilon^2 u_{tt}$ is regarded as small, and (B6) is recursively solvable to yield

$$\begin{aligned} u &= r - \epsilon^2 u_{tt} \\ &= r - \epsilon^2 r_{tt} + \epsilon^4 r_{tttt} - \cdots . \end{aligned} \tag{B10}$$

This formal development differs from the Taylor series for $r(t + \epsilon)$ in two respects unimportant for present purposes (the absence of odd powers of ϵ and the alternation of sign among the even powers) and in one important respect, the absence of $n!$ in the denominator of the ϵ^n term. It is thus clear that even the analyticity of $r(t)$ is insufficient to assure the convergence of (B10). For (B10) to have a radius of convergence greater than zero, the development of $r(t + \epsilon)$ must have an infinite radius of convergence, i.e., $r(t)$ must be an entire function. And even this is not quite enough, as can be seen by Fourier analysis of (B6).

5. More Appropriate Variables

Let us start our search for an asymptotic solution of (B4) by introducing a new coordinate system more appropriate than $\mathbf{x}$ for describing a nearly periodic but slowly drifting trajectory. To this end observe that the loops form an $(N-1)$-dimensional family, so that there exist $N-1$ independent functions of $\mathbf{x}$ each of which is constant on every loop; these can obviously also be chosen infinitely differentiable, since $\mathbf{f}^{(0)}$ is. [For instance, we may pick an infinitely differentiable $(N-1)$-dimensional hypersurface in $\mathbf{x}$ space which is nowhere tangent to $\mathbf{f}^{(0)}$, then pick any set of $N-1$ independent infinitely differentiable coordinates $\mathbf{Y}$ on the hypersurface, and then extend $\mathbf{Y}$ to the rest of $\mathbf{x}$ space so as to be constant on every loop.] Denote some arbitrary fixed set of such functions collectively by a vector function $\mathbf{Y}(\mathbf{x})$ with $N-1$ components. Thus each loop is characterized uniquely by the constant (vector) value assumed on it by $\mathbf{Y}(\mathbf{x})$.

Next, let $\Upsilon(\mathbf{x})$ be any infinitely differentiable angle-like function which varies monotonically around each loop; for definiteness say it increases with s. We then change variables from $\mathbf{x}$ to $\mathbf{y}$ and v

by the transformation

$$\mathbf{y} = \mathbf{Y}(\mathbf{x}), \qquad \upsilon = \Upsilon(\mathbf{x}). \tag{B11}$$

Of course, υ is determined only up to an arbitrary additive integer. The transformation clearly possesses an infinitely differentiable inverse which we denote by

$$\mathbf{x} = \mathbf{X}(\mathbf{y}, \upsilon), \tag{B12}$$

where $\mathbf{X}$ is, of course, periodic in υ (with period unity).

In terms of our new variables the system (B4) takes the form

$$\mathbf{y}_s = \epsilon\, \mathbf{g}(\mathbf{y}, \upsilon), \qquad \upsilon_s = \psi(\mathbf{y}, \upsilon), \tag{B13}$$

where (omitting the arguments $\mathbf{y}$, υ consistently)

$$\epsilon\mathbf{g} \equiv \mathbf{Y}_{\mathbf{x}}(\mathbf{X})\cdot\mathbf{f}(\mathbf{X}), \tag{B14}$$

$$\psi \equiv \Upsilon_{\mathbf{x}}(\mathbf{X})\cdot\mathbf{f}(\mathbf{X}); \tag{B15}$$

here $\mathbf{g}$ and ψ are infinitely differentiable formal series, their leading terms are $O(1)$ and that of ψ is definitely positive, and $\mathbf{g}$ is written accompanied by the factor ϵ because, by construction, $\mathbf{y}_s$ must vanish to zeroth order. These equations exhibit the gyration clearly: to zeroth order, as s increases $\mathbf{x}$ stays on a loop (since y is constant) and runs periodically around it (since υ increases monotonically); to first order the loop may drift (since $\mathbf{y}$ may vary).

We can hardly rest content with our system in the form (B13), however. For, the gyration it depicts may be very uneven, full of sudden accelerations and decelerations, since ψ may fluctuate wildly; further, and, in fact, more serious, the net rate of drift is not at all in evidence, since in the course of one gyration $\mathbf{g}$ may point in all different directions, even quite opposite to the direction that $\mathbf{y}$ is effectively drifting in, which can only be obtained by some sort of integration around a loop. Using a previous metaphor, we might say that we have introduced coordinates $\mathbf{y}$, υ in terms of which every trajectory does indeed look like a helix of small pitch (if we think of υ as the angle of a point on a circle), but only to zeroth order, and even there s does not increase uniformly around the loops.

But these defects can be remedied. As for one minor issue, it is not hard to see that if Υ had been chosen judiciously, $\psi^{(0)}$ would have come out independent of υ. It would only have been necessary to take

$$\Upsilon(\mathbf{x}) = \left[\int_{\mathbf{x}'}^{\mathbf{x}} \frac{d\mathbf{x}}{\mathbf{f}^{(0)}}\right] \Big/ S(\mathbf{x}'), \tag{B16}$$

where $\mathbf{x}'$ is some definite point on the loop through $\mathbf{x}$ (say the intersection with the hypersurface suggested before), S is the period (Sec. A6), and the line integral is taken along the loop (so that the ratio of vectors under the integral sign is well defined and scalar and, to be sure, nothing but ds). In carrying out our procedure in particular problems, it is in fact very desirable to start with such an Υ. We have not specified this to-be-so-called *judicious choice*, however, since for our abstract purposes it would make little difference, and that of dubious advantage; but we shall occasionally note where it would affect matters.

6. Nice Variables

The defects we have noticed in (B13) are all due to the dependence of $\mathbf{g}$ and ψ on υ. We therefore ask whether we can find an infinitely differentiable formal transformation to new variables which are similar to $\mathbf{y}$ and υ and satisfy equations similar to (B13), but without the odious dependence on the angle-like variable—and not only to lowest order but to all orders. We shall see that the answer is yes.

Denote the new sought variables by a vector $\mathbf{z}$ of $N - 1$ components and an angle-like variable ϕ, related to the old variables by

$$\mathbf{z} = \mathbf{Z}(\mathbf{y}, \upsilon), \qquad \phi = \Phi(\mathbf{y}, \upsilon), \tag{B17}$$

where $\mathbf{Z}$ and Φ are to be determined. Let the system (B13), and indirectly therefore (B4), transform into

$$\mathbf{z}_s = \epsilon\, \mathbf{h}(\mathbf{z}), \qquad \phi_s = \omega(\mathbf{z}), \tag{B18}$$

analogous to (B13) but without ϕ appearing as an argument on the right-hand sides; $\mathbf{h}$ and ω are also to be determined. Naturally $\mathbf{h}$ and ω, like $\mathbf{g}$ and ψ, will be infinite power series in ϵ, and it is only natural that $\mathbf{Z}$ and Φ be so too. Thus, the transformation (B17), unlike (B11), depends on the expansion parameter ϵ. But this is of no concern, especially since $\mathbf{Z}$ and Φ, and for that matter $\mathbf{h}$ and ω, are to be $O(1)$.

We must impose on $\mathbf{Z}$ and Φ periodicity conditions which express that υ is an angle-like variable and that ϕ is to be a similar one, namely,

$$\begin{aligned} \mathbf{Z}(\mathbf{y}, \upsilon + 1) &= \mathbf{Z}(\mathbf{y}, \upsilon), \\ \Phi(\mathbf{y}, \upsilon + 1) &= \Phi(\mathbf{y}, \upsilon) + 1. \end{aligned} \tag{B19}$$

It turns out that we may specify the dependence of $\mathbf{Z}$ and Φ on $\mathbf{y}$ for $\upsilon = 0$ with great freedom; for definiteness and simplicity we choose the initial conditions

$$\mathbf{Z}(\mathbf{y}, 0) = \mathbf{y}, \qquad \Phi(\mathbf{y}, 0) = 0. \tag{B20}$$

These express that $\mathbf{z}$ and ϕ are to equal $\mathbf{y}$ and v at one definite point on each loop, where $v = 0$.

With conditions (B18), (B19), and (B20) we now claim not only the existence but also the uniqueness of the formal transformation we are seeking.

7. Determination of Recursion Relations

The first step in determining $\mathbf{Z}$, Φ, $\mathbf{h}$, and ω is to substitute (B17) in (B18), use the chain rule of differentiation, and eliminate $\mathbf{y}_s$ and v_s by (B13), which yields

$$\epsilon\, \mathbf{Z}_{\mathbf{y}}\cdot\mathbf{g} + \mathbf{Z}_v\psi = \epsilon\, \mathbf{h}(\mathbf{Z}), \tag{B21}$$

$$\epsilon\, \Phi_{\mathbf{y}}\cdot\mathbf{g} + \Phi_v\psi = \omega(\mathbf{Z}). \tag{B22}$$

Let us assume temporarily that we know $\mathbf{h}$ and ω as functions of their arguments. These equations may then be construed, thanks to the factor ϵ appearing in the first term of each of them, as determining, order by order, the derivatives of $\mathbf{Z}$ and Φ with respect to v in terms of known quantities, while $\mathbf{y}$ plays the role of a parameter. Accordingly, we transfer the first terms to the right-hand side, divide by ψ, and integrate; in view of the initial conditions (B20) the results may be written

$$\mathbf{Z} = \mathbf{y} + \epsilon \int_0^v dv\, [\mathbf{h}(\mathbf{Z}) - \mathbf{Z}_{\mathbf{y}}\cdot\mathbf{g}]\, \frac{1}{\psi}, \tag{B23}$$

$$\Phi = \int_0^v dv\, [\omega(\mathbf{Z}) - \epsilon\, \Phi_{\mathbf{y}}\cdot\mathbf{g}]\, \frac{1}{\psi}. \tag{B24}$$

On these we next wish to impose the periodicity conditions (B19), which it suffices to use only for $v = 0$ [for, assuming that (B19) are satisfied in the previous orders, the integrands of (B23) and (B24) are periodic functions of v]. We obtain

$$\int_0^1 dv\, [\mathbf{h}(\mathbf{Z}) - \mathbf{Z}_{\mathbf{y}}\cdot\mathbf{g}]\, \frac{1}{\psi} = 0, \tag{B25}$$

$$\int_0^1 dv\, [\omega(\mathbf{Z}) - \epsilon\, \Phi_{\mathbf{y}}\cdot\mathbf{g}]\, \frac{1}{\psi} = 1. \tag{B26}$$

These are the conditions we employ to determine the functions $\mathbf{h}$ and ω, which we previously assumed known.

To this end, denote the integral in (B23) by $\mathbf{K}$, substitute $\mathbf{y} + \epsilon\mathbf{K}$ for the argument of $\mathbf{h}$ in (B25) and of ω in (B26), and expand $\mathbf{h}$ and ω in Taylor series around the argument $\mathbf{y}$. The factors $\mathbf{h}(\mathbf{y})$ and $\omega(\mathbf{y})$ in the leading terms of these expansions, being independent of v, come out of the integrals, and we obtain

$$\mathbf{h}(\mathbf{y}) \int_0^1 dv\, \frac{1}{\psi} = \int_0^1 dv \left[\mathbf{Z}_{\mathbf{y}}\cdot\mathbf{g} - \epsilon \left\{ \frac{\mathbf{h}(\mathbf{y} + \epsilon\mathbf{K}) - \mathbf{h}(\mathbf{y})}{\epsilon} \right\} \right] \frac{1}{\psi}, \tag{B27}$$

$$\omega(\mathbf{y}) \int_0^1 dv\, \frac{1}{\psi} = 1 + \epsilon \int_0^1 dv \left[\Phi_{\mathbf{y}}\cdot\mathbf{g} - \left\{ \frac{\omega(\mathbf{y} + \epsilon\mathbf{K}) - \omega(\mathbf{y})}{\epsilon} \right\} \right] \frac{1}{\psi}; \tag{B28}$$

the expressions enclosed in curly brackets are to be thought of as expanded in the obvious way, and are clearly $O(1)$. The factor $\int_0^1 dv/\psi$ can be developed as a series with a leading term which is $O(1)$ and definitely positive, due to the nature of ψ. Thus (B27) and (B28) provide recursion formulas for $\mathbf{h}(\mathbf{y})$ and $\omega(\mathbf{y})$.

8. Recursive Construction of Desired Functions

We are now prepared to prove the existence and uniqueness of formal series $\mathbf{Z}(\mathbf{y}, v)$, $\Phi(\mathbf{y}, v)$, $\mathbf{h}(\mathbf{y})$, and $\omega(\mathbf{y})$ satisfying (B17), (B18), (B19), and (B20). (Since it is $\mathbf{h}$ and ω as functions that are sought, it is of no account that we now denote their arguments by $\mathbf{y}$ instead of $\mathbf{z}$.) We proceed by induction. It suffices to prove, for each non-negative integer n, that if the series exist and are unique up to terms of order $n - 1$, then they do and are so up to terms of order n. (For they certainly exist and are unique up to order minus one, since they are required to start with terms of order zero!)

The argument is easy; we need merely appeal to the determining Eqs. (B23), (B24), (B27), (B28) we have derived, making sure to find $\mathbf{Z}$ before Φ or $\mathbf{h}$, and ω before Φ. Thus, we may first uniquely determine $\mathbf{Z}$ to order n by (B23), noting that the right-hand side can be obtained to that order in terms of lower order quantities, already known according to the induction hypothesis. Of course, it is supposed that $\mathbf{h}(\mathbf{Z})$ has been expanded in a Taylor series around its zeroth-order argument $\mathbf{y}$, the products and quotients expanded out, and the integration performed. Similarly, we may then uniquely obtain $\mathbf{h}$ from (B27), ω from (B28), and finally Φ from (B24).

As a matter of fact, since (B23) and (B27) do not involve Φ or ω, we can find $\mathbf{Z}$ and $\mathbf{h}$ to all orders before finding Φ and ω at all.

The construction of $\mathbf{Z}$ and Φ guarantees trivially that they satisfy the initial conditions (B20) to order n. At the same time, the construction of $\mathbf{h}$ and ω guarantees that (B25) and (B26) are satisfied to order n, and therefore, likewise the periodicity conditions (B19) for $v = 0$. But $\mathbf{Z}v$ as determined by (B21) is clearly periodic in v up to order n, since $\mathbf{Z}$ is periodic up to order $n - 1$ by induction hypothesis, and so the first periodicity condition holds in general. Similarly, it then follows that Φv is also periodic up to order n and hence that the second

periodicity condition holds in general. This completes the proof.

It is obvious by our construction that the formal series transformation from $\mathbf{y}$ and υ to $\mathbf{z}$ and ϕ expressed by (B17) is infinitely differentiable, and we can easily show that it possesses an infinitely differentiable inverse transformation. It is only necessary to obtain recursion relations for finding $\mathbf{y}$ and υ order by order, and the proof proceeds like the one just completed. The formula for $\mathbf{y}$ is merely (B23) with its last term shifted to the other side, and with the judicious choice we could have obtained the one for υ equally simply. As it is, we observe from (B24) and (B28) and from the positiveness of ψ that the function which gives ϕ to zeroth order as a function of $\mathbf{y}$ and υ can be inverted to give υ to zeroth order as a function of ϕ and $\mathbf{y}$ and hence of ϕ and $\mathbf{z}$; applying this inverse function to (B24) [i.e., using each side of (B24) in turn as argument of this inverse function and equating the results] and expanding leads to an equation whose only term of zeroth order on the right is just υ, which may then be solved for trivially to give the desired recursion relation.

9. Restatement of Formal Result

Let us restate what has been achieved, omitting the intermediate variables $\mathbf{y}$, υ and redefining $\mathbf{Z}$, Φ, and $\mathbf{X}$ accordingly. We have shown that, given the original infinitely differentiable system (B4) with all solutions nearly periodic, it is possible to find a transformation

$$\mathbf{z} = \mathbf{Z}(\mathbf{x}), \qquad \phi = \Phi(\mathbf{x}) \tag{B29}$$

and functions $\mathbf{h}(\mathbf{z})$, $\omega(\mathbf{z})$ such that the new variables satisfy (B18); moreover, the transformation possesses an inverse

$$\mathbf{x} = \mathbf{X}(\mathbf{z}, \phi), \tag{B30}$$

and $\mathbf{Z}$, Φ, $\mathbf{h}$, ω, and $\mathbf{X}$ are infinitely differentiable infinite series of non-negative powers of ϵ (the zeroth-order term of ω being definitely positive).

10. Proof That Series Represents True Solution Asymptotically

We now wish to prove that the solutions of (B18) provide asymptotically correct solutions of (B4) and also of any true system (B1) of which (B4) is the formal expansion. What this means, roughly speaking, is that if we are seeking the solution of (B1) or (B4) with a prescribed initial value at $s = 0$, and are content with an approximation to within some particular positive power of ϵ, then we may solve correspondingly approximate versions of (B18) instead, with initial values for $\mathbf{z}$ and ϕ obtained from that for $\mathbf{x}$ by a correspondingly approximate version of the transformation (B29), and convert the solutions so obtained into $\mathbf{x}$ space by a correspondingly approximate version of the inverse transformation (B30). And the claim is that we thereby obtain the desired approximate solution of (B1) or (B4), not only for s bounded but even for $s = O(1/\epsilon)$.

Owing to our desire for results valid for such large s, we must initially employ approximations up to two orders better than that of our conclusion, or than we would have needed to obtain the same conclusion restricted to bounded values of s. This is because we lose an order of accuracy, so to speak, in integrating over a large range of s, which we do twice. It is the estimation of the mobile phase angle that necessitates much of the extra accuracy; the drift of $\mathbf{z}$ alone is easier to follow, as will become clear.

To state precisely what we shall prove, let $\mathbf{x}^\dagger(s)$ be a solution of (B1) or alternatively of the system obtained from (B4) by truncating the formal series at order $n + 1$ ($n \geq 0$), so that

$$\mathbf{x}^\dagger_s = \mathbf{F}(\mathbf{x}^\dagger, \epsilon) \quad \text{or} \quad \mathbf{x}^\dagger_s = \mathbf{f}^{[n+1]}(\mathbf{x}^\dagger); \tag{B31}$$

these are of course genuine systems of differential equations, and not merely formal. Define $\mathbf{z}'(s)$ and $\phi'(s)$ as those solutions of the equations obtained from (B18) by appropriate analogous truncations of $\mathbf{h}$ and ω which have at $s = 0$ the values obtained from $\mathbf{x}^\dagger(0)$ by an appropriately truncated version of transformation (B29); specifically, $\mathbf{z}'(s)$ and $\phi'(s)$ are uniquely defined by the conditions

$$\mathbf{z}'_s = \epsilon\, \mathbf{h}^{[n]}(\mathbf{z}'), \qquad \phi'_s = \omega^{[n]}(\mathbf{z}'), \tag{B32}$$

$$\mathbf{z}'(0) = \mathbf{Z}^{[n]}(\mathbf{x}^\dagger(0)), \qquad \phi'(0) = \Phi^{[n]}(\mathbf{x}^\dagger(0)). \tag{B33}$$

Next, transform to obtain a curve in $\mathbf{x}$ space, defining $\mathbf{x}'(s)$ as the result of applying the appropriately truncated version of transformation (B30) to $\mathbf{z}'(s)$ and $\phi'(s)$:

$$\mathbf{x}' = \mathbf{X}^{[n-1]}(\mathbf{z}', \phi'). \tag{B34}$$

The claim now is that $\mathbf{x}'$ is a good approximation to $\mathbf{x}^\dagger$, specifically that

$$\mathbf{x}^\dagger(s) = \mathbf{x}'(s) + O(\epsilon^n), \tag{B35}$$

for s within a range of order $1/\epsilon$, so long as $\mathbf{x}^\dagger$ and $\mathbf{x}'$ stay in the domain of definition of $\mathbf{f}$. [For all we know yet this might limit s more severely, but in the next section it will be shown that they do stay in for $s = O(1/\epsilon)$.]

The proof proceeds in three stages. First we show that $\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)$ is close to $\mathbf{z}'$, then that $\Phi^{[n]}(\mathbf{x}^\dagger)$ is close to ϕ', and finally that $\mathbf{x}'$ is close to $\mathbf{x}^\dagger$, as desired. We need the three formal identities in $\mathbf{x}$

$$\mathbf{Z}_\mathbf{x}\cdot\mathbf{f} = \epsilon\,\mathbf{h}(\mathbf{Z}), \tag{B36}$$

$$\Phi_\mathbf{x}\cdot\mathbf{f} = \omega(\mathbf{Z}), \tag{B37}$$

$$\mathbf{x} = \mathbf{X}(\mathbf{Z}, \Phi), \tag{B38}$$

the first two of which express that (B29) transforms (B4) into (B18), and the last that (B30) inverts (B29). Or rather, we need their approximate versions

$$\mathbf{Z}^{[n+1]}_\mathbf{x}\cdot\mathbf{f}^{[n+1]} = \epsilon\,\mathbf{h}^{[n]}(\mathbf{Z}^{[n+1]}) + O(\epsilon^{n+2}), \tag{B39}$$

$$\Phi^{[n]}_\mathbf{x}\cdot\mathbf{f}^{[n+1]} = \omega^{[n]}(\mathbf{Z}^{[n]}) + O(\epsilon^{n+1}), \tag{B40}$$

$$\mathbf{x} = \mathbf{X}^{[n-1]}(\mathbf{Z}^{[n-1]}, \Phi^{[n-1]}) + O(\epsilon^n), \tag{B41}$$

the last of which, for instance, may be shown as follows: The difference between the left-hand side and the first term on the right is a well-defined infinitely differentiable function of $\mathbf{x}$ and ϵ. Its Taylor expansion around $\epsilon = 0$ vanishes identically up to terms of order $n - 1$, since it contains exactly the same terms up to that order as occur in the formal identity (B38). Thus, the quotient of the difference by ϵ^n is a continuous function of $\mathbf{x}$ and ϵ on a closed bounded domain and is therefore bounded independently of $\mathbf{x}$ and ϵ, which completes the proof. The same argument, but with n replaced everywhere by $n + 2$ or $n + 1$, applies to (B39) and (B40).

If we temporarily restrict ourselves to the second alternative in (B31), then (B39) evaluated at $\mathbf{x}^\dagger$ may be written

$$\{\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)\}_s = \epsilon\,\mathbf{h}^{[n]}(\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)) + O(\epsilon^{n+2}). \tag{B42}$$

We now wish to compare this with the first equation (B32), and invoke the approximation theorem of Appendix 1. The crux of the proof is to change the independent variable first, setting

$$\sigma = \epsilon\, s. \tag{B43}$$

The two equations we wish to compare become thereby (after division by ϵ)

$$\{\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)\}_\sigma = \mathbf{h}^{[n]}(\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)) + O(\epsilon^{n+1}), \tag{B44}$$

$$\mathbf{z}'_\sigma = h^{[n]}(\mathbf{z}'). \tag{B45}$$

Thus $\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger)$ and $\mathbf{z}'$ satisfy the same autonomous system of differential equations within $O(\epsilon^{n+1})$, and since by (B33) they also have the same initial values within $O(\epsilon^{n+1})$, according to the approximation theorem

$$\mathbf{Z}^{[n+1]}(\mathbf{x}^\dagger) = \mathbf{z}' + O(\epsilon^{n+1}) \tag{B46}$$

for bounded σ, i.e., for $s = O(1/\epsilon)$. (In this, we might as well replace the truncation index $n + 1$ on the left by n.) This completes the first stage, except for the remark that the case of the first alternative in (B31) may be treated in exactly the same way by using, instead of (B39), the obviously equally valid formula obtained from it by replacing the factor $\mathbf{f}^{[n+1]}$ by $\mathbf{F}$.

The second stage is even simpler. Restricting ourselves again to the second alternative (with the same trivial modification to cover the first alternative), we evaluate (B40) at $\mathbf{x}^\dagger$ and write it in the form

$$\begin{aligned}\{\Phi^{[n]}(\mathbf{x}^\dagger)\}_s &= \omega^{[n]}(\mathbf{Z}^{[n]}(\mathbf{x}^\dagger)) + O(\epsilon^{n+1}) \\ &= \omega^{[n]}(\mathbf{z}') + O(\epsilon^{n+1}) = \phi'_s + O(\epsilon^{n+1}),\end{aligned} \tag{B47}$$

where we have used (B46) and the second Eq. (B32). In view of the common initial values, asserted in the second Eq. (B33), a simple integration immediately yields

$$\Phi^{[n]}(\mathbf{x}^\dagger) = \phi' + O(\epsilon^n) \tag{B48}$$

for $s = O(1/\epsilon)$. (Again, we might as well replace the truncation index n on the left by $n - 1$.)

The third stage is the simplest of all. Evaluating (B41) at $\mathbf{x}^\dagger$ and using (B46) and (B48) yields

$$\mathbf{x}^\dagger = \mathbf{X}^{[n-1]}(\mathbf{z}', \phi') + O(\epsilon^n), \tag{B49}$$

which, in view of (B34), is precisely (B35). The proof is complete.

11. Proof that the True Solution Remains in Domain for Many Gyrations

We can now demonstrate that $\mathbf{x}^\dagger(s)$ actually remains interior to the given domain of $\mathbf{x}$ space over a large range in s of order $1/\epsilon$. For (B45) shows that there is some finite range of σ over which $\mathbf{z}'$ stays interior to the corresponding domain of $\mathbf{z}$ space. This corresponds by (B43) to some definite large range of s, to which we restrict the discussion from now on. [Since ϕ' is an angle-like variable it is subject to no restriction in magnitude, and to be sure it varies by $O(1/\epsilon)$, as we see from (B32).] From (B34) it is then clear that $\mathbf{x}'$ is defined and interior to the domain, so (B35) holds so long as $\mathbf{x}^\dagger$ stays inside. On the other hand, so long as (B35) holds, $\mathbf{x}^\dagger$ does in fact stay in. Pulling ourselves up by the bootstraps in this way, we arrive at the conclusion that $\mathbf{x}^\dagger$ stays inside as long as $\mathbf{x}'$ does (at least very nearly, though naturally one may reach the boundary a little before the other).

If this argument does not appear immediately

convincing (as may well be the case), it can be elaborated as follows. Knowing that $\mathbf{x}'$ stays well within the domain for some range of s of order $1/\epsilon$, that is, stays a finite distance δ away from the boundary, let us assume (for the sake of a *reductio ad absurdum*) that $\mathbf{x}^\dagger$ leaves the domain during that range. Consider the value of s nearest to zero for which $\mathbf{x}^\dagger$ and $\mathbf{x}'$ are at a distance of $\delta/2$. Since $\mathbf{x}^\dagger$ and $\mathbf{x}'$ are both within the domain up to that point, (B35) applies and shows that they are separated by only $O(\epsilon^n)$, which is much less than $\delta/2$. This contradiction proves that $\mathbf{x}^\dagger$, in fact, could not have left the domain during the range of interest.

C. RINGS AND THE REDUCED SYSTEM

1. Theorem of Phase Independence

At this point we introduce a simple but powerful tool, which we shall apply repeatedly to establish that functions are independent of the phase angle ϕ. Let $\mathbf{W}(\mathbf{z}, \phi)$ be a vector with any finite number M of components, a formal infinite series in powers of ϵ, $O(1)$, and a function of its indicated arguments $\mathbf{z}$ and ϕ, periodic in the latter (of course with period unity). Let $\mathbf{W}$ satisfy a formal differential equation

$$\mathbf{W}_\phi = \mathbf{A}(\mathbf{z}) + \epsilon\, \mathbf{B}(\mathbf{z}, \mathbf{W}, \mathbf{W}_\mathbf{z}), \tag{C1}$$

where $\mathbf{A}(\mathbf{z})$ and $\mathbf{B}(\mathbf{z}, \mathbf{w}, \mathbf{v})$ are vectors of M components, formal infinite series in powers of ϵ, and functions of their indicated arguments [$\mathbf{w}$ being a generic variable with M components, and $\mathbf{v}$ one with $M(N-1)$ components], and where $\mathbf{B}$ is $O(1)$, but no such assumption is made about $\mathbf{A}$. Then $\mathbf{W}$ is in fact independent of ϕ to all orders.

To prove this *theorem of phase independence*, we merely form the indefinite integral of (C1),

$$\mathbf{W}(\mathbf{z}, \phi) = \mathbf{W}(\mathbf{z}, 0) + \int_0^\phi d\phi\, [\mathbf{A}(\mathbf{z}) + \epsilon\, \mathbf{B}(\mathbf{z}, \mathbf{W}, \mathbf{W}_\mathbf{z})]; \tag{C2}$$

impose the periodicity condition on $\mathbf{W}$,

$$\int_0^1 d\phi\, [\mathbf{A}(\mathbf{z}) + \epsilon\, \mathbf{B}(\mathbf{z}, \mathbf{W}, \mathbf{W}_\mathbf{z})] = 0; \tag{C3}$$

and proceed by induction. Up to order minus one $\mathbf{W}$ vanishes and so is independent of ϕ. Suppose $\mathbf{W}$ is independent of ϕ up to terms of order $n-1$. Up to order n, then, the integrand of (C3) is independent of ϕ, and so equals its own integral, which vanishes. Therefore (C2) may be written

$$\mathbf{W}(\mathbf{z}, \phi) = \mathbf{W}(\mathbf{z}, 0) + O(\epsilon^{n+1}), \tag{C4}$$

which shows that $\mathbf{W}$ is independent of ϕ up to terms of order n, and thereby completes the induction. (And we have as a side result that $\mathbf{A} = O(\epsilon)$, and in fact that $\mathbf{A} + \epsilon\mathbf{B} = 0$ to all orders.)

It is clear that we could have allowed $\mathbf{B}$ to depend on higher derivatives of $\mathbf{W}$ with respect to $\mathbf{z}$, with no change in the argument or the conclusion. Of this we shall make no use, but there is a slight modification of the theorem which we shall use very frequently. Namely, let us specialize by assuming that $\mathbf{B}$ is a linear function of its last two arguments. We can then trivially generalize by allowing $\mathbf{W}$ to start with any (e.g., a negative) power of ϵ, instead of being $O(1)$, and obviously the same conclusion holds.

Another slight but useful modification is obtained by replacing $\mathbf{W}_\phi$ by $\mathbf{W}_s$ in (C1). Since

$$\mathbf{W}_s = \mathbf{W}_\phi\, \omega(\mathbf{z}) + \epsilon\, \mathbf{W}_\mathbf{z} \cdot \mathbf{h}(\mathbf{z}), \tag{C5}$$

the resulting equation can be immediately put back in the same form as (C1) with $\mathbf{A}$ and $\mathbf{B}$ appropriately modified, so the conclusion still holds.

It is convenient to observe here, what will be needed later and is obvious from (C5), that the operators of differentiation with respect to s and to ϕ commute with each other.

2. Arbitrariness of Nice Variables

It is clear that the nice variables $\mathbf{z}$ and ϕ obtained in part B are far from unique. For, the process by which we determined them involved some arbitrary choices, namely of the functions $\mathbf{Y}$ and Υ in Sec. B.5 and of the form of the initial conditions (B20). We address ourselves now to the investigation of exactly how much freedom there is finally in the choice of such desirable coordinates.

Suppose that we were to introduce new variables $\mathbf{z}^\dagger$ and $\phi^\dagger$, expressed as infinitely differentiable formal infinite series in non-negative powers of ϵ, with coefficients depending on $\mathbf{z}$ and ϕ, such that $\mathbf{z}^\dagger$ is periodic in ϕ and $\phi^\dagger$ changes by unity with ϕ, and such that the transformation is formally invertible to give similar expressions for $\mathbf{z}$ and ϕ in terms of $\mathbf{z}^\dagger$ and $\phi^\dagger$. We then may ask, first, under what reasonably general conditions $\mathbf{z}^\dagger$ and $\phi^\dagger$ will necessarily satisfy equations analogous to (B18).

An answer is easy to find. For $\mathbf{z}$ satisfies an autonomous system of equations, and this will certainly transform into an autonomous system for $\mathbf{z}^\dagger$ so long as $\mathbf{z}^\dagger$ is obtained by a transformation from $\mathbf{z}$ alone, i.e., so long as

$$\mathbf{z}^\dagger_\phi = 0. \tag{C6}$$

[Indeed, it is then immediate that $\mathbf{z}^\dagger_s = \epsilon\, \mathbf{h}^\dagger(\mathbf{z}^\dagger)$,

where $\mathbf{h}^\dagger(\mathbf{z}^\dagger) = \mathbf{z}^\dagger_{\mathbf{z}} \cdot \mathbf{h}(\mathbf{z})$.] Furthermore, ϕ may be thought of as satisfying the trivial first-order differential equation $\phi_s = \omega(\mathbf{z})$ if $\mathbf{z}$ is considered a parameter, and $\phi^\dagger$ will certainly satisfy a similar trivial equation so long as it differs from ϕ by only a "constant," which may of course depend on the "parameter" $\mathbf{z}$ (we cannot take $\phi^\dagger$ to be a more general linear function of ϕ, since the two angle variables must change by unity together), namely, so long as

$$\phi^\dagger_\phi = 1. \tag{C7}$$

[Indeed, it is then immediate that $\phi^\dagger_s = \omega^\dagger(\mathbf{z}^\dagger)$, where $\omega^\dagger(\mathbf{z}^\dagger) = \epsilon\, \phi^\dagger_{\mathbf{z}} \cdot \mathbf{h}(\mathbf{z}) + \omega(\mathbf{z})$; note that $\phi^\dagger_{\mathbf{z}}$ depends on $\mathbf{z}$ alone, since $\phi^\dagger_{\phi\mathbf{z}} = 0$.]

We next may ask, conversely, whether every transformation from $\mathbf{z}$, ϕ to new nice variables $\mathbf{z}^\dagger$, $\phi^\dagger$ in fact satisfies (C6) and (C7). The answer is yes; the conditions we have found are necessary as well as sufficient to preserve the form of (B18). For to begin with, suppose there exists a function $\mathbf{h}^\dagger$ such that

$$\mathbf{z}^\dagger_s = \epsilon\, \mathbf{h}^\dagger(\mathbf{z}^\dagger); \tag{C8}$$

then we may apply the theorem of phase independence and deduce (C6). And suppose also that there exists a function $\omega^\dagger$ such that

$$\phi^\dagger_s = \omega^\dagger(\mathbf{z}^\dagger); \tag{C9}$$

to this we cannot directly apply the theorem of phase independence (even though we know that $\mathbf{z}^\dagger$ depends on $\mathbf{z}$ only), because $\phi^\dagger$ is not periodic in ϕ, but we may rewrite (C9) as

$$[\phi^\dagger - \phi]_s = \omega^\dagger(\mathbf{z}^\dagger) - \omega(\mathbf{z}), \tag{C10}$$

to which we can, because $\phi^\dagger - \phi$ is periodic; we thereby deduce that

$$[\phi^\dagger - \phi]_\phi = 0, \tag{C11}$$

which is equivalent to (C7).

3. Rings, Phase Differences, and the Reduced System

The necessary and sufficient conditions (C6) and (C7) that new variables $\mathbf{z}^\dagger$, $\phi^\dagger$ be as nice as $\mathbf{z}$, ϕ enable us to introduce some simple concepts connected with the original system (B4) in $\mathbf{x}$ which are based on the existence of nice variables, but are independent of which ones they specifically are. These concepts will of course be defined only to all orders, and not in general exactly.

Consider the set of all points in $\mathbf{x}$ space which correspond to a common value of $\mathbf{z}$. Since ϕ is an angle variable, this set is a topological circle which we call a *ring*. Two points in $\mathbf{x}$ space with the same $\mathbf{z}$ we call *ringmates*. Now (C6) has the simple interpretation that $\mathbf{z}^\dagger$ is constant around any ring, which shows that the concept of rings and ringmates is invariant with respect to the choice of nice variables.

The particular value of the phase variable ϕ corresponding to a point $\mathbf{x}$ has of course no invariant significance. But if two points are ringmates, then their *phase difference* $\Delta\phi$ does have, for, by (C7),

$$\Delta\phi^\dagger = \int d\phi^\dagger = \int \phi^\dagger_\phi\, d\phi = \int d\phi = \Delta\phi, \tag{C12}$$

where the integrals are extended over a portion of the ring between the two points. Of course this phase difference is only defined modulo unity.

It is also clear that the knowledge of which points are ringmates and of the phase differences between those that are is all the information inherent in knowing some nice variables but independent of which ones they are. For from that knowledge one can construct nice variables.

Since $\mathbf{z}$ satisfies an autonomous system, it is obvious that if two solutions $\mathbf{x}(s)$ of the original system are ringmates at any one value of s, then they are so at every value of s. Also, their phase difference is always the same, since $(\Delta\phi)_s = \omega(\mathbf{z}) - \omega(\mathbf{z}) = 0$. It is therefore meaningful to apply the invariant concepts of rings and of phase difference not only to points in $\mathbf{x}$ space but also to entire solutions $\mathbf{x}(s)$ of (B4).

The ring constitutes a sophisticated or refined version of the loop, which is indeed a zeroth-order approximation to it. Each ring is characterized by its value of $\mathbf{z}$, just as each loop was characterized (in Sec. B.5) by its value of $\mathbf{y}$. That is, the rings form an $(N-1)$-dimensional space, and their slow drift as s varies is governed by the formal autonomous system

$$\mathbf{z}_\sigma = \mathbf{h}(\mathbf{z}), \tag{C13}$$

where we have reintroduced the change of independent variable (B43) to $\sigma = \epsilon s$ in order that what we shall call the *reduced system* (C13) might be in the same general form as the original system (B4) (though not necessarily in standard form). The phase now plays a subordinate role (in contrast to its previous rather dominant one), since it has no effect (to any finite order) on the drift.

Among other things, we have shown that the nearly recurrent system (B4) is splittable. The new

autonomous system $\mathbf{z}_s = \epsilon\,\mathbf{h}(\mathbf{z})$ contains one less variable.

If (C13) should itself happen to be nearly recurrent and in standard form, we can reapply the whole procedure to it and obtain a doubly reduced system, and so on. The arguments of Sec. B.11 can then be extended to show that the true solution $\mathbf{x}^\dagger(s)$ of (B31) remains in the relevant domain of $\mathbf{x}$, space for a range of s even larger than $O(1/\epsilon)$.

4. Roto-rate

It should at this point be apparent that all the information contained in knowledge of the rings and of phase differences on them is contained, in "capsule" form, in knowledge of $\mathbf{x}_\phi$. For, as we show in detail shortly, the rings are obtainable from $\mathbf{x}_\phi$ by integration of an autonomous system whose independent variable is the phase. It is noteworthy that in terms of $\mathbf{x}_\phi$ as a function[18] of $\mathbf{x}$ we may express our results entirely within $\mathbf{x}$ space, even though it was convenient to transform to other variables in deriving them; so expressed, furthermore, the results take on an esthetically satisfying uniqueness, since for any other nice coordinates $\mathbf{z}^\dagger$, $\phi^\dagger$ we have

$$\mathbf{x}_\phi = \mathbf{x}_{\mathbf{z}\dagger}\cdot\mathbf{z}^\dagger_\phi + \mathbf{x}_{\phi\dagger}\,\phi^\dagger_\phi = \mathbf{x}_{\phi\dagger} \tag{C14}$$

by (C6) and (C7). The function merits a specific name and symbol, and so we define the *roto-rate* $\mathbf{R}(\mathbf{x})$ by

$$\mathbf{R} \equiv \mathbf{x}_\phi, \tag{C15}$$

the name being supposed to suggest the defining characteristic of $\mathbf{R}$ as the rate of change (with respect to phase) of the point at $\mathbf{x}$ in rotating around its ring.

We next derive four properties of $\mathbf{R}$ which can be expressed independently of nice coordinates, rings, and phases, and show afterward that these properties characterize it uniquely. The first property is that the autonomous system (C15) is recurrent; this is obvious because every solution in fact runs around a ring. The second property (also obvious) is that the initial value recurs (for the first time) when the independent variable ϕ has increased by unity; this may be expressed by writing

$$\oint \frac{d\mathbf{x}}{\mathbf{R}} = 1, \tag{C16}$$

where the line integral is taken once around any integral curve of $\mathbf{R}$, and the ratio of vectors is defined, since precisely on such a curve the vectors are parallel (and, in fact, $d\mathbf{x} = \mathbf{x}_\phi\,d\phi = \mathbf{R}\,d\phi$). The third property is that to lowest (zeroth) order, $\mathbf{R}$ is parallel to $\mathbf{f}$ and unequal to zero, as seen immediately from (B4) and from (C5) with $\mathbf{W}$ replaced by $\mathbf{x}$. Finally, the fourth property is that

$$\mathbf{f}_\mathbf{x}\cdot\mathbf{R} = \mathbf{R}_\mathbf{x}\cdot\mathbf{f}, \tag{C17}$$

which expresses the consistency of (B4) with (C15) and is, in fact, a trivial rewriting of

$$\mathbf{f}_\mathbf{x}\cdot\mathbf{x}_\phi = \mathbf{f}_\phi = \dot{\mathbf{x}}_{s\phi} = \mathbf{x}_{\phi s} = \mathbf{R}_s = \mathbf{R}_\mathbf{x}\cdot\mathbf{x}_s, \tag{C18}$$

in which we have used the observation at the end of Sec. C.1.

To show that the roto-rate is uniquely characterized by these four properties, it suffices to show that for any (infinitely differentiable formal series) function $\mathbf{R}$ with these properties we can find nice variables $\mathbf{z}$ and ϕ such that (C15) holds. This is because for any other function $\mathbf{R}^\dagger$ with the four properties we could then find nice variables $\mathbf{z}^\dagger$ and $\phi^\dagger$ such that $\mathbf{R}^\dagger = \mathbf{x}_{\phi\dagger}$, whereupon (C14) would show immediately that $\mathbf{R} = \mathbf{R}^\dagger$.

We may construct the desired nice variables as solutions of

$$\mathbf{z}_\mathbf{x}\cdot\mathbf{R} = 0, \tag{C19}$$

$$\phi_\mathbf{x}\cdot\mathbf{R} = 1; \tag{C20}$$

that is, the $\mathbf{z}$ are constructed to be (of course independent) constants on the closed (first property!) integral curves of $\mathbf{R}$, and ϕ to be multi-valued. That ϕ is an angle variable, i.e., that its branches differ by integers, follows from (C16) (second property!),

$$\oint d\phi = \oint \phi_\mathbf{x}\cdot d\mathbf{x} = \oint \phi_\mathbf{x}\cdot\mathbf{R}\,d\mathbf{x}/\mathbf{R} = 1. \tag{C21}$$

The coordinates cannot be degenerate, since the $\mathbf{z}$ were chosen independent and, if ϕ were expressible as a function of $\mathbf{z}$, we would have $\phi_\mathbf{x}\cdot\mathbf{R} = 0$ as a consequence of (C19), contradicting (C20). We may therefore employ a standard representation of the unit dyadic in the derivation

$$\mathbf{R} = \mathcal{I}\cdot\mathbf{R} = (\mathbf{x}_\mathbf{z}\cdot\mathbf{z}_\mathbf{x} + \mathbf{x}_\phi\,\phi_\mathbf{x})\cdot\mathbf{R} = \mathbf{x}_\phi, \tag{C22}$$

which verifies (C15). It is now fairly easy to establish one of the main properties of nice coordinates, that $\mathbf{z}_s$ and ϕ_s are functions of $\mathbf{z}$ only[19]:

[18] Of course, $\mathbf{x}$ is to be treated as a function of ϕ and $\mathbf{z}$ through the transformation (B30) when the differentiation is performed, but the derivative may then be treated as a function of $\mathbf{x}$ by the inverse transformation (B29).

[19] In this paragraph we cannot appeal to the commutativity of the differentiations, which would enable us to bypass several intermediate steps in (C23), because we do not yet know that $\mathbf{z}$, ϕ are nice coordinates, that being, indeed, what we wish to establish.

$$(\mathbf{z}_s)_\phi = (\mathbf{z}_x\cdot\mathbf{f})_x\cdot\mathbf{R} = \mathbf{z}_{xx}:\mathbf{f}\,\mathbf{R} + \mathbf{z}_x\cdot\mathbf{f}_x\cdot\mathbf{R}$$
$$= \mathbf{z}_{xx}:\mathbf{R}\,\mathbf{f} + \mathbf{z}_x\cdot\mathbf{R}_x\cdot\mathbf{f} \qquad \text{(C23)}$$
$$= (\mathbf{z}_x\cdot\mathbf{R})_x\cdot\mathbf{f} = 0,$$

using successively (B4), (C15), the symmetry of the triadic $\mathbf{z}_{xx}$ in its last two vector positions, (C17) (fourth property!), and (C19); and the same for ϕ_s. This already shows that ϕ_s may be written as $\omega(\mathbf{z})$; to show that $\mathbf{z}_s$ may be written as $\epsilon\mathbf{h}(\mathbf{z})$ it remains only to note that $\mathbf{z}_x\cdot\mathbf{f} = O(\epsilon)$, in view of (C19) (third property!).

It does not appear obvious whether an explicit recursion formula to determine $\mathbf{R}$ in terms of $\mathbf{f}$ can be found. If so, the whole theory of this paper might be simplified and rendered less deep.

D. HEREDITARY PROPERTIES

1. Definition

We now have a systematic formal procedure which we can apply to any nearly recurrent formal system in N dependent variables, enabling us to "take out the fast gyration" and find a reduced system in $N - 1$ dependent variables which describes the drift of rings to all orders. There are, as we shall see, a number of properties of systems of differential equations which, if possessed by the original system (B4), are carried over or *inherited* by the reduced system (C13). We shall call a property of systems of differential equations *hereditary* if, whenever it holds for the original system (at least in the formal sense of expansions to all orders), it also holds for the reduced system. This definition, like that of local dependence, is metamathematical and need not be used with great strictness.

2. Inheritance of Splittability

As a first, fairly trivial, example of an hereditary property consider that of being splittable (Sec. A.7). If (B4) is splittable, and if the functionals which depend locally on $\mathbf{f}(\mathbf{x})$ and define the M new variables satisfying the new autonomous system are expansible in power series in ϵ, with zero-order leading terms, then the new autonomous system obviously also has all solutions nearly periodic and can be put into standard form by an appropriate change of variables. Introducing new nice variables, it can be seen by the methods of part C that the new angle variable is essentially the same as (and could have been chosen to be in fact the same as) the original angle variable ϕ, and thus that the reduced system of the new autonomous system is the new autonomous system of the original reduced system (C13). Therefore the original reduced system is splittable, and the decrease in number of dependent variables as a result of the splitting is the same as for the given splittable system (B4).

3. Inheritance of an Integral (Constant of Motion)

An important and simple hereditary property is the possession of an integral (independent of s). To see this let $I(\mathbf{x})$ be an (expansible) integral of (B4); that is, let

$$I_s = I_x\cdot\mathbf{f} = 0, \qquad \text{(D1)}$$

so that I is constant on every solution of (B4). By the theorem of phase independence we immediately deduce that I is a function of $\mathbf{z}$ only, and is therefore also an integral of the reduced system (C13).

It follows from this that (B4) cannot possess a complete set of integrals (N functionally independent integrals) all of which are expansible. For (C13) can of course have only $N - 1$ independent integrals. Any integral of (B4) which distinguishes between ringmates is inexpansible.

4. Inheritance of Invariant Measure

Another interesting hereditary property is possession of an invariant measure. Let $\rho(\mathbf{x})$ be the measure density of a measure which is invariant under the "flow" represented by (B4); that is, the measure of any region in $\mathbf{x}$ space is the integral of ρ over that region, and the measures of the images of that region for different values of s under the "motion" (B4) are all equal. The mathematical expression of this invariance is just the "time-independent" equation of continuity

$$\mathcal{J}:(\rho\,\mathbf{f})_\mathbf{x} = 0, \qquad \text{(D2)}$$

where the left-hand side is nothing but the divergence of $\rho\,\mathbf{f}$, written in accordance with our subscript-differentiation notation. This may also be written

$$\rho_s + \rho\,\mathcal{J}:\mathbf{f}_\mathbf{x} = 0. \qquad \text{(D3)}$$

Such an invariant measure is intrinsically independent of coordinate system. In $\mathbf{z}, \phi$ space it has the measure density $\tau(\mathbf{z}, \phi)$ obtained from ρ by multiplying by the Jacobian of $\mathbf{x}$ with respect to $\mathbf{z}, \phi$. We then have (D2) and (D3) in the forms

$$\mathcal{J}:(\tau\,\epsilon\,\mathbf{h})_\mathbf{z} + (\tau\,\omega)_\phi = 0, \qquad \text{(D4)}$$

$$\tau_s + \epsilon\,\tau\,\mathcal{J}:\mathbf{h}_\mathbf{z} = 0. \qquad \text{(D5)}$$

If ρ is expansible in a power series then of course so is τ (since the Jacobian certainly is), hence by the

theorem of phase independence τ is a function of $\mathbf{z}$ alone. Thus $(\tau\omega)_\phi = 0$, and (D4) becomes for the reduced system the precise analog of (D2) for the original system, which proves the inheritance.

5. General Integral Invariants

The last two hereditary properties discussed are essentially special cases of a more general one, possession of an integral invariant[20] of any degree m. They correspond in fact to the extreme cases $m = 0$ and $m = N$, whose particular simplicity seemed to merit separate treatment. We confine ourselves in this section to the case $m = 2$, which is perhaps uniquely characterized as being enough like the general case to permit immediate generalization[21] of all concepts, results, and non-notational features, yet simple enough to permit handling in our index-free notation with no further special conventions and no particular clumsiness. However, we consider not only absolute integral invariants but also relative integral invariants.

Let $\mathfrak{A}(\mathbf{x})$ be an antisymmetric dyadic[22] field in $\mathbf{x}$ space. Then $\mathfrak{A} : d\mathbf{x}\, d\mathbf{x}$ is a second-order differential form[23] whose integral over any surface imbedded in $\mathbf{x}$ space constitutes a sort of measure for such surfaces. The rate of change of this measure with respect to s, as (each point of) the surface varies with s according to (B4), is another measure similarly associated with the differential form

$$(\mathfrak{A} : d\mathbf{x}\, d\mathbf{x})_s = \mathfrak{A}_s : d\mathbf{x}\, d\mathbf{x} + \mathfrak{A} : (d\mathbf{f}\, d\mathbf{x} + d\mathbf{x}\, d\mathbf{f}) = (\mathfrak{A}_s + \mathfrak{A}\cdot\mathbf{f}_\mathbf{x} + {}_\mathbf{x}\mathbf{f}\cdot\mathfrak{A}) : d\mathbf{x}\, d\mathbf{x}. \qquad \text{(D6)}$$

The condition that $\mathfrak{A} : d\mathbf{x}\, d\mathbf{x}$ be a *relative integral invariant* is that (D6) vanish upon integration over every closed surface, the condition for which in turn is that there exist a single-valued vector field $\mathbf{V}(\mathbf{x})$ satisfying

$$\mathfrak{A}_s + \mathfrak{A}\cdot\mathbf{f}_\mathbf{x} + {}_\mathbf{x}\mathbf{f}\cdot\mathfrak{A} = \mathbf{V}_\mathbf{x} - {}_\mathbf{x}\mathbf{V}, \qquad \text{(D7)}$$

where the right-hand side is just the curl of $\mathbf{V}$. In this case the measure of any closed surface stays constant as the surface flows. The condition that $\mathfrak{A} : d\mathbf{x}\, d\mathbf{x}$ be an *absolute integral invariant* is that (D6) vanish upon integration over every (not necessarily closed) surface, namely that the left-hand side of (D7) vanish. In this case the measure of any surface at all stays constant as the surface flows.

6. Inheritance of Relative Invariants

Transforming to nice coordinates we have

$$\mathfrak{A} : d\mathbf{x}\, d\mathbf{x} = \mathfrak{A} : (\mathbf{x}_z\cdot dz + \mathbf{x}_\phi\, d\phi)(\mathbf{x}_z\cdot dz + \mathbf{x}_\phi\, d\phi) = \mathfrak{B} : dz\, dz + \mathbf{C}\cdot(dz\, d\phi - d\phi\, dz), \qquad \text{(D8)}$$

where there can of course be no $d\phi\, d\phi$ term (and its coefficient would vanish anyway), and where

$$\mathfrak{B} \equiv {}_z\mathbf{x}\cdot\mathfrak{A}\cdot\mathbf{x}_z\ , \quad \mathbf{C} \equiv \mathbf{x}_\phi\cdot\mathfrak{A}\cdot\mathbf{x}_z\ . \qquad \text{(D9)}$$

The condition that (D8) be a relative invariant, obtainable in the same way as (D7) or by transforming (D7), is that there exist single-valued $\mathbf{W}(\mathbf{x})$ and $T(\mathbf{x})$ such that

$$\mathfrak{B}_s + \mathfrak{B}\cdot\epsilon\,\mathbf{h}_z + \epsilon\,{}_z\mathbf{h}\cdot\mathfrak{B} - \mathbf{C}\,\omega_z + \omega_z\,\mathbf{C} = \mathbf{W}_z - {}_z\mathbf{W}, \qquad \text{(D10)}$$

$$\mathbf{C}_s + \mathbf{C}\cdot\epsilon\,\mathbf{h}_z = T_z - \mathbf{W}_\phi. \qquad \text{(D11)}$$

Let us denote the complete integral over phase (equivalently, the closed line integral around a ring, or the average value with respect to ϕ) by an asterisk. Then from (D11) we have immediately (since differentiations and integrations with respect to ϕ and to s commute)

$$\mathbf{C}^*_s + \mathbf{C}^*\cdot\epsilon\,\mathbf{h}_z = T^*_z. \qquad \text{(D12)}$$

This is just the condition [analogous to (D7), but of order one less] that $\mathbf{C}^*\cdot dz$ be a relative integral invariant of the reduced system. If $\mathfrak{A}$ and $\mathbf{V}$ are expansible in ϵ then so are $\mathbf{C}$ and T and so also $\mathbf{C}^*$ and T^*. Thus the inheritance of a relative invariant is proved.

Of course $\mathbf{C}^*\cdot dz$ might happen to be a trivial relative invariant, in the sense of its integral having a constant value independent of the closed path

---

[20] See De Donder [Th. De Donder, *Théorie des invariants intégraux* (Gauthier-Villars, Paris, 1927)] for the definitions, for a clear exposition of the general theory of integral invariants, and accordingly for a justification of some of the steps in this section. Strictly speaking, the hereditary property is possession of an integral invariant whose degree is less than the order of the system by a definite integer, since an invariant of degree m of the original system induces one of degree $m - 1$ of the reduced system. The word "essentially" above is to indicate that from this point of view the property of Sec. 3 is slightly anomolous.

[21] The important case $m = 1$ may be even more immediately obtained by the opposite process (here "degradation" rather than specialization), and moreover is illustrated by the theory of Hamiltonian systems in part E (as also is the case $m = 2$ for a relative invariant), where the invariant has the special form $\mathbf{p}\cdot d\mathbf{q}$.

[22] In the general case it should be a completely antisymmetric m-adic, which changes sign under the transposition of any two "vector locations" (in index notation a tensor of order m which changes sign under the interchange of any two indices).

[23] It should be noted that $d\mathbf{x}\, d\mathbf{x}$ (often written $d\mathbf{x} \wedge d\mathbf{x}$ in the literature) is, despite appearances, neither symmetric nor a dyad, but rather an antisymmetric dyadic. Indeed, since $\mathfrak{A}$ is antisymmetric, any symmetric component of $d\mathbf{x}\, d\mathbf{x}$ would contribute nothing to the differential form. The interchange of vector order reverses the orientation of the surface element $d\mathbf{x}\, d\mathbf{x}$ and with it the sign. In fact, on any surface determined by giving $\mathbf{x}$ as a function of two parameters u and v, we have $d\mathbf{x}\, d\mathbf{x} = (\mathbf{x}_u\,\mathbf{x}_v - \mathbf{x}_v\,\mathbf{x}_u)\, du\, dv$.

it is taken over. This is the case only if $\mathbf{C}^*$ is a gradient (i.e., a $\mathbf{z}$ derivative), so that its curl vanishes,

$$\mathbf{C}^*_{\mathbf{z}} - {}_{\mathbf{z}}\mathbf{C}^* = 0. \tag{D13}$$

But then we are amusingly compensated in that the phase integral of (D10) can be written

$$\mathfrak{B}^*_s + \mathfrak{B}^* \cdot \epsilon \mathbf{h}_{\mathbf{z}} + \epsilon\, {}_{\mathbf{z}}\mathbf{h} \cdot \mathfrak{B}^* = (\mathbf{W}^* + \mathbf{C}^* \omega)_{\mathbf{z}} - {}_{\mathbf{z}}(\mathbf{W}^* + \mathbf{C}^* \omega), \tag{D14}$$

which is just the condition [analogous to (D7)] that $\mathfrak{B}^* : d\mathbf{z}\, d\mathbf{z}$ be a relative integral invariant of the reduced system.

7. Inheritance of Absolute Invariants

There was an element of humbug in the just described inheritance of relative invariants. The fact is that any relative invariant of a system yields lower order relative invariants upon integration over closed sub-manifolds, so that if the system is splittable, then when the nonautonomous variables are integrated out, a relative invariant of the new autonomous subsystem is obtained. That the "niceness" of the nice variables was really irrelevant is apparent in that the theorem of phase independence was not invoked. With the inheritance of absolute invariants the situation is otherwise, and we are back in our subject proper.

If $\mathfrak{A} : d\mathbf{x}\, d\mathbf{x}$ is an absolute invariant, then transforming to nice variables and introducing $\mathfrak{B}$ and $\mathbf{C}$ by (D9) as before, we have (D10) and (D11) with the right-hand sides replaced by zero. From the latter we see, first, by the theorem of phase independence, that $\mathbf{C}$ depends only on $\mathbf{z}$, and second, as a consequence, that $\mathbf{C} \cdot d\mathbf{z}$ is an absolute invariant of the reduced system. From the former, we then see that $\mathfrak{B}$ depends only on $\mathbf{z}$, but it is not apparent what more can be deduced, unless $\mathbf{C} \cdot d\mathbf{z}$ is a trivial invariant, in which case $\mathfrak{B} : d\mathbf{z}\, d\mathbf{z}$ is easily seen as before to be a relative invariant (or unless either of the very special cases $\mathbf{C} = 0$ or $\omega_{\mathbf{z}} = 0$ obtains, when $\mathfrak{B} : d\mathbf{z}\, d\mathbf{z}$ is of course even an absolute invariant).

E. HAMILTONIAN SYSTEMS

1. Preparatory Transformations

The hereditary property with the perhaps most important applications is that of being in Hamiltonian form. Let the dependent variables consist of the $N = 2M$ canonically conjugate coordinates $\mathbf{q}$ and $\mathbf{p}$, and let the Hamiltonian $H(\mathbf{q}, \mathbf{p})$ be single-valued and time-independent. (Time-dependent systems, such as those with "slowly changing external parameters," are not excluded by this, since they can be converted to time-independent form by the well-known device of treating time and energy as a new additional pair of conjugate coordinates. See also Sec. B.4.) The $\mathbf{q}$ and $\mathbf{p}$ jointly satisfy the autonomous system constituted by Hamilton's equations

$$\mathbf{q}_t = H_{\mathbf{p}}, \qquad \mathbf{p}_t = -H_{\mathbf{q}}. \tag{E1}$$

In order to allow the widest latitude in applications, we do not require that (E1) be in standard form, but merely that there exist some (monotonic) transformation of the independent variable

$$t \leftrightarrow s \tag{E2}$$

and some transformation of the dependent variables

$$\mathbf{q}, \mathbf{p} \leftrightarrow \mathbf{x} \tag{E3}$$

which take the system into the standard form (B4) with all solutions nearly periodic. (The only reason for permitting the former transformation is to normalize the magnitude of $\mathbf{f}$; therefore, t and s will generally differ at most by a factor which is some power of ϵ.) Applying now our general theory, we make a further transformation of dependent variables

$$\mathbf{q}, \mathbf{p} \leftrightarrow \mathbf{x} \leftrightarrow \mathbf{z}, \phi \tag{E4}$$

to nice variables satisfying (B18). Here $\mathbf{x}$ has been a mere intermediary and is of no further interest.

2. Action Integral

It is well known that $\mathbf{p} \cdot d\mathbf{q}$ is a relative integral invariant (see Sec. D.5) of every Hamiltonian system; indeed, this property may be used to define Hamiltonian systems. In other words, the line integral $\oint \mathbf{p} \cdot d\mathbf{q}$ taken around any closed curve remains constant in the course of time if (every point of) the curve varies in accordance with the equations of motion (E1). In this way one may construct any number of constants of motion, which, however, are for the most part not useful, because not locally computable (see Sec. A.4). There is no way of telling what the family of closed curves chosen at one time will have become at a later time, except of course by fully solving the equations of motion, the very task to avoid or (to help) to effect which one would like to know constants of motion in the first place.

In the present situation there is, nevertheless, a family of curves admirably suited to the construction of a useful constant of motion; these are the

rings, which as a family remain invariant (since rings flow into rings; see Sec. C.3), and which are to an adequate degree locally computable (not local to a point, but local to the loop through a point). Accordingly we define the "phase integral of action"

$$J(\mathbf{z}) \equiv \oint_{\text{ring } \mathbf{z}} \mathbf{p}\cdot d\mathbf{q} = \int_0^1 d\phi\, \mathbf{p}\cdot\mathbf{q}_\phi, \qquad \text{(E5)}$$

in which $\mathbf{q}$ and $\mathbf{p}$ are to be thought of as functions of $\mathbf{z}$ and ϕ in accordance with (E4). It is perhaps worth giving the trivial direct proof of the constancy of J to all orders: since differentiation with respect to s commutes with differentiation or integration with respect to ϕ, so does differentiation with respect to t, whence

$$J_t = \oint (\mathbf{p}_t\cdot d\mathbf{q} - \mathbf{q}_t\cdot d\mathbf{p})$$

$$= -\oint (H_\mathbf{q}\cdot d\mathbf{q} + H_\mathbf{p}\cdot d\mathbf{p}) = -\oint dH = 0, \qquad \text{(E6)}$$

where in the first step we have integrated the second term by parts and in the next we have used (E1). (Incidentally, J is obviously also invariant under canonical transformations.)

The constancy of J to all orders clearly holds for ranges of the independent variable of order ϵ^{-n} for any n, so long as J remains defined.

3. Poisson Brackets

Let $\mathfrak{A}$ and $\mathfrak{B}$ be any polyadic functions of the state of our physical system, i.e., of $\mathbf{q}$, $\mathbf{p}$. Their *Poisson bracket* is defined (as standardly) by

$$[\mathfrak{A}, \mathfrak{B}] = \mathfrak{A}_\mathbf{q}\cdot_\mathbf{p}\mathfrak{B} - \mathfrak{A}_\mathbf{p}\cdot_\mathbf{q}\mathfrak{B}. \qquad \text{(E7)}$$

We shall need the formula

$$[\mathfrak{A}, \mathfrak{B}]_s = [\mathfrak{A}_s, \mathfrak{B}] + [\mathfrak{A}, \mathfrak{B}_s], \qquad \text{(E8)}$$

which is immediately equivalent to its analogue with t derivatives in place of s derivatives (multiply by t_s or s_t to go back or forth), a well known formula easily derived by straightforward calculation using nothing more than (E1).

Now, by (E8), (B18), and (E7) we have

$$[\mathbf{z}, \mathbf{z}]_s = [\epsilon\,\mathbf{h}, \mathbf{z}] + [\mathbf{z}, \epsilon\,\mathbf{h}]$$

$$= \epsilon\,\mathbf{h}_\mathbf{z}\cdot[\mathbf{z}, \mathbf{z}] + \epsilon\,[\mathbf{z}, \mathbf{z}]\cdot_\mathbf{z}\mathbf{h}, \qquad \text{(E9)}$$

so that by the theorem of phase independence the (obviously antisymmetric) dyadic $[\mathbf{z}, \mathbf{z}]$ is a function of $\mathbf{z}$ alone,

$$[\mathbf{z}, \mathbf{z}]_\phi = 0. \qquad \text{(E10)}$$

{Note that although there is no reason to suppose that $[\mathbf{z}, \mathbf{z}]$ is $O(1)$, the theorem is applicable anyway because (E9) is linear.} In the same way we have

$$[\phi, \mathbf{z}]_s = [\omega, \mathbf{z}] + [\phi, \epsilon\,\mathbf{h}]$$

$$= \omega_\mathbf{z}\cdot[\mathbf{z}, \mathbf{z}] + \epsilon\,[\phi, \mathbf{z}]\cdot_\mathbf{z}\mathbf{h}, \qquad \text{(E11)}$$

so that, again by the theorem of phase independence, which applies in view of (E10),

$$[\phi, \mathbf{z}]_\phi = 0. \qquad \text{(E12)}$$

We need the two formulas

$$[\mathbf{z}, \mathbf{z}]\cdot({}_\mathbf{z}\mathbf{p}\cdot\mathbf{q}_\phi - {}_\mathbf{z}\mathbf{q}\cdot\mathbf{p}_\phi) = 0, \qquad \text{(E13)}$$

$$[\phi, \mathbf{z}]\cdot({}_\mathbf{z}\mathbf{p}\cdot\mathbf{q}_\phi - {}_\mathbf{z}\mathbf{q}\cdot\mathbf{p}_\phi) = 1. \qquad \text{(E14)}$$

The short way to establish them is to observe that their common parenthetical factor is the Lagrange bracket of $\mathbf{z}$ and ϕ and to exploit the well-known relationship between the Poisson and the Lagrange brackets, in view of the fact that the summation indicated by the left-most dot in each of them runs over a set of variables complete except for the omission of ϕ, for which the Lagrange bracket would vanish anyway. Of course (E13) and (E14) can also be easily established by direct calculation. It is only necessary to expand them out by (E7) and the distributive law and then to make use of relations (coming from the chain rule for differentiation) between the derivatives of $\mathbf{q}$ and $\mathbf{p}$ with respect to $\mathbf{z}$ and ϕ and *vice versa*, namely

$$\mathbf{q}_\mathbf{z}\cdot\mathbf{z}_\mathbf{q} + \mathbf{q}_\phi\,\phi_\mathbf{q} = \mathbf{q}_\mathbf{q} = \mathcal{I} \qquad \text{(E15)}$$

and similar relations based on two-sided evaluations of $\mathbf{q}_\mathbf{p}$, $\mathbf{p}_\mathbf{q}$, $\mathbf{p}_\mathbf{p}$, $\mathbf{z}_\phi$, and ϕ_ϕ. (The two remaining similar relations coming from $\mathbf{z}_\mathbf{z}$ and $\phi_\mathbf{z}$ are not needed.)

We are now prepared to evaluate two particular Poisson brackets of interest. One is

$$[\mathbf{z}, J] = [\mathbf{z}, \mathbf{z}]\cdot J_\mathbf{z}$$

$$= [\mathbf{z}, \mathbf{z}]\cdot\int_0^1 d\phi\,({}_\mathbf{z}\mathbf{p}\cdot\mathbf{q}_\phi - {}_\mathbf{z}\mathbf{q}\cdot\mathbf{p}_\phi) = 0, \qquad \text{(E16)}$$

where the first equation is seen to hold by (E7) and the fact that J depends on $\mathbf{z}$ alone, the second by direct differentiation of (E5) followed by an integration by parts, and the last by (E13) after $[\mathbf{z}, \mathbf{z}]$ has been brought inside the integral by virtue of (E10). The other is

$$[\phi, J] = 1, \qquad \text{(E17)}$$

as seen in the same way using (E14) and (E12).

4. Nontriviality of Action Integral

We digress momentarily to point out that J can hardly be a trivial constant of motion. It certainly cannot be completely trivial, in the sense of being an outright constant, for then its Poisson bracket with any other quantity would vanish,

contradicting (E17). But suppose that J can be expressed as a function of H alone, which is obviously a kind of triviality since H is, of course, automatically a constant of motion. Now J_H cannot vanish, else we would have $[\phi, J] = [\phi, H]J_H = 0$, contradicting (E17) as before, so we may invert the relationship and express H as a function of J. Then

$$\mathbf{z}_t = [\mathbf{z}, H] = [\mathbf{z}, J]H_J = 0, \qquad \text{(E18)}$$

where the first equation is well known and obvious, and the last is a consequence of (E16). Thus the $\mathbf{z}$ provide a complete set of integrals of the system. Only in this extremely special case can J be a function of H alone.

5. Reduced Hamiltonian System

We continue with our program of showing the hereditary character of being Hamiltonian. To this end we wish to put the reduced system into Hamiltonian form. Through the reduction process itself we have gotten rid of (or "left behind") one variable, ϕ, and now that we have the nontrivial constant of motion J, we may use it to eliminate another variable. Thus we may hope to put the remaining system of $2M - 2$ variables into the form of a Hamiltonian system of $M - 1$ degrees of freedom.

Our actual procedure for obtaining this *reduced time-independent Hamiltonian* system is to make a canonical transformation from our original set of M pairs of canonically conjugate coordinates $(\mathbf{q}, \mathbf{p})$ to a new set of which one pair is (ϕ, J) and of which the remaining $M - 1$ pairs, to be denoted collectively by $(\mathbf{Q}, \mathbf{P})$, are also locally computable. Now the necessary and sufficient condition that the transformation to a new set of pairs of coordinates be canonical is that the Poisson bracket of any two new coordinates be zero if they come from different pairs and unity if they constitute a pair (in the right order). Because of (E17) it is not immediately excluded that we can find a canonical transformation of the kind described, but it is also not immediately obvious that we can. In Appendix 2 it is shown[24] that whenever one knows some of the new coordinates of a proposed prospective canonical transformation, and these coordinates satisfy all the requisite Poisson bracket conditions among themselves and have linearly independent derivatives with respect to the old coordinates, then one can indeed define the missing coordinates (in a way locally dependent on the Hamiltonian) so as to form a complete canonical transformation; furthermore the whole construction is obviously valid order by order when the coordinates involved are construed as series.

The condition of linear independence of the derivatives is certainly satisfied, for if we had, say,

$$J_{\mathbf{q}} = k\,\phi_{\mathbf{q}}, \qquad J_{\mathbf{p}} = k\,\phi_{\mathbf{p}}, \qquad \text{(E19)}$$

we could immediately deduce

$$[\phi, J] = k[\phi, \phi] = 0,$$

once again contradicting (E17). Accordingly, we may assume that we now have the sequence of transformations

$$\phi, \mathbf{z} \leftrightarrow \mathbf{q}, \mathbf{p} \leftrightarrow \Phi, J, \mathbf{Q}, \mathbf{P}, \qquad \text{(E20)}$$

where, in view of the discussion in Sec. A.3, we have introduced the new symbol Φ (not to be confused with the Φ of part B) for ϕ in its role as a member of the new set of variables, so that

$$\Phi = \phi \qquad \text{(E21)}$$

may be used freely except to substitute one subscript for another (but see below!). The new variables $\mathbf{Q}$ and $\mathbf{P}$ have been so constructed as to satisfy the correct bracket conditions

$$[\mathbf{Q}, \mathbf{Q}] = 0, \quad [\mathbf{Q}, \mathbf{P}] = \mathcal{I}, \quad [\mathbf{P}, \mathbf{P}] = 0 \qquad \text{(E22)}$$

among themselves and also the "cross-conditions"

$$0 = [\mathbf{Q}, \Phi] = [\mathbf{Q}, \phi] = \mathbf{Q}_{\mathbf{z}} \cdot [\mathbf{z}, \phi], \qquad \text{(E23)}$$

$$0 = [\mathbf{Q}, J] = \mathbf{Q}_{\phi}\,[\phi, J] + \mathbf{Q}_{\mathbf{z}} \cdot [\mathbf{z}, J] = \mathbf{Q}_{\phi}, \qquad \text{(E24)}$$

and the same with $\mathbf{Q}$ replaced throughout by $\mathbf{P}$. From (E24) we see that $\mathbf{Q}$ and $\mathbf{P}$ actually are independent of ϕ, like J, so that the transformation $\mathbf{z} \leftrightarrow J, \mathbf{Q}, \mathbf{P}$ is wholly independent of ϕ, and we may (and from now on do) obliterate completely the distinction between Φ and ϕ.

The new Hamiltonian is of course just the old one H itself, naturally expressed now as a function of the new canonical coordinates through the transformation (E20). In the new coordinates, Hamilton's equations take the form

$$\phi_t = H_J. \qquad \text{(E25)}$$

[24] A very similar theorem is demonstrated by Nordheim and Fues [L. Nordheim and E. Fues, *Handbuch der Physik*, edited by S. Flügge (Julius Springer, Berlin, 1927) B.5, Kap. 3, Ziff. 10], who assume that an integral is known and show how this can be used to reduce to a Hamiltonian system of one less degree of freedom; one finds a canonical transformation to new coordinates one of which is the known integral, in which case its conjugate coordinate obviously is ignorable. In our present application the ignorable coordinate ϕ has come first. It is not on account of this minor difference that a separate discussion is appended, but rather partly to keep this paper as self-contained as possible and partly because the method is so simple and seems to offer a slightly unconventional insight into the nature of the Poisson bracket conditions.

$$J_\iota = -H_\phi, \tag{E26}$$

$$\mathbf{Q}_\iota = H_\mathbf{P}, \qquad \mathbf{P}_\iota = -H_\mathbf{Q}. \tag{E27}$$

Since $\phi_\iota = \phi_s s_\iota = \omega s_\iota$, (E25) yields the familiar sort of relation between the frequency and the derivative of the Hamiltonian with respect to the action variable. Since $J_\iota = 0$, (E26) shows that H is independent of ϕ, which we might have known anyway from the fact that H is an integral of the system (see Sec. D.3). It is now clear that (E27) constitutes a one-parameter family (parametrized by J) of autonomous Hamiltonian systems.

6. Possibility of Repetition and Further Invariants

As a consequence of our ability to put the reduced system back into Hamiltonian form, we see immediately that if the reduced system itself has all solutions nearly periodic (naturally with a period longer than the period of the original system at least by an order of ϵ), then there exists also another constant of motion

$$J' \equiv \oint_{\text{ring } \mathbf{z}'} \mathbf{P}\cdot d\mathbf{Q} = \int_0^1 d\phi'\, \mathbf{P}\cdot\mathbf{Q}_{\phi'}, \tag{E28}$$

where $\mathbf{z}'$ and ϕ' are the new nice variables for the reduced system,

$$\mathbf{z} \leftrightarrow J, \mathbf{Q}, \mathbf{P} \leftrightarrow \mathbf{z}', \phi' \tag{E29}$$

As one should expect, it is by no means necessary to go through the complicated integration process involved in finding $\mathbf{Q}$ and $\mathbf{P}$ in order to compute J', for

$$J' = \int_0^1 d\phi'\, (\mathbf{P}\cdot\mathbf{Q}_{\phi'} + J\,\phi_{\phi'}) = \int_0^1 d\phi'\, \mathbf{p}\cdot\mathbf{q}_{\phi'}, \tag{E30}$$

since $\phi_{\phi'} = 0$ [the transformation (E29) being independent of ϕ] and the differential form $\mathbf{p}\cdot d\mathbf{q}$ is a canonical invariant.

Because $\mathbf{Q}$ and $\mathbf{P}$ are locally computable, so is J'. Furthermore, J' cannot depend on J only, nor even on J and H only except in the very trivial special case that we can find a complete set of integrals. Incidentally it may be noted that there is no chance here of finding still another integral by the well-known device of forming the Poisson bracket of two already known, since by a further canonical transformation we can introduce a set of canonical coordinates among which are the conjugate pairs (ϕ, J) and (ϕ', J'), which demonstrates that

$$[J, J'] = 0 \tag{E31}$$

in addition to three other such relations.

If the new reduced system should itself happen to have all solutions nearly periodic we can find still another integral J'' and ignore another angle variable ϕ'', and so on indefinitely, till the successive near periodicities run out. If the internal degrees of freedom run out first, we end up with a complete (asymptotic) description of the system in terms of M ignorable angle variables and their M locally determined conjugate action variables.

F. APPENDICES

1. Fundamental Approximation Theorem

For the sake of completeness we give here a brief outline of the well-known proof that two functions which satisfy nearly identical autonomous systems and have nearly equal initial values are themselves nearly equal for a bounded range of the independent variable. Let $\mathbf{x}(s)$ and $\mathbf{x}^\dagger(s)$ satisfy

$$\mathbf{x}_s = \mathbf{f}(\mathbf{x}), \qquad \mathbf{x}^\dagger_s = \mathbf{f}(\mathbf{x}^\dagger) + \boldsymbol{\gamma}(\mathbf{x}^\dagger), \tag{F1}$$

$$\mathbf{x}(0) = \mathbf{x}^\dagger(0) + \boldsymbol{\delta}, \tag{F2}$$

where $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are small quantities whose norms (maximized over $\mathbf{x}$ space in the case of $\boldsymbol{\gamma}$) we denote by γ and δ. (Any reasonable norm may be employed, and will be denoted by absolute value signs.) In the region of interest in $\mathbf{x}$ space $\mathbf{f}$ satisfies a Lipschitz condition,

$$|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}'')| \leq c\,|\mathbf{x}' - \mathbf{x}''| \tag{F3}$$

for an appropriate positive constant c and any $\mathbf{x}'$ and $\mathbf{x}''$ (if only because $\mathbf{f}_\mathbf{x}$ exists and is continuous on a closed bounded domain and therefore has all components bounded). To estimate the difference we have

$$\begin{aligned} |\mathbf{x}^\dagger - \mathbf{x}| &= \Big|\mathbf{x}^\dagger(0) - \mathbf{x}(0) \\ &\quad + \int_0^s ds\,[\mathbf{f}(\mathbf{x}^\dagger) + \boldsymbol{\gamma}(\mathbf{x}^\dagger) - \mathbf{f}(\mathbf{x})]\Big| \\ &\leq \delta + \int_0^s ds\,[c\,|\mathbf{x}^\dagger - \mathbf{x}| + \gamma], \end{aligned} \tag{F4}$$

for simplicity restricting ourselves to non-negative values of s. Now the maximum value $\mu(s)$ of $|\mathbf{x}^\dagger - \mathbf{x}|$ in the range from 0 to s satisfies

$$\begin{aligned} \mu &\leq \delta + \int_0^s ds\,[c\mu + \gamma] \\ &\leq \delta + s[c\mu + \gamma], \end{aligned} \tag{F5}$$

so that for $s \leq 1/2c$ (for instance) we have

$$\mu \leq (\delta + \gamma s)/(1 - cs) \leq 2\delta + 2\gamma s. \tag{F6}$$

In order to obtain an estimate for larger values of s we could apply (F6) over and over again, proceeding in steps of size $1/2c$ and using the final estimate at each step as the initial estimate corresponding to (F2) for the next step. In this way we would obtain an exponential estimate, which is, however, more elegantly and perhaps simply obtained by noting that an upper bound for $|\mathbf{x}^\dagger - \mathbf{x}|$ is provided by the function $\nu(s)$ which satisfies the equation obtained from (F4) by replacing $|\mathbf{x}^\dagger - \mathbf{x}|$ by ν and $\leq$ by $=$. Indeed, the difference $D \equiv \nu - |\mathbf{x}^\dagger - \mathbf{x}|$ satisfies

$$D \geq c \int_0^s ds\, D, \tag{F7}$$

$$\left(e^{-cs} \int_0^s ds\, D\right)_s \geq 0, \tag{F8}$$

whence upon integration from 0 to s it follows that the right-hand side of (F7) and *a fortiori* D are nonnegative. But it is trivial to solve explicitly for ν, and we find that

$$|\mathbf{x}^\dagger - \mathbf{x}| \leq \nu = (\delta + \gamma/c)e^{cs} - \gamma/c. \tag{F9}$$

That there can be no estimate radically better than this is obvious from even the simplest examples.

The standard uniqueness theorem is of course obtained as the special case $\boldsymbol{\gamma} = \boldsymbol{\delta} = 0$.

2. Completion of Canonical Transformation

The necessary and sufficient conditions that a transformation

$$\mathbf{q}, \mathbf{p} \leftrightarrow \mathbf{q}^\dagger, \mathbf{p}^\dagger \tag{F10}$$

be canonical are of course the fundamental Poisson bracket relations

$$[\mathbf{q}^\dagger, \mathbf{q}^\dagger] = 0, \quad [\mathbf{q}^\dagger, \mathbf{p}^\dagger] = \mathcal{I}, \quad [\mathbf{p}^\dagger, \mathbf{p}^\dagger] = 0. \tag{F11}$$

A consequence of these relations is that the derivatives of the components of $\mathbf{q}^\dagger$ and $\mathbf{p}^\dagger$ are linearly independent (and *a fortiori* that the components themselves are functionally independent), since for any linear combination of these derivatives which vanishes,

$$\mathbf{a}\cdot\mathbf{q}_\mathbf{q}^\dagger + \mathbf{b}\cdot\mathbf{p}_\mathbf{q}^\dagger = 0, \qquad \mathbf{a}\cdot\mathbf{q}_\mathbf{p}^\dagger + \mathbf{b}\cdot\mathbf{p}_\mathbf{p}^\dagger = 0, \tag{F12}$$

we have, by (F11) and (F12), that

$$\mathbf{a} = \mathbf{a}\cdot[\mathbf{q}^\dagger, \mathbf{p}^\dagger] = -\mathbf{b}\cdot[\mathbf{p}^\dagger, \mathbf{p}^\dagger] = 0 \tag{F13}$$

and similarly that $\mathbf{b} = 0$.

Suppose now that we are given, as functions of $\mathbf{q}$ and $\mathbf{p}$, some but not all of the components of $\mathbf{q}^\dagger$ and $\mathbf{p}^\dagger$ in a proposed transformation (F10). We wish to prove that we can construct the remaining components. Let $\mathbf{q}$, $\mathbf{p}$, $\mathbf{q}^\dagger$, $\mathbf{p}^\dagger$ each have M components. We denote the $N < 2M$ given components (of both $\mathbf{q}^\dagger$ and $\mathbf{p}^\dagger$, not necessarily paired) collectively by $\boldsymbol{\alpha}$, and require them to have linearly independent derivatives and to satisfy those components of relations (F11) which involve them alone. Since we can proceed by mathematical induction, it suffices to construct one of the remaining components of $\mathbf{q}^\dagger$ or $\mathbf{p}^\dagger$, say of $\mathbf{q}^\dagger$. We denote this component by q' and its conjugate by p'. We must construct q' to satisfy those components of relations (F11) which involve it and the $\boldsymbol{\alpha}$. These conditions on q' have in fact the form

$$[\boldsymbol{\alpha}, q'] = \mathbf{c}, \tag{F14}$$

where $\mathbf{c}$ is constant and, to be sure, has all components zero except for at most one of them, which equals -1 (that component, if it is among the $\boldsymbol{\alpha}$, which corresponds to p').

Conditions (F14) amount to a prescription of the directional derivative of q' in each of a number of directions, one for each component of $\boldsymbol{\alpha}$. These directions are linearly independent, since the derivatives of the components of $\boldsymbol{\alpha}$ are. We next inquire into the conditions for the compatibility of (F14), namely the conditions that the second derivatives of q' be the same independently of the order of differentiation. The dyadic second directional derivative of q' (which, if expanded out, contains first as well as second derivatives of q') is

$$[\boldsymbol{\alpha}, [\boldsymbol{\alpha}, q']] = [\boldsymbol{\alpha}, \mathbf{c}] = 0, \tag{F15}$$

since $\mathbf{c}$ is constant. Taking the directional derivatives in the reverse order gives

$$-[[\boldsymbol{\alpha}, q'], \boldsymbol{\alpha}] = 0 \tag{F16}$$

[which happens to be the transpose of (F15)]. This has the same second derivatives of q', so we obtain the consistency conditions by taking the difference

$$[\boldsymbol{\alpha}, [\boldsymbol{\alpha}, q']] + [[\boldsymbol{\alpha}, q'], \boldsymbol{\alpha}] = 0. \tag{F17}$$

Now, Jacobi's identity, written for two vectors and a scalar with due regard to the dyadic order, is

$$[\boldsymbol{\alpha}, [\boldsymbol{\alpha}', q']] + [[\boldsymbol{\alpha}, q'], \boldsymbol{\alpha}'] + [q', [\boldsymbol{\alpha}, \boldsymbol{\alpha}']] = 0, \tag{F18}$$

the middle term being the transpose of $[\boldsymbol{\alpha}', [q', \boldsymbol{\alpha}]]$ Taking $\boldsymbol{\alpha}'$ to be $\boldsymbol{\alpha}$, we see that (F17) is equivalent to

$$[[\boldsymbol{\alpha}, \boldsymbol{\alpha}'], q'] = 0, \tag{F19}$$

which is explicitly devoid of second derivatives of

q'. But $[\boldsymbol{\alpha}, \boldsymbol{\alpha}']$ is a constant dyadic, all of whose components are 0 or ± 1, in view of (F11), so (F19) holds identically.

Since the compatibility conditions are satisfied (tautologously), it follows from the general theory of partial differential equations that (F14) has a general solution q' with the freedom of an arbitrary function of $2M - N$ variables. From the geometrical point of view, indeed, we have shown that the N-dimensional vector space determined at each point by the directions of the directional derivatives $[\boldsymbol{\alpha}, \cdots]$ there, is the tangent space of a $(2M - N)$-dimensional family of N-dimensional hypersurfaces.[25] In (F14) we have a complete prescription of the derivatives of q' as a function on each hypersurface, but no restriction on the variation of q' from one hypersurface to another. Clearly q' is determined up to a constant (its "initial value" at any given point) on each hypersurface.

It remains only to show that the arbitrariness in q' permits us to choose it in such a way that the derivatives of q' and the $\boldsymbol{\alpha}$ will be linearly independent. This is merely a matter of counting dimensions, for suppose contrarily that

$$q'_{\mathfrak{q}} = \mathbf{k}\cdot\boldsymbol{\alpha}_{\mathfrak{q}}, \qquad q'_{\mathfrak{p}} = \mathbf{k}\cdot\boldsymbol{\alpha}_{\mathfrak{p}}. \tag{F20}$$

If p' is among the $\boldsymbol{\alpha}$, then since $[\boldsymbol{\alpha}, q']$ vanishes in all except that component, and since by (F20)

$$0 = [q', q'] = \mathbf{k}\cdot[\boldsymbol{\alpha}, q'], \tag{F21}$$

it follows that the corresponding component of $\mathbf{k}$ vanishes. Similarly, if any (other) two conjugate coordinates q'', p'' are among the $\boldsymbol{\alpha}$, the corresponding components of $\mathbf{k}$ must both vanish, else by means of (F20) we could solve for and eliminate the derivatives of either q'' or p'' and by (F11) obtain the contradiction

$$1 = [q'', p''] = 0.$$

Let the number of such pairs be L, so that the number of unpaired components of $\boldsymbol{\alpha}$ is $N - 2L$. Obviously

$$N - L = (N - 2L) + L \leq M, \tag{F22}$$

the right-hand equality holding only if p' is among the $\boldsymbol{\alpha}$. The number of possibly nonzero components of $\mathbf{k}$ is

$$N - 2L = 2(N - L) - N \leq 2M - N \tag{F23}$$

or $N - 2L - 1$ if p' is among the $\boldsymbol{\alpha}$; thus in either case it is strictly less than $2M - N$. But q' has the freedom of an arbitrary function of $2M - N$ variables, so its derivatives span a $(2M - N)$-dimensional linear space, contradicting (F20).

3. Iterated Near-Identity Mappings

There is a strong and intimate connection between the asymptotic theory of nearly recurrent systems and the asymptotic theory of iterated near-identity mappings, which has previously been treated less satisfactorily by Kruskal[5] and applied by Spitzer.[4] In this section we describe the latter theory and explore the connections.

A *near-identity mapping* of a space of vectors $\boldsymbol{\xi}$ into itself is a function, $\mathbf{T}(\boldsymbol{\xi})$ say, which can be written in the form

$$\mathbf{T}(\boldsymbol{\xi}) = \boldsymbol{\xi} + \epsilon\,\mathbf{D}(\boldsymbol{\xi}), \tag{F24}$$

where $\mathbf{D}(\boldsymbol{\xi})$ may be a series in powers of ϵ and is required to be $O(1)$. Associated with such a transformation (as with any) is the sequence of its iterates, the transformations $\mathbf{T}^n(\boldsymbol{\xi})$ defined recursively by

$$\mathbf{T}^1(\boldsymbol{\xi}) \equiv \mathbf{T}(\boldsymbol{\xi}), \qquad \mathbf{T}^{n+1}(\boldsymbol{\xi}) \equiv \mathbf{T}^n(\mathbf{T}(\boldsymbol{\xi})), \tag{F25}$$

where n runs over the positive integers.[26] We then ask for an asymptotic description (as $\epsilon \to 0$) of the discrete "trajectory" of a point $\boldsymbol{\xi}$ under iteration of $\mathbf{T}$, i.e., of the sequence of points $\mathbf{T}^n(\boldsymbol{\xi})$, $n = 0, 1, 2, \cdots$.

Obviously as $\epsilon \to 0$ the points get closer and closer together. For any fixed n, of course $\mathbf{T}^n(\boldsymbol{\xi}) \to \boldsymbol{\xi}$, but this does not imply that the trajectory as a whole collapses into the initial point, since for n taken approximately equal to c/ϵ (where c is a constant) we may expect $\mathbf{T}^n(\boldsymbol{\xi})$ to remain distinctly away from $\boldsymbol{\xi}$ and indeed to converge to a limit[27] different from $\boldsymbol{\xi}$. Rather, the discrete trajectory approximates better and better to a continuous (and smooth) curve denoted by $\boldsymbol{\Xi}^{(0)}(\sigma)$ and determined by

$$\boldsymbol{\Xi}^{(0)}(0) = \boldsymbol{\xi}, \qquad \boldsymbol{\Xi}^{(0)}_{\sigma} = \mathbf{D}(\boldsymbol{\Xi}^{(0)}), \tag{F26}$$

the latter being obtained from (F24) by replacing $\boldsymbol{\xi}$ by $\mathbf{T}^n(\boldsymbol{\xi})$ and then going to the limit after making the identifications

$$\sigma \equiv n\epsilon, \tag{F27}$$

[25] See Newcomb [W. Newcomb, Ann. Phys. **3**, 347 (1958)] for a heuristic description of the theory of such questions. In his terminology, we have shown that the "cross-bracket" of any two directions of differentiation is zero.

[26] Obviously (F24) can be formally inverted to yield an inverse transformation $\mathbf{T}^{-1}$, with iterates $\mathbf{T}^{-2}$, $\mathbf{T}^{-3}$, etc. With $\mathbf{T}^0(\boldsymbol{\xi}) \equiv \boldsymbol{\xi}$, (F25) becomes valid for all integers n.

[27] This limit can in fact be expressed as $\boldsymbol{\Xi}^{(0)}(c)$, see (F28).

$$\mathbf{\Xi}^{(0)}(\sigma) \equiv \lim_{\epsilon=0} \mathbf{T}^n(\xi). \tag{F28}$$

Except to zeroth order the points $\mathbf{T}^n(\xi)$ naturally need not lie on $\mathbf{\Xi}^{(0)}$. This suggests (what we have already presaged by the notation) that we seek a formal series $\mathbf{\Xi}(\sigma)$ such that the representation

$$\mathbf{T}^n(\xi) = \mathbf{\Xi}(\sigma) \tag{F29}$$

will be valid to all orders under the identification (F27), which there is no need to modify. And it is indeed easy to obtain the appropriate conditions (differential equation and initial condition) determining $\mathbf{\Xi}$ to higher orders. We could then develop a rigorous theory, with analogues of the results in the main body of this paper. It is simpler, however, to reduce the problem to the one that we have already exhaustively analyzed.

The crucial step is to imbed the sequence $\mathbf{T}^n(\xi)$ in a continuous trajectory satisfying an autonomous system of differential equations. That is, we seek a function $\mathbf{n}(\xi)$ such that the curve $\xi^\dagger(\xi, \sigma)$ defined as the solution of the system

$$\xi^\dagger_\sigma = \mathbf{n}(\xi^\dagger) \tag{F30}$$

determined by the initial condition

$$\xi^\dagger(\xi, 0) = \xi \tag{F31}$$

passes through the points in the sequence for σ given by (F27):

$$\xi^\dagger(\xi, n\epsilon) = \mathbf{T}^n(\xi). \tag{F32}$$

We shall show that $\mathbf{n}$ is determined (and uniquely) as a series to all orders by this condition (required to hold for arbitrary ξ).

Obviously it suffices to determine $\mathbf{n}$ merely so that (F32) holds with $n = 1$,

$$\xi^\dagger(\xi, \epsilon) = \mathbf{T}(\xi), \tag{F33}$$

since it will then automatically hold for all larger n by an induction based on the calculation

$$\begin{aligned}\xi^\dagger(\xi, n\epsilon + \epsilon) &= \xi^\dagger(\xi^\dagger(\xi, \epsilon), n\epsilon) \\ &= \xi^\dagger(\mathbf{T}(\xi), n\epsilon) = \mathbf{T}^n(\mathbf{T}(\xi)) = \mathbf{T}^{n+1}(\xi),\end{aligned} \tag{F34}$$

where the first equation is an instance of (A7), the second will follow from (F33), the third will be an instance of the induction hypothesis (F32), and the fourth is part of (F25). Now by a Taylor expansion

$$\begin{aligned}\xi^\dagger(\xi, \epsilon) &= \xi^\dagger(\xi, 0) + \epsilon\, \xi^\dagger_\sigma(\xi, 0) + \tfrac{1}{2}\epsilon^2\, \xi^\dagger_{\sigma\sigma}(\xi, 0) + \cdots \\ &= \xi + \epsilon\, \mathbf{n}(\xi) + \tfrac{1}{2}\epsilon^2\, \mathbf{n}_\xi(\xi)\cdot\mathbf{n}(\xi) + \cdots,\end{aligned} \tag{F35}$$

which by (F24) and (F33) may be written in the form

$$\mathbf{n} + \tfrac{1}{2}\epsilon\, \mathbf{n}_\xi\cdot\mathbf{n} + \cdots = \mathbf{D}, \tag{F36}$$

omitting the argument ξ everywhere. Inverting this we have immediately

$$\mathbf{n} = \mathbf{D} - \tfrac{1}{2}\epsilon\, \mathbf{D}_\xi\cdot\mathbf{D} + \cdots. \tag{F37}$$

The desired representation (F29) is now obviously obtained merely by taking

$$\mathbf{\Xi}(\sigma) \equiv \xi^\dagger(\xi, \sigma). \tag{F38}$$

It is at last a trivial matter to bring this theory under that of the main body of the paper. Treating n as a continuous variable and calling it s now, or in other words defining

$$s \equiv \sigma/\epsilon, \tag{F39}$$

and introducing a redundant angle variable θ equal to s up to an arbitrary additive integer, and finally defining[28]

$$\mathbf{x} \equiv (\xi, \theta), \tag{F40}$$

we have a system in standard form

$$\mathbf{x}_s = \mathbf{f} \equiv (\epsilon\, \mathbf{n}, 1). \tag{F41}$$

This is nearly recurrent, since to zeroth order ξ is constant and θ increases uniformly, returning to its initial value (up to an integer, which is all that has meaning for an angle value!) each time s increases by unity. Therefore the general theory applies. The introduction of nice variables is of course utterly trivial, being given by

$$\mathbf{z} = \xi, \quad \phi = \theta. \tag{F42}$$

And the reduced system is just (F30) with ξ instead of $\xi^\dagger$. The only point to this otherwise rather farcical merry-go-round is that the further developments of the theory, and especially the proof that the formal series is really an asymptotic representation of the true solution, is secured thereby.

We close with a brief description of the reverse procedure. Had we chosen to develop the mapping theory in full, we could have based the theory of nearly recurrent systems on it by choosing for ξ space a hypersurface in $\mathbf{x}$ space cutting across the loops. The mapping would then be obtained by starting from any point ξ in the hypersurface and following the solution of (B4) around until it first again intersects the hypersurface in a point to be called $\mathbf{T}(\xi)$. Thus we see that the two theories are completely equivalent.

[28] Thus $\mathbf{x}$ space is topologically the product of ξ space with a circle.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 4 JULY-AUGUST 1962

Erratum: Spin and Statistics with an Electromagnetic Field

[J. Math. Phys. 3, 50 (1962)]

DAVID G. BOULWARE
Harvard University, Cambridge, Massachusetts

THE proof of the positive frequency character of the series expansion of

$$\left\langle p' \left| e^{ip\xi}\left(\exp\left[-ie\int_0^{\xi^0} f(\xi,0,y)A^0(y)\,dy\right]\right)_+ \right| p''\right\rangle$$

on p. 54 is not correct. The general term should be $F_n(\xi, \xi, p', p'')$, where

$$F_n(\xi, x, p', p'') = \frac{(-ie)^n}{n!}\int_0^{x^0} dy_1 \cdots dy_n f(\xi,0y_1)\cdots f(\xi,0,y_n) \times e^{ip'x}\langle p' \,|(A^0(y_1)\cdots A^0(y_n))_+|\, p''\rangle$$

$$= -ie\int \langle p' \,|A^0|\, p'''\rangle\, dp''' \times \int_0^{x^0} e^{ip'(x-y)} f(\xi,0,y)\,dy F_{n-1}(\xi, y, p''', p'').$$

We assume

$$F_{n-1}(\xi, x, p''', p'') = \int dK\, dK' \theta(K^0)\theta(K'^0) \times e^{iKx}e^{iK'(\xi-x)}\bar{F}_{n-1}(K, K', p''', p'').$$

Then, $F_{n-1}(\xi, \xi, p''', p'')$ is a positive frequency function of ξ^0.

For $\xi^0 > x^0 > 0$, $F_1(\xi, x, p''', p'')$ satisfies this condition with

$$\bar{F}_1(K, K', p''', p'') = (e/2)\langle p''' \,|A^0|\, p''\rangle \times \left\{\delta(\mathbf{K} - \mathbf{p}'')\,\delta(\mathbf{K}' - \mathbf{p}'' + \mathbf{p}''')\,\delta(K'^0 - k)\right.$$
$$\times\left[\frac{\delta(K^0 - p'''^0 - k) - \delta(K^0 - p''^0)}{p''^0 - p'''^0 - k}\right] + \delta(K')\,\delta(\mathbf{K} - \mathbf{p}''') \times\left.\left[\frac{\delta(K^0 - p'''^0) - \delta(K^0 - p''^0 - k)}{p''^0 - p'''^0 + k}\right]\right\},$$

$$k = [(\mathbf{p}'' - \mathbf{p}''')^2]^{\frac{1}{2}}.$$

Then,

$$\bar{F}_n(\bar{K}, \bar{K}', p', p'') = (e/2) \times \int \langle p' \,|A^0|\, p'''\rangle\, dp'''\bar{F}_{n-1}(K, K', p''', p'') \times d^3k\, dK\, dK'\, \delta(\mathbf{p}' + \mathbf{k} + \mathbf{K}' - \mathbf{K}) \times\left\{\left(\frac{\delta(\bar{K}' - K' - k)\,\delta(\bar{\mathbf{K}} + \mathbf{K})}{K^0 - K'^0 - k^0 - p'^0}\right.\right.$$
$$\times\left.[\delta(\bar{K}^0 - p'^0 - k^0 - K'^0) - \delta(\bar{K}^0 - K^0)]\right) - \left(\frac{\delta(\bar{K}' - K')\,\delta(\bar{\mathbf{K}} - \mathbf{p}' - \mathbf{K}')}{K^0 + k^0 - p'^0 - K'^0}\right.$$
$$\times\left.\left.[\delta(\bar{K}^0 - p'^0 - K'^0) - \delta(\bar{K}^0 - k^0 - K^0)]\right)\right\}$$
$$= \theta(\bar{K}^0)\theta(\bar{K}'^0)\bar{F}_n(\bar{K}, \bar{K}', p', p'') \qquad k^0 = [\mathbf{k}^2]^{\frac{1}{2}}$$

By induction, $F_n(\xi, \xi, p', p'')$ is a positive frequency function of ξ^0.