JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 13, NUMBER 11 NOVEMBER 1972

Operational Statistics. I. Basic Concepts

D. J. Foulis and C. H. Randall
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01002
(Received 3 May 1971)

A generalized notion of a "sample space" is developed which allows for the simultaneous representation of the outcomes of a set of related "random experiments." Affiliated with each such generalized sample space is a so-called "logic," the elements of which are propositions that can be confirmed or refuted by observing the outcomes of the random experiments. Stochastic models for the experimental situation represented by a given sample space are introduced, and it is shown that such stochastic models induce generalized probability measures on the logic of this sample space.

1. INTRODUCTION

The purpose of the series of papers here begun is to erect a new mathematical foundation for an operational theory of probability and statistics based upon a generalization of the conventional notion of a sample space. In subsequent papers, we shall formally establish on this foundation the notion of a "physical system" and an affiliated "theory of measurement." This latter generalized theory of measurement should prove to be particularly useful in the developing behavioral sciences and in addition shed some light on the difficulties that surround the measuring process in quantum mechanics.[1] In other papers in this series, the matter of statistical inference will be considered in this nonclassical framework. In particular, it will be shown that the Bayesian inference strategy extends naturally to this more general situation. As a consequence of these considerations, four of the principal types of "probability" (frequency, credibility, rational betting and logical) put in an appearance and prove to be naturally and coherently related. Furthermore, the customary mathematical representation of the notion of conditioning resolves into a spectrum of such representations, thus formally exposing a variety of conditioning concepts.

In the present paper, we shall introduce our generalized sample spaces and show that there is a hierarchy of "propositional systems" affiliated with each such generalized sample space. Complete "stochastic models" for the empirical situation represented by a generalized sample space will be formally introduced as so-called weight functions, and it will be shown that these weight functions can be used to induce generalized probability measures (states) on the affiliated propositional systems.

The distinguishing feature of the empirical sciences is that propositions pertinent to such sciences are confirmed and refuted solely in terms of evidence secured as a consequence of the execution of physical operations. In this regard, we shall by no means restrict the term physical operation to apply only to traditional laboratory procedures. We are prepared, for instance, to regard test procedures on an assembly line, data gathering processes (such as opinion polling), pencil and paper operations (such as executing computational algorithms), and even procedures involving subjective approvals or disapprovals as bona fide physical operations.

Let us make our definition of a physical operation official: By a *physical operation*, we shall mean instructions that describe a well-defined, physically realizable, reproducible procedure and furthermore that specify what must be observed and recorded. In particular a physical operation must require that, as a consequence of each execution of the instructions, one and only one symbol from a specified set R be recorded as the *result* of that realization of the operation.[2] Carefully note, if we delete or add details to the instructions for a physical operation, and, in particular, if we modify the result set R in any way, we thereby define a *new* physical operation.

Since the early 1930's, when Kolmogorov laid the foundations for probability and statistics as we know them today,[3] it has become traditional to refer to the result set R for a physical operation $\mathfrak{D}$ as its *sample space*. A subset D of R is called an *event* for the operation $\mathfrak{D}$ and certain of these events—but not necessarily all of them—are decreed to be *observable events*. One usually assumes that the collection $\mathfrak{F}$ of all observable events forms a σ-field of subsets of the sample space R. The physical operation $\mathfrak{D}$ is then regarded as being mathematically represented by the ordered pair $(R, \mathfrak{F})$.

Suppose that the physical operation $\mathfrak{D}$ is represented, as above, by the ordered pair $(R, \mathfrak{F})$. One normally associates with each event $D \subseteq R$ a proposition $p(D)$ asserting that a result $d \in D$ was obtained as a consequence of executing $\mathfrak{D}$. In this way, the results $d \in D$ are regarded as *confirming* the proposition $p(D)$, while the results in $R \backslash D$, the set-theoretic complement of D in R, are regarded as *refuting* the proposition $p(D)$. Thus, $p(R \backslash D)$ plays the role of the negation or denial of the proposition $p(D)$. If D_1 and D_2 are events with $D_1 \subseteq D_2$, one naturally says that the proposition $p(D_1)$ *implies* the proposition $p(D_2)$, since any result that confirms $p(D_1)$ will automatically confirm $p(D_2)$. In this way, the set $\mathfrak{L}$ of all propositions of the form $p(D), D \subseteq R$, forms a proposition system isomorphic to the Boolean algebra $\mathcal{P}(R)$ of all subsets of R. The set $\mathfrak{L}_0$ of all propositions of the form $p(D)$, $D \in \mathfrak{F}$, forms a subsystem of $\mathfrak{L}$ that is isomorphic to the Boolean σ-algebra $\mathfrak{F}$. In this way, the observable events are made to correspond to those propositions that we regard as being operationally meaningful, i.e., the propositions in the system $\mathfrak{L}_0$. Stochastic models for the experimental situation at hand are now traditionally introduced as normed measures on $\mathfrak{F}$, or—what amounts to the same thing—on the propositional system $\mathfrak{L}_0$.

Suppose that $D_1, D_2, D_3, \cdots$ is a sequence of disjoint observable events for the operation $\mathfrak{D}$ such that $R = D_1 \cup D_2 \cup D_3 \cup \cdots$. Using this sequence, we can describe a "coarsened version" $\mathfrak{D}^*$ of the physical operation $\mathfrak{D}$ as follows: To execute $\mathfrak{D}^*$, we execute $\mathfrak{D}$, but we record the result of $\mathfrak{D}^*$ as that unique observable event D_n that contains the result r obtained from our execution of $\mathfrak{D}$. If $\mathfrak{F}^*$ denotes the σ-field generated in R by $D_1, D_2, D_3, \cdots$, then $\mathfrak{F}^*$ is a σ-subfield of $\mathfrak{F}$. By a slight abuse of notation, one can regard the coarsened operation $\mathfrak{D}^*$ as being mathematically represented by the pair $(R, \mathfrak{F}^*)$. If $\mathfrak{D}^*$ is a coarsened version of $\mathfrak{D}$ in the above sense, we say that $\mathfrak{D}$ is a *refinement* of $\mathfrak{D}^*$.

In the empirical arts and sciences, a well-founded study is frequently concerned not with a single physical operation, but rather with some coherent collection $\mathfrak{M}$ of physical operations—usually complete or exhaustive in some sense. We shall refer to such a collection $\mathfrak{M}$ of physical operations as a *manual*, since we might imagine it to be a manual or catalog of procedures. How might the Kolmogorov representation be extended to account for such a situation?

A traditional, and often implicit, answer to this question has been to presume that there exists a suitable "grand canonical operation" $\mathfrak{D}$ that simultaneously refines all of the operations in the manual $\mathfrak{M}$. The manual $\mathfrak{M}$ can then be represented by the appropriate collection of σ-subfields of the σ-field representing $\mathfrak{D}$. Perhaps the prototype for all such $\mathfrak{D}$ is the "grand canonical measurement" of classical mechanics. This "in principle" operation permits one to determine simultaneously the location and the momentum of all the particles of a physical system. It is essential for the determinism claimed for classical mechanics.

In quantum mechanics, the celebrated Heisenberg commutation rules reject both determinism and even an in principle possibility of a grand canonical measurement. Thus, in quantum mechanics, we are denied the convenience of a single classical sample space in terms of which we are always able to confirm or refute the propositions concerning the measurements of quantum mechanical observables.

Therefore, an extension of the Kolmogorov representation to any manual of physical operations by means of a grand canonical operation is prohibited by the tenets of quantum physics. The irreducible uncertainties commonly arising in the descriptive sciences would also suggest the absence of such a grand canonical operation. Incidentally, the existence of such a grand canonical operation is often implicit in the use of "mathematical models."

We are now ready to formulate our generalized definition of a sample space. Suppose that we have a given manual $\mathfrak{M}$ of physical operations. For $\mathfrak{D} \in \mathfrak{M}$ we let $R_{\mathfrak{D}}$ be the result set corresponding to $\mathfrak{D}$. In general, the various result sets $R_{\mathfrak{D}}$ will not be disjoint because of the accidental representation of different results corresponding to different operations by the same symbol. To obviate this difficulty, we introduce the notion of an *outcome* of a physical operation. Outcomes will presently be defined so as to satisfy certain axioms, and so our preliminary definition is only intended to be suggestive. Roughly, an outcome is a result so labeled or so constituted that the physical operation (or operations) capable of producing it can be discerned. [For instance, a result r of the physical operation $\mathfrak{D}$ could be converted into an outcome by replacing it by the ordered pair $(\mathfrak{D}, r)$.]

We are going to assume that the various result sets $R_{\mathfrak{D}}$ for the physical operations $\mathfrak{D}$ in the manual $\mathfrak{M}$ are actually outcome sets. We can now be quite specific about what we mean by this, namely, we are going to require that if $\mathfrak{D}, \mathfrak{P} \in \mathfrak{M}$ and if $R_{\mathfrak{D}} \subseteq R_{\mathfrak{P}}$, then $\mathfrak{D} = \mathfrak{P}$ and $R_{\mathfrak{D}} = R_{\mathfrak{P}}$. In particular, this requirement implies a one-to-one correspondence $\mathfrak{D} \leftrightarrow R_{\mathfrak{D}}$ between physical operations in the manual $\mathfrak{M}$ and their outcome sets. We are accordingly entitled to identify a given physical operation $\mathfrak{D}$ with its outcome set $R_{\mathfrak{D}}$, and it will prove to be mathematically convenient to do just that. Thus, we shall refer to a set of the form $R_{\mathfrak{D}}$, $\mathfrak{D} \in \mathfrak{M}$, as an *operation* (dropping the adjective "physical"). We denote by $\mathfrak{A}$ the set of all such operations; thus, $\mathfrak{A}$ is our mathematical representation for the manual $\mathfrak{M}$ of physical operations. Accordingly, we shall refer to $\mathfrak{A}$ as our *manual of operations*.

Let X denote the set-theoretic union of all of the operations in the manual, $\mathfrak{A}$, $X = \cup \{E \mid E \in \mathfrak{A}\}$. Thus, X is the set of all possible outcomes of all operations in our manual. We define a binary relation $\perp$ on the set X as follows: For $x, y \in X$, $x \perp y$ means that there exists an operation $E \in \mathfrak{A}$ such that both x and y belong to E, but $x \neq y$. Intuitively, $x \perp y$ means that the outcome x *operationally rejects* the outcome y in the sense that there exists an operation $E \in \mathfrak{A}$ for which x and y are mutually exclusive outcomes, i.e., if E was executed and x was obtained as the outcome, then y was not obtained as the outcome of this execution of E. (For an alternate interpretation of $\perp$, see Ref. 4.)

If $E \in \mathfrak{A}$ is an operation, we call a subset D of E an *event for* E. A subset D of X will be called an *event* provided that it is an event for some operation $E \in \mathfrak{A}$. An *orthogonal set* is defined to be a subset K of X

such that $x \perp y$ holds for all $x, y \in K$ with $x \neq y$. Evidently, every event is an orthogonal set. In general, there may be orthogonal sets that are not events. However, we have found that it is mathematically convenient to make the following assumption, which we call the *axiom of coherence:* If D is an orthogonal subset of X and if there exist two operations $E, F \in \mathfrak{A}$ such that $D \subseteq E \cup F$, then D is an event. It is not easy to provide the heuristics for the axiom of coherence at this stage of the game. Suffice it to say that all of the generalized sample spaces which we construct in the sequel will satisfy this axiom and that, without it, severe mathematical difficulties seem to arise. Roughly, the coherence axiom stipulates the existence of a "sufficient number of coherently related operations."

We summarize the above in the following definition: A *generalized sample space* is a triple $(X, \perp, \mathfrak{A})$ consisting of a nonempty set X, a symmetric relation $\perp$ on X such that $x \perp y$ implies $x \neq y$, and a collection $\mathfrak{A}$ of subsets of X satisfying the following conditions:

(i) $X = \cup\{E \mid E \in \mathfrak{A}\}$ (covering condition).
(ii) If $E \in \mathfrak{A}$, and if $x, y \in E$ with $x \neq y$, then $x \perp y$ (orthogonality condition).
(iii) If $E, F \in \mathfrak{A}$, and if $E \subseteq F$, then $E = F$ (irredundancy condition).
(iv) If $E, F \in \mathfrak{A}$, if $D \subseteq E \cup F$, and if $x \perp y$ holds for all $x, y \in D$ with $x \neq y$, then there exists $G \in \mathfrak{A}$ such that $D \subseteq G$ (coherence condition).

If $(X, \perp, \mathfrak{A})$ is such a generalized sample space, we refer to the sets E belonging to the collection $\mathfrak{A}$ as *operations* and we refer to $\mathfrak{A}$ as the *manual of operations.* A subset D of X such that $D \subseteq E$ for some operation $E \in \mathfrak{A}$ is called an *event.* A subset K of X such that $x \perp y$ holds for all $x, y \in K$ with $x \neq y$ is called an *orthogonal set.* Elements $x \in X$ are called *outcomes.* Outcomes x and y with $x \perp y$ are said to be *orthogonal* or to *operationally reject* each other.

Let $\mathfrak{D}$ be a single physical operation with result set R. Clearly, the generalized sample space corresponding to the manual $\mathfrak{M} = \{\mathfrak{D}\}$ consisting just of the single physical operation $\mathfrak{D}$ is $(X, \perp, \mathfrak{A})$, where $X = R$, $\perp$ is the relation of inequality, and $\mathfrak{A} = \{R\}$ is the collection consisting just of the single set R.

More generally, if $(X, \perp, \mathfrak{A})$ is the generalized sample space corresponding to a certain "empirical domain of discourse," then the operations $E \in \mathfrak{A}$ and the manner in which they intertwine constitute a mathematical description of the pertinent physical operations and the manner in which they are related.

2. PROPERTIES OF GENERALIZED SAMPLE SPACES

In what follows, we shall refer to generalized sample spaces simply as sample spaces. Thus, let $(X, \perp, \mathfrak{A})$ be a sample space.

Lemma 1: If D is an orthogonal subset of X and if $E_1, E_2, \ldots, E_n$ is a finite sequence of operations in $\mathfrak{A}$, then $D \subseteq E_1 \cup E_2 \cup \cdots \cup E_n$ implies that D is an event.

The proof of Lemma 1 is an easy exercise in mathematical induction making use of the coherence condition. A consequence of Lemma 1 and the covering condition is that every finite orthogonal subset of X is an event. In particular, using the orthogonality condition and the latter remark, we see that for $x, y \in X$, $x \perp y$ if and only if $x \neq y$ and there exists an operation $E \in \mathfrak{A}$ such that $x, y \in E$.

If D_1 and D_2 are events, we shall say that D_1 and D_2 are *compatible* if there exists an operation $E \in \mathfrak{A}$ such that $D_1 \cup D_2 \subseteq E$.

Lemma 2: Let $D_1, D_2, \ldots, D_n$ be a finite sequence of pairwise compatible events. Then $D_1 \cup D_2 \cup \ldots \cup D_n$ is an event.

Proof: Put $D = D_1 \cup D_2 \cup \cdots \cup D_n$. For each $i = 1, 2, \ldots, n$, choose an operation E_i with $D_i \subseteq E_i$. Then, $D \subseteq E_1 \cup E_2 \cup \cdots \cup E_n$. By Lemma 1, it will be sufficient to show that D is an orthogonal set. Thus, let $x, y \in D$ with $x \neq y$. Say $x \in D_i$ and $y \in D_j$, $1 \leq i, j \leq n$. By the assumption of pairwise compatibility, $D_i \cup D_j$ is an event, so that $x \perp y$ holds and the proof is complete.

Lemma 3: Every operation $E \in \mathfrak{A}$ is a maximal orthogonal set.

Proof: Suppose that $E \in \mathfrak{A}$, but that E is not a maximal orthogonal set. Then, there exists $x \in X$ with $x \notin E$ such that $E \cup \{x\}$ is an orthogonal set. By the covering condition, there exists an operation $F \in \mathfrak{A}$ with $x \in F$. Thus, $E \cup \{x\} \subseteq E \cup F$. By the coherence condition, $E \cup \{x\}$ is an event; hence, there exists an operation $G \in A$ with $E \cup \{x\} \subseteq G$. In particular, $E \subseteq G$, so the irredundancy condition forces $E = G$. This yields the contradiction $x \in E$ and proves the lemma.

The converse of Lemma 3 need not hold, as we now show by an example. Suppose that we have a counting device the output of which is a nonnegative integer. For each integer $n = 0, 1, 2, \cdots$, we describe a physical operation $\mathfrak{D}_n$ as follows: To execute $\mathfrak{D}_n$, read the output r of the counting device and record r as the result of $\mathfrak{D}_n$ if $r \leq n$. On the other hand, if $r > n$, record the result of $\mathfrak{D}_n$ as e_n. Thus, the result set for $\mathfrak{D}_n$ is $R_n = \{0, 1, 2, \ldots, n, e_n\}$. Clearly, $\mathfrak{D}_{n+1}$ is a refinement of $\mathfrak{D}_n$, since, if we know the result of $\mathfrak{D}_{n+1}$, then we automatically know the result of $\mathfrak{D}_n$. We now build a sample space $(X, \perp, \mathfrak{A})$ corresponding to the manual $\mathfrak{M} = \{\mathfrak{D}_n \mid n = 0, 1, 2, \cdots\}$ of physical operations by setting $\mathfrak{A} = \{R_n \mid n = 0, 1, 2, \cdots\}$, $X = \cup_{n=0}^{\infty} R_n$, and defining $x \perp y$ for $x, y \in X$ by the condition that $x \neq y$ and there exist a nonnegative integer n with $x, y \in R_n$. The set $E = \{0, 1, 2, \cdots\}$ of all nonnegative integers is a maximal orthogonal subset of X, but it does not appear as an operation in $\mathfrak{A}$. Intuitively, E corresponds to the physical operation $\mathfrak{D}_\infty$ of reading the counter and recording the output $r \in E$. However, we did not include $\mathfrak{D}_\infty$ in our manual $\mathfrak{M}$ of physical operations.

The above example shows that a maximal orthogonal subset E of a sample space $(X, \perp, \mathfrak{A})$ which is not an operation in $\mathfrak{A}$ might be construed as an "in principle" or as an "idealized" operation which, in some sense, is a "limiting case" of the operations available in the manual $\mathfrak{A}$. If D is an orthogonal subset of X which is not an event, then, by Zorn's lemma, D can be extended to a maximal orthogonal set E which cannot be an operation. It is natural to regard such a D as an

idealized event for the *idealized operation* E. This will be precisely our point of view in the sequel.

Lemma 1 suggests the following definition: The sample space $(X, \perp, \mathfrak{A})$ will be called σ-*coherent* provided that every orthogonal set D that is contained in the union of a countable sequence of operations is an event. If every orthogonal set is an event, so that there are no idealized operations, then we shall call the sample space $(X, \perp, \mathfrak{A})$ *completely coherent*. A slight extension of the argument in Lemma 2 shows that, in a σ-coherent sample space, the union of a countable sequence of pairwise compatible events is again an event.

An arbitrary subset A of the sample space $(X, \perp, \mathfrak{A})$ will be called an *evidence set*. If a physical operation corresponding to an operation $E \in \mathfrak{A}$ is executed and an outcome $e \in E$ is obtained as a consequence, then we shall say that *the evidence A has been secured* precisely when $e \in A$. We denote by $A^\perp$ the set of all outcomes $x \in X$ which operationally reject all of the outcomes $e \in A$ in terms of which the evidence A could be secured. In symbols, then, $A^\perp = \{x \in X \mid x \perp a \text{ for all } a \in A\}$. We define $A^{\perp\perp} = (A^\perp)^\perp$, $A^{\perp\perp\perp} = (A^{\perp\perp})^\perp$, etc. In the following lemma, we collect the basic facts about $A^\perp$. The proof is quite straightforward, using only the facts that $\perp$ is symmetric and that $x \perp y$ implies $x \neq y$.

Lemma 4: Let A and B be evidence sets for the sample space $(X, \perp, \mathfrak{A})$. Then:

(i) $A \cap A^\perp = \emptyset$ ($\emptyset$ denotes the empty set).
(ii) $A \subseteq B$ implies $B^\perp \subseteq A^\perp$.
(iii) $A \subseteq A^{\perp\perp}$.
(iv) $A^\perp = A^{\perp\perp\perp}$.
(v) $\emptyset^\perp = X$ and $X^\perp = \emptyset$.
(vi) If $\mathcal{C}$ is a collection of subsets of X, then $(\cup \{C \mid C \in \mathcal{C}\})^\perp = \cap \{C^\perp \mid C \in \mathcal{C}\}$.

If C is an evidence set such that $C = C^{\perp\perp}$, we say that C is a *closed* evidence set. Using part (v) of Lemma 4, we see that C is closed if and only if there is an evidence set B with $C = B^\perp$. From this and part (iv) of Lemma 4, it follows that the set-theoretic intersection of an arbitrary collection of closed evidence sets is again closed.

If A and B are evidence sets and if $A \subseteq B^\perp$, we shall say that A and B are *orthogonal* to each other and write $A \perp B$. Note that $A \perp B$ means that every outcome x in terms of which the evidence A could be secured operationally rejects every outcome y in terms of which the evidence B could be secured. Evidently, two events are orthogonal if and only if they are compatible and disjoint.

3. OPERATIONAL PROPOSITIONS

Sample spaces, as defined above, must be the source of the evidence in terms of which the propositions of empirical science are to be confirmed and refuted. In this connection, let us reaffirm our view regarding propositions as set forth in Ref. 2: *A proposition is well defined if and only if the exact conditions under which it is regarded as being confirmed, as well as those under which it is regarded as being refuted, are stipulated in terms of admissible evidence.*

Let $(X, \perp, \mathfrak{A})$ be a given sample space. In keeping with the above doctrine concerning propositions, we define an *operational proposition* over $(X, \perp, \mathfrak{A})$ to be an ordered pair (A, B) of evidence sets $A, B \subseteq X$, with the understanding that the proposition (A, B) is *confirmed* exactly when the evidence A is secured and *refuted* exactly when the evidence B is secured. Thus, an operational proposition (A, B) can be confirmed or refuted only by the expedient of executing an operation $E \in \mathfrak{A}$, recording its outcome $e \in E$, and checking to see if $e \in A$ or if $e \in B$. If $e \in A$, then the operational proposition (A, B) is confirmed by this execution of E, while if $e \in B$, then it is refuted by this execution of E. If neither $e \in A$ nor $e \in B$, then the operational proposition (A, B) is neither confirmed nor refuted by this execution of E.

We define Π_X to be the set of all operational propositions over $(X, \perp, \mathfrak{A})$. The *negation* of an operational proposition (A, B) is naturally defined to be the operational proposition $(A, B)' = (B, A)$. If $(A, B), (C, D) \in \Pi_X$, we say that (A, B) *implies* (C, D) and we write $(A, B) \leq (C, D)$ if and only if $A \subseteq C$ and $D \subseteq B$. Thus, $(A, B) \leq (C, D)$ means that every outcome confirming (A, B) confirms (C, D) and every outcome refuting (C, D) refutes (A, B). We say that the operational propositions (A, B) and (C, D) are *disjoint* in case $A \cap C = \emptyset$, that is, (A, B) and (C, D) can never be simultaneously confirmed by any outcome $x \in X$. An operational proposition (A, B) that is disjoint from its negation $(A, B)'$ is said to be *self-consistent*. Thus, the self-consistent operational propositions are precisely those that can never be simultaneously confirmed and refuted by an outcome $x \in X$. If (A, B), $(C, D) \in \Pi_X$ are such that $A \perp C$, then we say that (A, B) is *orthogonal* to (C, D) and write $(A, B) \perp (C, D)$. Evidently, $(A, B) \perp (C, D)$ means that every outcome $x \in X$ that confirms (A, B) operationally rejects every outcome $y \in X$ that could confirm (C, D) and conversely. If the operational proposition (A, B) is orthogonal to its negation $(A, B)'$, that is, if $A \perp B$, then we say that (A, B) is *orthoconsistent*.

Obviously, the system $(\Pi_X, \leq)$ is a complete lattice and the negation map $': \Pi_X \to \Pi_X$ is an anti-automorphism of period two on this lattice. The proposition system $(\Pi_X, \leq, \perp, ')$ will be called the *generalized operational logic* over $(X, \perp, \mathfrak{A})$. An operation $E \in \mathfrak{A}$ is said to *test* the operational proposition (A, B) in Π_X if and only if $E \subseteq A \cup B$; that is, E tests (A, B) precisely when every execution of E yields an outcome $e \in E$ which confirms or refutes (A, B). We say that $(A, B) \in \Pi_X$ is *testable* if there exists an operation $E \in \mathfrak{A}$ that tests (A, B). A collection of operational propositions is said to be *simultaneously testable* if there exists a single operation $E \in \mathfrak{A}$ that tests every operational proposition in the collection.

Let (A, B) be an operational proposition. We say that an outcome $x \in X$ *virtually confirms* (A, B) if it operationally rejects every outcome that could refute (A, B), that is, if $x \in B^\perp$. Similarly, if $x \in A^\perp$, that is, if x operationally rejects every outcome that could confirm (A, B), then we say that x *virtually refutes* (A, B). Notice that (A, B) is orthoconsistent if and only if every outcome that confirms (A, B) also virtually confirms (A, B). We define the *virtual negation* of the proposition $(A, B) \in \Pi_X$ by $(A, B)^\perp = (A^\perp, B^\perp)$. Thus, the virtual negation of (A, B) is that operational proposition which is confirmed by those outcomes that virtually refute (A, B) and is refuted by those outcomes that virtually confirm (A, B). If $(A, B)^\perp =$

$(A,B)'$, then we say that (A,B) is a *closed* operational proposition. Note that (A,B) is closed if and only if A is a closed evidence set and $B = A^\perp$. Clearly, every closed operational proposition is orthoconsistent and every orthoconsistent operational proposition is self-consistent. However, a closed operational proposition need not be testable. It can be shown that our decisions concerning the confirmation and refutation of the propositions associated with events in the non-classical sample space constructed in Ref. 2 lead to closed and testable operational propositions.

Consider, for a moment, the classical situation in which only a single physical operation $\mathfrak{D}$ with outcome set R is under consideration. The corresponding sample space is $(R, \neq, \mathfrak{A})$, where $\mathfrak{A} = \{R\}$ is the collection consisting of the single set R. Clearly, a testable operational proposition over $(R, \neq, \mathfrak{A})$ must be of the form (A,B), where $R = A \cup B$. The operational proposition (A,B) is self-consistent if and only if it is orthoconsistent, and it is orthoconsistent if and only if $A \cap B = \emptyset$. Thus, the testable, orthoconsistent operational propositions—in this classical case—are precisely those of the form $(D, R\backslash D)$, where D is an event. Note that $(D, R\backslash D)$ is automatically closed. Under such circumstances, the customary association of an event D with a proposition $p(D) = (D, R\backslash D)$ is quite natural—so much so, in fact, that such an event and proposition are rarely distinguished. Although it is less transparent in the general case, there is a reasonable extension of this correspondence between events and operational propositions to the situation represented by a generalized sample space.
Let $(X, \perp, \mathfrak{A})$ be any (generalized) sample space and set D be an event for this sample space. Our purpose is to associate with D an operational proposition $p(D) = (A,B)$ over $(X, \perp, \mathfrak{A})$ in such a way that the most salient features of the above classical correspondence are preserved. Naturally, we wish to have $D \subseteq A$ so that any outcome $d \in D$ will confirm $p(D)$. Also, if e is an outcome that operationally rejects every outcome $d \in D$, that is, if $e \in D^\perp$, then we wish to require that e refutes $p(D)$. Finally, we wish to stipulate that $p(D)$ be a closed operational proposition. This requires that $A = A^{\perp\perp}$ and that $B = A^\perp$.

Lemma 5: Let D be an event for the sample space $(X, \perp, \mathfrak{A})$ and let (A,B) be an operational proposition over $(X, \perp, \mathfrak{A})$ such that $D \subseteq A, D^\perp \subseteq B, A = A^{\perp\perp}$ and $B = A^\perp$. Then $A = D^{\perp\perp}$, $B = D^\perp$ and, if $E \in \mathfrak{A}$ is any operation with $D \subseteq E$, then E tests (A,B).

Proof: Using the facts in Lemma 4, we compute as follows: Since $D \subseteq A$, then $A^\perp \subseteq D^\perp$, so $D^{\perp\perp} \subseteq A^{\perp\perp} = A$. Since $D^\perp \subseteq B = A^\perp$, then $A = A^{\perp\perp} \subseteq D^{\perp\perp}$. Thus we have $A = D^{\perp\perp}$ and $B = A^\perp = D^{\perp\perp\perp} = D^\perp$, as required. Suppose that $D \subseteq E \in \mathfrak{A}$. Since E is an orthogonal set, then $E\backslash D \subseteq D^\perp = B$, so $E = D \cup (E\backslash D) \subseteq D \cup B \subseteq A \cup B$, that is, E tests (A,B). The proof is complete.

It follows from Lemma 5 that there is only one definition of $p(D)$ compatible with our requirements, namely $p(D) = (D^{\perp\perp}, D^\perp)$. Furthermore, if D is any event and if we define $p(D)$ in this fashion, then Lemma 5 shows that $p(D)$ is a testable closed operational proposition. It is convenient to extend the above definition of $p(D)$ in such a way that a closed operational proposition $p(A)$ will be affiliated with every evidence set $A \subseteq X$. Thus, we make the following definition: For $A \subseteq X$, $p(A) \in \Pi_X$ is defined by $p(A) = (A^{\perp\perp}, A^\perp)$.

4. COMPLETE STOCHASTIC MODELS

Let $(X, \perp, \mathfrak{A})$ be a sample space. We begin heuristically to introduce the notion of a "complete stochastic model" for the empirical situation represented by $(X, \perp, \mathfrak{A})$. If $x \in X$, denote by $\omega(x)$, $0 \leq \omega(x) \leq 1$, the "long-run relative frequency" of the occurrence of the outcome x as a consequence of the execution of an operation E for which $x \in E$. Implicit in this description of $\omega(x)$ is the supposition that this "long-run relative frequency" is independent of the choice of the operation E, provided only that $x \in E$. If, in a given experimental situation, such independence does not seem to obtain, the indication would be that the sample space $(X, \perp, \mathfrak{A})$ was incorrectly chosen and that it should be replaced by a more realistic sample space. Naturally, the above stipulations require that, for any operation $E \in \mathfrak{A}$, $\sum_{e \in E} \omega(e) = 1$.

The above considerations lead us to the following formal definition: By a *weight function* for the sample space $(X, \perp, \mathfrak{A})$, we mean a real-valued function ω defined on X such that $0 \leq \omega(x) \leq 1$ for all $x \in X$ and such that $\sum_{e \in E} \omega(e) = 1$ holds for all $E \in \mathfrak{A}$. We shall denote by $\Omega(X, \perp, \mathfrak{A})$ the set of all weight functions for the sample space $(X, \perp, \mathfrak{A})$. If $\omega \in \Omega(X, \perp, \mathfrak{A})$ and if D is an event for $(X, \perp, \mathfrak{A})$, we define $\omega(D) = \sum_{d \in D} \omega(d)$. Clearly, $0 \leq \omega(D) \leq 1$, and $\omega(D)$ can be interpreted as the long-run relative frequency with which the evidence D will be secured as a consequence of the execution of operations $E \in \mathfrak{A}$ for which $D \subseteq E$.

Now let $\omega \in \Omega(X, \perp, \mathfrak{A})$, let $E \in \mathfrak{A}$, and let $(A,B) \in \Pi_X$. We define $\omega_E(A,B) = \omega(E \cap A)$, noting that $E \cap A$ is an event. Thus, $\omega_E(A,B)$ can be interpreted as the long-run relative frequency with which the operational proposition (A,B) will be confirmed as a consequence of executions of the operation E.

Theorem 1: Let $\omega \in \Omega(X, \perp, \mathfrak{A})$; let $E, F \in \mathfrak{A}$; let (A,B) be an orthoconsistent operational proposition over $(X, \perp, \mathfrak{A})$ and suppose that both E and F test (A,B). Then $\omega_E(A,B) = \omega_F(A,B)$.

Proof: Since $A \perp B$, then $A \cap B = \emptyset$. Since $F \subseteq A \cup B$, then $F = (F \cap A) \cup (F \cap B)$, where the sets $F \cap A$ and $F \cap B$ are disjoint events. It follows that $1 = \omega(F) = \omega(F \cap A) + \omega(F \cap B)$. Put $D = (E \cap A) \cup (F \cap B)$. Since $A \perp B$ and since E and F are orthogonal sets, it follows that D is an orthogonal set. Since $D \subseteq E \cup F$, the coherence condition implies the existence of an operation $G \in \mathfrak{A}$ such that $D \subseteq G$. It follows that $\omega(D) \leq \omega(G) = 1$. Since A and B are disjoint, then $(E \cap A)$ and $(F \cap B)$ are disjoint events; hence, $\omega(E \cap A) + \omega(F \cap B) = \omega(D) \leq 1 = \omega(F) = \omega(F \cap A) + \omega(F \cap B)$. Consequently, $\omega(E \cap A) \leq \omega(F \cap A)$. A symmetric argument shows that $\omega(F \cap A) \leq \omega(E \cap A)$, and proves the theorem.

Let $\omega \in \Omega(X, \perp, \mathfrak{A})$ and let (A,B) be any orthoconsistent testable operational proposition over $(X, \perp, \mathfrak{A})$. We define $\omega(A,B)$ by $\omega(A,B) = \omega_E(A,B)$, where E is any test operation in $\mathfrak{A}$ for (A,B). By Theorem 1, $\omega(A,B)$ is well defined. Intuitively, $\omega(A,B)$ is the long-run relative frequency with which (A,B) will be confirmed as a consequence of the execution of test operations for (A,B). In particular, if D is any event,

then $p(D) = (D^{\perp\perp}, D^{\perp})$ is a testable orthoconsistent operational proposition, so that $\omega(p(D))$ is defined.

Corollary 1: Let $\omega \in \Omega(X, \perp, \mathfrak{A})$ and let D be an event for the sample space $(X, \perp, \mathfrak{A})$. Then, $\omega(p(D)) = \omega(D)$.

Proof: Since D is an event, there exists an operation $E \in \mathfrak{A}$ with $D \subseteq E$. By Lemma 5, E tests $p(D) = (D^{\perp\perp}, D^{\perp})$. Hence, $\omega(p(D)) = \omega_E(D^{\perp\perp}, D^{\perp}) = \omega(D^{\perp\perp} \cap E) = \omega(D)$, since $D^{\perp\perp} \cap E = D^{\perp\perp} \cap (D \cup (E\backslash D)) = D$. The corollary is proved.

We denote by $\Pi(X, \perp, \mathfrak{A})$ the set of all operational propositions of the form $p(D)$, where D is an event for $(X, \perp, \mathfrak{A})$. The system $(\Pi(X, \perp, \mathfrak{A}), \leq, \perp)$ will be called the *logic* of the sample space $(X, \perp, \mathfrak{A})$. If $\omega \in \Omega(X, \perp, \mathfrak{A})$ is any weight function, then ω induces a real-valued function defined on $\Pi(X, \perp, \mathfrak{A})$ by $\omega(p(D)) = \omega(D)$ for every event D (Corollary 1). The use of the same symbol ω to denote both the weight function and the induced function on $\Pi(X, \perp, \mathfrak{A})$ should cause no confusion—one can always tell from the context what is meant. A function induced on $\Pi(X, \perp, \mathfrak{A})$ in this manner will be called a *regular state* on the logic $\Pi(X, \perp, \mathfrak{A})$.

Let D_1 and D_2 be events for the sample space $(X, \perp, \mathfrak{A})$. Clearly, $D_1 \perp D_2$ if and only if $p(D_1) \perp p(D_2)$ in the logic $\Pi(X, \perp, \mathfrak{A})$. Thus, if $p(D_1) \perp p(D_2)$, the coherence condition implies that $D_1 \cup D_2$ is again an event for $(X, \perp, \mathfrak{A})$; hence, $p(D_1 \cup D_2) \in \Pi(X, \perp, \mathfrak{A})$. Since $D_1 \subseteq D_1 \cup D_2$, then, using part (ii) of Lemma 4 twice, we have $D_1^{\perp\perp} \subseteq (D_1 \cup D_2)^{\perp\perp}$. Similarly, $D_2^{\perp\perp} \subseteq (D_1 \cup D_2)^{\perp\perp}$. This implies that $p(D_1 \cup D_2)$ is an upper bound in the partially ordered set $(\Pi(X, \perp, \mathfrak{A}), \leq)$ for $p(D_1)$ and $p(D_2)$. Suppose that D_3 is an event for $(X, \perp, \mathfrak{A})$ and that $p(D_3)$ is also an upper bound for $p(D_1)$ and $p(D_2)$. Then we have $D_1^{\perp\perp} \subseteq D_3^{\perp\perp}$ and $D_2^{\perp\perp} \subseteq D_3^{\perp\perp}$; hence $D_1 \cup D_2 \subseteq D_1^{\perp\perp} \cup D_2^{\perp\perp} \subseteq D_3^{\perp\perp}$, and so $(D_1 \cup D_2)^{\perp\perp} \subseteq D_3^{\perp\perp\perp\perp} = D_3^{\perp\perp}$, that is, $p(D_1 \cup D_2) \leq p(D_3)$. This goes to show that $p(D_1 \cup D_2)$ is the least upper bound in $(\Pi(X, \perp, \mathfrak{A}), \leq)$ for $p(D_1)$ and $p(D_2)$.

The above considerations lead us to define $p(D_1) \oplus p(D_2) = p(D_1 \cup D_2)$ whenever $p(D_1) \perp p(D_2)$ in the logic $\Pi(X, \perp, \mathfrak{A})$. By an easy induction, based upon the above arguments, we see that if $p(D_1), p(D_2), \ldots, p(D_n)$ is a finite sequence of pairwise orthogonal propositions in the logic $\Pi(X, \perp, \mathfrak{A})$, then $p(D_1 \cup D_2 \cup \cdots \cup D_n)$ belongs to $\Pi(X, \perp, \mathfrak{A})$ and is effective as the least upper bound in $\Pi(X, \perp, \mathfrak{A})$ of the original sequence. Thus, we define $p(D_1) \oplus p(D_2) \oplus \cdots \oplus P(D_n) = p(D_1 \cup D_2 \cup \cdots \cup D_n)$.

Lemma 6: Let ω be a regular state on the logic $\Pi(X, \perp, \mathfrak{A})$. For each $p(D) \in \Pi(X, \perp, \mathfrak{A})$, $\omega(p(D))$ is a real number between 0 and 1. If $E \in \mathfrak{A}$, then $\omega(p(E)) = 1$. Also, $\omega(p(\emptyset)) = 0$. Finally, if $p(D_1), p(D_2), \ldots, p(D_n)$ is a finite sequence of pairwise orthogonal propositions in the logic $\Pi(X, \perp, \mathfrak{A})$, then $\omega(p(D_1) \oplus p(D_2) \oplus \cdots \oplus p(D_n)) = \omega(p(D_1)) + \omega(p(D_2)) + \cdots + \omega(p(D_n))$.

We omit the easy proof of Lemma 6. If the sample space $(X, \perp, \mathfrak{A})$ is σ-coherent, all of the above considerations can easily be extended to countably infinite sequences of mutually orthogonal propositions in the logic $\Pi(X, \perp, \mathfrak{A})$ and, in particular, regular states are countably additive over orthogonal sequences of propositions in this logic.

Lemma 7: Let ω be a regular state on the logic $\Pi(X, \perp, \mathfrak{A})$ and let D_1, D_2 be events for $(X, \perp, \mathfrak{A})$ such that $p(D_1) \leq p(D_2)$. Then, $\omega(p(D_1)) \leq \omega(p(D_2))$.

Proof: Since $p(D_1) \leq p(D_2)$, then $D_1^{\perp\perp} \subseteq D_2^{\perp\perp}$; hence, $D_2^{\perp\perp\perp} \subseteq D_1^{\perp\perp\perp}$, that is, $D_2^{\perp} \subseteq D_1^{\perp}$. Since D_2 is an event, there exists an operation $E \in \mathfrak{A}$ such that $D_2 \subseteq E$. Since E is an orthogonal set, it follows that $E\backslash D_2 \subseteq D_2^{\perp} \subseteq D_1^{\perp}$, that is, D_1 and $E\backslash D_2$ are orthogonal events. By the coherence condition, there exists an operation $F \in \mathfrak{A}$ such that $D_1 \cup (E\backslash D_2) \subseteq F$. It follows that $\omega(D_1) + \omega(E\backslash D_2) = \omega(D_1) + \omega(E) - \omega(D_2) \leq \omega(F)$. Since $\omega(E) = \omega(F) = 1$, the lemma is proved.

Suppose that Δ is a set of regular states on the logic $\Pi(X, \perp, \mathfrak{A})$. We shall say that Δ is a *full* set of regular states if the following condition holds: If $p(D_1)$, $p(D_2) \in \Pi(X, \perp, \mathfrak{A})$ are such that $\omega(p(D_1)) \leq \omega(p(D_2))$ holds for every $\omega \in \Delta$, then $p(D_1) \leq p(D_2)$. Thus, to say that Δ is a full set of regular states on $\Pi(X, \perp, \mathfrak{A})$ is to say that the implication relation $\leq$ on $\Pi(X, \perp, \mathfrak{A})$ can be recaptured merely from a knowledge of Δ.

Theorem 2: Let Δ be a set of regular states on the logic $\Pi(X, \perp, \mathfrak{A})$ of the sample space $(X, \perp, \mathfrak{A})$. Then Δ is a full set of regular states if and only if the following condition holds: Regarding Δ as a set of weight functions on $(X, \perp, \mathfrak{A})$, if $x, y \in X$ and if the condition $x \perp y$ fails, then there exists $\omega \in \Delta$ such that $\omega(x) + \omega(y) > 1$.

Proof: Suppose first that the given condition holds, but that Δ is not full. Then, there exist events D_1 and D_2 such that $\omega(D_1) \leq \omega(D_2)$ holds for all $\omega \in \Delta$, but $p(D_1) \not\leq p(D_2)$. Since $p(D_1) \not\leq p(D_2)$, we cannot have $D_2^{\perp} \subseteq D_1^{\perp}$; hence, there exists $x \in D_2^{\perp}$ and there exists $y \in D_1$ with $x \notin \{y\}^{\perp}$. By hypothesis, then, there exists $\omega \in \Delta$ such that $\omega(x) + \omega(y) > 1$. Now, $\omega(y) \leq \omega(D_1) \leq \omega(D_2)$. Since $x \in D_2^{\perp}$, then $\{x\}$ and D_2 are orthogonal events, so that the coherence condition implies the existence of an operation E with $\{x\} \cup D_2 \subseteq E$. It follows that $\omega(x) + \omega(D_2) \leq 1 < \omega(x) + \omega(y)$; hence, $\omega(D_2) < \omega(y)$. Since $y \in D_1$, this yields the contradiction $\omega(y) \leq \omega(D_1) \leq \omega(D_2) < \omega(y)$.

Conversely, suppose that Δ is full, but that the given condition fails. Then, there exist $x, y \in X$ such that $x \notin \{y\}^{\perp}$, but $\omega(x) + \omega(y) \leq 1$ holds for all $\omega \in \Delta$. Choose an operation $E \in \mathfrak{A}$ such that $y \in E$ and let B denote the event $B = E\backslash\{y\}$. We have $\omega(x) + \omega(y) \leq 1 = \omega(E) = \omega(B) + \omega(y)$; hence, $\omega(x) \leq \omega(B)$ holds for all $\omega \in \Delta$. Since Δ is full, it follows that $p(\{x\}) \leq p(B)$, that is, $\{x\}^{\perp\perp} \subseteq B^{\perp\perp}, B^{\perp} \subseteq \{x\}^{\perp}$. Since $y \in B^{\perp}$, then $y \in \{x\}^{\perp}$, contradicting $x \notin \{y\}^{\perp}$ and proving the theorem.

We shall call a sample space $(X, \perp, \mathfrak{A})$ a *Dacey Space*[4] if it has the following property: If $E \in \mathfrak{A}$, if $x, y \in X$, and if $E \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$, then $x \perp y$.

Theorem 3: Suppose that the logic $\Pi(X, \perp, \mathfrak{A})$ of the sample space $(X, \perp, \mathfrak{A})$ admits a full set of regular states Δ. Then $(X, \perp, \mathfrak{A})$ is a Dacey space.

Proof: Suppose that $E \in \mathfrak{A}$, that $x, y \in X$, and that $E \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$. Define events B and D by $B = E \cap \{x\}^{\perp}, D = E\backslash B$. Now E is the disjoint union of B and D; hence, $1 = \omega(E) = \omega(B) + \omega(D)$ holds for all $\omega \in \Delta$. Clearly, $B \subseteq \{x\}^{\perp}$ and $D \subseteq \{y\}^{\perp}$; hence, by the co-

herence condition, there exist operations $F, G \in \mathfrak{A}$ with $B \cup \{x\} \subseteq F$ and $D \cup \{y\} \subseteq G$. Thus, for all $\omega \in \Delta$, $\omega(B) + \omega(x) \leq \omega(F) = 1$ and $\omega(D) + \omega(y) \leq \omega(G) = 1$. Adding the latter inequalities gives $1 + \omega(x) + \omega(y) = \omega(B) + \omega(D) + \omega(x) + \omega(y) \leq 2$, that is, $\omega(x) + \omega(y) \leq 1$ for all $\omega \in \Delta$. Theorem 2 now implies $x \perp y$ and completes the proof.

The converse of Theorem 3 is false. Greechie[5] has given examples of finite Dacey spaces $(X, \perp, \mathfrak{A})$, whose logics are even orthomodular lattices,[6] but whose logics do not admit any regular states whatsoever!

Theorem 4: Let $(X, \perp, \mathfrak{A})$ be a Dacey space. Then, if $p(D) \in \Pi(X, \perp, \mathfrak{A})$, it follows that $(p(D))^{\perp} = (p(D))' \in \Pi(X, \perp, \mathfrak{A})$. In particular, if the logic $\Pi(X, \perp, \mathfrak{A})$ carries a full set of regular states, then it is closed under the negation mapping $p(D) \mapsto (p(D))'$.

Proof: Let D be an event for $(X, \perp, \mathfrak{A})$. We must find an event B such that $(D^{\perp}, D^{\perp\perp}) = (B^{\perp\perp}, B^{\perp})$, that is, such that $B^{\perp\perp} = D^{\perp}$. Since D is an event, there exists an operation $E \in \mathfrak{A}$ such that $D \subseteq E$. Put $B = E \backslash D$. Since E is an orthogonal set, then $B \subseteq D^{\perp}$. By Lemma 4, $B^{\perp\perp} \subseteq D^{\perp\perp\perp} = D^{\perp}$. It remains to show that $D^{\perp} \subseteq B^{\perp\perp}$. Thus, let $x \in D^{\perp}$, $y \in B^{\perp}$. It will suffice to show that $x \perp y$. Since $x \in D^{\perp}$, then, by Lemma 4, $D \subseteq D^{\perp\perp} \subseteq \{x\}^{\perp}$. Similarly, $B \subseteq \{y\}^{\perp}$. Since $E = D \cup B \subseteq \{x\}^{\perp} \cup \{y\}^{\perp}$, the hypothesis that $(X, \perp, \mathfrak{A})$ is a Dacey space implies $x \perp y$, and the theorem is proved.

Current research in "quantum logic" often requires that the set of admissible states be not only full, but also strong.[7] Thus, we define a set Δ of regular states on the logic $\Pi(X, \perp, \mathfrak{A})$ to be *strong* provided that whenever $p(D_1)$ and $p(D_2)$ are propositions in $\Pi(X, \perp, \mathfrak{A})$ such that, for every $\omega \in \Delta$, $\omega(p(D_1)) = 1$ implies $\omega(p(D_2)) = 1$, then $p(D_1) \leq p(D_2)$. Clearly, every strong set of regular states on $\Pi(X, \perp, \mathfrak{A})$ is automatically full. By an argument similar to that used to prove Theorem 2, one can prove the following theorem:

Theorem 5: Let Δ be a set of regular states on the logic $\Pi(X, \perp, \mathfrak{A})$. Then Δ is strong if and only if the following condition holds: Regarding Δ as a set of weight functions on $(X, \perp, \mathfrak{A})$, if $x, y \in X$ with $x \notin \{y\}^{\perp}$, there exists $\omega \in \Delta$ such that $\omega(x) = 1$ and $\omega(y) \neq 0$.

5. THREE EXAMPLES

We now give three examples to show that our techniques can handle not only classical probability theory, but the probabilities arising in quantum logics as well.

Example I

Let $\mathfrak{D}$ denote a single physical operation with result set R. Let $\mathfrak{F}$ be a given σ-field of subsets of R, the elements of $\mathfrak{F}$ being though of as observable events for $\mathfrak{D}$ in the usual sense. It will be convenient (although it is not necessary) for us to assume that if $x \in R$, then the set $\{x\}$ consisting of the single result x is an observable event. Let $E = \{M_1, M_2, \cdots\}$ denote a finite or countably infinite partition of R into pairwise disjoint nonempty elements $M_1, M_2, \cdots$ of the σ-field $\mathfrak{F}$. With each such E, we associate a coarsened version $\mathfrak{D}_E$ of the physical operation $\mathfrak{D}$ as follows: To execute $\mathfrak{D}_E$, execute $\mathfrak{D}$, but record the result of $\mathfrak{D}_E$ as the unique set M_n, $n = 1, 2, \cdots$, that contains the result of $\mathfrak{D}$ thereby obtained. We define $\mathfrak{A}$ to be the collection of all such partitions E of R with the understanding that each $E \in \mathfrak{A}$ is to be interpreted as the outcome set of the corresponding physical operation $\mathfrak{D}_E$.

Let $X = \cup \{E | E \in \mathfrak{A}\}$, so that $X = \{M \in \mathfrak{F} | M \neq \emptyset\}$. For $M, N \in X$, we define $M \perp N$ if and only if $M \cap N = \emptyset$. Thus, $M \perp N$ if and only if there exists $E \in \mathfrak{A}$ with $M, N \in E$ and $M \neq N$. One easily verifies that $(X, \perp, \mathfrak{A})$ is a σ-coherent sample space. Let $G = \{\{x\} | x \in R\}$, noting that G is a maximal orthogonal subset of X, but that—unless R is countable—G is not an operation in $\mathfrak{A}$. Thus, if R is uncountable, then G is an idealized operation for the sample space $(X, \perp, \mathfrak{A})$. Of course, the idealized operation G corresponds to the original physical operation $\mathfrak{D}$. In an obvious sense, the idealized operation G is a "limiting case" of the operations $E \in \mathfrak{A}$.

Clearly, the events for the sample space $(X, \perp, \mathfrak{A})$ are precisely the subsets D of X of the form $D = \{M_1, M_2, \cdots\}$ where $M_1, M_2, \cdots$ is a finite or countably infinite sequence of pairwise disjoint nonempty elements of the σ-field $\mathfrak{F}$. If $N \in X$, then, evidently, $N \in D^{\perp}$ if and only if N is disjoint from $\cup \{M | M \in D\}$ and $N \in D^{\perp\perp}$ if and only if N is a subset of $\cup \{M | M \in D\}$.

Let $G = \{\{x\} | x \in R\}$ be the (possibly) idealized operation in $(X, \perp, \mathfrak{A})$ corresponding to the physical operation $\mathfrak{D}$. Note that, for any event D for $(X, \perp, \mathfrak{A})$, G tests $p(D)$ in the sense that $G \subseteq D^{\perp\perp} \cup D^{\perp}$. In this sense, all of the propositions in the logic $\Pi(X, \perp, \mathfrak{A})$ are—at least "in principle"—simultaneously testable.

Define a mapping f: $\Pi(X, \perp, \mathfrak{A}) \to \mathfrak{F}$ by $f(p(D)) = \cup \{M | M \in D\}$ for each $p(D) \in \Pi(X, \perp, \mathfrak{A})$. One shows easily that f is an isomorphism of the logic $\Pi(X, \perp, \mathfrak{A})$ onto the Boolean σ-algebra $\mathfrak{F}$. In particular, for $p(D_1)$, $p(D_2) \in \Pi(X, \perp, \mathfrak{A})$, $p(D_1) \leq p(D_2)$ if and only if $f(p(D_1)) \subseteq f(p(D_2))$ and $p(D_1) \perp p(D_2)$ if and only if $f(p(D_1))$ is disjoint from $f(p(D_2))$. Also, if $p(D_1)$, $p(D_2), \cdots$ is a sequence of pairwise orthogonal propositions in $\Pi(X, \perp, \mathfrak{A})$, then $f(p(D_1) \oplus p(D_2) \oplus \cdots) = f(p(D_1)) \cup f(p(D_2)) \cup \cdots$.

Suppose now that $(R, \mathfrak{F}, \mu)$ is a probability space, that is, suppose that μ is a normed measure defined on the σ-field $\mathfrak{F}$. Define a weight function $\omega = \omega(\mu)$ on the sample space $(X, \perp, \mathfrak{A})$ by $\omega(M) = \mu(M)$ for all $M \in X$. As always, the weight function ω induces a regular state—also denoted by ω—on the logic $\Pi(X, \perp, \mathfrak{A})$. Clearly, for $p(D) \in \Pi(X, \perp, \mathfrak{A})$, we have $\omega(p(D)) = \mu(f(p(D)))$. It is easy to see that, in this way, one can set up a one-to-one correspondence $\omega \leftrightarrow \mu$ between regular states ω on the logic $\Pi(X, \perp, \mathfrak{A})$ and probability measures μ defined on the σ-field $\mathfrak{F}$.

Thus, classical probability theory can be subsumed by our theory of (generalized) sample spaces by the simple trick of promoting the classical observable events to the status of outcomes.

Example II

Quantum logic was born in 1936 with the publication of Birkhoff and von Neumann's ground breaking paper.[8] The appearance in 1963 of Mackey's book on the mathematical foundations of quantum mechanics[9] generated renewed interest in the study of abstract quantum logics. Mackey has shown[10] that a system

satisfying his Axioms[11] I–VI is mathematically equivalent to an orthocomplemented partially ordered set $\mathcal{L}$, closed under the formation of countable orthogonal suprema and equipped with a full strongly convex family of probability measures $\mathcal{S}$. Let $(\mathcal{L}, \mathcal{S})$ be any such system and denote the orthocomplementation on $\mathcal{L}$ by $e \mapsto e'$. Let $X = \{e \in \mathcal{L} \mid e \neq 0\}$ and define two elements $e, f \in X$ to be *orthogonal*, in symbols $e \perp f$, provided that $e \leq f'$. Let $\mathfrak{A}$ denote the family of all countable maximal orthogonal subsets E of X. Then $(X, \perp, \mathfrak{A})$ is a σ-coherent sample space. Let $\alpha \in \mathcal{S}$ and define a weight function ω_α on $(X, \perp, \mathfrak{A})$ by $\omega_\alpha(e) = \alpha(e)$ for all $e \in X$. Denote by Δ the set of all such weight functions ω_α as α runs through $\mathcal{S}$. Then Δ is a full set of regular states on the logic $\Pi(X, \perp, \mathfrak{A})$. Given any $e \in \mathcal{L}$, define an operational proposition $q(e) \in \Pi_X$ by $q(e) = (\{f \in X \mid f \leq e\}, \{g \in X \mid g \leq e'\})$. It is not difficult to verify that the operational proposition $q(e)$ so defined actually belongs to the logic $\Pi(X, \perp, \mathfrak{A})$ and that the mapping $q: \mathcal{L} \to \Pi(X, \perp, \mathfrak{A})$ is an isomorphism of $\mathcal{L}$ onto the logic $\Pi(X, \perp, \mathfrak{A})$. Furthermore, we have $\alpha(e) = \omega_\alpha(q(e))$ for all $e \in \mathcal{L}$.

Conversely, let $(X, \perp, \mathfrak{A})$ be any σ-coherent sample space and let Δ be any set of weight functions on $(X, \perp, \mathfrak{A})$ which is closed under the formation of pointwise countable convex combinations. Suppose, further, that Δ is a full set or regular states on the logic $\Pi(X, \perp, \mathfrak{A})$. It follows readily from the considerations in Sec. 4 of the present paper that, with $\mathcal{L} = \Pi(X, \perp, \mathfrak{A})$ and $\mathcal{S} = \Delta$, we obtain a system satisfying Mackey's conditions. In this way, we see that quantum logic can be subsumed by our theory of (generalized) sample spaces.

We now give a specific mathematical example pertinent to nonrelativistic quantum mechanics. Let $\mathcal{H}$

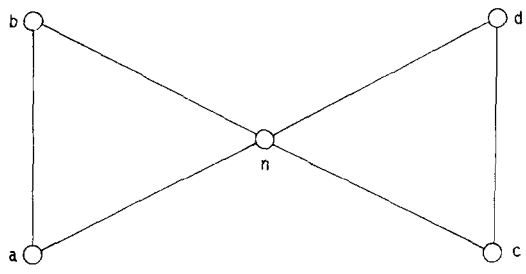


FIG. 1.

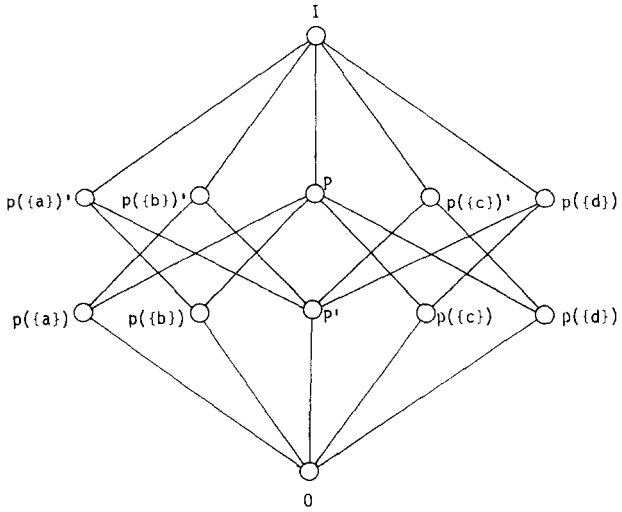


FIG. 2.

denote a separable infinite-dimensional complex Hilbert space. Denote by X the set of all normalized vectors $\Psi \in \mathcal{H}$. Say that two vectors Ψ and φ in X are orthogonal, in symbols $\Psi \perp \varphi$, if the inner product $\langle \Psi \mid \varphi \rangle$ vanishes. Let $\mathfrak{A}$ denote the collection of all maximal orthogonal subsets of X. Then $(X, \perp, \mathfrak{A})$ is a σ-coherent sample space. An event D for $(X, \perp, \mathfrak{A})$ is simply an orthogonal set of normalized vectors in $\mathcal{H}$. If D is such an event, define P_D to be the orthogonal projection onto the closed linear span of the vectors in D. One proves easily that if A and B are events for $(X, \perp, \mathfrak{A})$, then $p(A) \leq p(B)$ in the logic $\Pi(X, \perp, \mathfrak{A})$ if and only if $P_A = P_A P_B$. Also, $p(A) = p(B)'$ in the logic $\Pi(X, \perp, \mathfrak{A})$ if and only if $P_A = 1 - P_B$. Hence, the logic $\Pi(X, \perp, \mathfrak{A})$ is isomorphic to the complete orthocomplemented lattice of all orthogonal projection operators on the Hilbert space $\mathcal{H}$. This is especially interesting in view of von Neumann's interpretation[12] of orthogonal projection operators as quantum mechanical propositions. Using Gleason's theorem,[13] one can show that the weight functions on $(X, \perp, \mathfrak{A})$—or, what is the same thing, the regular states on $\Pi(X, \perp, \mathfrak{A})$—are in one-to-one correspondence with the von Neumann density operators on the Hilbert space $\mathcal{H}$.

Example III

Suppose we have a device that, from time to time, emits a particle and projects it along a linear scale. We consider two physical operations $\mathfrak{D}_1$ and $\mathfrak{D}_2$ defined as follows: To execute $\mathfrak{D}_1$, we look to see if there is a particle present. If there is not, we record the outcome of $\mathfrak{D}_1$ as the symbol n. If there is, we measure its position coordinate x. If $x \geq 1$, we record the outcome of $\mathfrak{D}_1$ as the symbol a, while if $x < 1$, we record the outcome of $\mathfrak{D}_1$ as the symbol b. Thus, the outcome of set $\mathfrak{D}_1$ is $R_1 = \{n, a, b\}$. To execute $\mathfrak{D}_2$, we look to see if there is a particle present. If there is not, we record the outcome of $\mathfrak{D}_2$ as the symbol n. If there is, we measure the x-component p_x of its momentum, recording the symbol c as the outcome of $\mathfrak{D}_2$ if $p_x \geq 1$ and the symbol d as the outcome of $\mathfrak{D}_2$ if $p_x < 1$. Thus, the outcome set for $\mathfrak{D}_2$ is $R_2 = \{n, c, d\}$. (The reason for our identification of the outcome n of $\mathfrak{D}_1$ with the outcome n of $\mathfrak{D}_2$ should be clear to the reader.)

We now build a sample space $(X, \perp, \mathfrak{A})$ reflecting the empirical "universe of discourse" represented by the manual $\{\mathfrak{D}_1, \mathfrak{D}_2\}$ of physical operations. We take $\mathfrak{A} = \{R_1, R_2\}$ and $X = R_1 \cup R_2 = \{a, b, n, c, d\}$. We define $\perp$ as usual, by specifying that, for $x, y \in X$, $x \perp y$ means $x \neq y$, but there exists $E \in A$ such that $x, y \in E$. Thus, for example, $a \perp b$, but it is false that $a \perp c$. The orthogonality relation $\perp$ on X can conveniently be depicted by the graph in Fig. 1. In this graph, the various outcomes $x \in X$ are represented by the nodes, and two nodes representing two orthogonal outcomes are connected by a line segment. Note that the operations R_1 and R_2 in the manual $\mathfrak{A}$ appear as maximal orthogonal sets.

There are exactly fourteen different events for the sample space $(X, \perp, \mathfrak{A})$. However, since $p(R_1) = p(R_2)$ and since $p(\{a, b\}) = p(\{c, d\})$, there are only twelve different propositions in the logic $\Pi(X, \perp, \mathfrak{A})$. Clearly, the proposition $P = p(\{a, b\}) = p(\{c, d\})$ is confirmed precisely when a particle has been observed in the course of an execution of either $\mathfrak{D}_1$ or $\mathfrak{D}_2$. The proposition $p(R_1) = p(R_2)$ is automatically confirmed by

any outcome. We denote it, in what follows, by the symbol I. Thus, $I = (X, \emptyset) \in \Pi(X, \perp, \mathfrak{a})$. The proposition $p(\emptyset)$ is never confirmed by any outcome. We denote it, in what follows, by the symbol O. Thus, $O = (\emptyset, X) = I' \in \Pi(X, \perp, \mathfrak{a})$.

We have discussed three of the propositions in the logic $\Pi(X, \perp, \mathfrak{a})$, namely, O, P and I. Clearly, $P' = p(\{n\}) \in \Pi(X, \perp, \mathfrak{a})$ is confirmed precisely when one of the operations $\mathfrak{D}_1$ or $\mathfrak{D}_2$ is executed, but no particle is observed. We can now display all twelve of the propositions in the logic $\Pi(X, \perp, \mathfrak{a})$, and the implication relations existing between them, by the diagram[14] in Fig. 2. Note that, for instance, $p(\{a\}) \leq P$ (if a particle with position coordinate $x \geq 1$ has been observed, then a particle has been observed), but $p(\{a\}) \not\leq (p(\{c\}))'$ (having observed a particle with position coordinate $x \geq 1$, we are not obliged to conclude that its momentum p_x must be less than 1).

Via Theorem 5 and Fig. 1, it can readily be verified that the logic $\Pi(X, \perp, \mathfrak{a})$ admits a strong set of regular states. Actually, $\Pi(X, \perp, \mathfrak{a})$ is an orthomodular lattice.[15] The subset $\Pi_1(X, \perp, \mathfrak{a})$ of $\Pi(X, \perp, \mathfrak{a})$ consisting of those propositions $p(D) \in \Pi(X, \perp, \mathfrak{a})$ for which R_1 is a test operation form a maximal Boolean sublattice of $\Pi(X, \perp, \mathfrak{a})$. (Note: $\Pi_1(X, \perp, \mathfrak{a})$ is the "left half" of Fig. 2.) An analogous relation exists between R_2 and the "right half" of Fig. 2.

Note that the least upper bound (or the join) of $p(\{a\})$ and $p(\{d\})$ in the lattice $\Pi(X, \perp, \mathfrak{a})$ is given by $p(\{a\}) \vee p(\{d\}) = P = p(\{a, b\}) = (\{a, b, c, d\}, \{n\})$. Thus, $p(\{a\}) \vee p(\{d\})$ is confirmed by any outcome that confirms at least one of the two propositions $p(\{a\})$ or $p(\{d\})$. Also, $p(\{a\}) \vee p(\{d\})$ is refuted precisely by those outcomes that refute both $p(\{a\})$ and $p(\{d\})$. To this extent, $p(\{a\}) \vee p(\{d\})$ behaves like the classical disjunction of the two propositions $p(\{a\})$ and $p(\{d\})$. However, it is important to notice that $p(\{a\}) \vee p(\{d\})$ fails to behave exactly like the classical disjunction since there are outcomes, namely b and c, which confirm $p(\{a\}) \vee p(\{d\})$, but which do not confirm either $p(\{a\})$ or $p(\{d\})$. However, the operational significance of this departure from classical logic is quite clear in the simple example at hand. We could, of course, form an operational proposition $(\{a, d\}, \{n\}) \in \Pi_X$ which would conform in all respects to the classical disjunction of $p(\{a\})$ and $p(\{d\})$; however, the operational proposition $(\{a, d\}, \{n\})$ fails to be testable!

[1] J. M. Jauch, E. P. Wigner, and M. M. Yanase, Nuovo Cimento **48B**, 145 (1967).
[2] C. H. Randall and D. J. Foulis, Amer. Math. Monthly **77**, 363 (1970).
[3] A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1956), 2nd ed. (German ed., 1933).
[4] J. C. Dacey, Caribbean J. Sci. Math. **1**, 51 (1969).
[5] R. J. Greechie, J. Combinatorial Theory **10**, 119 (1971).
[6] Garrett Birkhoff, *Lattice Theory*, AMS Colloq. Publ. XXV (Amer. Math. Soc., Providence, R.I., 1967), 3rd ed., p. 53.
[7] J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).
[8] Garret Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[9] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[10] Reference 9, p. 68.
[11] Reference 9, pp. 63–66.
[12] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1955), pp. 247–54.
[13] A. M. Gleason, J. Math. Mech. **6**, 885 (1953).
[14] Reference 6, p. 4.
[15] Reference 6, p. 53.

Some Theorems on Atomicity in Axiomatic Quantum Mechanics*

F. Jenč

Institute of Theoretical Physics, University of Marburg, 355 Marburg/Lahn, Federal Republic of Germany

(Received 18 February 1971)

Given a complete orthocomplemented lattice L and a set S of nonnegative real functions on L, sufficient conditions are established that S should fulfill in order that L be atomic. The conditions are investigated under which L may be represented by the lattice of all closed subspaces of a separable Hilbert space. (As is well known, the atomicity of L plays an important role here.) Some unsolved problems are pointed out. In axiomatic quantum mechanics, the lattice L may represent the set of propositions whereas the set of functions S represents the set of physical states. The conditions imposed on the pair (L, S) then have a simple and plausible physical interpretation; an important condition imposed on (L, S) is the existence of the "maximal" (i.e., maximally determined) states which appear in the theory as limit constructions.

1. INTRODUCTION

The validity of the covering law and the atomicity of the lattice of propositions in axiomatic quantum mechanics are sufficient conditions that the complete, orthocomplemented irreducible and orthomodular lattice of propositions in which any set of pairwise orthogonal elements is at most countable may be represented as the lattice of closed subspaces of a separable Hilbert space. This scheme corresponds to the generally accepted mathematical formulation of quantum mechanics.[1–3]

In a recent interesting paper, Jauch and Piron[4] tried to motivate the validity of the covering law and the atomicity of the lattice of propositions in axiomatic quantum mechanics using a nonprobabilistic formulation of quantum mechanics and a new definition of state. It seemed of interest to show that both of these conditions can be motivated in a similar way in the probabilistic formulation of axiomatic quantum mechanics without using the strong assumptions of Ref. 4. A report on this subject will be communicated in a forthcoming paper.[5]

In the course of the work it seemed interesting to prove several theorems which are presented in this paper. These theorems are not employed in the axiomatic system of Ref. 5; however, they may be of general interest in quantum axiomatics as well as in the theory of functions over lattices.

In Theorems 1, 2, and 3, the atomicity of the lattice L is proved by imposing certain conditions on the completion of S (in a uniform structure generated in S by L), where S is a set of functions on L which may be thought of as representing the states. Theorem 4

any outcome. We denote it, in what follows, by the symbol I. Thus, $I = (X, \emptyset) \in \Pi(X, \perp, \mathfrak{a})$. The proposition $p(\emptyset)$ is never confirmed by any outcome. We denote it, in what follows, by the symbol O. Thus, $O = (\emptyset, X) = I' \in \Pi(X, \perp, \mathfrak{a})$.

We have discussed three of the propositions in the logic $\Pi(X, \perp, \mathfrak{a})$, namely, O, P and I. Clearly, $P' = p(\{n\}) \in \Pi(X, \perp, \mathfrak{a})$ is confirmed precisely when one of the operations $\mathfrak{D}_1$ or $\mathfrak{D}_2$ is executed, but no particle is observed. We can now display all twelve of the propositions in the logic $\Pi(X, \perp, \mathfrak{a})$, and the implication relations existing between them, by the diagram[14] in Fig. 2. Note that, for instance, $p(\{a\}) \leq P$ (if a particle with position coordinate $x \geq 1$ has been observed, then a particle has been observed), but $p(\{a\}) \not\leq (p(\{c\}))'$ (having observed a particle with position coordinate $x \geq 1$, we are not obliged to conclude that its momentum p_x must be less than 1).

Via Theorem 5 and Fig. 1, it can readily be verified that the logic $\Pi(X, \perp, \mathfrak{a})$ admits a strong set of regular states. Actually, $\Pi(X, \perp, \mathfrak{a})$ is an orthomodular lattice.[15] The subset $\Pi_1(X, \perp, \mathfrak{a})$ of $\Pi(X, \perp, \mathfrak{a})$ consisting of those propositions $p(D) \in \Pi(X, \perp, \mathfrak{a})$ for which R_1 is a test operation form a maximal Boolean sublattice of $\Pi(X, \perp, \mathfrak{a})$. (Note: $\Pi_1(X, \perp, \mathfrak{a})$ is the "left half" of Fig. 2.) An analogous relation exists between R_2 and the "right half" of Fig. 2.

Note that the least upper bound (or the join) of $p(\{a\})$ and $p(\{d\})$ in the lattice $\Pi(X, \perp, \mathfrak{a})$ is given by $p(\{a\}) \vee p(\{d\}) = P = p(\{a, b\}) = (\{a, b, c, d\}, \{n\})$. Thus, $p(\{a\}) \vee p(\{d\})$ is confirmed by any outcome that confirms at least one of the two propositions $p(\{a\})$ or $p(\{d\})$. Also, $p(\{a\}) \vee p(\{d\})$ is refuted precisely by those outcomes that refute both $p(\{a\})$ and $p(\{d\})$. To this extent, $p(\{a\}) \vee p(\{d\})$ behaves like the classical disjunction of the two propositions $p(\{a\})$ and $p(\{d\})$. However, it is important to notice that $p(\{a\}) \vee p(\{d\})$ fails to behave exactly like the classical disjunction since there are outcomes, namely b and c, which confirm $p(\{a\}) \vee p(\{d\})$, but which do not confirm either $p(\{a\})$ or $p(\{d\})$. However, the operational significance of this departure from classical logic is quite clear in the simple example at hand. We could, of course, form an operational proposition $(\{a, d\}, \{n\}) \in \Pi_X$ which would conform in all respects to the classical disjunction of $p(\{a\})$ and $p(\{d\})$; however, the operational proposition $(\{a, d\}, \{n\})$ fails to be testable!

[1] J. M. Jauch, E. P. Wigner, and M. M. Yanase, Nuovo Cimento **48B**, 145 (1967).
[2] C. H. Randall and D. J. Foulis, Amer. Math. Monthly **77**, 363 (1970).
[3] A. N. Kolmogorov, *Foundations of the Theory of Probability* (Chelsea, New York, 1956), 2nd ed. (German ed., 1933).
[4] J. C. Dacey, Caribbean J. Sci. Math. **1**, 51 (1969).
[5] R. J. Greechie, J. Combinatorial Theory **10**, 119 (1971).
[6] Garrett Birkhoff, *Lattice Theory*, AMS Colloq. Publ. XXV (Amer. Math. Soc., Providence, R.I., 1967), 3rd ed., p. 53.
[7] J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).
[8] Garret Birkhoff and J. von Neumann, Ann. Math. **37**, 823 (1936).
[9] G. W. Mackey, *Mathematical Foundations of Quantum Mechanics* (Benjamin, New York, 1963).
[10] Reference 9, p. 68.
[11] Reference 9, pp. 63–66.
[12] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton U.P., Princeton, N.J., 1955), pp. 247–54.
[13] A. M. Gleason, J. Math. Mech. **6**, 885 (1953).
[14] Reference 6, p. 4.
[15] Reference 6, p. 53.

Some Theorems on Atomicity in Axiomatic Quantum Mechanics*

F. Jenč

Institute of Theoretical Physics, University of Marburg, 355 Marburg/Lahn, Federal Republic of Germany

(Received 18 February 1971)

Given a complete orthocomplemented lattice L and a set S of nonnegative real functions on L, sufficient conditions are established that S should fulfill in order that L be atomic. The conditions are investigated under which L may be represented by the lattice of all closed subspaces of a separable Hilbert space. (As is well known, the atomicity of L plays an important role here.) Some unsolved problems are pointed out. In axiomatic quantum mechanics, the lattice L may represent the set of propositions whereas the set of functions S represents the set of physical states. The conditions imposed on the pair (L, S) then have a simple and plausible physical interpretation; an important condition imposed on (L, S) is the existence of the "maximal" (i.e., maximally determined) states which appear in the theory as limit constructions.

1. INTRODUCTION

The validity of the covering law and the atomicity of the lattice of propositions in axiomatic quantum mechanics are sufficient conditions that the complete, orthocomplemented irreducible and orthomodular lattice of propositions in which any set of pairwise orthogonal elements is at most countable may be represented as the lattice of closed subspaces of a separable Hilbert space. This scheme corresponds to the generally accepted mathematical formulation of quantum mechanics.[1–3]

In a recent interesting paper, Jauch and Piron[4] tried to motivate the validity of the covering law and the atomicity of the lattice of propositions in axiomatic quantum mechanics using a nonprobabilistic formulation of quantum mechanics and a new definition of state. It seemed of interest to show that both of these conditions can be motivated in a similar way in the probabilistic formulation of axiomatic quantum mechanics without using the strong assumptions of Ref. 4. A report on this subject will be communicated in a forthcoming paper.[5]

In the course of the work it seemed interesting to prove several theorems which are presented in this paper. These theorems are not employed in the axiomatic system of Ref. 5; however, they may be of general interest in quantum axiomatics as well as in the theory of functions over lattices.

In Theorems 1, 2, and 3, the atomicity of the lattice L is proved by imposing certain conditions on the completion of S (in a uniform structure generated in S by L), where S is a set of functions on L which may be thought of as representing the states. Theorem 4

deals with the representation of L as the lattice of closed subspaces of a Hilbert space.

One of the main conditions implying atomicity of L is the existence of the so-called "maximal" states. The concept of a maximal state is, from the physical point of view, a strong idealization (cf. Footnote 31 and Ref. 5). The probabilistic formulation of quantum axiomatics in which the state is defined as a functional on L permits us to consider the maximal states as abstract constructs obtained from the realizable physical states as idealized limit states. They are obtained through the completion of S in the uniform structure compatible with the "weak" topology of S. This weak topology (generated by L) has a simple physical interpretation[6]: The limit states are probability distributions which can be approximated for any finite set of measurements by the physically realizable states.

It is easy to see from the definition (cf. Theorem 1) that only pure states (extremal points) can be maximal. Since a pure state represents a pure statistical ensemble (again an idealization), one could start with a set S of physically realizable states which contains no pure states. The pure states are obtained by using the Krein–Milman theorem in the completion of S.

Taking different subsets of $\tilde{S}$ (the completion of S) as the set of states for the theory leads to different mathematical schemes. [cf. conditions (B), (M) of Theorem 1, (B_1), (M_1) of Theorem 2, and (M_2) of Theorem 3]. If one restricts the states of the theory to probability measures[7] a structure compatible with the Hilbert space quantum mechanics in principle results. If one admits a larger set of "limit states", mathematical schemes of different types are obtained. States which are not probability measures (nonnormal states) are of interest in quantum statistics.[8] (Infinite additivity of a state function is, of course, an abstraction which cannot be verified through physical experiments.)

2. THEOREMS ON ATOMICITY

Theorem 1[9–15]: Let L be an orthocomplemented complete lattice. Let S be a convex set of functions on L with values in the closed interval $[0, 1]$ such that for $\alpha \in S$, $\alpha(0) = 0$ $\alpha(1) = 1$. Let $\tilde{S}$ be the completion of S embedded in the cube $[0, 1]^L$ endowed with the uniform structure given by the product of the uniform structures of $\{[0, 1]_{a \in L}\}$. This uniform structure is compatible with the so-called weak topology of S for which the family of neighborhoods, $U_{a,\epsilon}$, forms a subbase:

$$U_{a,\epsilon}(\alpha_0) \equiv \{\alpha \in S \,|\, |\alpha(a) - \alpha_0(a)| < \epsilon,\ \epsilon > 0,\ a \in L\}. \tag{1}$$

Let the following assumptions be valid:

(A) to every $a \in L$, $a \neq 0$, there exists an $\alpha \in \tilde{S}$ such $\alpha(a) = 1$.

(B) $\alpha \in \tilde{S}$, $\alpha(a_\gamma) = 1$ for a subset, $\{a_\gamma\}$, of L implies $\alpha(\bigvee a_\gamma) = 1$.

(C) If $\alpha \in \tilde{S}$, $a, b \in L$, $a \leq b$, then $\alpha(a) = 1$ implies $\alpha(b) = 1$.[16]

(M) Let R^+ be the set of all extremal points of $\tilde{S}$. Then all elements of R^+ are maximal elements of $\tilde{S}$ $(\leq_E)$ with the (not necessarily antisymmetric) partial ordering relation $\leq_E$ being defined as follows: For $\alpha, \beta \in \tilde{S}$: $\alpha \leq_E \beta \iff \alpha(a) = 1$ implies $\beta(a) = 1$.

In defining $E_1(\alpha) \equiv \{a \in L \,|\, \alpha(a) = 1\}$, we can rewrite this definition in the following form:

$$\text{For } \alpha, \beta \in \tilde{S}, \quad \alpha \leq_E \beta \iff E_1(\alpha) \subseteq E_1(\beta). \tag{2}$$

(We want to stress once more that, in this context, by a "partial ordering" relation we mean a reflexive and transitive relation, which, in general, need not be antisymmetric, i.e., $E_1(\alpha)$ does not determine α uniquely in a general case. An element m of $\tilde{S}$ is then maximal in $\tilde{S}(\leq_E)$ if and only if $m \leq_E \alpha$, $\alpha \in \tilde{S}$, implies $\alpha = m$.)

We define further: $S_1^+(a) \equiv \{\alpha \in \tilde{S} \,|\, \alpha(a) = 1\}$, $R_1^+(a) = \{m \in R^+ \,|\, m(a) = 1\}$.

Then the following statements are true:

(I) R^+ is not void. For any $a \neq 0$, $R_1^+(a)$ is not void.

(II) L is atomic.

(III) There is a one-to-one correspondence between the elements m of R^+ and the atoms e of L, which can be defined as:

$$e_m \leftrightarrow m_e \quad \text{if } E_1(m) = \{a \in L \,|\, e_m \leq a\}. \tag{3}$$

Any extremal point m is uniquely determined by the set $E_1(m)$.

(IV) $\tilde{S} = \overline{co} R^+$, $S_1^+(a) = \overline{co} R_1^+(a)$, where the symbol $\overline{co}$ denotes the closed convex hull.

(V) $E_1(m)$ is an ultrafilter in L for any extremal point m of $\tilde{S}$. It also follows that assumption (M) is more stringent than the assumption of atomicity of L [(A), (B), and (C) being valid].

Proof of Theorem 1: S is a convex set. Taking S as the generator of the positive cone[13] C_S in $\mathfrak{L}$, we define a partially ordered linear space $\mathfrak{L}$ as a direct sum:

$$\mathfrak{L} = C_S + (-C_S). \tag{4}$$

Let us denote the elements of $\mathfrak{L}$ by x, y, etc. The elements of L are functionals on S defined by $a(\alpha) = \alpha(a)$. With the natural definition $x(a) = {}_i\sum c_i \alpha_i(a)$ for $x = {}_i\sum c_i \alpha_i$, L becomes a total family of linear functionals on $\mathfrak{L}$ defined by $a(x) = x(a)$. [For $y = c_1 x_1 + c_2 x_2$, $a(y) = y(a) = c_1 x_1(a) + c_2 x_2(a) = c_1 a(x_1) + c_2 a(x_2)$.] Therefore, $\mathfrak{L}$ becomes a locally convex Hausdorff linear topological space by introducing the weak topology τ' generated by the subbase of neighborhoods:

$$U_{a,\epsilon}(x_0) \equiv \{x \in \mathfrak{L} \,|\, |a(x) - a(x_0)| < \epsilon,\ \epsilon > 0\}. \tag{5}$$

τ' induces in S a topology τ, the weak topology of S. Its subbase is obtained by taking only elements of S in (5).

$\mathfrak{L}$ is a uniform space with a unique translationally invariant uniformity $\mathfrak{W}'$, generated by the subbase of vicinities $\mathfrak{W}'_{a,\epsilon}$,

$$\mathfrak{W}'_{a,\epsilon} = \{(x, y) \in \mathfrak{L} \times \mathfrak{L} \,|\, |a(x) - a(y)| < \epsilon,\ \epsilon > 0\}.$$

Therefore, we can form a completion of $\mathfrak{L}$ in $\mathfrak{W}'$, say $\tilde{\mathfrak{L}}$, which also is a locally convex Hausdorff linear topological space. In the same way, we obtain a completion of S, say $\tilde{S}$, in the uniformity $\mathfrak{W}$ induced in S by

$\mathfrak{W}'$. Let us now denote the topologies in $\tilde{\mathfrak{L}}$ and $\tilde{S}$ by $\tilde{\tau}'$ and $\tilde{\tau}$, respectively.

$\mathfrak{L}$ is dense[17] in $\tilde{\mathfrak{L}}(\tilde{\tau}')$ and S is dense in $\tilde{S}(\tilde{\tau})$.

In the following text we denote the elements of $\tilde{S}$ by $\alpha, \beta, \ldots$ and the elements of $\tilde{\mathfrak{L}}$ by $x, y, \cdots$.

The elements of L are naturally continuous functionals on $\mathfrak{L}$. The seminorms generating the locally convex topology in $\tilde{\mathfrak{L}}$ are obtained by the unique continuous extension of the seminorms generating the topology τ' in $\mathfrak{L}$.[13] These extensions are equal to $|a(x)|$, where $a(x)$ is the continuous extension of the functional $a \in L$ from $\mathfrak{L}$ to $\tilde{\mathfrak{L}}$. Extending L on $\tilde{\mathfrak{L}}$ in this way, we may consider any $a \in L$ as a continuous linear functional on $\tilde{\mathfrak{L}}$. $\tilde{\tau}'$ is then again generated by the family of linear functionals L.[18] We have $\alpha(1) = 1$, $\alpha(0) = 0$ for any $\alpha \in \tilde{S}$. $\tilde{S}$ evidently is a convex set. According to proposition (A) of Theorem 1, $S_1^+(a) \neq \emptyset$ for any $a \in L$, $a \neq 0$. The continuity of $a \in L$ also implies that $S_1^+(a)$ is closed in $\tilde{\mathfrak{L}}$ and closed in $\tilde{S}$ for any $a \in L$.

One can prove that $\tilde{S}$ is compact in $\tilde{\tau}'$.[19] This can easily be verified by embedding $\tilde{S}$ into the space $\mathfrak{F}$, i.e., the space of all functions on L having their values in the interval $[0, 1]$. $\mathfrak{F}$ may be represented as the (Tychonoff) cube $[0, 1]^L$, i.e., as the direct product of $\{[0, 1]_a\}$, $a \in L$. In the direct product topology (this topology "extends" τ to $\mathfrak{F}$), $\mathfrak{F}$ is a compact uniform space; since $\tilde{S}$ is complete in $\mathfrak{F}$, it is closed and hence compact.[11] This result completes the proof since the closure and the completion of S in $\mathfrak{F}$ and in $\tilde{\mathfrak{L}}$ are identical.

It follows that $S_1^+(a)$ is compact in $\tilde{\tau}$ for any $a \in L$. Let us denote by R^+ the set of all extremal points of $\tilde{S}$. Then we conclude from the Krein–Milman theorem[13] that:

(a) $R^+ \neq \emptyset$,

(b) $\tilde{S} = \overline{co} R^+$,

(c) $R_1^+(a) \neq \emptyset$ for any $a \in L$, $a \neq 0$, with $R_1^+(a) \equiv \{m \in R^+ | m(a) = 1\}$,

and

(d) $S_1^+(a) = \overline{co} R_1^+(a)$ for $a \neq 0$.

Now it is easy to conclude the proof. The set $E_1(\alpha)$ is evidently nonvoid, and we have from (B) and (C) that, for any $\alpha \in \tilde{S}$, $E_1(\alpha)$ is a filter in L. Let us denote by $\wedge[E_1(\alpha)]$ the greatest lower bound of $E_1(\alpha)$. Assumption (B) also implies that

$$\wedge[E_1(\alpha)] \in E_1(\alpha) \text{ and thus } E_1(\alpha) = \{a \in L | \wedge[E_1(\alpha)] \leq a\}.$$

It follows from assumption (M) that, for every $m \in R^+$, $\wedge[E_1(m)]$ is an atom in L. Since if $\wedge[E_1(m)]$ is not an atom, there exists an $a \in L$, $0 < a < \wedge[E_1(m)]$ and an $\alpha \in \tilde{S}$, $\alpha \neq m$, with $\alpha(a) = 1$. Thus $a \in E_1(\alpha)$, whence [from (C)] $\wedge[E_1(m)] \in E_1(\alpha)$. Since $a < \wedge[E_1(m)]$, $E_1(m)$ is a proper subset of $E_1(\alpha)$ which contradicts (M). Thus $\wedge[E_1(m)]$ is an atom.

Since, for $a \neq 0$, there exists an extremal point m, $m(a) = 1$, there also exists an atom $e \leq a$, to any $a \in L$, $a \neq 0$, namely $e = \wedge[E_1(m)]$. This proves that L is atomic.

It is evident that, for an extremal point m, $E_1(m)$ is an ultrafilter. Since if $E_1(m)$ is a proper subset of a filter F in L, then there exists a $b \in F$ such that b is not comparable with $e \equiv \wedge[E_1(m)]$. Hence $b \wedge e = 0$, which is impossible. However, not every ultrafilter corresponds to a state, since generally the intersection of all elements of F can be equal to 0, i.e., F need not be generated by an atom.

We have seen that, to any extremal point m of $\tilde{S}$, there exists an atom e such that $E_1(m) = \{a \in L | e \leq a\}$. On the other hand, to any atom e, there exists an m for which $m(e) = 1$ and

$$E_1(m) = \{a \in L | e \leq a\}$$

since for any other $b \in L$, $e \wedge b = 0$.

Any m is uniquely determined by the set $E_1(m)$ [assumption (M)]. Thus there exists a one-to-one mapping of R^+ onto the set of atoms of L. Assumption (M) is evidently more stringent than the assumption of atomicity of L. Since if L atomic, the set $E_1(\alpha)$ is maximal in $\tilde{S}(\leq_E)$ for those $\alpha \in \tilde{S}$ for which there exists an atom e with the property $\alpha(e) = 1$ [owing to (B) and (C)]. However, one cannot prove, without further assumption, that to any extremal point of $\tilde{S}$ such an atom exists.

Remark: It is clear that the conclusions of the theorem remain valid if the condition (B) is replaced by the following weaker condition:

(B′) If, for an extremal point m of $\tilde{S}$, $m(a_\gamma) = 1$ for every element of a subset of L, $\{a_\gamma\}$, then $m(\bigvee a_\gamma) = 1$.

Theorem 2: Let $\tilde{S}_p$ denote the set of all probability measures in $\tilde{S}$. Let the condition (A) of Theorem 1 be valid and the premises (B), (C), and (M) be replaced by the following weaker conditions:

(B_1) $\alpha \in \tilde{S}_p$, $\alpha(a_\gamma) = 1$ for a subset $\{a_\gamma\}$ of L implies $\alpha(\bigvee a_\gamma) = 1$.

(C_1) $\alpha \in \tilde{S}_p$, $a, b \in L$, $a \leq b$, and $\alpha(a) = 1$ implies $\alpha(b) = 1$.

(M_1) Any element of $R^+ \cap \tilde{S}_p$ is a maximal element of $\tilde{S}(\leq_E)$.

However, let us add the following assumption:

(H) If $R_1^+(a)$ is not void, it contains a probability measure, i.e., $R_1^+(a) \cap \tilde{S}_p \neq \emptyset$.

Then the conclusions (I), (II), and (IV) of Theorem 1 are valid and the conclusions (III) and (V) are changed as follows:

(III_1) There is a one-to-one correspondence between the elements m of $R^+ \cap \tilde{S}_p$ and the atoms e of L, which can be defined as $e_m \leftrightarrow m_e$ if $E_1(m) = \{a \in L | e_m \leq a\}$. Any extremal point m of $\tilde{S}$ which is a probability measure is uniquely determined by the set $E_1(m)$.

(V_1) $E_1(m)$ is an ultrafilter for any extremal point m of $\tilde{S}$ which is a probability measure. Assumption (M_1) is more stringent than the assumption of atomicity of L (the other conditions being valid).

Theorem 3: Let all the premises of Theorem 2 except (M_1) be valid and let us replace (M_1) by the following assumption:

(M_2) all elements of $R_1^+ \cap \tilde{S}_p$ are maximal elements of $\tilde{S}_p(\leq_E)$. Then all conclusions of Theorem 2 remain valid [if (M_1) is replaced by (M_2) in (V_1)]. (Note that $m \in R^+ \cap \tilde{S}_p$ is now uniquely determined by $E_1(m)$ only in $\tilde{S}_p$.)

Proofs of Theorems 2 and 3: The proofs follow from the proof of Theorem 1. The main difference is that, via the premises of Theorems 2 and 3, the sets $E_1(\alpha)$ are proved to be filters only if α is a probability measure. The relevant extremal points in the proofs are then elements of $\tilde{S}_p$.

Theorem 4: Let $\mathfrak{B}(H)$ be the algebra of all bounded linear operators in a separable Hilbert space H. Let S_H be the set of all normal positive linear functionals of norm one on $\mathfrak{B}(H)$. Let $\mathcal{P}(H)$ be the lattice of all orthogonal projectors in H. Let $\mathcal{S}$ be a set of positive linear functionals of norm one on $\mathfrak{B}(H)$. The functions $|P(\alpha)| = |\alpha(P)|$, $P \in \mathcal{P}(H)$, $\alpha \in \mathcal{S}$ form a total family of seminorms on $\mathcal{S}$ and generate a uniform structure in $\mathcal{S}$ (cf. Proof of Theorem 1). Let $\tilde{\mathcal{S}}$, the completion of $\mathcal{S}$ in this uniform structure, be equal to $\tilde{S}_H$, the set of all positive linear functionals of norm one on $\mathfrak{B}(H)$.

Then the following statements are valid:

(a) If there exists a countably infinite set of pairwise orthogonal elements in L, the structure of (L, S) of Theorem 1 is incompatible with the structure of $(\mathcal{P}(H), \mathcal{S})$ ($\dim H = \infty$); if every set of pairwise orthogonal elements in L is finite, the structure of (L, S) is compatible with the structure of $(\mathcal{P}(H), \mathcal{S})$ ($\dim H < \infty$).

(b) The structure of the pair (L, S) of Theorem 3 is compatible with the structure of $(\mathcal{P}(H), \mathcal{S})$ [i.e., (L, S) can be, under additional conditions, represented as the familiar Hilbert space scheme even if $\dim H = \infty$].

We need a few Lemmas to prove Theorem 4.

Proof of part (a) of Theorem 4:

Lemma 1: S_H is not compact in the weak topology generated by $\mathcal{P}(H)$.

Proof: Let us denote by τ_p the weak topology[14] generated by $\mathcal{P}(H)$ in $\mathfrak{B}'$ the dual of $\mathfrak{B}(H)$. The proposition "S_H is τ_p-compact" is equivalent to the proposition "S_H is τ_p-closed in $\mathfrak{B}'$" since S_H is a subset of the w^*-compact[20] and hence τ_p-compact unit ball of $\mathfrak{B}'$. S_H is w^*-dense[21,22] and hence τ_p-dense in $\tilde{S}_H$, the set of all positive linear functionals of norm one on $\mathfrak{B}(H)$. $\tilde{S}_H$ is compact in the w^*-topology[23(a)] and hence is compact in τ_p. Thus $\tilde{S}_H$ is τ_p-closed and is the closure[24,25] of S_H. Thus $\tilde{S}_H$ is the compactification of S_H, i.e., the completion of S_H in the uniform structure compatible with τ_p.

On the other hand, it is well known (and may be seen also from Lemma 2) that there exist nonnormal functionals in $\tilde{S}_H$. Hence $S_H \neq \tilde{S}_H$; $\tilde{S}_H$ is not τ_p-closed and noncompact. QED

For completeness, we recapitulate some known facts in the following lemma:

Lemma 2:

(1) All extremal points of S_H are extremal points of $\tilde{S}_H$.

(2) There exists an extremal point of $\tilde{S}_H$ in $\tilde{S}_H \dot{-} S_H$, i.e., a nonnormal extremal point.

Proof:

(1) the proof follows from the lemma[26(a)] [cf. also Refs. 23(b), 26(b)]: Let $\mathfrak{A}$ be a *-subalgebra of $\mathfrak{B}(H)$ containing unity and φ a positive linear functional majorized by the functional $\omega_x(T) = \langle x | Tx \rangle$. Then there exists an element T' in the commutant of $\mathfrak{A}$ such that $\varphi(T) = \omega_{T'x}(T) = \langle T'x | TT'x \rangle$. [For $\mathfrak{A} = \mathfrak{B}(H)$, the commutant is the set of scalars.]

(2) A cyclic representation of $\mathfrak{B}(H)$ is irreducible if and only if its generating positive functional is extremal.[23(b)] The only irreducible representations of $\mathfrak{B}(H)$ in a separable Hilbert space are isomorphisms of the algebra $\mathfrak{B}(H)$ itself (Type 1) or of the quotient algebra $\mathfrak{B}(H)/I_0$ (Type 2), where I_0 is the norm-closed two-sided ideal of all compact operators.[23(c)] A concrete representation of Type 2 was constructed by Calkin.[27] The positive functional generating a representation of Type 2 is not normal, since it annulates all projectors of finite rank and hence cannot be of the form

$$\omega(T) = \langle x | Tx \rangle, \qquad x \in H.$$

There is a one-to-one correspondence $m_e \leftrightarrow e_m$ between the atoms e of $L = \mathcal{P}(H)$ (i.e., projectors of rank one) and extremal positive functionals in S_H of the type $m(T) = \langle x | Tx \rangle$ (the "pure states"), defined by the equation $E_1(m_e) = \{a \in L \,|\, e_m \leq a\}$. Thus we cannot have a one-to-one correspondence between the atoms of $\mathcal{P}(H)$ and *all* extremal points of $\tilde{S}_H$ defined by the same relation. Indeed we see the following:

Lemma 3: Condition (B) is not satisfied in $\tilde{S}_H$.

Proof: Let us denote an extremal point of $\tilde{S}_H$, which generates an irreducible representation of $\mathfrak{B}(H)$ of Type 2, by ψ. ψ is a nonnormal functional. $\psi(P) = 1$ for any projector $P \in \mathcal{P}(H)$ belonging to a subspace whose orthogonal complement has finite dimension (finite additivity of ψ). Let us take such a projector P and a sequence of projectors of finite rank P_i, $P_i \leq P$, $\bigvee_i P_i = P$. Let P_i' and P' denote the orthogonal complements of P_i and P, respectively. Then $\psi(P_i') = 1$, for every i. However, $\bigwedge_i P_i' = P'$ is a projector of finite rank so that $\psi(\bigwedge_i P_i') = 0$. Thus (B) is not fulfilled in $\tilde{S}_H$. Hence our axiomatic formulation is not compatible with Hilbert space quantum mechanics if $\dim H = \infty$.

We remark in concluding that the axiomatic scheme of Theorem 1 is compatible with Hilbert space quantum mechanics if the dimension of H is finite, since then $S_H = \tilde{S}_H$ and condition (B) is fulfilled in $\tilde{S}_H$. This conclusion can be seen as follows: If $\dim H < \infty$, any positive functional on $\mathfrak{B}(H)$ is normal since in this case $\mathfrak{B}(H)$ is a finite-dimensional Banach space.

Proof: In a finite-dimensional space, any locally convex topology is equivalent to the norm topology.[23(d)] Hence any norm-continuous functional on $\mathfrak{B}(H)$ is ultraweakly continuous. Since every positive functional on $\mathfrak{B}(H)$ is norm continuous, it is ultraweakly continuous and hence normal.[26(b)]

It is clear that L can be represented by $\mathcal{P}(H)$ in this case only if there exists at most a finite set of pairwise orthogonal elements in L.

Proof of part (b) of Theorem 4: It is easy to check that all premises and conclusions of Theorem 3 are valid with $L = \mathcal{P}(H)$ and $\tilde{\mathcal{S}} = \tilde{S}_H$ since then[1] $\mathcal{S}_p = S_H$. Hence the structure of (L, S) is compatible with the structure of $(\mathcal{P}(H), \mathcal{S})$.

Remark: For $\mathcal{S}$ we could take, e.g., the set of all "mixed" states of S_H if we wish to start with a set of "physically realizable" states which contains no pure states (corresponding to the idealization of pure ensembles).

3. COMMENTS AND REMARKS

We note that the orthocomplementation of L is not needed in the proof of Theorem 1. It is, however, indispensable in Theorems 2 and 3 for the definition of $\tilde{S}_p$.

In proving Theorems 1, 2, and 3, we did not impose the condition of finite additivity on the elements of S. In physical applications, however, we would postulate this property (or even σ-additivity). In that case, all elements of $\tilde{S}$ would be additive functions on L.

An example of a couple (L, S) satisfying the premises of Theorem 1 is provided for instance by a physically plausible axiomatic system, constructed in analogy to Pool's axiomatic system,[28] with the following changes: S need not be σ-convex but only convex, the elements of S need not be probability measures but are additive functions on L, and L is a complete orthocomplemented lattice. The lattice operations are uniquely defined by the relations

$$S_1(\underset{\gamma}{\wedge} a) = \underset{\gamma}{\cap} S_1(a_\gamma) \quad \text{and} \quad S_0(\underset{\gamma}{\vee} a) = \underset{\gamma}{\cup} S_0(a_\gamma),$$

where $S_0(a) \equiv \{\alpha \in S \mid \alpha(a) = 0\}$. [In Pool's axiomatic system, $a \leq b$, $b \leq a \Longleftrightarrow a = b$, and $a \leq b \Longleftrightarrow S_1(a) \subseteq S_1(b) \Longleftrightarrow S_0(b) \subseteq S_0(a)$.] L is then orthomodular so that every $\alpha \in S$ is an increasing function on L and, furthermore,[29] $a \leq b \Longleftrightarrow \alpha(a) \leq \alpha(b)$ for every $\alpha \in S$. Hence we can prove (cf. Proof of Theorem 1) that any $\alpha \in \tilde{S}$ is an increasing function on L. We have then the equivalence

$$a \leq b \Longleftrightarrow S_1^+(a) \subseteq S_1^+(b) \Longleftrightarrow \alpha(a) \leq \alpha(b)$$

for every $\alpha \in \tilde{S}$.

The validity of assumptions (A), (B), and (C) follows easily. Assumption (M) evidently does not contradict the other conditions.

In the physical interpretation (cf. Refs. 4 and 5) the lattice operation $\wedge$ plays a role analogous to the logical conjunction and the partial ordering relations $\leq$ in L plays the role of implication [we note, however, that $\leq$ does not denote a logical implication in $L(\leq)$ but an empirical one]. This interpretation motivates the validity of (B) for any $\alpha \in \tilde{S}$ (the completion of S) which may be accepted as representing a state in the abstract setting of the theory (if the propositions a_γ are true for every γ then the proposition $\underset{\gamma}{\wedge} a_\gamma$ is true).[30] We note that if S represents the physically relizable states, $\tilde{S}$ is the greatest set whose elements could at all be meaningfully taken as the abstract limit states. In Theorem 1, we postulate the validity of (B) for $\tilde{S}$, i.e., we take all elements of $\tilde{S}$ as potential "states"; in Theorem 3 we considered only states which are probability measures. Theorem 2 is evidently not of much interest from the standpoint of the physicist, since conditions (B) and (C) are postulated in $\tilde{S}_p$, whereas the maximality in $\tilde{S}$ is postulated for elements of $R^+ \cap \tilde{S}_p$. The atomicity of L is deduced by imposing certain conditions on the completion of S (in the uniformity generated by L). This procedure has the following physical significance: We postulate that the structure of (L, S) is such as to admit the introduction of some limit concepts which naturally arise as useful abstractions in our way of physical thinking, such as the concept of a maximal state, i.e., maximally determined state.[31]

The results of Theorems 1 and 3 may then by physically interpreted as follows: We can introduce into our abstract scheme the concept of maximal state only if we use an atomic L, i.e., if we introduce the idealized "atomic" propositions in L. In Ref. 5, the connection between the atomicity of L and the maximality of states is further analyzed.[32]

It is interesting to note that the conclusions of Theorem 2 are valid whether we postulate the maximality of elements of $R^+ \cap \tilde{S}_p$ in $\tilde{S}$ or in $\tilde{S}_p$. The main difference between Theorems 2 and 3 is that, in the first case, $m \in \tilde{S}_p$ is uniquely determined by $E_1(m)$ in $\tilde{S}$ whereas in the second case it is uniquely determined by $E_1(m)$ only in $\tilde{S}_p$. Whether (L, S) of Theorem 2 can be represented as a Hilbert space scheme can not be definitively decided since little is known on the properties of nonnormal functionals in $\tilde{S}_H$. As was seen in the proof of Theorem 4, condition (B) is not generally valid in $\tilde{S}_H$. On the other hand, this condition holds in $\tilde{S}_{Hp}$. (We have $\tilde{S}_{Hp} = S_H$!) It is, however, not clear whether $E_1(\alpha)$ is a filter for a nonnormal α, whether the extremal points of S_H are maximal in $\tilde{S}_H$, or what other maximal elements exist in $\tilde{S}_H$. One can easily prove that extremal points of S_H are incomparable [in $\tilde{S}_H(\leq_E)$] with nonnormal extremal points of Type 2.

Proof: For any extremal point m of S_H there exist projectors of finite rank for which $m(P) = 1$, $m(1 - P) = 0$. All such projectors are annulated by any nonnormal functional m' of Type II, i.e., $m'(1 - P) = 1$. However, little is known about the properties of nonnormal extremal points of Type 1. Thus we cannot prove the maximality of pure states of S_H in $\tilde{S}_H$ nor make a comparison of nonnormal functionals in $\tilde{S}_H$ $(\leq_E)$ in general.

Thus two mathematical problems arise in the Hilbert-space theory:

(1) Is the set $E_1(\alpha)$ a filter for a nonnormal functional in $\tilde{S}_H$?

(2) What are the maximal elements of $\tilde{S}_H$ $(\leq_E)$?

Note added in manuscript: To make the physical interpretation of the premises of Theorems 1–3 more plausible, we note that the somewhat unphysical condition of completeness of L may be replaced by the following weaker condition: (X) to every set $\{a_i\} \subseteq L$ of pairwise orthogonal elements a_i of L, the lowest upper bound $\underset{i}{\vee} a_i$ exists.

Condition (X) implies, together with the orthocomplementation and the separability of L the completeness of L (cf. Zierler[33(a)]). The physical motivation of (X) is evidently much clearer than that of the completeness of L, the meaning of (X) being the following: To any countable set of mutually exclusive propositions, the proposition "a_1 or a_2 or $\cdots$" is an admissible physical proposition.

Similarly the condition (B_1) in Theorems 2 and 3 may be replaced by the following condition pertaining only, to finite sets of L:

(Y) $a, b \in L$, $\alpha \in \tilde{S}_p$, $\alpha(a) = 1$, $\alpha(b) = 1$ implies $\alpha(a \wedge b) = 1$

or alternatively $\alpha(a) = \alpha(b) = 0$ implies $\alpha(a \vee b) = 0$.[33(b)]

The latter condition may be loosely interpreted as follows:

If, in a state α, one gets with certainty a negative answer to both of the questions a and b, one also gets with certainty a negative answer to the question "a or b".

ACKNOWLEDGMENTS

The author thanks Professor J. M. Jauch and Professor C. Piron for helpful discussions and critical comments during his stay at the University of Geneva. Thanks are due to Professor W. A. Bingel for his kind hospitality during author's stay at the University of Göttingen and to Professor G. Ludwig for his kind interest in this work.

---

* This work was partly supported by a Battelle Advanced Study Center Fellowship at the University of Geneva.

1 J. M. Jauch, *Foundations of Quantum Mechanics* (Addison-Wesley, Reading, Mass., 1968).

2 V. S. Varadarajan, *Geometry of Quantum Theory* (Van Nostrand, Princeton, N.J., 1968), Vol. I.

3 C. Piron, Helv. Phys. Acta **37**, 439 (1964).

4 J. M. Jauch and C. Piron, Helv. Phys. Acta **42**, 842 (1969).

5 F. Jenč, to appear.

6 R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).

7 The following conventions are used in the rest of the paper. We denote the greatest lower bound by $\wedge$, the least upper bound by $\vee$, and the complement of a by a'. 0 and 1 denote the least and greatest element of L, respectively. Orthogonality of two elements $a, b \in L$, $a \perp b$, means as usual $a \leq b'$. By probability measure on L we mean a nonnegative real function α on L with values in the interval $[0, 1]$, such that (i) $\alpha(0) = 0$, (ii) $\alpha(1) = 1$, and (iii) $\alpha(\vee_i a_i) = \sum_i \alpha(a_i)$ if the elements a_i are pairwise orthogonal ($i = 1, 2, \cdots$).
As a filter F we denote a subset of a bounded lattice L, which satisfies the following conditions: (1) 0 is not an element of F; (2) $a, b \in F$ implies $a \wedge b \in F$; (3) $a \in F$ implies $x \in F$ for any element x of L greater than a ($a \leq x$).

8 G. Emch, J. Math. Phys. **7**, 1413 (1966).

9 All the necessary concepts and the theorems from general topology and the theory of linear topological vector spaces will be found in Refs. 10–14. Lattice theory is treated in Ref. 15.

10 J. L. Kelley, *General Topology* (Van Nostrand, Princeton, N.J., 1955).

11 N. Bourbaki, *Elêments de mathêmatique, Livre III. Topologie gênêrale* (Hermann, Paris, 1961), Chaps. 1, 2.

12 E. Čech, *Topological Spaces* (Publish. House Czech. Acad. Sci., Prague, 1966).

13 H. H. Schaefer, *Topological Vector Spaces* (Macmillan, New York, 1966).

14 G. Köthe, *Topological Linear Spaces. I* (Springer, Berlin, 1969).

15 L. Szász, *Einführung in die Verbandstheorie* (Verlag Ungar. Akad. Wiss., Budapest, 1962).

16 As will be seen in the proof of Theorem 1, this condition is true if every $\alpha \in S$ is increasing on L [i.e., $\alpha \in S$, $a \leq b$ implies $\alpha(a) \leq \alpha(b)$].

17 The completion of $\mathfrak{L}$ or S is defined only up to isomorphism[13]; let us denote the isomorphic images of $\mathfrak{L}$ and S by $\mathfrak{L}_1$ and S_1, respectively. Then it is easy to see that a completely equivalent problem is obtained by taking for L_1 the corresponding functionals on $\mathfrak{L}_1$ or S_1 (i.e., taking on the same values on S_1 as the functionals of L have on S).

18 Since S is dense in $\tilde{S}$ in $\tilde{\tau}'$, there exists, to any $\alpha \in \tilde{S}$, a generalized sequence, $\alpha_d \in S$, $\alpha_d \overset{\tilde{\tau}'}{\to} \alpha$ so that we have $\alpha(a) = a(\alpha) = a(\lim_d \alpha_d) = \lim_d a(\alpha_d) = \lim_d \alpha_d(a)$. Thus if every $\alpha \in S$ is increasing on L, then every $\alpha \in \tilde{S}$ is increasing on L and, for $a \leq b$, $\alpha(a) = 1$ implies $\alpha(b) = 1$.

19 J. Gunson, Commun. Math. Phys. **6**, 262 (1967).

20 The w^*-topology is the weak topology[13,14] generated in the dual of $\mathfrak{B}(H)$ by $\mathfrak{B}(H)$. We have $\tau_p \leq w^*$-topology (i.e., τ_p is weaker than the w^*-topology).

21 J. M. Fell, Trans. Amer. Math. Soc. **94**, 365 (1960).

22 This result is an easy consequence of Theorem I of Ref. 21.

23 M. A. Neumark, *Normierte Algebren* (Deutscher Verlag der Wiss., Berlin, 1959): (a) p. 278; (b) p. 277, Theorem 2, and p. 278, I; (c) p. 305, Corollary of V; (d) p. 64, I.

24 $\tilde{S}_H$ is the set of states in Segal's sense (cf. Ref. 25).

25 I. E. Segal, *Mathematical Problems of Relativistic Physics* (Benjamin, New York, 1963).

26 J. Dixmier, *Les algèbres d'operateurs dans l'espace Hilbertien (Les algèbres de v. Neumann)* (Gauthier-Villars, Paris, 1969): (a) p. 48, Lemme 1; (b) p. 51, Theorème 1.

27 J. W. Calkin, Ann. Math. **42**, 839 (1941).

28 J. C. T. Pool, Commun. Math. Phys. **9**, 118 (1968).

29 J. C. T. Pool, thesis (State University of Iowa, 1963) (Rept. No. SUI-63, 17).

30 The physical interpretation of (A) and (C) is clear: (A) means that for any proposition $a \in L$ there exists a state $\alpha \in \tilde{S}$ such that a is true in the state α. (C) means that if a implies b, then if a is true in the state $\alpha \in \tilde{S}$, b is true in the state α.

31 Since, in the set of admissible states, a maximal state m is uniquely determined by the maximal set $E_1(m)$, the physical situation it describes cannot be further specified by any information. It seems worthwhile to remark that hypothesis (H) of Theorems 2 and 3 is a weakening of the postulate that, to any $a \in L$, there exists a maximal pure state m with $m(a) = 1$. The latter postulate might be physically interpreted as follows: To any proposition $a \in L$, there exists a set of propositions $E_1(m)$ (including a), which represents a maximal specification of a physical situation m of the system that cannot be further refined by any propositions concerning the system [the state m being uniquely determined[31] by $E_1(m)$ and vice versa].

33 N. Zierler, Pacific J. Math. **11**, 1151 (1961): (a) Lemma 1, 11; (b) Lemma 1, 12.

Approximate Functional Integral Methods in Statistical Mechanics. I. Moment Expansions*

Armand Siegel and Terence Burke†
Department of Physics, Boston University, Boston, Massachusetts 02215
(Received 19 April 1971)

In this paper, four distinct ideas are combined, which under a wide range of circumstances can give very rapidly converging series expansions for functional integrals. (1) Expansion of the functional being integrated in functional Taylor series. In the familiar case arising in quantum statistical mechanics, that of the Wiener integral of $\exp[-\int V(x(t))dt]$, $V(x)$ being the perturbing potential, this is equivalent to expanding the characteristic functional of the probability functional of $V(x(t))$ in central moments of $V(x(t))$. The lowest-order term of the series is the approximation obtained by Feynman and Hibbs through a variational method. (2) Transfer of the harmonic term of the potential, when the functional integral is the quantum-statistical density matrix (Green's function of the Bloch equation), to the weighting function. This transforms the functional integral from a Wiener to an Uhlenbeck–Ornstein integral. The formal expressions for the terms of the expansion are somewhat more complicated, but they can be worked out, and the result is a great improvement in the speed of convergence of the series with decreasing temperature and/or decreasing relative magnitude of the anharmonic part of the potential. (3) "Reservation of variables" in the integration. This amounts to breaking the averaging process down into an average over subsets of the distributed random function (conditional average), followed by an averaging of these averages. Any step of this kind (it may be repeated within the subsets, etc.) gives an improvement of accuracy. (4) When the quantity being evaluated through functional integration is the partition function, the device introduced by Feynman and Hibbs, of interchanging the functional integration with the integration of the Green's function over the equated initial and final configuration-space points, may be combined with the above techniques. This eliminates one integration in the terms of the expansion and seems to improve accuracy at the same time. The general series obtained is correlated with more conventional operator techniques of quantum-mechanical perturbation theory, in order to answer the perennial question, does the path-integral method bring with it anything that could not be derived by other methods? It is in some sense a Feynman–Dyson expansion of the Green's function, but one that is further modified mathematically in a way characteristic only of the path-integral point of view, and which, moreover, improves its accuracy. It thus appears unlikely that the result is merely one of standard type disguised as a functional integral result. Sample numerical calculations are given to assay the accuracy of the methods, which are shown to compare very favorably with the traditional approximation of finite subdivision of the time interval.

1. INTRODUCTION AND SUMMARY OF RESULTS

A. Introductory Formulation

In this and subsequent papers we aim to give a detailed presentation of methods of relatively recent exploitation in the evaluation of functional or path integrals.[1] These methods consist of various combinations of four devices, which can give cumulative improvement in the accuracy of the approximation. One of our important sources is in the work of Feynman and Hibbs in the tenth and eleventh chapters of their book on path integrals in quantum and statistical mechanics.[2] It is in their treatment that one can find, in a brief table (p. 286), evidence of the surprising accuracy of a certain approximation they proposed. Yet the expression of this approximation is only the first term of one of the series we discuss here, and can, moreover, be improved even without taking any higher terms. It therefore seems in order, particularly in view of the variety of devices that can be employed, to present a systematic development and formulation of such methods at this time.

We start with a summary of the formalism of functional integrals, somewhat adapted to our own purposes. A fully explicit, "constructive" even though somewhat symbolic, definition of the integral of a functional $f\{x(\cdot)\}$ of a function $x(t)$ is[2–6]

$$I(f) = \int f\{x(\cdot)\}\, P\{x(\cdot)\} \prod_{0\le t'\le t} dx(t'). \tag{1}$$

The integral is a multiple one, with respect to the continuum of variables $x(t')$ obtained by letting t' range through the interval $0 \le t' \le t$; the limits of integration are usually $\pm\infty$. The functional $P\{x(\cdot)\}$ is a weighting functional, and $\prod$ denotes a product over t' of differentials $dx(t')$.

To become physically and mathematically interesting, the above has to be specialized. Our concern is mainly with $I(f)$ as a solution of a certain partial differential equation, or something closely related to such a solution. The equation is that of Bloch or Schrödinger in dimensionless units (for details of the elimination of units, see Appendix A),

$$\frac{\partial\psi}{\partial t} = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - V(x)\right)\psi = H\psi, \tag{2}$$

real t giving the Bloch equation and imaginary t the Schrödinger equation. The functional integral

$$K(x_1(\beta), x_0(0)) = \int \exp\left\{-\int_0^\beta \left[\frac{1}{2}\left(\frac{dx(t')}{dt'}\right)^2 + V(x(t'))\right]\right\} dt' \times \delta(x(\beta) - x_1(\beta))\delta(x(0) - x_0(0)) \times \prod_{0\le t'\le\beta} \frac{dx(t')}{(2\pi\delta t')^{1/2}}, \tag{3}$$

is the Green's function at $t = \beta$ of Eq. (2) for a unit source located at x_0 when $t = 0$. Here the weighting functional is given by

$$P\{x(\cdot)\} = \exp\left[-\int_0^\beta \frac{1}{2}\left(\frac{dx(t')}{dt'}\right)^2 dt'\right] \prod_{0\le t'\le\beta,} (2\pi\delta t')^{-1/2}, \tag{4}$$

$\delta t'$ being the limit of a subdivision of the t interval in the mathematical process underlying the symbolic form (1). The functional being integrated is

$$f\{x(\cdot)\} = \exp\left(-\int_0^\beta V(x(t'))dt'\right) \times \delta(x(\beta) - x_1(\beta))\delta(x(0) - x_0(0)). \tag{5}$$

A somewhat new notation has been introduced in Eq. (3), which will be much utilized for compactness of expression in our ensuing discussion: that of appending subscripts to a function at a particular value of the independent variable. If $x(t)$ is the value at time t of the function x, functional integration involves integrating over a range of such values [$(-\infty, \infty)$ in the present case] for each fixed t. But one would like a way of denoting a particular, fixed (or parametrically variable) value of $x(t)$ which would distinguish it from a value of the different variable $x(t')$ (where $t \ne t'$). This we do by retaining the identifying argument t while at the same time fixing the value by a subscript.

Thus $x_a(t)$ denotes the value x_a (some number) of $x(t)$ regarded as a variable for fixed t. The quantities $x_a(t)$ and $x_a(t')$, when $t \neq t'$, are (for example) *numerically equal* but distinguished by virtue of being values taken on by two different variables. [This feature is not utilized in Eq. (3), but will be later on.]

Thus $K(x_1(\beta), x_0(0))$ denotes explicitly and quite compactly the value taken on by the Green's function at time β at the point x_1, for a source at time zero located at x_0. But the greatest usefulness of the new notation lies in the fact that it can convey this meaning in the context of an equation in which $x(0)$ and $x(\beta)$ play the role of variables of integration as well as taking on the special values x_0 and x_1.

Essentially all important forms of the Wiener integral (Bloch equation) and Feynman integral (Schrödinger equation) can be obtained from Eq. (3).

When the weighting functional is real, we shall assume it to be normalized to unity for the appropriate interval of t, say,

$$\int P\{x(\cdot)\}\prod_{t_1 \leq t' \leq t_2} dx(t') = 1, \tag{6}$$

and it has the properties of a probability density in the space of functions $x(\cdot)$. The functional integral over this interval would then be an expectation value

$$I(f) = E_{[0,t]}[f]; \tag{7}$$

the symbol $[0,t]$ of the interval $0 \leq t' \leq t$ is appended as an often useful indication.

Conditional expectation values are often resorted to; in fact, the above functional integrals are really conditional expectation values, because the probability density is in fact always specified as a conditional one. The Wiener process (and, at least formally, the Feynman "process") is Markovian, and $P\{x(\cdot)\}$ is the limit of the probability of a Markov chain. For an interval $0 \leq t' \leq t$,

$$P\{x(\cdot)\} = \lim_{n\to\infty} P(x(t)\,|\,x(t_n))P(x(t_n)\,|\,x(t_{n-1}))\cdots \times P(x(t_1)\,|\,x(0)), \tag{8}$$

where

$$P(x(t')\,|\,x(t)) = [2\pi(t'-t)]^{-1/2} \times \exp\{-[x(t')-x(t)]^2/2(t'-t)\} \tag{8'}$$

and $x(0), x(t_1), x(t_2), \ldots, x(t_n), x(t)$ are values at a finite set of successive time instants of the function $x(t)$. The limit $n \to \infty$ is taken for increasingly fine subdivision. The product of conditional probability densities on the right-hand side is the joint probability density of $x(t_1), x(t_2), \ldots, x(t)$ *given* $x(0)$; hence, in reality, $P\{x(\cdot)\}$ is also a conditional probability density or, explicitly, $P\{x(\cdot)\,|\,x(0)\}$.

The result of Eq. (3) is unchanged if the factor $\delta(x(0) - x_0(0))$ and the integration over $x(0)$ are omitted while using the probability conditional on $x(0) = x_0(0)$. In fact, the Green's function is a conditional value:

$$\begin{aligned} K(x_1(\beta), x_0(0)) &= \int P\{x(\cdot)\,|\,x(0)\}\ \exp\left(-\int_0^\beta V(x(t'))dt'\right) \\ &\quad \times \delta(x(\beta) - x_1(\beta))\prod_{0<t'\leq\beta} dx(t') \\ &= E_{(0,\beta]}\left[\exp\left(-\int_0^\beta V(x(t'))dt'\right)\right. \\ &\quad \left.\times\, \delta(x(\beta) - x_1(\beta))\,|\,x_0(0)\right]. \end{aligned} \tag{9}$$

The omission of integration over $x(0)$ is indicated in the subscript to the Π sign and in the interval notation appended to E.

The *quantum mechanical partition function* of a particle in the potential $V(x)$ at a temperature β^{-1} is

$$Z(\beta) = \int K(x_0(\beta), x_0(0))dx_0. \tag{10}$$

In writing this formula, we are able to take full advantage of our new notational device. Equal values of $x(t)$ at different times, $x_0(0)$ and $x_0(\beta)$, are denoted by equal subscripts without losing the distinct time indications.

B. Methods Used and Outline of Results Obtained

The techniques discussed in this paper consist of: (A) the reduction of the infinite-dimensional functional integral expression, which is of symbolic value only, to a series of explicit integrals of finite order; (B) and (C), two devices which improve the accuracy of the results, to be further described below; and (D) the use of a new probability distribution in the case of the partition function.

The result (A) is obtained by Taylor expansion of the exponential in Eq. (9) after adding and subtracting the expectation value of the exponent, *in* the exponent.[1] The resulting series generalize the result obtained by Feynman and Hibbs (Ref. 2, Sec. 10-3 and Chap. 11) in terms of a variational argument, their result being the first term of the series. The convenience of the series lies in the fact that each term can be evaluated in terms of a finite number of quadratures.

The series may be compared with that of Gel'fand and Yaglom [Ref. 3, Eq. (4.10)]. The latter is, in effect, an expansion of the functional being averaged, Eq. (5), with respect to the random function $x(\cdot)$. Ours is an expansion with respect to the random function $V(x(\cdot)) - E[V(x(\cdot))]$.

Somewhat more closely related is the series used by Kac (Ref. 4, p. 168) to prove that the Wiener integral of $\exp[-\int V(x(t'))dt']$ satisfies the Bloch equation. However, Kac's expansion does not lend itself to accurate practical evaluation, because (cf. Sec. IIIA below for details of our method) it does not separate out the factor $\exp\{E[-\int V(x(t'))\,dt']\}$ beforehand; hence the higher terms are ordinary, not central, moments, which decrease more slowly than the latter. In the expansion according to central moments, the leading term $\exp\{E[-\int V(x(t'))dt']\}$ is already in many cases an excellent approximation to the exact result, as was realized (cf. the first paragraph of this article) by Feynman and Hibbs.

We can also compare series methods in general with the classic method of approximation based on the rigorous definition of the functional integral. This uses a finite number, say N, of intervals for the time subdivision, i.e., stops at a finite stage of the limit process of Eq. (8). In this way an N-fold integral is obtained, but not a series; if an improvement is desired, the calculation must be repeated. Various refinements of this technique have been given.[7,8] We shall see that the lowest-order approximation of the series method is as accurate as the result obtained with quite large values of N by the time-subdivision method. It also has the advantage of the natural analytic structure of its terms.

The special devices referred to under (B) and (C) are:
(B) The quadratic term in $V(x)$, or part of it, may be split off and incorporated in the weighting function (4).[1] The result is again a Markovian probability-density functional, whose univariate propagator $P(x(t') \mid x(t))$ is the well-known Uhlenbeck–Ornstein probability density. Functional integrals with this weighting, Uhlenbeck–Ornstein integrals, are not much more difficult to carry out than Wiener integrals, because of the simplicity of form of the Uhlenbeck–Ornstein function.[9] If the quadratic term is appreciable, a considerable gain in accuracy is obtainable. This is evident from the basic fact that the Uhlenbeck-Ornstein integral has in effect a more advanced starting point, this advance over the Wiener integral being measurable by the fact that unlike the latter it is *exact* if the potential is purely harmonic. An additional property to be expected from the Uhlenbeck–Ornstein weighting is much greater accuracy at low temperatures ($\beta >> 1$). This is because of the greatly reduced dispersion of Uhlenbeck–Ornstein paths compared to Wiener paths (for details, see Sec. IIA).

(C) As Feynman and Hibbs (Ref. 1, Chap. 10) proposed, we may reserve a parameter of the functional $x(\cdot)$, namely its time average $\bar{x}$, from the expectation value at first, integrating over $\bar{x}$ only after carrying out operation A, the expansion in central moments. This improves the accuracy because it corresponds to carrying out operation A with respect to subensembles of lower dispersion, which are then recombined after the performance of the various operations defined above. (If we did this with respect to a complete set of parameters defining the path, instead of just $\bar{x}$, we would, in fact, have an *exact* result.)

(D) In a rather different category from the preceding, because it applies only to the partition function, belongs another technique due to Feynman and Hibbs. They observed that the equating of $x(\beta)$ to $x(0)$ in the partition function [Eq. (10)] gives rise to a family of paths of a remarkable and useful periodicity property. We show that this can be formulated in terms of a new probability space. The integration over the equated end points $x_0(0)$ and $x_0(\beta)$ is carried out *before* the operation E in Eq. (10). This operation combined with E can be manipulated into the form of a new kind of expectation value, so that the partition function is itself an expectation value, apart from a factor.

2. METHODS FOR FUNCTIONAL INTEGRALS OF THE FORM $\exp[-F(x(\cdot))]$ IN GENERAL

A. Expansion of the Integral of an Exponential Functional in Terms of Central Moments of the Exponent ("Operation A")

We evaluate the quantity

$$I = E\left[\exp\left(-\int V(x(t'))\,dt'\right)\right]. \tag{11}$$

No specification need to made of the form of the probability weighting. Also, no specification need be made, as yet, of the time interval involved. Certain conditions on the expectation value, such as a fixed $x(0)$, fixed $\bar{x}$, etc., may be implicit in the expectation operator E.

We now multiply and divide by

$$\exp\left[-E\left(\int V(x(t'))dt'\right)\right] \tag{12}$$

to get

$$I = \exp\left[-E\left(\int V(x(t'))dt'\right)\right] E\left(\exp\left\{-\left[\int V(x(t'))dt' - E\left(\int V(x(t'))dt'\right)\right]\right\}\right). \tag{13}$$

We also expand the second exponential and take the expectation values term by term. The resulting terms are *central moments* of the random functional $\int V(x(t'))dt'$. Putting

$$V_c(x(t)) = V(x(t)) - E[V(x(t))] \tag{14}$$

and

$$M^{(k)}[V_c] = E\left[\left(\int V_c(x(t'))dt'\right)^k\right], \tag{14'}$$

we have

$$I = \exp\left(-E\left[\int V(x(t'))dt'\right]\right) \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} M^{(k)}[V_c]. \tag{15}$$

The point of this expression is that all terms involve only a finite number of integrations. This may be seen as follows: Apart from multiplicative factors such as powers of $E[\int V_c(x(t'))dt']$, each moment consists of terms like

$$\begin{aligned} &E\left[\left(\int V_c(x(t'))dt'\right)^l\right] \\ &= E\left[\int\cdots\int V_c(x(t_1))\,V_c(x(t_2))\cdots V_c(x(t_l))dt_1\cdots dt_l\right] \\ &= \int\cdots\int E\left[V_c(x(t_1))\,V_c(x(t_2))\cdots V_c(x(t_l))\right]dt_1\cdots dt_l. \end{aligned} \tag{16}$$

But the expectation value inside the integral sign is given by

$$\begin{aligned} &E[V_c(x(t_1))\,V_c(x(t_2))\cdots V_c(x(t_l))] \\ &\quad = \int\cdots\int P(x(t_1),x(t_2),\cdots,x(t_l))\,V_c(x(t_1)) \\ &\qquad \times V_c(x(t_2))\cdots V_c(x(t_1))dx(t_1)dx(t_2)\cdots dx(t_l). \end{aligned} \tag{17}$$

QED

(The probability density P contains the same implicit conditions as the expectation E.)

Of course, the practical usefulness of the expansion depends on the possibility of evaluating the probability density. In the cases we shall be concerned with, the probabilities are Gaussian, Chapman–Kolmogorov conditions are satisfied, and the quantities held fixed in the conditions are linear functionals of $x(\cdot)$; hence the evaluation of P is always straightforward.

The *convergence properties* of the expansion may be discussed as follows. For the method to work at all, the first factor of I in Eq. (13) must be finite. But the theorem $E[e^y] \geq e^{E[y]}$, where y is any random variable, guarantees that this factor is finite if the functional integral exists at all. As for the second factor in Eq. (13), we can state that the term-by-term sum of the expectation values of the expanded exponential surely converges if the potential function is bounded from below, for then $E[\int V(x(t'))dt']$ is bounded from below; it is bounded from above if the functional integral itself exists, as follows from the theorem cited earlier in this paragraph; and the exponent in the second factor of Eq. (13) is bounded from above. Let the (finite) upper bound be called B.

Then the terms in the expansion of the operand in the said factor are majorized term by term by those in the expansion of e^B. Since the latter has a term-by-term sum of expectation values which converges to $E[e^B] = e^B$, the expectation value in question must likewise converge. A more delicate discussion will be needed for potentials unbounded from below, but the present one will suffice to permit the application of our methods to many commonly used interatomic and intermolecular potentials.

To summarize what has been accomplished so far: The series expansion converts an infinite-fold integral into a series of integrals of finite order, convergent under the conditions stated. Increasingly accurate estimation of the functional integral is obtained by adding successive terms, rather than by redoing the previous approximation.

More general functionals: The above may be regarded as a special case of evaluation value of a functional through its functional Taylor series. Let there be given a random function $f(t)$, which may be a function, say $f(t) = \phi(x(t))$, of another random function $x(t)$.

We work through characteristic functions, i.e., Fourier transforms. Denote the characteristic function of $P\{f\}$ by $\hat{P}\{\hat{f}\}$, i.e., $f(t)$ is the Fourier conjugate function to $f(t)$. Then

$$\hat{P}\{\hat{f}\} = E\left[\exp\left(i\int\hat{f}(t)f(t)\,dt\right)\right] = \exp\left(i\int\hat{f}(t)E[f(t)]dt\right) \times E\left[\exp\left(i\int\hat{f}(t)\{f(t) - E[f(t)]\}\,dt\right)\right] \quad (18)$$

Expanding the exponential following the E sign, we have

$$\hat{P}\{\hat{f}\} = \exp\left(i\int\hat{f}E[f]dt\right)\sum_n \frac{1}{n!}\int\cdots\int[i\hat{f}(t_1)]\cdots \times [i\hat{f}(t_n)]\mu_n(t_1,\cdots,t_n)dt_1\cdots dt_n, \quad (19)$$

where μ_n is the nth central moment of $f(t)$. Inverting the Fourier transform, we obtain

$$P\{f\} = \sum\frac{1}{n!}\int dt_1\cdots dt_n\cdot\mu_n\cdot\int[i\hat{f}(t_1)]\cdots[i\hat{f}(t_n)] \times \exp(-i\int\{f - E[f]\}\hat{f}dt)\prod_t\frac{d\hat{f}(t)}{2\pi} \quad (20)$$

$$= \sum\frac{1}{n!}\int dt_1\cdots dt_n\cdot\mu_n\cdot\frac{(-\delta)^n}{\delta f(t_1)\cdots\delta f(t_n)} \times \delta\{f(t) - E[f(t)]\}. \quad (21)$$

In Eq. (21) the second integral is a functional integral, of which $\Pi_t d\hat{f}(t)$ is the multiple differential. In the third member, the "fraction" immediately following μ_n is a functional derivative, and the term following this is a functional Dirac delta function.

Now consider the expectation value of a functional $F\{f(\cdot)\}$:

$$E[F\{f\}] = \int P\{f\}F\{f\}\prod_t df(t). \quad (22)$$

If (21) is inserted for $P\{f\}$ in (22), integration by parts transfers the functional derivative to $F\{f\}$, and the integral over $\Pi df(t)$ can be carried out at once because the delta function has been freed of differentiation. The result is

$$E[F\{f\}] = \sum\frac{1}{n!}\int\cdots\int dt_1\cdots dt_n\cdot\mu_n \cdot\frac{\delta^n}{\delta f(t_1)\cdots\delta f(t_n)}F\{f\}|_{f=E[f]} \quad (23)$$

Now specialize to $f(t) = -V(x(t))$ and $F\{f\} = \exp(-\int V_c dt)$. The expansion (15) is obtained immediately.

B. The Green's Function as an Uhlenbeck–Ornstein Integral

To obtain the Uhlenbeck–Ornstein distribution in its standard form, it is necessary to choose the units somewhat differently than is done in arriving at Eq. (2). The dimensionless form of the partial differential equation becomes (see Appendix B)

$$\frac{\partial\psi}{\partial t} = \left(\frac{\partial^2}{\partial x^2} - \frac{x^2}{4} - U(x)\right)\psi, \quad (24)$$

in which $U(x)$ does not contain any quadratic part, this having been split off to form the $(-x^2/4)$ term.

One further functional transformation is needed before we can have the Uhlenbeck–Ornstein integral in the required standard form: Put

$$\phi(x) = e^{-t/2}e^{-x^2/4}\psi(x). \quad (25)$$

Then (24) becomes

$$\frac{\partial\phi}{\partial t} = [-H_{FP} + U(x)]\phi, \quad (26)$$

where

$$H_{FP} = -\frac{\partial}{\partial x}\left(x + \frac{\partial}{\partial x}\right) \quad (27)$$

is the "Fokker–Planck operator." The Green's function of this operator alone, i.e., the solution of

$$\frac{\partial}{\partial t}P(x(t)|x(0)) = H_{FP}P(x(t)|x(0)) \quad (28)$$

such that $P(x(t)|x(0)) \underset{t\to 0}{\to}\delta(x - x_0)$, is the Uhlenbeck–Ornstein function

$$P_{UO}(x(t)|x(0)) = [2\pi(1 - e^{-2t})]^{-1/2} \times \exp-\left(\frac{(x(t) - x(0)e^{-t})^2}{2(1 - e^{-2t})}\right). \quad (29)$$

This function is a conditional probability normalized to unity; hence the Green's function of Eq. (26) is of the form (9), with V replaced by U and with the conditional functional probability density formed from Eq. (8) with P_{UO} used in the product on the right-hand side.

To express the Green's function of Eq. (24) as an Uhlenbeck–Ornstein integral, we have to express it in terms of the Green's function of Eq. (26). In view of the transformation (25), the Green's function of Eq. (24), which we distinguish by a subscript ψ, is given by

$$K_\psi(x_1(\beta),x_0(0)) = e^{-\beta/2}e^{-x_1^2/4}k_\phi(x_1(\beta),x_0(0)), \quad (30)$$

where k_ϕ is the Green's function of Eq. (26) that satisfies the initial condition

$$k_\phi(x_1(t),x_0(0)) \to e^{-x_1^2/4}\,\delta(x_1 - x_0) = e^{-x_1^2/4}\delta(x_1 - x_0). \quad (31)$$

Due to the linearity of the differential equation and the fact that the factor $\exp(-x_0^2/4$ in the last form of the above equation is not subject to the differential operation of eq. (26), $k_\phi(x_1(t), x_0(0))$ is equal to $\exp(-x_0^2/4)$ times K_ϕ, the Green's function for the more usual source condition $\delta(x_1 - x_0)$. Thus

$$K_\psi(x_1(\beta), x_0(0)) = \exp(-\tfrac{1}{2}\beta + \tfrac{1}{4}(x_1^2 - x_0^2))K_\phi(x_1(\beta), x_0(0))$$
$$= \exp(-\tfrac{1}{2}\beta + \tfrac{1}{4}(x_1^2 - x_0^2))$$
$$\times E^{UO}_{(0,\beta)}\left[\exp\left(-\int_0^\beta U(x(t'))dt'\right)\delta(x(\beta) - x_1(\beta))\,|\,x_0(0)\right]. \quad (32)$$

In forming the partition function of the system from the Green's function of Eq. (24).

$$Z(\beta) = \int K_\psi(x_0(\beta), x_0(0))dx_0, \quad (33)$$

it must be realized that, for a given temperature, the t (and β) of Eq. (2) and that of Eq. (24) are different due to the different units used.

Advantage of the Uhlenbeck–Ornstein over the Wiener Integral: Since the terms of the expanded exponential functional integral are essentially the central moments of a functional of $x(\cdot)$, rapidity of convergence is obtainable by reducing the *dispersion* of $x(\cdot)$; the central moments are, virtually by definition, the measure of the dispersion of a probability distribution. This is the reason for the superiority of the method of transferring the quadratic part of the potential to the distribution function (making it the Uhlenbeck–Ornstein distribution).

For consider the Wiener distribution. The proportionality of the variance of $x(t)$, given $x(0)$, to t means a very wide range of values in the probability ensemble when t becomes large. Of course, we deal with paths fixed at *both* ends when we evaluate Green's functions, and so we should be more precise and speak in terms of the distribution of $x(t')$, given $x(0)$ *and* $x(t)$, where $0 < t' < t$. This is given by Ref. 10,

$$P(x(t')\,|\,x(0), x(t)) = \frac{P(x(t)\,|\,x(t'))P(x(t')\,|\,x(0))}{P(x(t)\,|\,x(0))}. \quad (34)$$

For the Wiener process, this is, if we put $x(0) = 0$,

$$\left[\frac{2\pi t}{t'(t-t')}\right]^{-1/2} \exp\left(\frac{x(t)^2}{2t} - \frac{[x(t') - x(t)]^2}{2(t-t')} - \frac{x(t')^2}{2t'}\right), \quad (35)$$

which may be put in the form

$$\left[\frac{2\pi t}{t'(t-t')}\right]^{-1/2} \exp\left(-\frac{[x(t') - (t'/t)x(t)]^2}{2(t'/t)(t-t')}\right). \quad (36)$$

This shows that $x(t')$ has mean $(t'/t)x(t)$ [which lies for all t' on the straight line joining $x(0)$ to $x(t)$] and variance $t'(t-t')/t$. The latter vanishes at zero time and at time t, with a maximum $t/4$ at the midpoint of the time interval. Hence, even with end points fixed, the dispersion of $x(t')$ values is of order t over the major part of the time interval, and increases without limit as t increases.

With the Uhlenbeck–Ornstein distribution the situation is entirely different. Here the dispersion is *bounded for all t*. To show this, it is sufficient to observe that the dispersion of $x(t)$, given fixed initial point only, is bounded; Eq. (34) above shows that the dispersion with both end points fixed is governed by that with initial point fixed. Indeed, the dispersion of $x(t)$ under the latter condition is, from Eq. (16), $1 - e^{-2t}$, which is never greater than unity. This striking difference between the Uhlenbeck–Ornstein and Wiener distributions is, of course, due to the *damping* in the Langevin processes underlying the former, and is vividly illustrated in the physical processes which, historically, motivated their introduction.[11–13].

A detailed account of the clustering of Uhlenbeck–Ornstein paths with both end points fixed about the mean path will be found in a paper by Siegel.[10] The situation is illustrated for the Wiener and Uhlenbeck–Ornstein distributions in Figs. 1 and 2, respectively.

The advantage of the Uhlenbeck–Ornstein paths over the Wiener paths increases with the value of β in the functional integral and with the coefficient of the quadratic term in the potential relative to the magnitude of the remainder. If the quadratic term is relatively small, the remainder of the potential becomes very large after reduction to the standardized units, and essentially magnifies the dispersion, nullifying the effect of the damping. On the other hand, no matter how small the quadratic term, for sufficiently

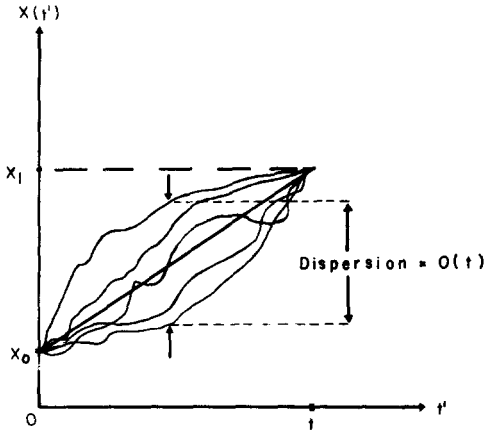


FIG. 1. Dispersion of Wiener paths with both ends fixed. The straight line $x(t') = x_0 + (t'/t)(x_1 - x_0)$ corresponds to the average $E[x(t')\,|\,x_0(0), x_1(0)]$, the classical path.

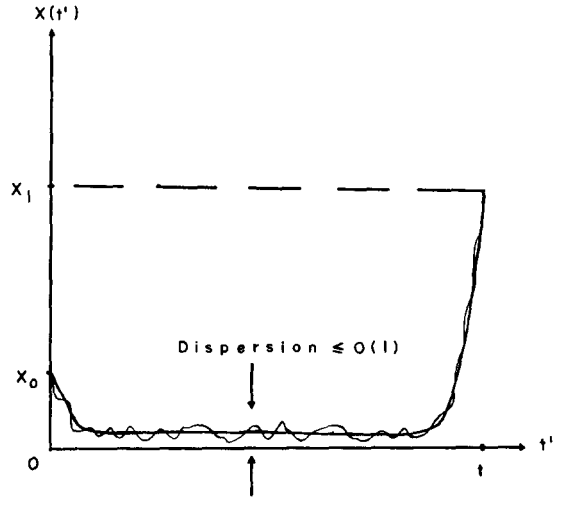


FIG. 2. Dispersion of Uhlenbeck–Ornstein paths with both ends fixed. The smooth curve is $E[x(t')\,|\,x_0(0), x_1(t)]$; the equation for this curve and that for the dispersion may be found in Siegel, Ref. 10.

large values of β the reduction of dispersion does come into play, so that the inclusion of the quadratic term in the distribution is to be preferred at low temperatures, other things being equal.

C. Reserving a Variable before Performing "Operation A"

By reserving a variable (some parameter of the path, or distributed quantity) from an operation we mean holding it fixed during that operation and completing the expectation value subsequently to the operation by averaging over the previously fixed variable. The concept of the reserving operation in its relation to other operations will be clearer in terms of a condensed notation which we now introduce. If the variable is ξ, and the entire space of sample functions is $S, S-\xi$ is the space with ξ excluded, and we put $E_{S-\xi}$ for the conditional expectation value with ξ fixed:

$$E_{S-\xi} f\{x(\cdot)\} = \int P(x(\cdot)\,|\,\xi) f(x\{\cdot\}) \prod_t dx(t). \tag{37}$$

For the expectation value over ξ we put E_Ξ : If g is a function $g(\xi)$,

$$E_{(\xi)} g(\cdot) = \int P(\xi) g(\xi) d\xi. \tag{38}$$

A petty consistency impels us to parenthesize ξ in $E(\xi)$ in order to recognize that this operation is, unlike $E_{S-\xi}$, not a function of ξ (ξ having been integrated out).

With these notations, the expectation value over S is the convolution

$$E = E_{(\xi)} E_{S-\xi}. \tag{39}$$

A certain defeat of consistency may be observed here in that the ξ in the last expression would, to agree with previous usage, have to be replaced by a dot.

If we put

$$\mathfrak{V} = \int V(x(t'))dt', \tag{40}$$

then "operation A" consists of the transformation

$$E(e^{-\mathfrak{V}}) = e^{-E(\mathfrak{V})} E(e^{-[\mathfrak{V}-E(\mathfrak{V})]}), \tag{41}$$

whose effectiveness as an approximation is a function of the smallness of the dispersion of $\mathfrak{V}$; when the dispersion of $\mathfrak{V}$ is small, $|\mathfrak{V} - E(\mathfrak{V})|$ is small and $e^{-[\mathfrak{V}-E(\mathfrak{V})]}$ is close to one over "most of" the probability space: This means that $e^{-E(\mathfrak{V})}$ is then a good approximation and, what is also important computationally, the series expansion of $E(e^{-[\mathfrak{V}-E(\mathfrak{V})]})$ will converge rapidly.

When the variable ξ is reserved from operation A, E is decomposed according to the convolution (39), and operation A is carried out with respect to $E_{S-\xi}$ only:

$$E(e^{-\mathfrak{V}}) = E_{(\xi)}(\exp[-E_{S-\xi}(\mathfrak{V})] \times E_{S-\xi}(\exp\{-[\mathfrak{V} - E_{S-\xi}(\mathfrak{V})]\})). \tag{42}$$

This will usually improve the degree of approximation, in view of the argument above, since subensembles of S with fixed ξ will tend to have lower dispersion than S. In fact, if a series of parameters of the sample function is reserved from operation A, the zero-order term approaches the exact result.

A price is paid, of course, since reserving ξ means another integration to perform; moreover, once past the first term of the series, the integrand is a product, and the integration rapidly becomes complicated. As rule of thumb, if an accurate zero-order result is desired, it is necessary to reserve variables; if an expansion seems preferable, it would be better to keep it simple by not reserving variables.

D. A Comprehensive Operator Formalism

The use of the Uhlenbeck–Ornstein distribution as described in Sec. IIB can be regarded as an operation, on a par with the reservation of variables.

To make this explicit, we have to show the relationship between the expectation value of the functional of $\int V(x(t'))dt'$ using the Wiener weighting and that of the functional of $\int U(x(t'))dt'$ using the Uhlenbeck-Ornstein weighting. A bit of (straightforward) work is needed here, partly because different units were used in the dimensionless Bloch/Schrödinger equation (2) and the dimensionless Fokker–Planck equation (28) in order that each would be in standard form. In the first place, let us distinguish the quantities involved in the two ψ equations, (2) and (24), by using primes for the variables, functions, and parameters of the latter. Thus Eq. (2) remains

$$\frac{\partial \psi}{\partial t} = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - V(x)\right)\psi \tag{43}$$

while Eq. (24) becomes

$$\frac{\partial \psi'}{\partial t'} = \left(\frac{\partial^2}{\partial x'^2} - \frac{x'^2}{4} - U'(x')\right)\psi'. \tag{44}$$

Primes will also be used for the units used to effectuate the change to dimensionless form, to distinguish those of the Uhlenbeck–Ornstein case from those of the Wiener/Feynman case (Appendices A and B); thus we have ω' and l' former case, ω and l in the latter.

If $K(x(t), x_0(0))$ and $K'(x'(t'), x'_0(0))$ are the Green's functions of Eqs. (43) and (44), respectively, with formally equivalent initial conditions, the first reducing to $\delta(x'(t') - x_0)$ as $t' \to 0$, then one can show that

$$l^{-1/2} K(x(t), x_0(0)) = l'^{-1/2} K'(x'(t'), x'_0(0)). \tag{45}$$

We use this equation to relate expectation values, through Eqs. (9) and (32), with respect to the two distributions, of the two functionals. First, to have a more straightforward result, we eliminate the delta functions from Eqs. (9) and (32). In the case of Eq. (9) this is done as follows: By putting [see Eq. (8) for notation]

$$P(x(t), x(t_n), x(t_{n-1}), \cdots, x(t_1)\,|\,x(0)) = P(x(t)\,|\,x(0)) P(x(t_n), x(t_{n-1}), \cdots, x(t_1)\,|\,x(0), x(t)), \tag{46}$$

we find that (Ref. 4, p. 172)

$$E^W_{(0,\beta]}\left[\exp\left(-\int_0^\beta V(x(\tau))d\tau\right)\delta(x(\beta) - x_1)\,|\,x_0(0)\right] = P(x_1(\beta)\,|\,x_0(0)) E_{(0,\beta)}\left[\exp\left(-\int_0^\beta V(x(\tau))d\tau\right)\Big| \; x_0(0), x_1(\beta)\right]. \tag{47}$$

On the right-hand side, as indicated by the interval subscript, the expectation value is taken over the (infinite) set of variables $x(\tau)$ for the *open* interval $0 < t' < \beta$. [Note that we are now using τ as the variable of integration in the exponent instead of t'; this is to avoid confusion with the use of the prime in this section to distinguish the variables, etc., of Eq. (2).] Since an exactly similar expression to this holds for the Uhlenbeck–Ornstein expectation in terms of primed variables, Eq. (32) becomes

$$K'(x_1'(\beta'), x_0'(0)) = \exp[-\tfrac{1}{2}\beta' + \tfrac{1}{4}(x_1'^2 - x_0'^2)]\, P^{UO}(x_1'(\beta')\,|\,x_0'(0)) \times E^{UO}_{(0,\beta')}\left[\exp\left(-\int_0^{\beta'} U(x'(\tau'))\,d\tau'\right)\middle|\, x_0'(0), x_1'(\beta')\right]. \tag{48}$$

Equations (9), (45), (47), and (48) enable us to relate the expectation values as follows:

$$E^{W}_{(0,\beta)}\left[\exp\left(-\int_0^{\beta} V(x(\tau))\,d\tau\right)\middle|\, x_0(0), x_1(\beta)\right] = \sqrt{\frac{l}{l'}}\exp\left(\frac{-\beta'}{2} + \frac{1}{4}(x_1'^2 - x_0'^2)\right) \times \frac{P^{UO}(x_1'(\beta')\,|\,x_0'(0))}{P^{W}(x_1(\beta)\,|\,x_0(0))} \times E^{UO}_{(0,\beta)}\left[\exp\left(-\int_0^{\beta'} U'(x'(\tau'))\,d\tau'\right)\middle|\, x_0'(0), x_1'(\beta)\right]. \tag{49}$$

The implication of Eq. (49) for the operator formalism is as follows: Let $Q(x)$ be the quadratic part of $V(x)$, while $U(x)$ is the nonquadratic part,[14] so that

$$V(x) = U(x) + Q(x). \tag{50}$$

Now, the product of an energy and a "time" (θ, t or t') is invariant with respect to the transformations of variables in Appendixes A and B, so we can make the replacement

$$\int_0^{\beta'} U'(x'(\tau'))\,d\tau' \to \int_0^{\beta} U(x(\tau))\,d\tau \tag{51}$$

in the exponent following the $E^{UO}_{(0,\beta')}$ sign in Eq. (49), with the understanding that U, x, τ, and β are regarded as functions of U', x', τ', and β' for purposes of taking the Uhlenbeck–Ornstein expectation value. Equation (49) may then be written

$$E^{W}(e^{-\mathfrak{R}}) = \gamma\, E^{UO}(e^{-\mathfrak{U}}) \tag{52}$$

where γ is the coefficient of $E^{UO}_{(0,\beta')}$ in Eq. (49) and the gothic letter $\mathfrak{U}$ stands for the time integral of U, a notation parallel to that of Eq. (40). Since, with a parallel notation for Q

$$\mathfrak{V} = \mathfrak{R} + \mathfrak{U}, \tag{53}$$

Eq. (52) means that as an *operator* equation, we have

$$E^{W} e^{-\mathfrak{Q}} = \gamma E^{UO}. \tag{54}$$

This equation shows in explicit fashion how the *transfer of the part* of the potential to the Wiener distribution function gives rise to the Uhlenbeck–Ornstein distribution.

"Operation A" can be carried out after the transfer of $e^{\mathfrak{Q}}$, so that we have

$$E(e^{-\mathfrak{V}}) = \gamma E^{UO}(e^{-\mathfrak{U}}) = \gamma \exp[-E^{UO}(\mathfrak{U})] \times E^{UO}(\exp\{-[\mathfrak{U} - E^{UO}(\mathfrak{U})]\}), \tag{55}$$

corresponding to Eq. (42), but with more rapidly converging terms on expanding the exponential, due to the use of the Uhlenbeck–Ornstein potential.

The *reservation of variables* is equally possible after transferring the quadratic part of the potential, viz.,

$$E(e^{-\mathfrak{V}}) = \gamma E^{UO}_{(\xi)}\exp[E^{UO}_{S-\xi}(\mathfrak{U})]E^{UO}_{S-\xi}(\exp\{-[\mathfrak{U} - E^{UO}_{S-\xi}(\mathfrak{U})]\}). \tag{56}$$

Our approximation techniques for the functional integral of $e^{-\mathfrak{V}}$ are summed up in Eqs. (41), (42), (55), and (56). Of these, Eq. (56) is presumably the most powerful, giving the most rapidly converging series on expanding the exponential operand.

E. Explicit Results for the Green's Function with Wiener and Uhlenbeck–Ornstein Probabilities, with No Variables Reserved

We take up the Wiener probability case first. In order to use Eqs. (11) and (15), we transform Eq. (9) through Eq. (47). This gives

$$K(x_1(\beta), x_0(0)) = P(x_1(\beta)\,|\,x_0(0)) \times \exp\left(-E\left[\int_0^{\beta} V(x(t'))\,dt'\,|\,x_0(0), x_1(\beta)\right]\right) \times \sum_{k=0}^{\infty}\frac{(-)^k}{k!}M^{(k)}[V_c\,|\,x_0(0), x_1(\beta)]. \tag{57}$$

We have suppressed the subscript $(0,\beta)$ on E, and it is implicit that probabilities and expectation values are taken with the Wiener measure. $P(x_1(\beta)\,|\,x_0(0))$ is given by Eq. (8′). It should be noted that the first term of the series in (57) vanishes identically, by definition of V_c; it is the first *central* moment of the random functional $\int V(x(t'))dt'$. It is, of course, as we have said, the smallness of central moments in general, and the vanishing of the first one in particular, that makes for the accuracy of the method.

For the exponent we have, by permuting E with the t' integral,

$$E\left[\int_0^{\beta} V(x(t'))\,dt'\,|\,x_0(0), x_1(\beta)\right] = \int_0^{\beta} E[V(x(t'))|\,x_0(0), x_1(\beta)]\,dt' = \int_0^{\beta}\int_{-\infty}^{\infty} V(x(t'))\,P(x(t')\,|\,x_0(0), x_1(\beta))\,dx(t')\,dt'. \tag{58}$$

The probability density in this last integrand has already been obtained [Eq. (36)] for $x_0 = 0$. If $x_0 \neq 0$, the formula (36) need only be altered by subtracting x_0 from both $x(t')$ and $x(t)$. We then obtain from Eq. (58)

$$E\left[\int_0^{\beta} V(x(t'))\,dt'\,|\,x_0(0), x_1(\beta)\right] = \int_0^{\beta} dt'\left(\frac{2\pi\beta}{t'(\beta - t')}\right)^{-1/2}$$

$$\times V(x(t'))\exp\left(-\frac{\{x(t')-E[x(t')|x(0),x(\beta)]\}^2}{2t'(\beta-t')/\beta}\right), \tag{59}$$

where

$$E[x(t')|x(0),x(\beta)]=[(\beta-t')x(0)+t'x(\beta)]/\beta, \tag{60}$$

which is the "classical path" (i.e., average path) of a force-free Brownian particle starting at $x(0)$ and ending at $x(\beta)$.

The corresponding expression with the quadratic part of the potential transferred to the weighting function is obtained with $U(x(t'))$ instead of $V(x(t'))$ and the Uhlenbeck–Ornstein distribution function[10]

$$P(x(t')|x_0(0),x_1(\beta))=(2\pi\sigma)^{-1/2}\exp(-\{x(t')-E[x(t')|x_0(0),x_1(\beta)]\}^2/2\sigma), \tag{61}$$

where

$$\sigma=[2\sinh t'\sinh(\beta-t')]/\sinh\beta, \tag{62}$$

$$E[x(t')|x(0),x(\beta)]=[x(0)\sinh(\beta-t')+x(\beta)\sinh t']/\sinh\beta, \tag{63}$$

which is again a "classical path", this time for a linearly damped motion; see the paper by Siegel[10] for details. It will be noted that the transition from the Wiener to the UO process is effected by two simple steps: (a) In the probability density function of Eq. (59), replace t' and $\beta-t'$ by their hyperbolic sines throughout, and (b) use double the variance thus obtained [cf. Appendices A and B, expressions for l in Eqs. (A3) and (B2)].

The partition function can be evaluated straightforwardly by equating x_1 and x_0 and integrating over this variable. But a more accurate approximation can be obtained with less work by a method that will be described in Sec. III.

F. The Functional Series in Terms of Propagators of the Zero-Order Hamiltonian. Relation to the Feynman–Dyson Expansion

The kth moment in Eq. (57) is evaluated by means of the probability density

$$P(x(t_k),x(t_{k-1}),\cdots,x(t_1)|x_0(0),x_1(\beta)). \tag{64}$$

When this combined with the first factor on the right-hand side of this equation, we have, for ordered times $0<t_1<t_2\ldots<t_k<\beta$,

$$\begin{aligned}P(x_1(\beta)|x_0(0))\,P(x(t_k),x(t_{k-1}),\cdots,x(t_1)|x_0(0),x_1(\beta))&=P(x_1(\beta),x(t_k),x(t_{k-1}),\cdots,x(t_1)|x_0(0))\\&=P(x_1(\beta)|x(t_k))\,P(x(t_k)|x(t_{k-1})),\cdots,P(x(t_1)|x_0(0)).\end{aligned} \tag{65}$$

Thus we find

$$\begin{aligned}K(x_1(\beta),x_0(0))=&\exp\left(-E\left[\int_0^\beta V(x(t'))dt'|x_0(0),x_1(\beta)\right]\right)\\&\times\sum_{k=0}^{\infty}(-)^k\int_0^\beta\cdots\int_0^{t_3}\int_0^{t_2}dt_1\cdots dt_n\\&\times\Big[\int\cdots\int P(x_1(\beta)|x(t_k)\,dx(t_k))\\&\times\bar V_c(x(t_k))\,P(x(t_k)|x(t_{k-1}))\,dx(t_{k-1})\,V_c(x(t_{k-1}))\cdots\\&dx(t_1)\,V_c(x(t_1))P(x(t_1)|x_0(0))dx(t_1)\cdots dx(t_k)\Big].\end{aligned} \tag{66}$$

The sum on the right-hand side is exactly of the form of the Feynman–Dyson series, but with the unexpected feature that the static potential energy $V(x)$ is replaced by a time-dependent potential energy equal to the deviation of $V(x)$ from its mean, at time t, taken with respect to the zero-order path probability.

Since this is so, there should be a way of deriving the series which uses the customary mode of derivation of the Feynman–Dyson expansion. The following is in fact such a derivation. Starting from the Bloch equation (2), we put

$$\Psi(x,t)=\exp\left(-\int_0^t\tilde V(t')dt'\right)\psi(x,t), \tag{67}$$

where we have put for brevity

$$\tilde V(t)=E[V(x(t))|x_0(0),x_1(\beta)]. \tag{68}$$

The function Ψ satisfies the equation

$$\frac{\partial\Psi}{\partial t}=-\left(-\frac{1}{2}\frac{\partial^2}{\partial x^2}+V(x)-\tilde V(t)\right)\Psi, \tag{69}$$

i.e., a Bloch equation with precisely the above-mentioned deviation of $V(x)$ from its zero-order averaged value at time t as potential energy.

The Green's function for Eq. (2) is then obtainable from Eq. (67) as $\exp[-\int_0^t\tilde V(t')dt']$ times the Feynman–Dyson series for the Green's function of Eq. (69); but this is just Eq. (66).

We have thus partly reduced the derivation of (66) to non-path-integral methods. However, path-integral methods remain essential to the definition of the exponential factor in (67) and, more importantly, to the definition of the potential energy in (69). The latter is a potential function only in a very unusual sense, having built into it some of the properties of the zero-order dynamics, through the propagator used in obtaining $\tilde V(t)$ and also being dependent on the initial and final points $x_0(0)$ and $x_1(\beta)$.

Even if the perturbative derivation just given is accepted (and it is perfectly logical despite its unusual features), the path-integral point of view remains invaluable as a source of motivation, as well as for the interpretation it provides (Sec. IC), of the reasons for the accuracy of the path-integral series. However, one can, in a general way, appreciate the accuracy of the expansion also from the perturbative point of view: By subtracting $\tilde V(t)$ from the potential energy, we have greatly decreased its effective value, so that the perturbation series for Eq. (69) will converge much faster than that for Eq. (2).

It should be mentioned that the considerations of this section apply equally well when the zero-order Hamiltonian includes the quadratic part of the potential, being associated with the $\partial^2/\partial x^2$ term in (69), so that we would have $U(x)$ instead of $V(x)$ throughout.

3. A SPECIAL METHOD FOR THE PARTITION FUNCTION: CYCLICAL STOCHASTIC PROCESSES

A. How the Cyclical Process Arises

We find the partition function from Eq. (10) with the Wiener process (harmonic part of potential retained).

Using Eqs. (9) and (47) for the integrand of this equation, we have explicitly, in terms of the distribution function,

$$Z(\beta) = \int dx_0 \int \exp\left(-\int_0^\beta V(x(t'))\,dt'\right) \times P^W(x_0(\beta), \cdots, x(t'), \cdots | x_0(0))\, {\prod}_{0<t'<\beta} dx(t'). \quad (70)$$

In the argument of P the expression $\cdots, x(t'), \cdots$ stands for the set of all $x(t)$ with $0 < t < \beta$.

If the Uhlenbeck–Ornstein process is used (transfer of harmonic potential to the distribution), we have in the same way from Eqs. (33) and (32)

$$Z(\beta) = e^{-\beta/2} \int dx_0 \int \exp\left(-\int_0^\beta U(x(t'))\,dt'\right) \times P^{UO}(x_0(\beta), \cdots, x(t'), \cdots | x_0(0))\, {\prod}_{0<t'<\beta}\, dx(t'). \quad (71)$$

Both of these forms for the partition function can be studied in terms of the general integral

$$Z(\beta) = \int dx_0 \int \exp\left(-\int_0^\beta F(x(t'))\,dt'\right) P(x_0(\beta), \cdots, x(t'), \cdots | x_0(0))\, {\prod}_{0<t'<\beta} dx(t'), \quad (72)$$

where F is a sufficiently well-behaved function and P is the distribution functional of a Markov process, $x(\cdot)$.

The conditional probability with equal $x(0)$ and $x(\beta)$ in Eq. (72) is the probability of a modified process, now to be described. The reason for introducing this process will be given later (last paragraph of this section).

In the modified process $x_0(\beta)$, instead of being a parameter distinct from the variables of integration in Eq. (72), may be regarded as a random variable on the same footing with them, so that (with one simple additional feature, to be described) the integration over x_0 becomes part of a new kind of expectation value. Let the time interval $[0, \beta]$ be bent back on itself so that the point β coincides with the point 0. In effect we have a circle (Fig. 3) of circumference β. On this circle define a random process $x(t)$ whose probability distribution $\tilde{P}\{x(\cdot)\}$ is invariant to rigid displacement (rotation) of the function $x(\cdot)$:

$$\tilde{P}\{x(\cdot + \tau)\} = \tilde{P}\{x(\cdot)\} \quad (73)$$

for all τ. The probability distribution of $x(\cdot)$ shall be defined by the set of all possible m-point distributions, $m = 1, 2, \ldots,$

$$\begin{aligned}\tilde{P}(x_m(t_m), x_{m-1}(t_{m-1}), \cdots, x_2(t_2), x_1(t_1)) &= C \cdot P(x_m(t_m) | x_{m-1}(t_{m-1})) \\ &\times P(x_{m-1}(t_{m-1}) | x_{m-2}(t_{m-2})) \cdots \\ &\times P(x_1(t_1) | x_m(t_m - \beta)),\end{aligned} \quad (74)$$

where $P(x(t_{k+1}) | x(t_k))$, $t_{k+1} > t_k$, is the conditional probability distribution of a stationary Markov process. C is a normalizing constant. The equal subscripts (m) of the variables $x(t_m)$ and $x(t_m - \beta)$ are to be noted. This, according to our convention, makes the variables equal. Since for such a process the basic link probability $P(x(t_{k+1}) | x(t_k))$ is invariant to time displacement, the above definition is invariant to time displacement and consistent with the postulated property (73). It is consistent with the necessary periodicity (of period β) of the postulated process, in particular.

Another consistency requirement that $\tilde{P}$ must satisfy is that integrating out any variable must yield a probability of the postulated form for the remaining variables. That this is so follows from Eq. (74) directly with the use of the Markov property

$$\int P(x(t_{k+1}) | x(t_k))\, dx(t_k)\, P(x(t_k) | x(t_{k-1})) = P(x(t_{k+1}) | x(t_{k-1})), \quad (75)$$

for all of the variables except x_m. It follows for x_m too, with the added use of the time-displacement invariance of the basic link probabilities.

The evaluation of the normalization constant goes as follows: By integrating over all variables in (74), the application of (75) to all the integrations except that over x_m yields

$$C \int P(x_m(t_m) | x_m(t_m - \beta))\, dx_m = 1. \quad (76)$$

In the Wiener case the integrand is a constant [see Eq. (8')], and the process is not normalizable; the cyclic Wiener probability is invariant to vertical displacement of the sample function. But in the Uhlenbeck–Ornstein case [Eq. (29)] the integral does exist, and we find

$$C = 1 - e^{-\beta} \quad (77)$$

The function P in Eq. (72) is the general distribution function whose m-point representative in terms of the basic link probabilities is the product of P's on the right-hand side of (74). The integration over x_0 completes the set of integrations denoted by the product sign so that the result is a complete expecation value. We can drop the subscripts to obtain the general expression

$$Z(\beta) = C^{-1} \int \exp\left(-\int_0^\beta F(x(t'))\,dt'\right) \times \tilde{P}\{x(\cdot)\}\, {\prod}_{0<t'\le\beta} dx(t'). \quad (78)$$

The argument of $\tilde{P}$ is the function $x(t)$ defined over $(0, \beta]$.

When $\tilde{P}$ is normalized, we can write $Z(\beta)$ in terms of the cyclic expectation value $\tilde{E}$:

$$Z(\beta) = C^{-1}\, \tilde{E}\left[\exp\left(-\int_0^\beta F(x(t'))\,dt'\right)\right]. \quad (79)$$

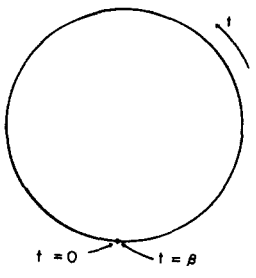


FIG. 3. Schematic representation of the time axis in a cyclical stochastic process.

All the methods of Sec. II are applicable to $\tilde{E}[\cdots]$ in this case: expansion in central moments, reservation of variables, and transfer of the quadratic term.

In the Wiener case, where $\tilde{P}$ is not normalizable, Eq. (78) is still valid, with the definition (74). C may be assigned an arbitrary value. $\tilde{E}[\cdots]$ may be assigned the formal meaning of integration with the weighting $\tilde{P}\{x(\cdot)\}$, but the result of the integration does not exist for all the operations of Sec. II.

The utility of the cyclic process is apparent from a comparison of Eqs. (72) and (78). Equation (72) is the integral over x_0 of a functional integral; Eq. (78) is just a functional integral alone. As will be seen, the functional integral in the latter is of very much the same type as in the former. Thus the effect is to eliminate the integration over x_0, an added complication which cannot be suppressed in any other way. In addition to the analytical simplification, the result using the cyclic process turns out to be more accurate, as can be shown by comparing numerical results in the lowest order of the series.

The foregoing discussion constitutes an explicit formulation and generalization of the highly intuitive line of argument by which Feynman and Hibbs showed that due to the equating of $x(0)$ and $x(\beta)$, the time integration in the exponent of Eq. (57) can be carried out trivially.

B. The Partition Function via the Cyclic Probability with No Variables Reserved

To take up the Wiener weighting first: Since the cyclic probability is not normalizable, we arbitrarily assign C, which cancels out anyway, the value 1. We shall show that in effect operation A by itself cannot be carried out in this case. Since [Eq. (8′)]

$$\tilde{P}(x(t')) = (2\pi\beta)^{-1/2}, \tag{80}$$

we have

$$\tilde{E}[V(x(t))] = (2\pi\beta)^{-1/2}\int_{-\infty}^{\infty} V(x)\,dx \equiv a, \tag{81}$$

where a is finite or infinite. If a is infinite, the method fails and there is nothing more to say. If the potential is integrable, a is finite. Suppose then, first, that it is nonvanishing. Then the moments diverge: E.g., we have for the first central moment

$$(2\pi\beta)^{-1/2}\int[V(x)-a]dx = a - a\cdot\infty = -\infty, \tag{82}$$

and operation A fails. If, on the other hand, $a = 0$, $V(x) - a = V(x)$, and the "central" moments reduce to ordinary moments, the over-all exponential factor of the series is just unity; operation A has a null effect, the resulting expansion being only the standard Feynman–Dyson one.

If the Uhlenbeck–Ornstein weighting is used, operation A can be carried out by itself. We have

$$\begin{aligned}\tilde{P}(x_1(t_1)) &= C^{-1}P^{UO}(x_1(t_1)\,|\,x_1(t_1-\beta))\\ &= \left(\frac{(1-e^{-\beta})^2}{2\pi(1-e^{-2\beta})}\right)^{1/2}\exp\left(-\frac{(1-e^{-\beta})^2}{2(1-e^{-2\beta})}\,x^2\right),\end{aligned} \tag{83}$$

from (29) and (77). This is independent of t_1. The partition function is, from (71), (72), (73), (74), and (77),

$$Z(\beta) = e^{-\beta/2}(1-e^{-\beta})^{-1}\,\tilde{E}\left[\exp\left(-\int_0^{\beta}U(x(t'))dt'\right)\right]. \tag{84}$$

{The coefficient of $\tilde{E}[\cdots]$ is $[2\sinh(\beta/2)]^{-1}$, the known partition function of the harmonic oscillator. Thus we have the standard result if $U(x) = 0$.} Operations A and C may be carried out with respect to the operation $\tilde{E}$. Using A, we have, to zero order,

$$Z(\beta) \cong Z_0(\beta) \equiv \left(2\sinh\frac{\beta}{2}\right)^{-1/2}\exp\left(-\tilde{E}\left[\int_0^{\beta}U(x(t'))dt'\right]\right), \tag{85}$$

where

$$\begin{aligned}\tilde{E}\left[\int_0^{\beta}U(x(t'))\,dt'\right] &= \beta\left(\frac{\tanh(\beta/2)}{2\pi}\right)^{1/2}\\ &\quad\times\int_{-\infty}^{\infty}\exp\left[-\frac{x^2}{2}\tanh\left(\frac{\beta}{2}\right)\right]\,U(x)dx.\end{aligned} \tag{86}$$

In Sec. IV we shall compare this numerically with the results of other approximations.

C. Partition Function via Cyclic Probability, Reserving $\bar{x}$

We first take up the case of Wiener weighting. Because of the nonnormalizability we put $C = 1$. Equations (71), (72), and (78) give

$$Z(\beta) = \int\exp\left(-\int_0^{\beta}V(x(t'))\,dt'\right)\tilde{P}\{x(\cdot)\}\prod_{0<t'\le\beta}dx(t'). \tag{87}$$

In order to reserve $\bar{x}$ in a reasonably straightforward way, we define

$$\tilde{P}(x(\cdot),\bar{x}) = \tilde{P}\{x(\cdot)\}\,\delta\left(\beta^{-1}\int_0^{\beta}x(t')\,dt' - \bar{x}\right), \tag{88}$$

from which the path probability conditional on $\bar{x}$ can be obtained through the usual relation

$$\tilde{P}(x(\cdot)\,|\,\bar{x}) \equiv \tilde{P}(x(\cdot),\bar{x})/\tilde{P}(\bar{x}). \tag{89}$$

We can then calculate probability distributions conditional on $\bar{x}$, and write

$$Z(\beta) = \int\tilde{P}(\bar{x})\,\tilde{E}\left[\exp\left(-\int_0^{\beta}V(x(t'))dt'\right)|\,\bar{x}\right]d\bar{x}. \tag{90}$$

The equivalence of this form to that of Eq. (87) is readily demonstrated with the use of the definitions (88) and (89). If the conditional probability distribution is normalizable, we can carry out operation A on the expectation value $\tilde{E}[\cdots]$ and get

$$\begin{aligned}Z(\beta) = \int\tilde{P}(\bar{x})\exp\left(-\tilde{E}\left[\int_0^{\beta}V(x(t'))\,dt'\,|\,\bar{x}\right]\right)\\ \times\sum\frac{(-1)^k}{k!}M^{(k)}[V_c\,|\,\bar{x}]\,d\bar{x}.\end{aligned} \tag{91}$$

Is the conditional probability normalizable? If the expression for it on the right-hand side of (89) exists, it is normalizable and, in fact, normalized to unity automatically by virtue of the relationship

$$\int\tilde{P}(x(\cdot),x)\prod_{0<t'\le\beta}dx(t') = \tilde{P}(\bar{x}). \tag{92}$$

Hence all that remains is to calculate $\tilde{P}(\bar{x})$ and show it to exist. From Eq. (92),

$$\tilde{P}(\bar{x}) = \int \tilde{P}(x(\cdot))\delta(\beta^{-1}\int_0^\beta x(t')\,dt' - x) \prod_{0<t'\le\beta} dx(t'). \quad (93)$$

From Eq. (74),

$$\tilde{P}(x(\cdot)) = P(x(\cdot))_{x(0)=x(\beta)}, \quad (94)$$

where the right-hand side denotes the basic (non-cyclic) probability for a path with equal end points. Parametrizing the delta function in (93) and separating the integration over $x(\beta)$ from that over the other path variables, we obtain

$$\tilde{P}(\bar{x}) = \int dx(\beta)\,\frac{1}{2\pi}\int_{-\infty}^{\infty} d\lambda e^{i\lambda\bar{x}} \int P(x(\cdot))_{x(0)=x(\beta)} \times \exp\left(-\frac{i\lambda}{\beta}\int_0^\beta x(t')\,dt'\right) \prod_{0<t'\le\beta} dx(t'). \quad (95)$$

The inner path integral has been calculated by Feynman and Hibbs [Ref. 2, Eq. (3.62)], for the slightly more general case $x(0) \neq x(\beta)$. Since we are now using Wiener weighting, we put $\omega = 0$ in their formula and, of course, omit their imaginary factor i. We obtain an integral over λ

$$(2\pi)^{-3/2}\beta^{-1/2}\exp[-i\lambda(\bar{x} - x(\beta)) - \beta\lambda^2/24],$$

whence

$$\tilde{P}(\bar{x}) = (2\pi\beta)^{-1/2}, \quad (96)$$

and the normalizability of $\tilde{P}(x(\cdot))$ is established.

The present case is the only application of operation A given by Feynman and Hibbs, who find the zero-order result. Continuing with our methods, we need next the one-point conditional path probability which we calculate from the joint probability via

$$\tilde{P}(x(t)\,|\bar{x}) \equiv P(\bar{x}(t),\bar{x})/\tilde{P}(\bar{x}). \quad (97)$$

Again by parametrizing the δ function as in Eq. (95),

$$P(x(t),x) = (\sqrt{3}/\pi\beta)\exp\{-(6/\beta)[x(t) - x]^2\}, \quad (98)$$

which is indeed independent of t as was shown by Feynman and Hibbs, using a somewhat complicated argument. On interchanging the $\tilde{E}$ operation and $\int_0^\beta dt'$ in Eq. (91), we obtain a zeroth-order approximation to the partition function in the case of Wiener weighting; this is the result obtained by Feynman and Hibbs (Ref. 2, p. 285),

$$Z_W(\beta) \simeq (2\pi\beta)^{-1/2}\int_{-\infty}^{\infty} d\bar{x}\,\exp[-\beta W_W(\bar{x})] \quad (99)$$

which has just the form of the classical partition function with an "effective potential"

$$W_W(\bar{x}) = \tilde{E}[V(y\,|\bar{x}] = \left(\frac{6}{\pi\beta}\right)^{1/2}\int_{-\infty}^{\infty} dy\,V(y) \times \exp\left(-\frac{6}{\beta}(y - \bar{x})^2\right). \quad (100)$$

(In this section we distinguish between Wiener and Uhlenbeck–Ornstein weighting.) $W_W(\bar{x})$ is a Gauss transform of the true potential and in the high temperature limit $\beta \to 0$, $W_W(\bar{x}) \to V(\bar{x})$, and the system is described by the classical partition function. In Eq. (100) we have replaced $x(t')$ by y as an integration variable. If we introduce the Fourier transform of $V(y)$ by

$$V(k) = \int_{-\infty}^{\infty} dy e^{-iky}V(y), \quad (101)$$

then $W_W(\bar{x})$ may be written in the more compact form

$$W_W(\bar{x}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,V(k)e^{ik\bar{x}}e^{-k^2\beta/24}. \quad (102)$$

In the case of the Uhlenbeck–Ornstein weighting the corresponding results have an identical form although with more complicated coefficients. Parametrizing the δ function in Eq. (95) the path integral is again a standard one [Ref. 2, Eq. (3.66)] and the zeroth-order approximation to the partition function is

$$Z_{UO}(\beta) \simeq \left(\frac{\beta}{8\pi[\cosh(\beta) - 1]}\right)^{1/2}\int_{-\infty}^{\infty} d\bar{x} \times \exp\{-\beta[W_{UO}(\bar{x}) + \tfrac{1}{4}\bar{x}^2]\}, \quad (103)$$

where the effective potential is

$$W_{UO}(\bar{x}) = \left(\frac{\beta/2}{2\pi[(\beta/2)\coth(\beta/2) - 1]}\right)^{1/2}\int_{-\infty}^{\infty} dy\,U(y) \times \exp\left(-\frac{\beta}{4}\frac{(y - \bar{x})^2}{[(\beta/2)\coth(\beta/2) - 1]}\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\,U(k)e^{ik\bar{x}} \times \exp\left(\frac{-k^2[(\beta/2)\coth(\beta/2) - 1]}{\beta}\right). \quad (104)$$

D. Higher Terms in the Expansion

The use of the theory of Gaussian probabilities allows us to derive an explicit expression for the nth term in the expansion (91) in both the Wiener and Uhlenbeck–Ornstein cases. In either case we can write in terms of cylindrical probabilities

$$S_n(\bar{x}) \equiv [(-1)^n/n!]M^{(n)}[V_c\,|\bar{x}] = (-1)^n\int_0^\beta dt_n\int_0^{t_n} dt_{n-1}\cdots \times \int_0^{t_2} dt_1\cdots\int dx_1\cdots dx_n\,V_c(x_1)\cdots V_c(x_n) \times \tilde{P}(x_n(t_n),\cdots,x_1(t_1),\bar{x})/\tilde{P}(\bar{x}). \quad (105)$$

One can consider $\bar{x}$ as a random variable in the same way as $x_1(t_1),\cdots,x_n(t_n)$, and, since the process is Gaussian, we can express $\tilde{P}$ by the characteristic function

$$\tilde{P}(x_n(t_n),\cdots,x_1(t_1),\bar{x}) \equiv P(x_n(t_n),\cdots,x_1(t_1),\bar{x}\,|x_n(t_n - \beta)) = \frac{1}{(2\pi)n+1}\int dk_1\cdots dk_{n+1}\exp\left(-i\sum_{i=1}^{n-1}k_i(x_i - \langle x\rangle)\right) \times \exp[-ik_{n+1}(\bar{x} - \langle\bar{x}\rangle)]\exp\left(-\frac{1}{2}\sum_{ij}^{n}k_ik_j\sigma_{ij}\right), \quad (106)$$

where in the Wiener case

$$\langle x_i\rangle = \langle\bar{x}\rangle = x_n,$$
$$\sigma_{ij} = \sigma_{ji} = \langle[x_i - \langle x_i\rangle][x_j - \langle x_j\rangle]\rangle = \min[t_i + (\beta - t_n), t_j + (\beta - t_n)], \quad (107)$$

$$\sigma_{nn} = \beta, \quad \sigma_{n\,n+1} = \beta/2, \quad \sigma_{n+1\,n+1} = \beta/3,$$
$$\sigma_{n+1\,i} = [t_i + (\beta - t_n)]\{1 - [t_i + (\beta - t_n)]/2\beta\}.$$

Integrating (106) with respect to k_n and k_{n+1} and using (107), we obtain, after considerable manipulation, the following compact form for the expansion terms in the case of Wiener weighting:

$$S_n^W(\bar{x}) = \frac{(-1)^n}{(2\pi)^n} \int_0^\beta dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1\, dk_1 \cdots dk_n \times V_c(k_1)\cdots V_c(k_n) \exp\left(i\bar{x}\sum_{i=1}^n k_i\right) \exp\left(-\frac{\beta}{24}\sum_{i=1}^n k_i^2\right) \times \exp\left\{-\sum_{i<j}^n k_i k_j \frac{\beta}{2} \times \left[\frac{1}{6} - \frac{(t_j - t_i)}{\beta}\left(1 - \frac{(t_j - t_i)}{\beta}\right)\right]\right\}, \quad (108)$$

where

$$V_c(k) = V(k) - \delta(k)\int dk' V(k') \exp(ik'\bar{x}) \exp(-k'^2\beta/24 = V(k) - 2\pi\,\delta(k) W_W(\bar{x}). \quad (109)$$

The Uhlenbeck–Ornstein case is treated in a similar manner; it is somewhat more complicated but equally straightforward. Via the properties of the Uhlenbeck–Ornstein distribution,[13] an expression of the same form as (108) is obtained

$$S_n^{UO}(\bar{x}) = \frac{(-1)^n}{(2\pi)^n} \int_0^\beta dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} dt_1 \times \int dk_1 \cdots dk_n\, V_c(k_1)\cdots V_c(k_n) \exp \bar{x} \sum_{i=1}^n k_i \times \exp\left\{-\frac{1}{2\beta}\left[\left(\frac{\beta}{2}\right)\coth\left(\frac{\beta}{2}\right) - 1\right]\sum_{i=1}^n k_i^2\right\} \times \exp\left(-\sum_{i<j}^n k_i k_j \frac{1}{\beta}\left\{\frac{\beta}{2}\left[\cosh(t_j - t_i) \times \coth\left(\frac{\beta}{2}\right) - \sinh(t_j - t_i)\right] - 1\right\}\right), \quad (110)$$

where in this case

$$V_c(k) = V(k) - 2\pi\delta(k) W_{UO}(\bar{x}). \quad (111)$$

Equations (108) and (110) have an identical structure. This can be seen by introducing functions referring to the Wiener and Uhlenbeck–Ornstein cases,

$$\phi^W(t) = \frac{\beta}{2}\left[\frac{1}{6} - \frac{t}{\beta}\left(1 - \frac{t}{\beta}\right)\right], \quad (112)$$

$$\phi^{UO}(t) = \frac{1}{\beta}\left\{\frac{\beta}{2}\left[\cosh(t)\coth\left(\frac{\beta}{2}\right) - \sinh(t)\right] - 1\right\}. \quad (113)$$

We can then express S_n in a form applicable to both the Wiener and Uhlenbeck–Ornstein cases by using the appropriate functions $\phi(t)$ and $V_c(k_n)$,

$$S_n(\bar{x}) = \frac{(-1)^n}{(2\pi)^n} \int_0^\beta dt_n \cdots \int_0^{t_2} dt_1 \int dk_1 \cdots dk_n V_c(k_1)\cdots V_c(k_n) \times \exp\left(i\bar{x}\sum_{i=1}^n k_i\right)\exp\left(-\tfrac{1}{2}\phi(0) \times \sum_{i=1}^n k_i^2 - \sum_{i<j}^n k_i k_j \phi(t_j - t_i)\right). \quad (114)$$

Equation (114) involves multiple t and k integrations, and except for small values of n it would not be practical to use it directly to calculate corrections to the zeroth-order estimate of the partition function. However, for each $S_n(\bar{x})$ it is possible to obtain a simple expansion in terms of the derivatives of $W(\bar{x})$. Expand the last exponential of (114) in a Taylor series; the δ function in (109) or (111) cancels off all terms in this expansion which do not involve k_i, $i = 1, \ldots, n$. The remaining terms can be expressed in terms of derivatives of $W(\bar{x})$ since

$$W^{(p)}(\bar{x}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} dk (ik)^p e^{ik\bar{x}} \exp[-\tfrac{1}{2}\phi(0)k^2], \quad (115)$$

thus removing the k integrations. The remaining t integrations over the ϕ functions can in principle be computed and have been calculated in the first few cases. For example,

$$S_2(\bar{x}) = -[W'(\bar{x})]^2 \int_0^\beta dt_2 \int_0^{t_2} dt_1 \phi(t_2 - t_1) + [W''(\bar{x})]^2 \cdot \tfrac{1}{2}\int_0^\beta dt_2 \int_0^{t_2} dt_1 \phi^2(t_2 - t_1) + \cdots, \quad (116)$$

$$S_3(\bar{x}) = -[W''(\bar{x})]^3 \int_0^\beta dt_3 \int_0^{t_3} dt_2 \int_0^{t_2} dt_1 \times \phi(t_2 - t_1)\phi(t_3 - t_2)\,\phi(t_3 - t_1) + \cdots. \quad (117)$$

If $W(\bar{x})$ is a smooth function of $\bar{x}$, which will be the case since it is defined as a Gauss transform of $V(\bar{x})$, an expansion in its derivatives will give a good approximation to each $S_n(\bar{x})$. The particularly simple case of a one-dimensional oscillator (in which derivatives higher than the second vanish) is treated numerically in the next section. A different expansion for the partition function, which corresponds to a reordering of the expansion (91) in which each $S_n(\bar{x})$, has been expanded in derivatives of $W(\bar{x})$, can be obtained by an operator technique and will be described in the second paper of this series.

4. NUMERICAL RESULTS

In this section we present the results of the application of the approximation methods described in this paper to the extremely simple case of a quantum particle in a harmonic potential. This problem has the advantage that one can immediately obtain an exact solution for the partition function and related quantities in order to make an accurate estimate of the accuracy of the different terms in the expansions. For this reason approximate methods of calculating functional integrals are usually tested on this potential; see, for example, the work of Fosdick,[8] in which the functional integral is approximated by an n-dimensional integral which is then calculated numerically.

The exact partition function for the quantum harmonic oscillator is (Ref. 2, Chap. 10)

$$Z = [2\sinh(\beta/2)]^{-1}. \quad (118)$$

(The parameter β in our units corresponds to $\hbar\omega/kT$ in Feynman and Hibbs; $\beta \to 0$ is the classical limit, $\beta \to \infty$ the quantum limit, i.e., $T \to 0$.) With the usual definition of the free energy F, then the quantity

$$f \equiv 2F/\hbar\omega = (2/\beta)\ln[2\sinh(\beta/2)] \to 1 \text{ in limit } \beta \to \infty. \quad (119)$$

TABLE I. The relative errors δ_0, δ_2, and δ_3 obtained by the application of the moment expansion to zeroth, second, and third order to the estimation of the free energy of a quantum particle in a harmonic potential.

β_1/β	β	0.1	0.5	1.0	2.0	4.0
0	δ_0	1.508×10^{-8}	3.166×10^{-5}	8.271×10^{-3}	6.117×10^{-3}	3.606×10^{-2}
	δ_2	2.393×10^{-12}	1.252×10^{-7}	1.295×10^{-4}	3.661×10^{-4}	6.920×10^{-3}
	δ_3	7.106×10^{-16}	9.304×10^{-10}	3.864×10^{-6}	4.443×10^{-5}	3.653×10^{-3}
0.3	δ_0	1.249×10^{-8}	2.620×10^{-5}	6.830×10^{-3}	5.009×10^{-3}	2.860×10^{-2}
	δ_2	1.803×10^{-12}	9.425×10^{-8}	9.717×10^{-5}	2.718×10^{-4}	4.992×10^{-3}
	δ_3	4.872×10^{-16}	6.370×10^{-10}	2.632×10^{-6}	2.969×10^{-5}	2.248×10^{-3}
0.6	δ_0	6.176×10^{-9}	1.293×10^{-5}	3.350×10^{-3}	2.397×10^{-3}	1.253×10^{-2}
	δ_2	6.273×10^{-13}	3.269×10^{-8}	3.340×10^{-5}	9.038×10^{-5}	1.517×10^{-3}
	δ_3	1.192×10^{-16}	1.551×10^{-10}	6.313×10^{-7}	6.731×10^{-6}	4.289×10^{-4}
0.9	δ_0	5.442×10^{-10}	1.136×10^{-6}	2.911×10^{-4}	2.002×10^{-4}	9.235×10^{-4}
	δ_2	1.641×10^{-14}	8.509×10^{-10}	8.566×10^{-7}	2.195×10^{-6}	3.156×10^{-5}
	δ_3	9.254×10^{-19}	1.194×10^{-12}	4.745×10^{-9}	4.617×10^{-8}	2.217×10^{-6}

We have made approximate calculations of f and of the relative error

$$\delta = (f_{\text{approx}} - f)/f. \tag{120}$$

The lowest-order approximation with Wiener weighting was calculated for the harmonic oscillator by Feynman and Hibbs (Ref. 2, Sec. 10.3):

$$f_0^W = (2/\beta)\ln(\beta) + \beta/24. \tag{121}$$

Calculations of f_0^W for different values of β are given in Feynman and Hibbs (Ref. 2 p. 286). One can introduce Uhlenbeck–Ornstein weighting by splitting the potential into two parts. In nonreduced units,

$$V(y) = \tfrac{1}{2}m\omega^2 y^2 = \tfrac{1}{2}m\omega_1^2 y^2 + \tfrac{1}{2}m\omega_2^2 y^2. \tag{122}$$

The first part is treated exactly, contributing the Uhlenbeck–Ornstein weighting in the functional integrals; the second part of the potential is subjected to the approximation methods discussed in this paper. ($\omega_2 = \omega$ reduces to the case of Wiener weighting; $\omega_1 = \omega$ corresponds to the exact calculation for the harmonic oscillator.)

Since $V(y)$ is a quadratic function of y, it turns out that $W(\bar{x})$ is also a quadratic function of $\bar{x}$. From the discussion at the end of Sec. IIID we see that, for each term in the expansion (91), $S_n(\bar{x})$ involves only $W''(\bar{x})$ and higher derivatives, and thus each term in the moment expansion will be independent of $\bar{x}$. The $\bar{x}$ integration in the zeroth-order expression (103) can be performed exactly, and we can write the nth approximation to the harmonic oscillator as

$$\begin{aligned} Z_n &= Z_0\,(1 + S_2 + S_3 + \cdots + S_n), \\ f_n &= f_0 - (2/\beta)\ln(1 + S_2 + S_3 + \cdots + S_n). \end{aligned} \tag{123}$$

Using the results of Sec. IIID up to third order in the moment expansion, we obtain

$$f_0^{UO} = \frac{1}{\beta}\left\{\left(\frac{\beta_2}{\beta}\right)^2\left[\left(\frac{\beta_1}{2}\right)\coth\left(\frac{\beta_1}{2}\right) - 1\right] + \ln\left(\frac{2\beta^2[\cosh(\beta_1 - 1]}{\beta_1^2}\right)\right\}, \tag{124}$$

$$S_2^{UO} = \frac{1}{4}\left(\frac{\beta_2}{\beta}\right)^4 \frac{1}{\sinh^2(\beta_1/2)} \times \left[\frac{\beta_1^2}{8} - \sinh^2\left(\frac{\beta_1}{2}\right) + \frac{\beta_1}{8}\sinh(\beta_1)\right], \tag{125}$$

$$\begin{aligned} S_3^{UO} = -\frac{1}{12}\left(\frac{\beta_2}{\beta}\right)^6 \Bigg\{ & \frac{4\cosh(\beta_1/2)}{\beta_1\sinh^3(\beta_1/2)} \\ & \times \left[\frac{\beta_1^2}{8} - \sinh^2\left(\frac{\beta_1}{2}\right) + \frac{\beta_1}{8}\sinh\left(\frac{\beta_1}{2}\right)\right]^2 \\ & - \frac{8}{\beta_1^2}\left[\frac{\beta_1}{2}\coth\left(\frac{\beta_1}{2}\right) - 1\right]^2 \\ & \times \left[\frac{\beta_1^2}{8} - \sinh^2\left(\frac{\beta_1}{2}\right) + \frac{\beta_1}{8}\sinh\left(\frac{\beta_1}{2}\right)\right] \\ & + \frac{8}{\beta_1^2}\sinh^2\left(\frac{\beta_1}{2}\right)\left[\frac{\beta_1}{2}\coth\left(\frac{\beta_1}{2}\right) - 1\right]^3 \\ & - \frac{3\beta_1\cosh(\beta_1)\cdot\cosh(\beta_1/2)}{8\sinh(\beta_1/2)} \\ & + \frac{\sinh^3(\beta_1/2)}{\sinh(\beta_1/2)} - \frac{3}{\beta_1}\sinh(\beta_1) + \frac{3\beta_1^2}{16\sinh^2(\beta_1/2)}\Bigg\}, \end{aligned} \tag{126}$$

where

$$\beta_1 = \hbar\omega_1/kT, \quad \beta_2 = \hbar\omega_2/kT, \quad \beta_1^2 + \beta_2^2 = \beta^2. \tag{127}$$

The expressions (124)–(126) have been calculated and the corresponding errors S_1, S_2, and S_3 in the estimation of f are presented in Table I for different values of the quantum parameter β and for values of the ratio β_1/β between 0 (Wiener weighting) and 1 (exact calculation). From this table it is seen that the moment expansion provides an extremely accurate estimate of f, the accuracy increasing rapidly as $\beta \to 0$ (the approximation becomes poorer for $\beta \gtrsim 5$ as the higher terms become more significant). When applied to the harmonic potential, the approximation method of Fosdick[8] required the calculation of more than 3000 terms to match the accuracy obtained by going just to second order in the moment expansion.

In the spirit of this paper the numerical calculations have been restricted to the simplest possible case to which the formalism is applicable, that of a particle in a one-dimensional harmonic potential. This case has the advantage that it is one of the few cases where the appropriate functional integrals can be evaluated exactly and thus an estimate of the accuracy of the various approximations is possible. It has been shown that in this case the first few terms of the moment expansion give an extremely accurate estimate of the partition function and the corresponding free energy with much less labor than existing computational methods.

Finally, we refer to the possible advantage of using Uhlenbeck–Ornstein weighting without reservation of variables (see Sec. IIIB)—relative to Wiener weighting with reservation of $\bar{x}$. Obviously the former will give good results only for not too small harmonic content of the potential. This is because, in view of the non-normalizability of the cyclic path probability for the Wiener weighting, having two small a harmonic content means approaching too closely to this singular case, with undesirable results on the accuracy of the calculation. We have done some computations to compare these two methods. Typically, for $\beta \cong 3.5$, and by assuming the perturbation as usual to be harmonic, the Uhlenbeck–Ornstein weighting without reservation of variables in the lowest order becomes increasingly more accurate than the Wiener weighting with reservation of $\bar{x}$ once the ratio of perturbing potential to the harmonic potential included in the weighting function goes below 0.3. Insofar as this is a test of a real situation, say that of a truly anharmonic perturbation, this implies a perturbation in this same ratio to a presumed, harmonic component. When the much greater mathematical simplicity of the results without reservation of variables is taken into account, this might imply a substantial advantage. It should be noted that this method tends to work best in the quantum region.

CONCLUDING REMARKS

This paper is the first in a series which presents a new approach to the calculation of those functional integrals that appear in problems of quantum statistical mechanics and which approach, it is hoped, will lead to a powerful method of performing practical calculations. The present paper has mostly been restricted to a description of the mathematical basis of the approximations in the simplest case of a single particle in a one-dimensional potential. In later papers we shall consider more realistic situations—the extension to three-dimensional systems is simple and to many-particle systems difficult, because of the problem of statistics.

APPENDIX A

The one-dimensional Bloch–Schrödinger equation is

$$\frac{\partial \Psi}{\partial \theta} = \left(\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} - \mathcal{U}(y)\right)\Psi \equiv -\mathcal{H}\Psi, \tag{A1}$$

in which $\hbar$ is Planck's constant and m is the mass of a particle in a potential $\mathcal{U}$. The Bloch equation is (A1) as it stands, with $\theta = 1/kT$, whereas (A1) becomes the Schrödinger equation under the transformation $\theta = i\tau/\hbar$ (τ = time, T = absolute temperature, and k is Boltzmann's constant.) Wiener measure goes into Feynman measure under this latter transformation; while we speak throughout in the language of the Bloch equation and Wiener measure, our results are, at least formally, transformable into the Schrödinger–Feynman situation.

Equation (A1) goes into the dimensionless form (2) under the following substitutions:

$$x = y/l, \quad t = \hbar\omega\theta, \quad H(x) = \mathcal{H}(y(x))/\hbar\omega, \quad \psi(x,t) = \sqrt{l}\,\Psi(y(x),\theta), \tag{A2}$$

which make use of the basic length and energy parameters

$$\hbar\omega = \text{energy of the ground state of (A1)}, \tag{A3}$$

$$l = (\hbar/m\omega)^{1/2} = \text{width of the ground-state wavefunction of (A1)}.$$

APPENDIX B

In order to obtain the standard form of the Uhlenbeck–Ornstein distribution, we must first redefine ω of Eq. (A3) as the frequency associated with the harmonic part only of $\mathcal{U}(y)$; i.e., we assume

$$\mathcal{U}(y) = \frac{K}{2}\,y^2 + \frac{1}{3!}\,\mathcal{U}'''(0)\,y^3 + \cdots, \tag{B1}$$

and define

$$\omega = (K/m)^{1/2}, \quad l = (\hbar/2m\omega)^{1/2}. \tag{B2}$$

Although the definitions (B2) both differ from (A3), we still transform variables and functions according to (A2). This leads to the alternative dimensionless form (24) of the Bloch–Schrödinger equation, where the modified potential $U(x)$ is defined by

$$U(x) = \mathcal{U}(y(x))/\hbar\omega - x^2/4. \tag{B3}$$

---

* Work supported in part by the U.S. Air Force Office of Scientific Research, Grant No. AF-AFOSR 557-67.

† Present address: Materials Science Division, Army Materials and Mechanics Research Center, Watertown, Massachusetts.

[1] A preliminary account of this work was presented at the conference on Stochastic Processes in Chemical Physics, held at the University of California, San Diego, March, 1968. See T. Burke and A. Siegel, Advan. Chem. Phys. **15**, 355 (1969).

[2] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[3] I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. **1**, 48 (1960).

[4] M. Kac, *Probability and Related Topics in the Physical Sciences* (Interscience, New York, 1959).

[5] S. G. Brush, Rev. Mod. Phys. **33**, 79 (1961). This article contains comprehensive references.

[6] E. W. Montroll, Commun. Pure Appl. Math. **5**, 415 (1952).

[7] R. H. Cameron, Duke Math. J. **18**, 111 (1951).

[8] L. D. Fosdick, J. Math. Phys. **3**, 1251 (1962).

[9] D. Falkoff, Ann. Phys. (N.Y.) **4**, 325 (1958).

[10] A. Siegel, Phys. Rev. **106**, 609 (1957).

[11] N. Wiener, J. Math. & Phys. **2**, 131 (1923).

[12] R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain*, Amer. Math. Soc. Colloq. Pub. 19 (Amer. Math. Soc., New York, 1934).

[13] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930) [reprinted in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York, 1954)].

[14] In the present context, $U(x)$ is the nonquadratic part of the potential in "Wiener–Feynman" units (Appendix A), and therefore is to be distinguished from the $U(x)$ of Eq. (24), which is the nonquadratic part of the potential in "Uhlenbeck–Ornstein" units, denoted in the present section by $U'(x')$.

All Stationary Vacuum Metrics with Shearing Geodesic Eigenrays

J. Kóta and Z. Perjés
Central Research Institute for Physics, Hungarian Academy of Sciences, Budapest 114, Hungary
(Received 17 November 1970; Revised Manuscript Received 4 January 1971)

The general solution of the field equations of stationary vacuum gravitational fields possessing geodesic eigenrays with nonvanishing shear is obtained. Nontrivial solutions exist only if the eigenrays do not rotate. The resulting metrics fall into two classes: either there is a functional dependence among the field quantities (this class belongs to the Papapetrou solutions), or the quantity γ^0, which in the shear-free case has been interpreted as the central mass, is uniquely determined. This latter class consists of two space-times. The curvature invariants vanish in the $r \to \infty$ limit for both solutions; however, the metrics exhibit singular behavior in this limit.

1. INTRODUCTION

The class of gravitational fields possessing geodesic rays[1] is very conveniently treated by the spin coefficient technique developed by Newman and Penrose in 1962.[2] Newman and Tamburino[1] have shown how the metrics can be obtained in explicit form using spin coefficients. They calculated metrics for which the rays have nonvanishing shear, and, surprisingly, they learned that this class cannot be considered as a generalization from the nonshearing case. A later result of Unti and Torrence[3] indicated that the class of metrics with shearing geodesic rays is rather poor, in the sense that solutions exist only if the rays are either hypersurface orthogonal or cylindrical.

If the space–time contains a Killing vector field, the gravitational equations can be reformulated in a three-dimensional space V_3 associated with the trajectories of the Killing motion.[4,5] (For notations see Ref. 5. This paper will hereafter be referred to as P.) For stationary space–times (timelike Killing field) an $SU(2)$ spin coefficient method has been developed in P. The field equations in $SU(2)$ spinor base can be solved exactly if the eigenrays are geodesics of V_3. (The notion of eigenrays will be elucidated below.) As shown in P, the gravitational fields with nonshearing geodesic eigenrays are of Petrov type D; they have been thoroughly studied in previous papers. Therefore, we can anticipate new results only for metrics with eigenrays of nonvanishing shear.

In this paper all stationary vacuum metrics with shearing geodesic eigenrays are constructed in explicit form. The findings resemble in many respects those of Newman and Tamburino. The dropping of the nonshearing condition leads to a rather restricted set of metrics which does not contain the shear-free Kerr solution as a limiting case. Nevertheless, there are some solutions with shearing eigenrays. To demonstrate this, we first list some results of P.

Let the coordinate $x^0 = t$ be chosen as the arc of the trajectories of motion. The line element is then of the form

$$d\tilde{s}^2 = -f^{-1}ds^2 + f(dt + \omega_i dx^i)^2, \tag{1}$$

with all functions independent of t. ds $(ds = \sqrt{g_{ik}dx^i dx^k})$ stands for the line element of the three-dimensional background space V_3.

One can introduce in V_3 a complex basic vector "triad" $z^i_{\mathbf{p}} = (l^i, m^i, \bar{m}^i)$, $\mathbf{p} = 0, +, -$ with the orthogonality properties $l_i l^i = m_i \bar{m}^i = 1$, $l_i m^i = m_i m^i = 0$. The direction of the real unit vector l^i is conveniently fixed by the relation

$$G_+ \equiv G_i m^i = 0. \tag{2}$$

Here, G_i is a complex 3-vector determined by the gravitational field as follows:

$$G_i \overset{\text{DEF}}{=} \mathcal{E}_{,i}/2\mathrm{Re}\mathcal{E}, \quad \mathcal{E}_{,i} = f_{,i} + i\epsilon_{ijk}\omega^{,j}\sqrt{g}\, f^2. \tag{3}$$

We notice that the triad projections $G_+ = G_i m^i$ and $\bar{G}_- = \bar{G}_i \bar{m}^i$ are related by $(\bar{G}_+) = \bar{G}_-$, and we have a corresponding equation $(\bar{G}_-) = \bar{G}_+$.

Equation (2) defines a congruence of curves with the tangent vector l^i. The curves are called the eigenrays of the gravitational field. We now take the coordinate $x^1 = r$ to be the arc length of the eigenrays. Thus we have for the base vectors

$$l^i = \delta^i_1, \quad m^i = \omega\delta^i_1 + \xi^a\delta^i_a, \quad a = 2, 3. \tag{4}$$

The coordinate transformations

$$r' = r + r^0(x^a), \tag{5}$$

$$x^{a'} = x^{a'}(x^b), \quad b = 2, 3, \tag{6}$$

are still permissible.

The quantity[6] $\epsilon = m_{i;j}\bar{m}^i l^j$ is made zero by a complex rotation of m^i. There is still a freedom in the choice of the triad:

$$l'^i = l^i, \quad m'^i = e^{iC^0}m^i. \tag{7}$$

Here, C^0 is an arbitrary real function of the coordinates x^a.

Geodesic eigenrays are characterized by vanishing of the complex rotation coefficient $\kappa = m_{i;j}l^i l^j$. The field equations for geodesic eigenrays are taken from P:

$$D\omega = \bar{\rho}\omega + \sigma\bar{\omega}, \tag{8a}$$

$$D\xi^a = \bar{\rho}\xi^a + \sigma\bar{\xi}^a, \tag{8b}$$

$$D\sigma = (\rho + \bar{\rho})\sigma, \tag{8c}$$

$$D\rho = \rho^2 + \sigma\bar{\sigma} + G_0\bar{G}_0, \tag{8d}$$

$$DG_0 = (2\rho + G_0 - \bar{G}_0)G_0, \tag{8e}$$

$$D\tau = \rho\tau - \bar{\sigma}\bar{\tau} + \bar{G}_0 G_-, \tag{8f}$$

$$\delta\bar{\omega} - \bar{\delta}\omega = \bar{\tau}\bar{\omega} - \tau\omega + \bar{\rho} - \rho, \tag{8g}$$

$$\delta\bar{\xi}^a - \bar{\delta}\xi^a = \bar{\tau}\bar{\xi}^a - \tau\xi^a, \tag{8h}$$

$$\delta\rho - \bar{\delta}\sigma = -2\sigma\tau + \bar{G}_+ G_0, \tag{8i}$$

$$\delta\tau + \bar{\delta}\bar{\tau} = \rho\bar{\rho} - \sigma\bar{\sigma} + 2\tau\bar{\tau} - G_0\bar{G}_0 + G_-\bar{G}_+, \tag{8j}$$

$$\bar{\delta}G_0 - DG_- = -\rho G_- + \bar{G}_0 G_-, \tag{8k}$$

$$\delta G_0 = -\sigma G_- - G_0\bar{G}_+, \tag{8l}$$

$$\delta G_- = (\bar{\rho} - \rho)G_0 + \bar{\tau}G_- - G_-\bar{G}_+. \tag{8m}$$

In our coordinate system the scalar differential operators appearing in the field equations are of the form $D = \partial/\partial r$, $\delta = \omega\partial/\partial r + \xi^a\partial/\partial x^a$. From Eq. (8c) it follows that the phase of the complex shear σ is independent of r. Therefore, by the triad freedom (7) we make σ real and positive. Thus the triad is completely fixed.

2. THEOREM ON SHEARING GEODESIC EIGENRAYS

When trying to integrate the field equations (8) for nonvanishing σ one is led to many separate cases, most of which do not contain any solution since the calculation ends at some prohibitive relation. It is desirable to recognize such cases directly, without the lengthy integration procedure, from the field equations. Information can be gained immediately from the field equations by an operation used by Newman and Penrose in their proof of the Goldberg–Sachs theorem[2]: One takes appropriately chosen derivatives of the Newman–Penrose equations and eliminates second-order terms by the commutators of the scalar differential operators. Consecutive effectuation of a procedure of this kind provides the proof of our main theorem[7]:

Theorem: Geodesic eigenrays in a curved vacuum stationary space-time cannot have coexisting shear and curl. If the eigenrays do shear ($\sigma \neq 0$), then they also diverge ($\rho + \bar{\rho} \neq 0$), and one has $\delta\sigma = \delta|G_0| = \rho - \bar{\rho} = 0$ and

$$\rho\bar{\rho} - \sigma\bar{\sigma} - G_0\bar{G}_0 = 0. \tag{9}$$

The proof takes a more concise form by the use of the operators

$$\delta_\pm \overset{\text{DEF}}{=} R(\delta \pm i\bar{\delta}). \tag{10}$$

Here, R is the "luminosity distance" satisfying[2] $DR = -[(\rho + \bar{\rho})/2]R$. By definition (10) the operators D and $\delta_\pm$ commute as follows:

$$D\delta_\pm - \delta_\pm D = -i(a \mp \sigma)\delta_\mp, \quad a \overset{\text{DEF}}{=} \operatorname{Im}\rho. \tag{11}$$

We now observe that the only pair of field equations from which new first-order relations can be obtained is (8e) and (8l). It is easy to show for $\sigma \neq 0$ that, except in singular points, G_0 is nonzero; for the assumption $G_0 \equiv 0$, by Eq. (8l) leads to $G_+ = G_- = G_0 = 0$ and thus the curvature invariants Ψ_A vanish [see Eqs. (81) in P]. The above formulation of the theorem, however, excludes the case $\Psi_A = 0$ (the flat space).

Thus we may write, after dividing by G_0,

$$D\ \ln G_0 = 2\rho + G_0 - \bar{G}_0, \tag{8e'}$$

$$-\delta\ \ln G_0 = \bar{G}_+ + \sigma G_-/G_0. \tag{8l'}$$

Taking the mixed derivatives and summing the equations thus obtained, we can eliminate the second-order terms by use of the identity $(D\delta - \delta D)\ \ln G_0 = (\bar{\rho}\delta - \sigma\bar{\delta})\ \ln G_0$ [cf. (72a) in P]. The first-order terms are substituted from the field equations (8) such that, as a result, we have

$$\sigma(\bar{\delta}\ \ln G_0 + \bar{\delta}\ \ln\sigma - G_- - 2\tau) = 0. \tag{12}$$

Considering the case when $\sigma \neq 0$, we have to prove that $2ia = \rho - \bar{\rho} = 0$ or, which is equivalent, that there exists a coordinate system in which $\omega = 0$.

Acting on Eq. (12) with the D operator and subtracting the δ derivatives of appropriate radial equations such that the second-order terms are canceled again, we obtain

$$3\gamma^2\bar{\delta}\rho + (\gamma^2 + \sigma^2)\delta\sigma + (\gamma^2 - \sigma^2)(\bar{\delta}\bar{\rho} + 2\bar{\sigma}\bar{\tau}) = 0, \qquad \gamma \overset{\text{DEF}}{=} |G_0|. \tag{13}$$

Equation (13) can be made homogeneous by substitution of (8i):

$$\gamma(3\bar{\delta}\rho + 2\delta\sigma + \bar{\delta}\bar{\rho}) + 2\sigma\delta\gamma = 0. \tag{13'}$$

In the $\delta_\pm$ notation, (13') can easily be split up into components which are mutually orthogonal in the complex plane:

$$\gamma\delta_\pm(\rho + \bar{\rho}) - i\gamma\delta_\mp a \pm i\delta_\mp(\sigma\gamma) = 0. \tag{14}$$

Denoting $\Delta^2 = \sigma^2 + \gamma^2 - a^2$, we obtain from the commutators (11) and from (8)

$$[D - (\rho + \bar{\rho})]\,\delta_\pm(\rho + \bar{\rho}) = 4\Delta\delta_\pm\Delta - i(a \mp \sigma)\delta_\mp(\rho + \bar{\rho}),$$

$$[D - (\rho + \bar{\rho})]\delta_\pm\mathbf{v} = \mathbf{v}\,\delta_\pm(\rho + \bar{\rho}) - i(a \mp \sigma)\delta_\mp\ \mathbf{v} = \begin{pmatrix}\Delta\\ a\\ \sigma\\ \gamma\end{pmatrix} \tag{15}$$

Repeated application of the operator $D - (\rho + \rho)$ on Eq. (14) and use of (15) yields the following series of first-order equations:

$$(3a \pm 2\sigma)\delta_\pm(\rho + \bar{\rho}) + 4i\Delta\delta_\mp\Delta - 2i(a \mp \sigma)\delta_\mp a = 0, \tag{16}$$

$$[(2a \pm 4\sigma)(a \mp \sigma) - 4\Delta^2]\delta_\mp(\rho + \bar{\rho}) + i(3a \pm 4\sigma)4\Delta\delta_\pm\Delta = 0, \tag{17}$$

$$[a(a^2 - \sigma^2) - \Delta^2(5a \mp 4\sigma)]\delta_\mp(\rho + \bar{\rho}) + i[(2a \mp \sigma)(a \pm \sigma) - \Delta^2]4\Delta\delta_\pm\Delta = 0. \tag{18}$$

(17) and (18) are homogeneous in $\delta_\pm(\rho + \bar{\rho})$ and $\delta_\pm\Delta$ with the determinants

$$D_{\substack{u\\l}} = 4\Delta^4 - \Delta^2(8\sigma^2 - 5a^2 \mp 2a\sigma) + (a^2 - \sigma^2)(a^2 - 4\sigma^2 \pm 2a\sigma). \tag{19}$$

Here u and l label the determinants of the equations with the upper and lower signs, respectively. The simultaneous vanishing of both D_u and D_l means $\gamma = 0$ (flat space). We may still have either of the determinants vanishing. Let us consider, for example, the case where

$$D_u = 0 \tag{20}$$

and

$$\delta_+(\rho + \bar{\rho}) = \delta_-\Delta = 0. \tag{21}$$

Operating again with $D-(\rho+\bar{\rho})$ and using (15), from Eqs. (21) we derive the first-order relations

$$4\Delta\delta_+\Delta - i(a-\sigma)\delta_-(\rho+\bar{\rho}) = 0,$$
$$\Delta\delta_-(\rho+\bar{\rho}) - i(a+\sigma)\delta_+\Delta = 0. \tag{22}$$

Hence we immediately have

$$\delta_+\Delta = \delta_-(\rho+\bar{\rho}) = 0 \tag{23}$$

if the determinant of Eqs. (22) is different from zero. The condition

$$\det \equiv 4\Delta^2 + a^2 - \sigma^2 = 0, \tag{24}$$

when substituted back into (17), leads to the same result.

In a similar fashion, for the alternate case $D_l = \delta_-(\rho+\bar{\rho}) = \delta_+\Delta = 0$, we obtain

$$\delta_-\Delta = \delta_+(\rho+\bar{\rho}) = 0. \tag{23'}$$

We thus conclude that for $\gamma \neq 0$ Eqs. (17) and (18) always yield

$$\delta_+(\rho+\bar{\rho}) = \delta_-(\rho+\bar{\rho}) = \delta_+\Delta = \delta_-\Delta = 0 \tag{25}$$

or, in terms of the δ operator,

$$\delta(\rho+\bar{\rho}) = \delta\Delta = 0. \tag{26}$$

Assuming for the moment that $\rho+\bar{\rho}$ is nonzero, the commutator [Eq. (72b) of P]

$$\delta\bar{\delta} - \bar{\delta}\delta = \bar{\tau}\bar{\delta} - \tau\delta + (\bar{\rho}-\rho)D, \tag{27}$$

when acting on $\rho+\bar{\rho}$ and Δ, gives $a=0$. We now show that $\rho+\bar{\rho}$ actually cannot vanish. Equations (25) together with (14) and (16) ensure that $\delta\sigma = \delta\gamma = \delta a = 0$.

Let us denote the phase of G_0 by χ. Our starting relation (12) is then written in the form

$$i\bar{\delta}\chi - G_- = 2\tau. \tag{28}$$

Letting the operator δ act on (28), taking the real part and subtracting (8j), we get

$$(\rho-\bar{\rho})^2 + 2(\rho\bar{\rho} - \gamma^2 - \sigma^2) = 0. \tag{29}$$

This equation, when compared with (8d), tells us that $\rho+\bar{\rho}$ is nonvanishing and that we have $\rho\bar{\rho} - \gamma^2 - \sigma^2 = 0$. This completes the proof of our central theorem.

The "radial" equations (8a)–(8e) containing the operator D are now readily integrated to yield

$$\sigma/\sigma^0 = \gamma/\gamma^0 = -\rho = 1/2r, \tag{30}$$

$$G_0 = -(\gamma^0/2r)(r\gamma^0 - iQ)/(r\gamma^0 + iQ), \tag{31}$$

$$\xi^a = (1/\sqrt{2r})(A^a r^{\sigma^0/2} + iB^a r^{-\sigma^0/2}). \tag{32}$$

Here Q, A^a, B^a are real integration "constants", depending only on $x^2 = x$ and $x^3 = y$; σ^0 and γ^0 are positive numbers subject to

$$\sigma^{02} + \gamma^{02} = 1. \tag{33}$$

The coordinate freedom (5) has been used in (30) to fix the origin of r. From $\delta\rho = 0$ we get $\omega = 0$ in this coordinate system. The r dependence of the complex scalar quantity $\mathcal{E}$ appearing in (3) can be obtained from the definition of G_0 (see P):

$$G_0 = D\mathcal{E}/2\mathrm{Re}\mathcal{E}. \tag{34}$$

Now G_0 is given by (31), and thus $\mathcal{E}$ takes the form

$$\mathcal{E} \equiv f + i\varphi = [f^0/(r\gamma^0 + iQ)] + i\varphi^0, \tag{35}$$

with f^0, φ^0 real functions of x, y.

We are now faced with the following remnants of the original field equations (8):

$$\delta\mathcal{E} = 0, \tag{36a}$$

$$\mathrm{Im}[(\delta - \bar{\tau})\bar{\xi}^a] = 0, \tag{36b}$$

$$2\sigma\tau = \bar{G}_+ G_0, \tag{36c}$$

$$\tau(\sigma^2 - \gamma^2) = 0. \tag{36d}$$

In accordance with Eq. (36d), the metrics split up into classes with either $\tau = 0$ or $\gamma = \sigma$. This bifurcate logics of the field equations must be dealt with by treating both of the classes separately in the following sections.

3. METRICS WITH $\tau = 0$

For this class from Eq. (36c) we get $\bar{G}_+ = 0$ or $\bar{\delta}\mathcal{E} = 0$. Thence $\varphi^0 = 0$ and the quantities f^0, Q are in fact constants. This gives rise to a functional relationship between the quantities f and φ.

On substituting the expression (32) for ξ^a into $\mathrm{Im}\delta\bar{\xi}^a = 0$ [Eq. (36b)] we find that the operators $A \overset{\mathrm{DEF}}{=} A^a\partial_a$ and $B \overset{\mathrm{DEF}}{=} B^a\partial_a$ commute:

$$[A, B] = 0. \tag{37}$$

The only coordinate freedom in V_3 is now (6). The quantities A^a, B^a behave as two-component vectors under the transformations (6). Since at regular space–time points A^a and B^a are linearly independent, we can make them tangential to the coordinate curves x and y, respectively:

$$A^a = \delta^a_2, \qquad B^a = \delta^a_3. \tag{38}$$

With this choice of the coordinates we have

$$\xi^2 = (1/\sqrt{2r})r^{\sigma^0/2}, \qquad \xi^3 = (i/\sqrt{2r})r^{-\sigma^0/2}. \tag{39}$$

The 3-vector ω_i in the line element (1) will now be evaluated. The relation required at this point is taken from P:

$$\omega_{i,j} - \omega_{j,i} = \epsilon_{ijk}\varphi^{;k}\sqrt{g}\, f^{-2}. \tag{40}$$

The coordinate t can be shifted without disturbing (1): $t' = t + F(r, x^a)$. By this transformation we are able to put $\omega_r = 0$, and we still have

$$t' = t + t^0(x, y). \tag{41}$$

Equation (40) takes the form

$$\omega_{x,r} = \omega_{y,r} = 0, \quad \omega_{x,y} - \omega_{y,x} = 2\gamma^0 Q/f^0. \tag{42}$$

From (41) and (42) we obtain

$$\omega_i = (0, 0, -2(\gamma^0 Q/f^0)x). \tag{43}$$

With the aid of formulas (4) and (39) we may calculate the background metric $g_{ij} = l_i l_j + m_i \bar{m}_j + \bar{m}_i m_j$. Consequently, Eqs. (35) and (43) yield the full four-dimensional line element (1) for the class with $\tau = 0$:

$$d\tilde{s}^2 = -\frac{r^{2\gamma^0} + Q^2}{f^0 r^{\gamma^0}}(dr^2 + r^{1-\sigma^0}dx^2 + r^{1+\sigma^0}dy^2) + \frac{f^0 r^{\gamma^0}}{r^{2\gamma^0} + Q^2}\left(dt - 2\frac{\gamma^0 x}{f^0}Q dy\right)^2. \tag{44}$$

This space-time is stationary and axially symmetric. In addition, as observed above, the invariant f is a function of φ. Therefore the line element (44) represents a particular Papapetrou solution.[8]

4. METRICS WITH $\sigma = \gamma$

Equation (33) precludes the flat space limit by determining uniquely the constants σ^0 and γ^0:

$$\sigma^0 = \gamma^0 = 1/\sqrt{2}. \tag{45}$$

The remaining field equations (36a)–(36c) can be worked out after calculating the quantity $G_- = \bar{\delta} \ln f$. (36c) immediately yields τ. From (36a), (36b) we get

$$\delta\mathcal{E} = 0 \rightarrow \begin{cases} A\varphi^0 = 0, & AQ = (B - QA)\ln f^0, \\ B\varphi^0 = Af^0, & BQ = Q(B - QA)\ln f^0, \end{cases} \tag{46a}$$

$$\mathrm{Im}[(\delta - \bar{\tau})\xi^a] = 0 \rightarrow 2[A, B] = 2Q(A \ln f^0)A - (A \ln f^0)B - (B \ln f^0)A. \tag{46b}$$

The problem becomes somewhat simpler if instead of A and B we use the operators

$$\alpha = \sqrt{f^0}A, \quad \beta = \sqrt{f^0}(B - QA). \tag{47}$$

In terms of the α and β operators, taking proper linear combinations of the field Eqs. (46), we can write

$$[\alpha, \beta] = -(\alpha Q)\alpha, \tag{48a}$$

$$\beta Q = 0, \tag{48b}$$

$$\alpha Q = \beta \ln f^0, \tag{48c}$$

$$\alpha f^0 = \beta \varphi^0, \tag{48d}$$

$$\alpha \varphi^0 = 0. \tag{48e}$$

Letting the commutator (48a) act on Q, and taking account of (48b), (48c), we find that

$$\beta(f^{0^{-2}}\beta f^0) = 0. \tag{49}$$

This relation, when compared with (48b), tells us that $f^{0^{-2}}\beta f^0$ is a functional of Q if Q is not constant. But let us consider first the case when Q is constant. For such metrics the operators α and β commute; therefore, the coordinates can be chosen to have

$$\alpha = \partial_2, \quad \beta = \partial_3 - Q\partial_2. \tag{50}$$

Equations (48c)–(48e) are easily integrated to yield

$$f^0 = P(x + Qy), \quad \varphi^0 = Py. \tag{51}$$

Here, P is a constant of integration and the origin of coordinates has been shifted to make the constant terms vanish. There exists another solution with f^0 and φ^0 constant, but this latter metric has vanishing τ and thus belongs to the class which has been discussed in Sec. 3.

The most important field quantities obtained by use of (51) are

$$\mathcal{E} \equiv f + i\varphi = P(x + ir\gamma^0 y)/(r\gamma^0 + iQ),$$
$$\xi^2 = (2rf^0)^{-1/2} r^{\gamma^0/2}, \quad \xi^3 = i(2rf^0)^{-1/2} r^{-\gamma^0/2},$$
$$g_{ij} = \begin{bmatrix} 1 & & \\ & f^0 r^{1-\gamma^0} & \\ & & f^0 r^{1+\gamma^0} \end{bmatrix}. \tag{52}$$

The calculation of the four-dimensional line element terminates with the evaluation of ω_i, using formula (35). As a result we have

$$d\tilde{s}^2 = -(f^0/f)(r^{1-\gamma^0}dx^2 + r^{1+\gamma^0}dy^2) + 2dr(dt - 2\gamma^0 Qy\,dx) + f(dt - 2\gamma^0 Qy\,dx)^2. \tag{53}$$

A glance at the line element convinces us that for Q vanishing, $\partial/\partial y$ is a Killing vector. Investigation of the curvature invariants Ψ_A (cf. P) shows that this space-time has true singularity at $r = 0$ and becomes flat in the limit $r \to \infty$, $f < \infty$. However, the behavior of the metric is rather awkward; it remains regular for $r \to \infty$ only if x or y also goes properly to infinity.

Consider now the case when the quantity Q does depend on the coordinates. We want to integrate the simultaneous equations (48) without committing ourselves to any particular coordinate system. According to Eq. (49), we may write

$$f^{0^{-2}}\beta f^0 = q(Q). \tag{54}$$

Next we act with (48a) on f^0:

$$\beta\{f^{0^{-3}}[\alpha f^0 + (q'/q)\beta f^0]\} = 0 \rightarrow \alpha f^0 = qp(Q)f^{0^3} - (q'/q)\beta f^0. \tag{55}$$

(Prime stands for d/dQ.) From (55) and (48d) it follows that

$$\beta[\varphi^0 - \tfrac{1}{2}pf^{0^2} + (q'/q)f^0] = 0 \rightarrow \varphi^0 = \tfrac{1}{2}pf^{0^2} - (q'/q)f^0 + s(Q). \tag{56}$$

Here, the functionals p, q, and s are arbitrary for the moment.

The latter expression for φ^0 is now placed in Eq. (48e):

$$[\tfrac{1}{2}p'f^{0^2} - (q'/q)'f^0 + s' + (pf^0 - q'/q)^2 f^0]\beta f^0 = 0. \tag{57}$$

(48c) shows that βf^0 cannot vanish. This condition means that f^0 still varies once Q is fixed. Therefore, the parenthesized quantity of eq. (57) is equal to zero

and can be regarded a polynomial equation in f^0. We make the coefficients of the various f^0 powers vanish:

$$p = 0, \quad s' = 0, \quad (q'/q)' - (q'/q)^2 = 0. \tag{58}$$

It is possible to set $s = 0$ because φ is defined only up to an arbitrary constant term. Straightforward integration yields

$$q = 1/(aQ + b), \tag{59}$$

with a, b real constants. We are now in position to put down the relation connecting f^0, Q, and φ[cf. Eq. (56)]

$$\varphi^0 = [a/(aQ + b)]f^0. \tag{60}$$

Upon the above considerations, the quantities f^0 and Q are appropriate candidates for independent coordinates. Nevertheless, some simplification of the final results is achieved by the following choice of the coordinate system:

$$Q = x/y, \quad f^0 = (ax + by)/y^2. \tag{61}$$

Calculation of the metric can be performed in a similar way to that used in previous examples. The results are summarized in the following:

$$\begin{aligned} &\varphi^0 = a/y, \quad f = (ax + by)/(x^2\gamma^{-\gamma^0} + y^2\gamma\gamma^0), \\ &d\tilde{s}^2 = -(f^0/f)(r^{1-\gamma^0}dx^2 + r^{1+\gamma^0}dy^2) \\ &\quad + 2dr[dt - \gamma^0(x^2/y)dy] + f[dt - \gamma^0(x^2/y)dy]^2. \end{aligned} \tag{62}$$

The curvature invariants vanish for $r \to \infty$ except in such directions in which f becomes unbounded. Curvature singularities exist at $r = 0$, $y = 0$, $ax + by = 0$, and in the exceptional $r \to \infty$ limit.

1 E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962).
2 E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
3 T. W. Unti and R. J. Torrence, J. Math. Phys. **7**, 535 (1966).
4 Z. Perjés, Commun. Math. Phys. **12**, 275 (1969).
5 Z. Perjés, J. Math. Phys. **11**, 3383 (1970).
6 The $SU(2)$ spin coefficients which we use here should not be confused with the corresponding $SL(2, C)$ ones. We distinguish the latter by a wavy line.[5]

7 We would like to thank B. Kent Harrison for helpful comments and suggestions concerning this theorem and many other results presented in this paper.

8 A. Papapetrou, Ann. Physik **12**, 309 (1953).

Scalar-Tensor Theory of Gravitation in a Lyra Manifold

W. D. Halford
Mathematics Department, Massey University, Palmerston North, New Zealand

(Received 7 June 1971)

A closed-form exact solution to the field equations of a scalar–tensor theory, formally similar to the Brans–Dicke theory, is obtained. It is shown that the present theory predicts the same effects, within observational limits, as the Einstein theory.

1. INTRODUCTION

Recently[1] a new scalar–tensor theory of gravitation, based on a modified Riemannian manifold,[2] was proposed. This theory may be regarded formally as a special case of the Brans–Dicke theory,[3] but is significantly different from the latter in that the scalar field is characterized by the function $x^0 = x^0(x^\alpha)$, where the x^α are coordinates in the four-dimensional Lyra manifold, and the tensor field is identified with the metric tensor $g_{\alpha\beta}$ of the manifold.

The field equations given by Sen and Dunn[1] for the combined scalar and tensor fields are

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega(x^0)^{-2}x^{0,\alpha}x^{0,\beta} + \tfrac{1}{2}\omega(x^0)^{-2}g^{\alpha\beta}x^{0,\nu}x^0{}_{,\nu} \\ = -8\pi G(x^0)^{-2}T^{\alpha\beta}, \end{aligned} \tag{1.1}$$

where $\omega = \frac{3}{2}$, $T_{\alpha\beta}$ is the energy-momentum tensor of the field, and R is the usual Riemann curvature scalar. It was pointed out that these equations are identical with the Brans–Dicke equations *viz.*

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega\phi^{-2}\phi^{,\alpha}\phi^{,\beta} + \tfrac{1}{2}\omega\phi^{-2}g^{\alpha\beta}\phi^{,\nu}\phi_{,\nu} \\ = -8\pi\phi^{-1}T^{\alpha\beta} + \phi^{-1}(\phi^{,\alpha;\beta} - g^{\alpha\beta}\Box\phi) \end{aligned} \tag{1.2}$$

if the scalar function ϕ satisfied the condition

$$\phi_{,\alpha;\beta} - g_{\alpha\beta}\Box\phi = 0 \tag{1.3}$$

and $\omega = \frac{3}{2}$. It should be added that the gravitational "constant" must be redefined as well. Furthermore, Sen and Dunn gave only a series-type solution to the static vacuum field equations.

In this paper I shall enlarge upon the discussion presented by Sen and Dunn. Specifically, an exact solution to the static vacuum field equations is obtained in closed form and the equations of motion of a test particle in the vicinity of a point mass are discussed. From these and from the linearized form of the field equations, it becomes clear that the observable predictions of the present scalar–tensor theory, at least to second order in small quantities, are identical with those of the Einstein theory in a Riemann manifold.

2. STATIC SPHERICALLY SYMMETRIC FIELD

This section details the general solution to the vacuum field equations of the scalar–tensor theory in the Lyra manifold in the static spherically symmetric case. The solution appears in closed form and reduces to the Schwarzschild solution as a special case.

The field equations in the matter-free region surrounding a point mass are

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega(x^0)^{-2}x^{0,\alpha}x^{0,\beta} \\ + \tfrac{1}{2}\omega(x^0)^{-2}g^{\alpha\beta}x^{0,\nu}x^0{}_{,\nu} = 0, \end{aligned} \tag{2.1}$$

where $\omega = \frac{3}{2}$. We shall assume an isotropic metric

and can be regarded a polynomial equation in f^0. We make the coefficients of the various f^0 powers vanish:

$$p = 0, \quad s' = 0, \quad (q'/q)' - (q'/q)^2 = 0. \tag{58}$$

It is possible to set $s = 0$ because φ is defined only up to an arbitrary constant term. Straightforward integration yields

$$q = 1/(aQ + b), \tag{59}$$

with a, b real constants. We are now in position to put down the relation connecting f^0, Q, and φ[cf. Eq. (56)]

$$\varphi^0 = [a/(aQ + b)]f^0. \tag{60}$$

Upon the above considerations, the quantities f^0 and Q are appropriate candidates for independent coordinates. Nevertheless, some simplification of the final results is achieved by the following choice of the coordinate system:

$$Q = x/y, \quad f^0 = (ax + by)/y^2. \tag{61}$$

Calculation of the metric can be performed in a similar way to that used in previous examples. The results are summarized in the following:

$$\begin{aligned} &\varphi^0 = a/y, \quad f = (ax + by)/(x^2\gamma^{-\gamma^0} + y^2\gamma\gamma^0), \\ &d\tilde{s}^2 = -(f^0/f)(r^{1-\gamma^0}dx^2 + r^{1+\gamma^0}dy^2) \\ &\quad + 2dr[dt - \gamma^0(x^2/y)dy] + f[dt - \gamma^0(x^2/y)dy]^2. \end{aligned} \tag{62}$$

The curvature invariants vanish for $r \to \infty$ except in such directions in which f becomes unbounded. Curvature singularities exist at $r = 0$, $y = 0$, $ax + by = 0$, and in the exceptional $r \to \infty$ limit.

1 E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962).
2 E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
3 T. W. Unti and R. J. Torrence, J. Math. Phys. **7**, 535 (1966).
4 Z. Perjés, Commun. Math. Phys. **12**, 275 (1969).
5 Z. Perjés, J. Math. Phys. **11**, 3383 (1970).
6 The $SU(2)$ spin coefficients which we use here should not be confused with the corresponding $SL(2, C)$ ones. We distinguish the latter by a wavy line.[5]

7 We would like to thank B. Kent Harrison for helpful comments and suggestions concerning this theorem and many other results presented in this paper.

8 A. Papapetrou, Ann. Physik **12**, 309 (1953).

Scalar-Tensor Theory of Gravitation in a Lyra Manifold

W. D. Halford
Mathematics Department, Massey University, Palmerston North, New Zealand

(Received 7 June 1971)

A closed-form exact solution to the field equations of a scalar–tensor theory, formally similar to the Brans–Dicke theory, is obtained. It is shown that the present theory predicts the same effects, within observational limits, as the Einstein theory.

1. INTRODUCTION

Recently[1] a new scalar–tensor theory of gravitation, based on a modified Riemannian manifold,[2] was proposed. This theory may be regarded formally as a special case of the Brans–Dicke theory,[3] but is significantly different from the latter in that the scalar field is characterized by the function $x^0 = x^0(x^\alpha)$, where the x^α are coordinates in the four-dimensional Lyra manifold, and the tensor field is identified with the metric tensor $g_{\alpha\beta}$ of the manifold.

The field equations given by Sen and Dunn[1] for the combined scalar and tensor fields are

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega(x^0)^{-2}x^{0,\alpha}x^{0,\beta} + \tfrac{1}{2}\omega(x^0)^{-2}g^{\alpha\beta}x^{0,\nu}x^0{}_{,\nu} \\ = -8\pi G(x^0)^{-2}T^{\alpha\beta}, \end{aligned} \tag{1.1}$$

where $\omega = \frac{3}{2}$, $T_{\alpha\beta}$ is the energy-momentum tensor of the field, and R is the usual Riemann curvature scalar. It was pointed out that these equations are identical with the Brans–Dicke equations *viz.*

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega\phi^{-2}\phi^{,\alpha}\phi^{,\beta} + \tfrac{1}{2}\omega\phi^{-2}g^{\alpha\beta}\phi^{,\nu}\phi_{,\nu} \\ = -8\pi\phi^{-1}T^{\alpha\beta} + \phi^{-1}(\phi^{,\alpha;\beta} - g^{\alpha\beta}\Box\phi) \end{aligned} \tag{1.2}$$

if the scalar function ϕ satisfied the condition

$$\phi_{,\alpha;\beta} - g_{\alpha\beta}\Box\phi = 0 \tag{1.3}$$

and $\omega = \frac{3}{2}$. It should be added that the gravitational "constant" must be redefined as well. Furthermore, Sen and Dunn gave only a series-type solution to the static vacuum field equations.

In this paper I shall enlarge upon the discussion presented by Sen and Dunn. Specifically, an exact solution to the static vacuum field equations is obtained in closed form and the equations of motion of a test particle in the vicinity of a point mass are discussed. From these and from the linearized form of the field equations, it becomes clear that the observable predictions of the present scalar–tensor theory, at least to second order in small quantities, are identical with those of the Einstein theory in a Riemann manifold.

2. STATIC SPHERICALLY SYMMETRIC FIELD

This section details the general solution to the vacuum field equations of the scalar–tensor theory in the Lyra manifold in the static spherically symmetric case. The solution appears in closed form and reduces to the Schwarzschild solution as a special case.

The field equations in the matter-free region surrounding a point mass are

$$\begin{aligned} R^{\alpha\beta} - \tfrac{1}{2}g^{\alpha\beta}R - \omega(x^0)^{-2}x^{0,\alpha}x^{0,\beta} \\ + \tfrac{1}{2}\omega(x^0)^{-2}g^{\alpha\beta}x^{0,\nu}x^0{}_{,\nu} = 0, \end{aligned} \tag{2.1}$$

where $\omega = \frac{3}{2}$. We shall assume an isotropic metric

$$ds^2 = -e^{2p}dt^2 + e^{2q}[dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)], \tag{2.2}$$

where $p = p(r)$, $q = q(r)$ and $x^1 = r$, $x^2 = \theta$, $x^3 = \phi$ are spherical coordinates and $x^4 = t$. In the spherically symmetric field $x^0{}_{,2} = x^0{}_{,3} = x^0{}_{,4} = 0$ and Eqs. (2.1), with the metric (2.2) reduce to the three independent equations

$$2p' + 2q' + 2rp'q' + r(q')^2 + r(h')^2 = 0, \tag{2.3}$$

$$rp'' + rq'' + r(p')^2 + p' + q' - r(h')^2 = 0, \tag{2.4}$$

$$2rq'' + r(q')^2 + 4q' - r(h')^2 = 0, \tag{2.5}$$

where we have put

$$x^0 = e^{h\sqrt{2/\omega}}. \tag{2.6}$$

From (2.3), (2.4), and (2.5) we obtain the equations

$$rp'' + r(p')^2 + 2p' + rp'q' = 0, \tag{2.7}$$

$$rp'' - rq'' + r(p')^2 - r(q')^2 + p' - 3q' = 0. \tag{2.8}$$

The successive substitutions

$$p = \log v, \quad v > 0, \tag{2.9}$$

$$v' = x \tag{2.10}$$

put (2.7) in the form

$$(x'/x) + q' + (2/r) = 0, \tag{2.11}$$

provided $x \neq 0$; this is guaranteed since we do not require $p = \text{const}$. Equation (2.11) possesses the integral

$$v' = x = k_1 e^{-q}/r^2, \tag{2.12}$$

where k_1 is an arbitrary constant.

The substitution

$$q = \log w, \quad w > 0, \tag{2.13}$$

together with (2.9), brings (2.8) into the form

$$v''/v + v'/rv - (w''/w + 3w'/rw) = 0. \tag{2.14}$$

Now (2.7) and (2.9) yield

$$v''/v + 2v'/rv + v'q'/v = 0$$

or, using (2.13),

$$v''/v + 2v'/rv + (v'/v)(w'/w) = 0. \tag{2.15}$$

Equations (2.14) and (2.15) yield

$$wv'/r + v'w' + v(w'' + 3w'/r) = 0$$

or

$$r^2wv' + d(r^3vw')/dr = 0. \tag{2.16}$$

Now $r^2wv' = r^2v'e^q = k_1$ using (2.12) and (2.13), so that (2.16) gives upon integration

$$r^3vw' = -k_1r + k_2, \tag{2.17}$$

where k_2 is an arbitrary constant. This result may be written, with the aid of definitions (2.9) and (2.13), in the form

$$r^3q'e^pe^q = -k_1r + k_2. \tag{2.18}$$

Likewise (2.12) implies

$$r^3p'e^pe^q = k_1r. \tag{2.19}$$

From (2.18) and (2.19) we then have

$$q' = -p'(1 - k_3/r), \tag{2.20}$$

where $k_3 = k_2/k_1$ is an arbitrary constant.

Via (2.20) Eq. (2.7) becomes

$$p'' + (2/r)p' + (k_3/r)(p')^2 = 0, \tag{2.21}$$

which is a Riccati-type equation. By means of the substitution

$$p' = rs'/k_3s, \quad s = s(r), \tag{2.22}$$

we can reduce (2.21) to

$$rs'' = 3s' = 0, \tag{2.23}$$

which is easily integrated to give

$$s' = k_4/r^3, \quad k_4 = \text{arbitrary const}, \tag{2.24}$$

$$s = (k_5/r^2) + k_6, \tag{2.25}$$

where $k_5 = -\frac{1}{2}k_4$ and k_6 are arbitrary constants. Hence (2.22) gives

$$p' = k_7/(1 + k_8r^2), \tag{2.26}$$

where $k_7/k_3k_5 = -2k_1/k_2$ and $k_8 = k_6/k_5$ are arbitrary constants. In order to integrate (2.26), we need to know what sign k_8 takes. We can associate a sign with k_8 by appeal to the physics, as follows.

Since (2.7) and (2.8) are independent of h they must hold, in particular, for $h = \text{const}$, i.e., for $x^0 = \text{const}$, in which case the field equations (2.1) are just the Einstein field equations. The isotropic Schwarzschild solution

$$p = \log[(1 - B/r)/(1 + B/r)], \tag{2.27}$$

$$q = 2\log(1 + B/r), \tag{2.28}$$

$B = \text{arbitrary const}$

should be a particular solution to Eq. (2.26). This will be so if

$$k_8 = -1/B^2 \quad (< 0). \tag{2.29}$$

With this sign for k_8 we can now integrate (2.26), obtaining

$$p = \log\left(\frac{1 - r/B}{1 + r/B}\right)^{Bk_1/k_2} + \log k_9, \tag{2.30}$$

where k_9 is an arbitrary constant; or

$$e^p = k_{10}\left(\frac{1 - B/r}{1 + B/r}\right)^{Bk_1/k_2} = k_{10}\left(\frac{1 - B/r}{1 + B/r}\right)^{1/\lambda}, \tag{2.31}$$

where B and $k_{10} = (-1)^{Bk_1/k_2}$. k_9 are arbitrary constants, and we have put

$$\lambda = k_2/Bk_1 \quad (= \text{arbitrary const}). \tag{2.32}$$

Equations (2.20) and (2.26) yield, with proper arrangement of constants,

$$q' = -p' + 2B^2r^{-3}/(1 - B^2/r^2) \tag{2.33}$$

which has the integral

$$q = -p + \log(1 - B^2/r^2) + \log k_{11}, \tag{2.34}$$

where k_{11} is an arbitrary constant. Hence, using (2.31), we have

$$e^q = k_{11}e^{-p}(1 - B^2/r^2) = k_{12}(1 - B^2/r^2) \times [(1 - B/r)/(1 + B/r)]^{-1/\lambda},$$

where $k_{12} = k_{11}/k_{10}$ is an arbitrary constant, or in another form

$$e^q = k_{12}(1 + B/r)^2\left(\frac{1 - B/r}{1 + B/r}\right)^{1-1/\lambda}. \tag{2.35}$$

From (2.20) we have

$$q'' = -p''(1 - k_3/r) - k_3p'/r^2, \tag{2.36}$$

which with (2.5) leads to

$$r(h')^2 = -2(r - k_3)p'' + r(1 - k_3/r)^2(p')^2 - (4 - 2k_3/r)p'. \tag{2.37}$$

But (2.31) gives $p' = 2B/\lambda(r^2 - B^2)$, $p'' = -4Br/\lambda(r^2 - B^2)^2$, and substitution of these in (2.37) yields, after some algebra,

$$(h')^2 = 4B^2(1 - \lambda^2)/\lambda^2(r^2 - B^2)^2. \tag{2.38}$$

Hence

$$h = \log\left[k_{13}\left(\frac{1 - B/r}{1 + B/r}\right)^{-C_1/\lambda}\right], \tag{2.39}$$

where $C_1^2 = 1 - \lambda^2$ (C_1 = arbitrary const) and k_{13} is an arbitrary constant. From (2.6) we have, therefore,

$$x^0 = k_{14}\left(\frac{1 - B/r}{1 + B/r}\right)^{-C/\lambda}, \tag{2.40}$$

where $k_{14} = k_{13}^{\sqrt{2/\omega}}$ and we have put

$$\omega C^2 = 2C_1^2 = 2 - 2\lambda^2. \tag{2.41}$$

So we have, finally, the *general solution* to the field equations (2.1) in isotropic, closed form:

$$e^p = e^{p_0}\left(\frac{1 - B/r}{1 + B/r}\right)^{1/\lambda}, \tag{2.42a}$$

$$e^q = e^{q_0}(1 + B/r)^2\left(\frac{1 - B/r}{1 + B/r}\right)^{(\lambda-1)/\lambda}, \tag{2.42b}$$

$$x^0 = x_0^0\left(\frac{1 - B/r}{1 + B/r}\right)^{-C/\lambda}, \tag{2.42c}$$

where

$$2\lambda^2 = 2 - \omega C^2, \qquad \omega = \tfrac{3}{2} \tag{2.43}$$

and $p_0 (= \log k_{10}), q_0 (= \log k_{12}), x_0^0 (= k_{14}), B, C$ (and hence λ) are arbitrary constants.

By putting $C = 0$, $\lambda = 1$, $p_0 = q_0 = 0$, the above solution reduces to the ordinary Schwarzschild solution to Einstein's field equations.

It should be noted that if λ is to be real, a natural but not essential restriction, then $|C| < \sqrt{2/\omega} = 2/\sqrt{3}$.

3. MOTION OF A TEST PARTICLE

We shall next consider the equations governing the motion of a test particle in the field represented by Eqs. (2.1), where $\omega = \frac{3}{2}$.

In the Lyra manifold the autoparallels of the affine connection are

$$\ddot{x}^\mu + \left\{{\mu \atop \alpha\beta}\right\}\dot{x}^\alpha\dot{x}^\beta + \tfrac{1}{2}x^0(\delta^\mu_\alpha\phi_\beta + \delta^\mu_\beta\phi_\alpha - g_{\alpha\beta}\phi^\mu)\dot{x}^\alpha\dot{x}^\beta - \tfrac{1}{2}x^0(\phi_\alpha - \overset{\circ}{\phi}_\alpha)\dot{x}^\mu\dot{x}^\alpha = 0, \tag{3.1}$$

where $\dot{x}^\mu \equiv dx^\mu/ds$, $\ddot{x}^\mu \equiv d^2x^\mu/ds^2$, and

$$\overset{\circ}{\phi}_\alpha = 2(x^0)^{-2}x^0{}_{,\alpha} \tag{3.2}$$

The connection between the vector ϕ_α and $\overset{\circ}{\phi}_\alpha$ is, for the field of equations (2.1), given by

$$2\phi_\alpha + \overset{\circ}{\phi}_\alpha = 0 \tag{3.3}$$

Using (3.2) and (3.3), the equations of motion (3.1) become

$$x^0\ddot{x}^\mu + x^0\left\{{\mu \atop \alpha\beta}\right\}\dot{x}^\alpha\dot{x}^\beta + \tfrac{1}{2}x^0{}_{,\alpha}\dot{x}^\mu\dot{x}^\alpha + \tfrac{1}{2}g_{\alpha\beta}g^{\mu\nu}x^0{}_{,\nu}\dot{x}^\alpha\dot{x}^\beta = 0, \tag{3.4}$$

where the $g_{\alpha\beta}$ are given by the metric (2.2).

Since $x^0 = x^0(r)$ from (2.42c), we obtain Eqs. (3.4) explicitly as

$$\ddot{r} + [q' + (x^0)^{-1}(dx^0/dr)]\dot{r}^2 + [-r^2q' - r + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)r^2]\dot{\theta}^2 + [-r^2q' - r + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)r^2]\sin^2\theta\dot{\phi}^2 + [p' - \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]e^{2p-2q}\dot{t}^2 = 0, \tag{3.5}$$

$$\ddot{\theta} + [2r^{-1} + 2q' + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{\theta} - \sin\theta\cos\theta\dot{\phi}^2 = 0, \tag{3.6}$$

$$\ddot{\phi} + [2r^{-1} + 2q' + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{\phi} + 2\cot\theta\dot{\theta}\dot{\phi} = 0, \tag{3.7}$$

$$\ddot{t} + [2p' + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{t} = 0, \tag{3.8}$$

where the prime denotes differentiation with respect to r.

Equation (3.8) possesses the first integral

$$\dot{t} = k_1e^{-2p}(x^0)^{-1/2}, \tag{3.9}$$

where k_1 is a constant. Using (2.42), we can write this in the form

$$\dot{t} = \text{const} \times [(r + B)/(r - B)]^{(4-C)/2\lambda}$$

or

$$\dot{t} = k[(B + r)/(B - r)]^A, \tag{3.10}$$

where k is a constant and $A = (4 - C)/2\lambda$ is also a constant. If we take $\dot{t} = 1$ when $r = 0$, then $k = 1$ and (3.10) becomes

$$dt/ds = [(B + r)/(B - r)]^A. \tag{3.11}$$

Thus the proper time interval Δs measured by an observer moving with the test particle in the free region surrounding the point mass is related to the coordinate time interval Δt measured by an observer at the point mass ($r = 0$) by

$$\Delta s = [(B - r)/(B + r)]^A \Delta t. \tag{3.12}$$

Planetary motion about the sun may be considered to take place in the plane $\theta = \frac{1}{2}\pi$; that such motion will always be confined to this plane may be seen from Eq. (3.6). Then we get Eq. (3.7) in the form

$$\ddot{\phi} + [2r^{-1} + 2q' + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{\phi} = 0, \tag{3.13}$$

which possesses the first integral

$$r^2\dot{\phi} = k_2 e^{-2q}(x^0)^{-1/2}, \tag{3.14}$$

where k_2 is a constant. Using Eq. (2.42b), we can put (3.14) in the form

$$\frac{d\phi}{ds} = \frac{Kr^2}{(r^2 - B^2)^2}\left(\frac{r - B}{r + B}\right)^D, \tag{3.15}$$

where K is a constant and $D = (4 + C)/2\lambda$ is also a constant. A third equation for planetary motion is the metric (2.2), in the form

$$\dot{r}^2 + r^2\dot{\phi}^2 - e^{2p-2q}\dot{t}^2 - e^{-2q} = 0, \tag{3.16}$$

where we have remembered that $\theta = \frac{1}{2}\pi$. Then Eqs. (3.9), (3.14), and (3.16) provide us with enough information to predict a perihelic shift, provided that we have values of the constants B, C, and k on hand.

4. PERIHELIC SHIFT

It is instructive to obtain an expression for the advance of the perihelion of a planet about the sun using the series solution given by Sen and Dunn to the field equations (2.1). This will be an alternative to using the exact solution in Eqs. (3.9), (3.14), and (3.16).

The metric of the geometry is given by

$$ds^2 = e^{\nu}dt^2 - e^{\lambda}dr^2 - r^2d\theta^2 - r^2\sin^2\theta\, d\phi^2, \tag{4.1}$$

where $\lambda = \lambda(r)$, $\nu = \nu(r)$

$$e^{\nu} = D + C\phi(r), \tag{4.2}$$

$$e^{\lambda} = Ar^4[\phi'(r)]^2/[D + C\phi(r)], \tag{4.3}$$

$$\phi(r) = \sum_{n=0}^{\infty} a_n r^{-n}. \tag{4.4}$$

D, C, A are arbitrary constants and the coefficients a_n are given by

a_0 arbitrary, $\quad Aa_1^2 = D + Ca_0 \quad (a_1 \neq 0)$,

$a_2 = 0$, $\quad a_3$ arbitrary,

a_n, $n > 3$, are determined by a certain recurrence relation in terms of a_0 and a_3.

Also

$$x^0 = k\cdot\exp\int\{-[4/(\omega r^2) + (2/\omega r)(\phi''/\phi')]\}^{1/2}, \tag{4.5}$$

where k is a constant. Retaining only a few terms, we write (4.2), (4.3), and (4.5) as

$$e^{\nu} = C(b_0 + a_1r^{-1} + a_3r^{-3} + a_4r^{-4}), \tag{4.6}$$

$$e^{\lambda} = Ae^{-\nu}(a_1^2 + 6a_1a_3r^{-2} + 8a_1a_4r^{-3} + 9a_3^2r^{-4}), \tag{4.7}$$

$$(x^0)^{-1}(dx^0/dr) = 2c_0[r^{-2} + (a_4/a_3)r^{-3} - \{(a_1a_4^2 + 3a_3^3)/2a_1a_3^2\}r^{-4}], \tag{4.8}$$

where $b_0C = D + Ca_0$ and $c_0^2 = 2a_3/a_1$.

The equations of motion (3.4), with the metric (4.1), are explicitly

$$\begin{aligned}\ddot{r} &+ [\tfrac{1}{2}\lambda' + (x^0)^{-1}(dx^0/dr)]\dot{r}^2 \\ &+ e^{-\lambda}[-r + \tfrac{1}{2}r^2(x^0)^{-1}(dx^0/dr)]\dot{\theta}^2 \\ &+ e^{-\lambda}[-r + \tfrac{1}{2}r^2(x^0)^{-1}(dx^0/dr)]\sin^2\theta\dot{\phi}^2 \\ &+ e^{\nu-\lambda}[\tfrac{1}{2}\nu' - \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{t}^2 = 0,\end{aligned} \tag{4.9}$$

$$\ddot{\theta} + [2r^{-1} + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{\theta} - \sin\theta\cos\theta\dot{\phi}^2 = 0, \tag{4.10}$$

$$\ddot{\phi} + [2r^{-1} + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{\phi} + 2\cot\theta\dot{\theta}\dot{\phi} = 0, \tag{4.11}$$

$$\ddot{t} + [\nu' + \tfrac{1}{2}(x^0)^{-1}(dx^0/dr)]\dot{r}\dot{t} = 0. \tag{4.12}$$

We observe that Eq. (4.9)–(4.12) are identical, to first order in $1/r$, with the Einstein equations of motion; this is so because the term in x^0 is of second order in $1/r$, by (4.8).

Putting $\theta = \frac{1}{2}\pi$ for planetary motion, Eqs. (4.11) and (4.12) have the respective integrals

$$\dot{t}e^{\nu} = k_1(x^0)^{-1/2}, \tag{4.13}$$

$$r^2\dot{\phi} = k_2(x^0)^{-1/2}, \tag{4.14}$$

where k_1 and k_2 are constants. A third equation of motion, in lieu of (4.9), is the metric (4.1) with $\theta = \frac{1}{2}\pi$:

$$e^{\lambda}\dot{r}^2 + r^2\dot{\phi}^2 - e^{\nu}\dot{t}^2 + 1 = 0. \tag{4.15}$$

Substitution of (4.13) and (4.14) in (4.15) results in

$$\dot{r}^2 + H^2r^{-2}e^{-\lambda} - (k_1^2/k_2^2)H^2e^{-\lambda}e^{-\nu} + e^{-\lambda} = 0, \tag{4.16}$$

where we have put

$$H = r^2\dot{\phi}. \tag{4.17}$$

By putting $u = r^{-1}$ and making the substitution $d/ds = Hu^2d/d\phi$, we can write (4.16) in the form

$$(du/d\phi)^2 + e^{-\lambda}u^2 - (k_1^2/k_2^2)e^{-\lambda}e^{-\nu} + e^{-\lambda}H^{-2} = 0, \tag{4.18}$$

where $e^{-\lambda}$, $e^{-\nu}$, and H^{-2} are functions of u, given to third order by

$$e^{-\lambda} = 1 + (Aa_1^2)^{-1}[Ca_1u - 6Aa_1a_3u^2 - (5Ca_3 + 8Aa_1a_4)u^3], \tag{4.19}$$

$$e^{-\lambda}e^{-\nu} = (Aa_1^2)^{-1}[1 - 6(a_3/a_1)u^2 - 8(a_4/a_1)u^3], \tag{4.20}$$

$$H^{-2} = k_2^{-2}\{1 - 2c_0u + (2c_0^2 - c_0a_4/a_3)u^2 - [4c_0^3/3 - 2c_0^2a_4/a_3 - (c_0a_1a_4^2 + 3c_0a_3^3)/3a_1a_3^2]u^3\}. \tag{4.21}$$

To obtain the last expression, we have used (4.5) and (4.8). Substituting in (4.18) we have, to third order in u, the following orbit equation:

$$(du/d\phi)^2 = \alpha_0 + \alpha_1 u + \alpha_2 u^2 + \alpha_3 u^3, \tag{4.22}$$

where

$$\alpha_0 = -\frac{1}{k_2^2}\left(1 - \frac{k_1^2}{Aa_1^2}\right), \tag{4.23}$$

$$\alpha_1 = -\frac{1}{k_2^2}\left(\frac{Ca_1}{Aa_1^2} - 2c_0\right), \tag{4.24}$$

$$\alpha_2 = -\frac{1}{k_2^2}\left(k_2^2 + \frac{6k_1^2 a_3}{Aa_1^3} - \frac{6a_3}{a_1} - \frac{2c_0 Ca_1}{Aa_1^2} + 2c_0^2 + \frac{c_0 a_4}{a_3}\right), \tag{4.25}$$

$$\alpha_3 = -\frac{1}{k_2^2}\left(\frac{k_2^2 Ca_1}{Aa_1^2} + \frac{8k_1^2 a_4}{Aa_1^3} - \frac{4Ca_3}{Aa_1^2} - \frac{7a_4}{a_1} + \frac{68c_0^3}{3} - \frac{c_0 Ca_4}{Aa_1 a_3} + \frac{c_0 a_1 a_4^2 + 3c_0 a_3^3}{3a_1 a_3^2}\right). \tag{4.26}$$

We refer to Møller[4] for the method we have used to calculate the perihelion advance $\Delta\phi$ from Eq. (4.22). We have

$$\Delta\phi = \tfrac{3}{2}\pi(u_1 + u_2)\alpha_3, \tag{4.27}$$

where u_1 and u_2 are the roots of the quadratic $\alpha_2 u^2 + \alpha_1 u + \alpha_0 = 0$, so that

$$u_1 + u_2 = -\alpha_1/\alpha_2 = \left[2c_0 - \frac{Ca_1}{Aa_1^2}\right] \Big/ \left[k_2^2 + \frac{6k_1^2 a_3}{Aa_1^3} - \frac{6a_3}{a_1} - \frac{2c_0 Ca_1}{Aa_1^2} + 2c_0^2 + \frac{c_0 a_4}{a_3}\right], \tag{4.28}$$

where $c_0^2 = a_3/2a_1$. Equations (4.26)–(4.28) give

$$\Delta\phi = \tfrac{3}{2}\pi\,(-Ca_1/Aa_1^2 k_2)^2 \times (E + 1), \tag{4.29}$$

where E is an expression all terms of which involve a_3 or a_4 (to this level of approximation).

In the Einstein case we have $Ca_1 = -2\mu$, $Aa_1^2 = 1$, $c_0 = a_3 = a_4 = 0$, $k_2 = h$, $\alpha_3 = 2\mu$, where $2\mu = \frac{1}{2}GM$ and h is the usual "areal" constant. Then

$$(\Delta\phi)_{\text{Einstein}} = \tfrac{3}{2}\pi\,(-Ca_1/Aa_1^2 k_2)^2 = 6\pi\mu^2/h^2, \tag{4.30}$$

in accordance with the usual theory.

Thus (4.29) and (4.30) yield

$$\Delta\phi/(\Delta\phi)_{\text{Einstein}} \sim E + 1, \tag{4.31}$$

which shows that there is no difference between the predictions of the present scalar-tensor theory and Einstein's general relativity theory when we are satisfied with an approximation to the second order in $1/r$; this is the usual order of accuracy for predictions to be experimentally detectable at present.

5. LINEARIZED FIELD EQUATIONS

We shall set

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}, \tag{5.1}$$

where $\eta_{ii} = 1$, $\eta_{44} = -1$, $i = 1, 2, 3$, and also set

$$x^0 = x_0^0 + \xi. \tag{5.2}$$

We shall neglect squared terms in $h_{\alpha\beta}$ and ξ. It will be helpful to define

$$\gamma_{\alpha\beta} = h_{\alpha\beta} - \tfrac{1}{2}\eta_{\alpha\beta}h \tag{5.3}$$

and

$$\sigma_\alpha = \gamma_{\alpha\beta,\mu}\eta^{\beta\mu}, \tag{5.4}$$

where

$$h = \eta^{\alpha\beta}h_{\alpha\beta} \quad \text{to first order.} \tag{5.5}$$

Then the field equations (1.1), to first order in $h_{\alpha\beta}$ and ξ, become

$$\Box\gamma_{\alpha\beta} - \sigma_{\alpha,\beta} - \sigma_{\beta,\alpha} + \eta_{\alpha\beta}\sigma_{\mu,\nu}\eta^{\mu\nu} = 16\pi G(x_0^0)^{-2}T_{\alpha\beta}, \tag{5.6}$$

where to first order $\Box\gamma_{\alpha\beta} = \eta^{\mu\nu}\gamma_{\alpha\beta,\mu,\nu}$. Introduce the coordinate conditions $\sigma_\alpha = 0$, and then (5.6) reduces to

$$\Box\gamma_{\alpha\beta} = 16\pi G(x_0^0)^{-2}T_{\alpha\beta}, \tag{5.7}$$

which is identical to the linearized form of the Einstein field equations if the gravitational constant G^* of the present theory is related to the Newtonian constant G by

$$G^* = G(x_0^0)^{-2}. \tag{5.8}$$

Hence the weak-field solution to the field equations (1.1) is just that of the Einstein case. In such a field the constants of the solution (2.42) take the values

$$p_0 = q_0 = 0, C = 0, \lambda = 1, B = \tfrac{1}{2}G^*M, x^0 = x_0^0. \tag{5.9}$$

One therefore expects in such a circumstance that, at least to second order in $1/r$, the predictions of this scalar-tensor theory for perihelion shift, bending of light rays, and gravitational redshift will not differ from those of the Einstein theory, as we have shown above by considering explicitly the equations of motion. When strong gravitational fields are present terms of third order in $1/r$ should be significant and then physical differences will be apparent between the two theories.

ACKNOWLEDGMENTS

I wish to thank Dr. D. K. Sen for his helpful comments. This work was begun in Toronto, and I am grateful to the National Research Council of Canada for a grant covering the early aspects of the program.

[1] D. K. Sen and K. A. Dunn, J. Math. Phys. **12**, 578 (1971).
[2] G. Lyra, Math. Z. **54**, 52 (1951).
[3] C. Brans and R. H. Dicke, Phys. Rev. **124**, 925 (1961).
[4] C. Møller, *The Theory of Relativity* (Oxford U.P., Oxford, 1952), pp. 350–52.

Gauge Invariant Decomposition of Yang-Mills Potentials*

Richard P. Treat
Physics Department, West Virginia University, Morgantown, West Virginia, 26506
(Received 15 March 1972)

The Yang–Mills (YM) potentials are decomposed into an isovector part and a part which transforms nonhomogeneously under local gauge transformations. Two decompositions are shown; one of them is based on a gauge-invariant version of the transversality condition, and the other arises from a gauge-invariant modification of the Lorentz condition. The latter is Lorentz as well as gauge invariant. The gauge invariance of the decompositions is obtained at the expense of locality since the separate parts of the decomposed potential are functionals of the full YM potential. The transverse-longitudinal decomposition is used to throw the YM sourceless field equations into a gauge-invariant Hamiltonian form. Static fields in the Hamiltonian formulation are discussed. The decompositions are used to construct massive, gauge-invariant but nonlocal Lagrangians. A Lorentz and gauge-invariant nonlocal interaction of the YM field with a spinor–isospinor field is formed. The transverse-longitudinal decomposition is used to investigate the geometric structure of a configuration space Ω of YM potentials. The nonexistence of submanifolds of Ω orthogonal to the gauge-invariant manifolds $X \in \Omega$ is proved in contradistinction to the electromagnetic case. A Green's functional for the Yang–Mills field is represented explicitly by an infinite power series of functionals and is shown to be self-adjoint.

1. INTRODUCTION

The nonhomogeneous transformation of the electromagnetic vector potentials under local gauge transformations is the cause of difficulties in the quantum theory of the field. To eliminate these difficulties, a number of studies have been devoted to reformulating the theory in terms of gauge invariant quantities. Relevant to the problems considered here are the investigations of Belinfante[1] and Rohrlich and Strocchi.[2] They have, by averaging over paths of the path-dependent potentials of DeWitt[3] and Mandelstam,[4] constructed potentials which are invariant under local gauge transformations. The vector potential of Belinfante satisfies a tranversality condition and is thus not Lorentz invariant. Rohrlich and Strocchi[2] have derived a vector potential which is a gauge and also a Lorentz invariant functional and which satisfies the Lorentz condition.

The nonhomogeneous transformation law of the Yang–Mills[5] potentials is likewise troublesome. The analysis carried out here is directed to an end similar to that of Refs. 1 and 2 in that we seek to identify a homogeneously transforming part of the YM potential which would appear in the theory as a physically significant dynamical variable. However, the problem is approached here from a point of view which is different from that held in Refs. 1 and 2. Nevertheless, their results are recovered when the YM field becomes Abelian, that is, when the structure constants of the isotopic spin group $SU(2)$ vanish. Although not emphasized by them, the work of Belinfante and Rohrlich and Strocchi yields gauge-invariant decompositions of the electromagnetic vector potentials into a gauge scalar and a part which transforms nonhomogeneously under a local gauge transformation. Here, for the non-Abelian case, we develop analogous decompositions which separate the YM potentials into an isovector part and a "longitudinal" part which transforms nonhomogeneously under local gauge transformations. The longitudinal part of the potential is carried along in the theory, in somewhat the same way as in Schwinger's method of group parameters,[6] and is not supposed to have direct physical significance. In one of the decompositions the isovector part of the potential satisfies a gauge invariant version of the transversality condition, and in the other decomposition the isovector part satisfies a gauge invariant "Lorentz condition."[7,8]

The paper begins with a specification of the mathematical and physical framework in which the analysis is carried out. The transverse-longitudinal decomposition of the gauge potentials which is based on a transversality condition is then defined and investigated. Utilizing this decomposition, we write the dynamical equations for the uncoupled YM field in the Hamiltonian form and then, with the aid of the decomposition, restate some known facts[9] on static fields.

We then study the structure of a configuration space Ω consisting of YM potentials which are bounded and which at spatial infinity are of order $|x^\alpha|^{-2}$ and which satisfy certain other special conditions.[10] This analysis starts within the framework delineated by Loos[10,11] and then shows some additional structure of Ω which can be inferred from the transverse-longitudinal decomposition.

Next a gauge and Lorentz invariant decomposition of the potentials based on a "Lorentz condition" is given. We indicate how the isovector part of the potential, which is a functional of the total potential, can be used to introduce gauge invariant, but spatially nonlocal terms in the free YM Lagrangian. We then display an isoscalar nonlocal interaction of the YM potential with a spinor–isospinor field.

Finally, in the Appendix a Green's functional is examined which is essentially the non-Abelian generalization of the Coulomb Green's function and which plays a central role in the transverse-longitudinal decomposition. This same functional also arises in the work of Ref. 6 and in a study of charged states of the YM field.[10] A series representation of this Green's functional is given, and its self-adjointness is shown.

2. PRELIMINARIES

Space–time is taken to be the Minkowski space where time–space coordinates are denoted by $x^\kappa, \kappa, \lambda, \cdots = 0, 1, 2, 3$. Spatial coordinates are designated by $x^\alpha, \alpha, \beta, \cdots = 1, 2, 3$. The Minkowski metric has the signature $+,-,-,-$, so the spatial part of the metric is negative definite. In the Minkowski space we consider YM potentials $b_{\kappa i}$, where $i, j, \cdots = 1, 2, 3$ indicate components in the Lie algebra space of the isotopic spin group $SU(2)$. The spatial components $b_{\alpha i}$ are assumed to be members of the configuration space Ω as defined in Ref. 10. Ω is defined as the space of real-valued functions, which are subject to the conditions that a real constant B exists such that

$$|b_{\alpha i}| \leq B, \qquad |\partial_\beta b_{\alpha i}| \leq B,$$
$$|x^\beta| > R, \qquad |x^\beta|^2 |b_{\alpha i}| \leq B, \qquad |x^\beta|^3 |\partial_\gamma b_\alpha| \leq B. \tag{2.1}$$

There are a number of other conditions cited in Ref.

8 which are required of potentials belonging to Ω. These further conditions have been used to prove existence of charged states and we do not make explicit use of them here.

It is sometimes convenient to use potentials $\Gamma_{\kappa i}^{j}$ (parameters of connections for isovectors) rather than the $b_{\kappa i}$. The two are related according to

$$\Gamma_{\alpha i}{}^{j} = -b_{\alpha}{}^{k} C_{ki}{}^{j}, \tag{2.2}$$

where $C_{ij}{}^{k}$ are the structure constants of $SU(2)$. The Cartan metric is constructed by the prescription

$$g_{ij} = C_{ik}{}^{n} C_{jn}{}^{k}. \tag{2.3}$$

The g_{ij} is used to raise and lower indices and is negative definite for $SU(2)$, thus co- and the corresponding contra-variant vectors in the Lie algebra space differ by a sign. The bases in the Lie algebra space are chosen such that g_{ij} has the diagonal form. The bases are thus arbitrary up to the local group of transformations which is orthogonal with respect to g_{ij}, which is thus just the adjoint representation of $SU(2)$.

The gauge field constructed from the potentials $b_{\kappa i}$ is given by

$$B_{\kappa\lambda}{}^{i} = \partial_{\kappa} b_{x}{}^{l} - \partial_{\lambda} b_{\kappa}{}^{i} - C_{jk}{}^{i} b_{\kappa}{}^{j} b_{\lambda}^{k}. \tag{2.4}$$

For brevity the isotopic spin indices are sometimes suppressed, and then we write

$$B_{\kappa\lambda} = \partial_{\kappa} b_{\lambda} - \partial_{\lambda} b_{\kappa} - b_{\kappa} \times b_{\lambda}. \tag{2.5}$$

The field constructed from the potentials $\Gamma_{ki}{}^{j}$ is given by

$$\Phi_{\kappa\lambda i}{}^{j} = \partial_{\kappa} \Gamma_{\lambda i}^{j} - \partial_{\lambda} \Gamma_{\kappa i}^{j} - \Gamma_{\kappa i}^{k} \Gamma_{\lambda k}^{j} + \Gamma_{\lambda i}^{k} \Gamma_{\kappa k}^{j}. \tag{2.6}$$

Under local gauge transformations represented by matrices, $S_{i}^{j}(x^{\kappa})$, the potentials $\Gamma_{\kappa i}^{j}$ transform as

$$\Gamma'^{j}_{\kappa i} = S^{-1}{}_{i}^{k} \Gamma_{\kappa k}{}^{m} S^{j}_{m} \tag{2.7}$$

while $\Phi_{\kappa\lambda i}{}^{j}$ is a tensor and transforms as

$$\Phi'_{\kappa\lambda i}{}^{j} = S_{i}^{-1\,k} \Phi_{\kappa\lambda k}{}^{m} S^{j}_{m} \tag{2.8}$$

Under an infinitesimal local gauge transformation the change in the potentials $b_{\kappa i}$ is given by

$$\delta b_{\kappa i} = -\nabla_{\kappa} \eta_{i}, \tag{2.9}$$

where η_{i} is an isovector, and

$$\eta_{i}{}^{j} = \eta^{k} C_{ki}{}^{j} \tag{2.10}$$

is the generator of the local gauge transformation $S^{j}_{i}(x^{\kappa})$. In (2.9) ∇_{K} is the symbol for gauge differentiation. For a covariant isovector,

$$\nabla_{\kappa} \eta^{i} = \partial_{\kappa} \eta^{i} + \Gamma_{\kappa j}^{i} \eta^{j}, \tag{2.11}$$

or

$$\nabla_{\kappa} \eta^{i} = \partial_{\kappa} \eta^{i} - C_{kj}{}^{i} b_{\kappa}^{k} \eta^{j}. \tag{2.12}$$

For brevity we sometimes suppress the Lie algebra indices, and then the covariant derivative of a contravariant vector is written as

$$\nabla_{\kappa} \eta = \partial_{\kappa} \eta - b_{\kappa} \times \eta. \tag{2.13}$$

3. GAUGE-INVARIANT TRANSVERSE-LONGITUDINAL DECOMPOSITION OF THE YANG–MILLS POTENTIALS

The longitudinal part of the space components of the potential is defined as the solution b^{*}_{α} of the differential equations

$$\partial_{\alpha} b^{*}_{\beta} - \partial_{\beta} b^{*}_{\alpha} - b^{*}_{\alpha} \times b^{*}_{\beta} = 0, \tag{3.1}$$

$$\partial_{\alpha} b^{*\,\alpha} - b_{\alpha} \times b^{*\,\alpha} = 0, \tag{3.2}$$

where algebra indices are suppressed. The transverse part of b_{α} is defined as the difference between b_{α} and b^{*}_{α} or, equivalently, as

$$b^{T}_{\alpha} = b_{\alpha} - b^{*}_{\alpha}. \tag{3.3}$$

It follows from (3.2) and (3.3) that

$$\partial_{\alpha} b^{T\alpha} - b_{\alpha} \times b^{T\alpha} = 0. \tag{3.4}$$

We regard b^{*}_{α} and b^{T}_{α} as functionals of the independent variable b_{α}, the functional relations being determined by the differential equations (3.1), (3.2), and the linear relation (3.3).

The variation δb^{*}_{α} resulting from the variation δb_{α} satisfies the differential equations

$$\nabla^{*}_{\alpha} \delta b^{*}_{\beta} - \nabla^{*}_{\beta} \delta b^{*}_{\alpha} = 0, \tag{3.5}$$

$$\nabla_{\alpha} \delta b^{*\,\alpha} = \nabla^{*}_{\alpha} \delta b^{\alpha}, \tag{3.6}$$

where ∇^{*}_{α} is a symbol for gauge covariant differentiation with respect to the longitudinal part of the potential. The derivatives ∇_{α} and ∇^{*}_{α} of isovectors are again isovectors because b_{α} and b^{*}_{α} both transform as parameters of a linear connection. However, it is only the full potential b_{α} which gives the definition of parallel displacement.[12] Equation (3.5) has a solution if and only if δb^{*} can be expressed in the form

$$\delta b^{*}_{\beta} = \nabla^{*}_{\beta} \chi. \tag{3.7}$$

Using (3.7) in (3,6) shows that χ satisfies

$$\nabla_{\alpha} \nabla^{*\,\alpha} \chi = \nabla^{*}_{\alpha} \delta b^{\alpha}. \tag{3.8}$$

We assume the existence of a fundamental solution $\mathfrak{D}_{xx'i}{}^{i'}$ of the differential equation

$$\nabla_{\alpha} \nabla^{*\alpha} \mathfrak{D}_{xx'i}{}^{i'} = \delta_{i}^{i'}(x - x'), \tag{3.9}$$

which satisfies the boundary conditions $\mathfrak{D}_{xx'i}{}^{i'} \to 0$ as $|x^{\alpha}|, |x'^{\alpha}| \to \infty$. In (3.9), xx' is short for x^{α}, x'^{α}, the prime on the index i specifying a component at the spatial point x'^{α}, while the unprimed i specifies a component in the local algebra space associated with the unprimed coordinate x^{α}. On the right-hand side of (3.9), $\delta_{i}{}^{i'}(x - x')$ is short notation for $\delta_{i}{}^{i'} \delta(x^{1} - x'^{1}) \delta(x^{2} - x'^{2}) \delta(x^{3} - x'^{3})$. In (3.9) the index notation on $\mathfrak{D}$ means that $\mathfrak{D}_{xx'i}{}^{i'}$ is a function of the two spatial points x and x' with components in the algebra spaces at x and x'. We shall also make use of the notations

$$\partial_{\beta'} = \frac{\partial}{\partial x'^{\beta}}, \qquad b_{\beta'} = b_{\beta}(x'),$$

$$\nabla_{\beta'} = \frac{\partial}{\partial x'^{\beta}} - \Gamma_{\beta}(x').$$

The above notations have been used by DeWitt.[13]

By using $\mathfrak{D}$ satisfying (3.9), we express χ formally as

$$\chi = \int \mathfrak{D}_{xx'} \nabla^*_{\beta'} \delta b^{\beta'} dx', \tag{3.10}$$

where the variations δb^β of b^β belong to the configuration space Ω and algebra indices are suppressed. Integration by parts yields

$$\chi = \int \nabla^*_{\beta'} \mathfrak{D}_{xx'} \delta b^{\beta'} dx' + \int \partial_{\beta'} (\mathfrak{D}_{xx'} \delta b^{\beta'}) dx'.$$

If

$$|x^\alpha| |\mathfrak{D}_{xx'}| \leq B, \quad |x^\alpha| \to \infty, \tag{3.11}$$

the integrals (3.10) and (3.11) exist and the surface integral of (3.11) vanishes and we obtain

$$\chi = - \int \nabla^*_{\beta'} \mathfrak{D}_{xx'} \delta b^{\beta'} dx', \tag{3.12}$$

and, consequently,

$$\delta b^*_\alpha = - \int \nabla^*_\alpha \nabla^{*\beta'} \mathfrak{D}_{xx'} \delta b_{\beta'} dx'. \tag{3.13}$$

From (3.13) the functional derivative of b^*_α is seen to be

$$\frac{\delta b^*_\alpha}{\delta b_{\beta'}} = - \nabla^*_\alpha \nabla^{*\beta'} \mathfrak{D}_{xx'}. \tag{3.14}$$

The functional derivative of b^T_α is

$$\frac{\delta b^T_\alpha}{\delta b_{\beta'}} = \delta^{\beta'}_\alpha (x - x') + \nabla^*_\alpha \nabla^{*\beta'} \mathfrak{D}_{xx'}, \tag{3.15}$$

as follows from the definition of b^T_α and (3.14).

To decompose the time component of b_κ, we define a part b^*_0 as the solution of

$$\partial_0 b^*_\alpha - \partial_\alpha b^*_0 - b^*_0 \times b^*_\alpha = 0. \tag{3.16}$$

We suppose that a solution $b^*_0[b]$ of Eq. (3.16) exists, which allows us to define a remainder φ of the time component b_0 by the linear relation

$$b_0 = \varphi + b^*_0. \tag{3.17}$$

To obtain a representation of the variation δb^*_0, which will be needed later, we take the variation of (3.16) to obtain

$$\nabla^*_0 \delta b^*_\alpha - \nabla^*_\alpha \delta b^*_0 = 0. \tag{3.18}$$

A solution of (3.16) is given by

$$\delta b^*_0 = \nabla^*_0 \chi. \tag{3.19}$$

With the aid of (3.12) we see that

$$\delta b^*_0 = - \nabla^*_0 \int \nabla^*_{\beta'} \mathfrak{D}_{xx'} \delta b^{\beta'} dx' \tag{3.20}$$

To examine the transformation properties of the potentials we consider local infinitesimal gauge transformations, which are of the form

$$\delta b_\alpha = - \Delta_\alpha \eta, \tag{3.21}$$

where it is required[8] that η have the asymptotic behavior $|x^\alpha| |\eta| \leq B$ as $x \to \infty$.

To show the transformation properties of b^*_α and b^T_α under local gauge transformations, we consider the integral expression (3.13). For an infinitesimal gauge transformation δb_α is given by (3.21), yielding

$$\delta b^*_\alpha = - \nabla^*_\alpha \int \nabla^{*\beta'} \nabla_{\beta'} \mathfrak{D}_{xx'} \eta(x') dx' - \nabla^*_\alpha \int \partial_{\beta'} [\nabla^{*\beta'} \mathfrak{D}_{xx'} \eta(x')] dx'.$$

For $|x^\alpha|^2 |\mathfrak{D}_{xx'}| \leq B$, $|x^\alpha| |\eta| \leq B$ as $|x^\alpha| \to \infty$, the last integral vanishes. By (3.9) it follows that

$$\delta b^*_\alpha = - \nabla^*_\alpha \eta(x). \tag{3.22}$$

Taking the variation of each member of (3.3) and use of (3.20) and (3.22) yields the isovector transformation rule

$$\delta b^T_\alpha = b^T_\alpha \times \eta. \tag{3.23}$$

The same procedure which was used to obtain the transformation properties of b^*_α and b^T_α, when applied to b^*_0 and φ, show that under a local gauge transformation generated by η that

$$\delta b^*_0 = - \nabla^*_0 \eta, \quad \delta \varphi = \varphi \times \eta. \tag{3.24}$$

Thus, φ is the isovector part of the time component of the potential.

The nonhomogeneous transformation law for b^*_κ implies that Γ^*_κ transforms under a finite transformation S as

$$\Gamma^*_\kappa = S^{-1} \Gamma_\kappa S - S^{-1} \partial_\kappa S. \tag{3.25}$$

The infinitesimal transformations shown to hold for φ and b^T_α imply that the time-space components

$$\Phi_i{}^j = - \varphi^k C_{ki}{}^j, \quad \Gamma^T_{\alpha i}{}^j = - b^{Tk}_\alpha C_{ki}{}^j \tag{3.26}$$

transform as mixed isotensors under infinitesimal and also finite local gauge transformations. That is, under a local gauge transformation,

$$\Phi' = S^{-1} \Phi S, \quad \Gamma'^T_\alpha = S^{-1} \Gamma^T_\alpha S. \tag{3.27}$$

Thus, we have decomposed the potentials (or parameters of connection Γ_κ) into a tensor and a part that transforms as a linear connection.

This type of decomposition has a counterpart in differential geometry. It is well known that if a tensor field is added to a given linear connection the sum is also a linear connection.[14] We are dealing here with the converse of this situation, where we have decomposed a given connection, presumably by a prescription which gives a unique separation, and find that one part is a tensor and that the other part transforms as a linear connection, where, referring to (3.1) and (3.16), we see that the nonhomogeneously transforming part has been defined in such a way that it is integrable. That is to say,

$$\partial_\alpha \Gamma^*_\beta - \partial_\beta \Gamma^*_\alpha - [\Gamma^*_\alpha, \Gamma^*_\beta] = 0. \tag{3.28}$$

This means that Γ^*_β can be transformed away by a local gauge transformation. Assigning a definite function Γ^*_β compatible with (3.28) is equivalent to fixing a gauge. We do not attempt to do this, but instead carry the integrable part of the potentials along in the theory in the spirit of Schwinger's method of

"group parameters,"[6] thereby maintaining a system of equations which are obviously gauge invariant, but separating out the part (Φ, Γ_α^T), or equivalently (φ, b_α^T), of the potentials which should be physically relevant.

The full potentials b_α are to be regarded at this point as independent variables whereas b_α^T, b_α^*, and b_0^* are functionals of b_α. The vector part φ of b_0 can be taken as independent as long as the YM field equations are not imposed. However, when the field equations are in effect, φ then becomes a functional of the spatial potentials b_α.

This decomposition into a transverse and longitudinal potential may be regarded as a generalization to non-Abelian gauge fields of the gauge invariant theory of the electromagnetic field given by Belinfante.[1] Belinfante showed that by averaging over path dependent potentials one could obtain a path independent and at the same time, a gauge independent theory of free and coupled electromagnetic fields. We have bypassed introducing path dependent potentials and averaging over paths, by decomposing the potentials of the YM theory directly into a tensor and an integrable part, where the integrable part carries the "divergence" of the full potential. Our result reduces to that obtained by Belinfante, as is easily verified, by setting the structure constants equal to zero to achieve the reduction to the Abelian case.

4. ON THE GEOMETRY OF THE CONFIGURATION SPACE

We now look at certain aspects of the structure of the configuration space Ω which follow from the assumption that the transverse-longitudinal decomposition exists and is unique. Following Loos[11] we introduce in the "Euclidean" metric

$$g_{\alpha\beta} g_{ij} \delta(x - x'). \tag{4.1}$$

This metric makes Ω a Hilbert space. With (4.1) the square of the distance between two neighboring points in Ω is given by

$$\| \delta b \|^2 = \int\int g_{\alpha\beta'} g_{ij'} \delta(x - x') \delta b^{\alpha i} \delta b^{\beta' j'} dx dx', \tag{4.2}$$

which yields, after carrying out one of the x integrations,

$$\| \delta b \|^2 = \int \delta b_{\alpha i}(x) \delta b^{\alpha i}(x) dx. \tag{4.3}$$

Let X_0 denote the manifold of integrable potentials b_α^* in Ω. Then, consider a set of potentials, say $\Xi(b^*)$, which consists of those potentials b_α which have a fixed longitudinal part b_α^* and a transverse part which satisfies

$$\nabla^{*\alpha} b_\alpha^T = 0. \tag{4.4}$$

In a finite-dimensional flat space, a flat submanifold is defined by a set of linear algebraic equations with constant coefficients. Similarly, in the infinite-dimensional Hilbert space Ω the linear differential equation (4.4) for fixed b_α^* defines a flat submanifold $\Xi(b^*)$. As we will show below, elements $-\nabla_\alpha^* \eta$ for fixed $b^*{}_\alpha$ are orthogonal to $\Xi(b^*)$ at all $b_\alpha \in \Xi(b^*)$ provided η satisfies the conditions for $x \to \infty$ required of generators of local gauge transformations. The manifolds $\Xi(b^*)$ intersect X_0 at the points $b_\alpha = b_\alpha^*$, which is to say at the points of $\Xi(b^*)$ where $b_\alpha^T = 0$. We are using this point of intersection to label the manifold $\Xi(b^*)$ as our notation indicates. The union of all manifolds $\Xi(b^*)$, $b_\alpha^* \in X_0$ is the whole configuration space Ω.

The set of potentials $b_\alpha \in \Xi(b^*)$ and the set of potentials $b_\alpha + \delta b_\alpha$ in a neighboring manifold $\Xi(b^* + \delta b^*)$ are related by a gauge transformation. For, to move from $b_\alpha^* \in X_0$ to a neighboring point $b_\alpha^* + \delta b_\alpha^* \in X_0$, the longitudinal part of the potential is changed according to

$$\delta b_\alpha^* = -\nabla_\alpha^* \eta, \tag{4.5}$$

which insures that the new potential $b_\alpha^* + \delta b_\alpha^*$ is also integrable. In order for the transverse potential $b_\alpha^T + \delta b_\alpha^T$ to belong to $\Xi(b^* + \delta b^*)$, it is necessary and sufficient that

$$\delta b_\alpha^T = b_\alpha^T \times \eta, \tag{4.6}$$

as follows from (4.4). The change δb_α is given by the sum of (4.5) and (4.6) so that

$$\delta b_\alpha = -\nabla_\alpha \eta, \tag{4.7}$$

which shows that elements of neighboring $\Xi(b^*)$ manifolds are related by a gauge transformation.

A gauge invariant manifold is defined as a subset of points in Ω which are related to each other by a local gauge transformation. It is evident that the set X_0 of integrable potentials is an invariant manifold. However, it is a special one inasmuch as it passes through the origin of Ω. Let us call the set of gauge invariant manifolds X. The X manifolds intersect the transverse manifolds $\Xi(b^*)$. This is clear because we can start at a point on $\Xi(b^*)$ and by performing all possible gauge transformations of this point generate a gauge invariant manifold.

The manifold X_0 intersects the $\Xi(b^*)$ orthogonally. To see this we form the inner product of vectors tangent to X_0 and $\Xi(b^*)$. The element $-\nabla_\alpha^* \eta$ is tangent to X_0 at $b_\alpha = b^*{}_\alpha$. An element δb_α^T is tangent to $\Xi(b^*)$ if

$$\nabla^{*\alpha} \delta b_\alpha^T = 0. \tag{4.8}$$

Variations δb_α^T satisfying (4.8) deserve a special notation since an arbitrary variation δb_α^T satisfies

$$\nabla^{*\alpha} \delta b_\alpha^T = -b_\alpha^T \times \delta b^{*\alpha},$$

as can be seen by taking the variation of (3.4). We denote variations satisfying (4.8) by $(\delta b_\alpha^T)_{b^*}$.

The inner product of the two tangent vectors is by (4.1),

$$(\delta b^*, (\delta b^T)_{b^*}) = \int \delta b_\alpha^* (\delta b^{T\alpha})_{b^*} dx = \int \partial^\alpha (\eta \times (\delta b_\alpha^T)_{b^*}) dx = 0, \tag{4.9}$$

where we have used (4.8) and (4.5). The last integral vanishes because of the asymptotic behavior required of members of Ω. The other gauge invariant manifolds X are not generally orthogonal to the $\Xi(b^*)$ at their intersection. To see this we form the inner product of the elements $(\delta b_\alpha^T)_{b^*}$ and $-\nabla_\alpha \eta$ which are tangent to $\Xi(b^*)$ and X at b_α. This gives

$$(-\nabla^{\alpha}\eta, (\delta b^{T}_{\alpha})_{b*}) = \int (\nabla^{\alpha}(\delta b^{T}_{\alpha})_{b*})\eta^{dx} = -\int \Gamma^{T\alpha} b^{T}_{\alpha}\eta dx, \tag{4.10}$$

which does not always vanish.

It is tempting to try to find a submanifold of Ω which is orthogonal to the gauge invariant manifolds X. However, as has been shown, such a manifold does not exist.[15] We show another proof of this in the framework of the transverse-longitudinal decomposition.

First we note that any infinitesimal element δb_{α} can be expressed as the sum

$$\delta b_{\alpha} = (\delta b^{T}_{\alpha})_{b*} + \nabla^{*}_{\alpha}\delta\psi; \tag{4.11}$$

$(\delta b^{T}_{\alpha})_{b*}$ satisfies (4.8) and $\delta\psi$ is an undetermined infinitesimal isotopic vector field. This is a slight generalization of the statement that a vector field can be represented as the sum of longitudinal and transverse fields. We wish to construct an element orthogonal to X at b_{α}. A necessary and sufficient condition for the orthogonality, inferred by inserting δb_{α} in place of $(\delta b^{T}_{\alpha})_{b*}$ in (4.10) is

$$\nabla^{\alpha}\delta b_{\alpha} = 0. \tag{4.12}$$

Imposing this condition on (4.11) shows that $\delta\psi$ must satisfy

$$\nabla^{\alpha}\nabla^{*}_{\alpha}\delta\psi = \Gamma^{T\alpha}(\delta b^{T}_{\alpha})_{b*}. \tag{4.13}$$

Using the fundamental solution $\mathfrak{D}_{xx'}$, we obtain from (4.13) the functional derivative equation

$$\left(\frac{\delta\psi}{\delta b_{\alpha'}}\right)_{b*} = \mathfrak{D}_{xx'}\Gamma^{T\alpha'}, \tag{4.14}$$

where the functional differentiation is taken keeping b^{*}_{α} fixed, so that the b^{*}_{α} appear as parameters in this equation. The solution ψ of (4.14), if it would exist, defines a functional on $\Xi(b^{*})$. Figure 1 is shown to illustrate the situation.

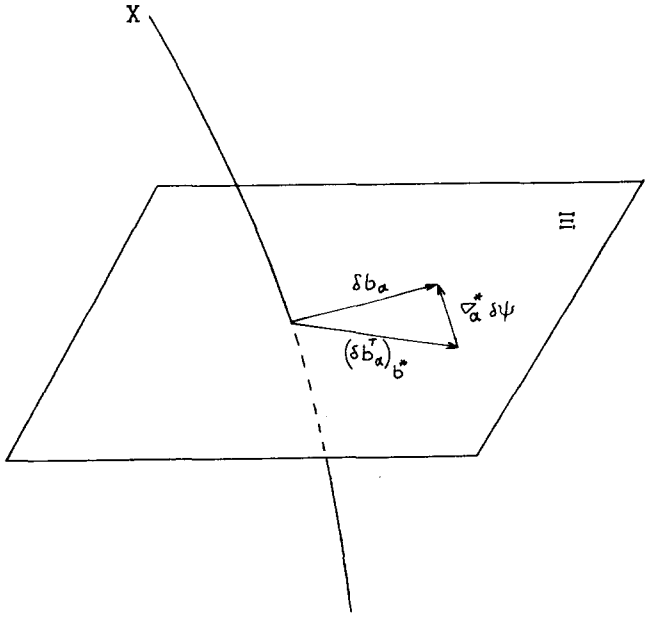


FIG. 1. Illustration of the construction of δb_{α} orthogonal to a gauge invariant manifold at a point.

The vector field δb_{α} in a neighborhood of b_{α} defines a manifold, which is tangent to the δb_{α}, and which is orthogonal to manifolds X. The construction of the tangent field δb_{α} in a neighborhood of a point is possible if and only if the solution $\psi[b^{T}, b^{*}]$ of the functional derivative equation (4.14) exists in the neighborhood of b_{α}.

In order that (4.14) has a solution it is necessary that the alternating functional derivative of the right-hand side vanishes. Thus it is necessary that

$$\frac{\delta}{\delta b^{T}_{\alpha_2}}(\mathfrak{D}_{xx_1}\Gamma^{T\alpha_1}) - \frac{\delta}{\delta b^{T}_{\alpha_1}}(\mathfrak{D}_{xx_2}\Gamma^{T\alpha_2}) \tag{4.15}$$

vanish.

To compute the functional derivative of $\mathfrak{D}_{xx'}$ we need the expression (A6), derived in the Appendix, generalized to hold on any $\Xi(b^{*})$ manifold. The correct expression for the functional derivative is obtained by replacing the partial derivative in the right-hand side of (A6) by the longitudinal covariant derivative ∇^{*}_{α}, which gives

$$\frac{\delta\mathfrak{D}_{xx'}{}^{ii'}}{\delta b^{T}_{\alpha_2 i_1}} = \mathfrak{D}_{xx_1}{}^{ii_1} C^{k_1}{}_{i_1}{}^{j_1}\nabla^{*\alpha_1}\mathfrak{D}_{x_1x'j_1}{}^{i'}. \tag{4.16}$$

Use of

$$\frac{\delta\Gamma^{T\alpha_1}{}_{i_1}{}^{j_1}}{\delta b^{T}_{\alpha_2 i_2}} = C_{k_1 i_1}{}^{j_1} g^{\alpha_1\alpha_2} g^{i_2 k_1}\delta(x_1 - x_2) \tag{4.17}$$

in conjunction with (4.16) shows that the integrability condition is not satisfied and, hence, that the functional ψ does not exist. We infer from this that there does not exist a manifold orthogonal to gauge invariant manifolds in a neighborhood of b_{α}.

5. HAMILTON EQUATIONS OF MOTION AND THE STATIC YM FIELD

A. The Hamilton Equations

The principle of stationary action, for the Lagrangian density

$$\mathfrak{L} = \tfrac{1}{4}B_{\kappa\lambda i}B^{\kappa\lambda i}, \tag{5.1}$$

yields the YM field equation

$$\nabla_{\kappa}B^{\kappa}{}_{\lambda i} = 0. \tag{5.2}$$

The momentum conjugate to b_{α} is defined as

$$\pi_{\alpha} = \frac{\partial\mathfrak{L}}{\partial(\partial^{0}b^{\alpha})} = B_{0\alpha}, \tag{5.3}$$

where we are now suppressing algebra indices. The definition of π_{α} and the transverse-longitudinal decomposition show that we can write the momentum as

$$\pi_{\alpha} = \nabla^{*}_{\alpha}b^{T}_{\alpha} - \nabla_{\alpha}\varphi. \tag{5.4}$$

From the YM constraint equation we have

$$\nabla_{\alpha}B^{\alpha}{}_{0} = 0 \tag{5.5}$$

and from the decomposition, it follows that

$$\nabla_{\alpha}\pi^{\alpha} = 0, \tag{5.6}$$

$$\nabla^*_\alpha \pi^\alpha = b^T_\alpha \times \pi^\alpha. \tag{5.7}$$

By (5.7), (5.4), and use of the functional $\mathfrak{D}_{xx'}$, we obtain

$$\varphi = \int \mathfrak{D}_{xx'} b^T_{\alpha'} \times \pi^{\alpha'} dx', \tag{5.8}$$

where we have used the transversality condition and the commutivity of ∇^*_0 and ∇^*_α. Equation (5.8) shows, as is well known,[16] that the time component of the potential should not be regarded as an independent variable.

The spatial components of the YM equations, with the aid of (5.3) and the decomposition of the time component of the potential, acquire the form

$$\nabla^*_0 \pi_\alpha = \varphi \times \pi_\alpha - \nabla_\beta B^\beta{}_\alpha. \tag{5.9}$$

The Hamiltonian density is now defined as

$$\mathcal{H} = \tfrac{1}{2}\pi_{\alpha i}\pi^{\alpha i} - \tfrac{1}{4}B_{\alpha\beta i}B^{\alpha\beta i} + \pi^\alpha_i \nabla_\alpha \varphi^i. \tag{5.10}$$

Let us take variations of the Hamiltonian

$$H = \int \mathcal{H} dx \tag{5.11}$$

on a fixed $\Xi(b^*)$ manifold. That is, we keep b^*_α fixed and take the variation of H in directions tangent to $\Xi(b^*_\alpha)$ with time held constant. For fixed b^*_α we obtain

$$\delta H = \int\left(-\varphi \times \pi^\alpha + \nabla_\beta B^{\beta\alpha} + \int \pi^{\beta 1} \nabla_{\beta 1} \frac{\delta\varphi}{\delta b^T_\alpha} dx'\right) \delta b^T_\alpha dx. \tag{5.12}$$

Imposing the constraint equation (5.6) then yields

$$\frac{\delta H}{\delta b^T_\alpha} = -\varphi \times \pi^\alpha + \nabla_\beta B^{\beta\alpha}. \tag{5.13}$$

Functional differentiation of (5.11) with respect to π_α shows that

$$\frac{\delta H}{\delta \pi^\alpha} = \pi_\alpha + \nabla_\alpha \varphi. \tag{5.14}$$

Collecting results we have the Hamilton equations

$$\nabla^*_0 \pi^\alpha = -\frac{\delta H}{\delta b^T_\alpha}, \tag{5.15}$$

$$\nabla^*_0 b^T_\alpha = \frac{\delta H}{\delta \pi^\alpha}. \tag{5.16}$$

These equations must be supplemented with the definitions of b^*_κ and φ (Eq. 5.8) and the constraint of π_α [Eq. (5.6)]. It is understood in (5.15) that the functional derivative is taken tangent to an arbitrary $\Xi(b^*)$ manifold at a fixed time.

There are integrability conditions to consider for the Hamilton equations. Taking the divergence $\nabla^{*\alpha}$ of (5.16) yields

$$\nabla^*_0(\nabla^{*\alpha} b^T_\alpha) = \nabla^{*\alpha}\pi_\alpha + \nabla^{*\alpha}\nabla_\alpha \varphi. \tag{5.17}$$

According to the definition of φ, (5.17) becomes

$$\nabla^*_0(\nabla^{*\alpha} b^T_\alpha) = \nabla^\alpha \pi_\alpha. \tag{5.18}$$

If the constraint equation is satisfied, the right-hand side vanishes and we obtain the integrability condition

$$\nabla^*_0(\nabla^{*\alpha} b^T_\alpha) = 0. \tag{5.19}$$

The implication of (5.19) is that if $b_\alpha = b^T_\alpha + b^*_\alpha$ belongs at some time t to a manifold $\Xi[b^*(t)]$ then at a later time $t + dt$, $b_\alpha(t + dt)$ will lie in the manifold $\Xi[b^* + (db^*/dt)dt]$. Therefore, the tranversality condition on b^T_α will be satisfied at $t + dt$. Thus, if the tranversality condition is imposed at one time, the Hamiltonian dynamics insures that the condition is satisfied at all times. This means that we do not need to acknowledge explicitly in the Hamiltonian equations the functional relation (Eq. 3.13) between b^T_α and the full potential b_α. Rather we can choose b^*_α at t as an arbitrary integrable potential, which merely fixes the initial manifold, choose initial conditions on b^T_α such that the gauge invariant transversality condition is satisfied on $\Xi[b^*(t)]$, and then the transversality condition is automatically satisfied at later times.

In cases where b^*_α would be taken independent of time, the Hamiltonian dynamics leaves the potentials in a fixed $\Xi(b^*)$ manifold. For example, the transverse gauge $b^*_\alpha = 0$, where $\partial_\alpha b^{T\alpha} = 0$, is of this type. The gauge $b_3 = 0$ investigated in Ref. 16 requires a time-dependent b^*_α in order to maintain the relation

$$b^T_3 + b^*_3 = 0. \tag{5.20}$$

The constraint (5.6) on π_α is also compatible with the Hamilton equations. If π_α is gauge covariant divergenceless at an initial time, the evolution in time of π_α determined by the Hamilton equations maintains this condition.

B. Transverse Momentum and the Static YM Field

The transverse momentum is defined as

$$\pi^T_\alpha = \nabla^*_0 b^T_\alpha \tag{5.21}$$

and has the property

$$\nabla^{*\alpha}\pi^T_\alpha = 0. \tag{5.22}$$

Substituting

$$\pi_\alpha = \pi^T_\alpha - \nabla_\alpha \varphi \tag{5.23}$$

in the constraint equation and using (5.22) shows that the time-component φ satisfies

$$\nabla^\alpha \nabla_\alpha \varphi = -b^{T\alpha} \times \pi^T_\alpha. \tag{5.24}$$

Let us assume the existence of a fundamental solution $\mathcal{E}_{xx'}$ which satisfies

$$\nabla^\alpha \nabla_\alpha \mathcal{E}_{xx'}{}_i{}^{i'} = \delta^{i'}_i(x - x'). \tag{5.25}$$

Then formally

$$\varphi = -\int \mathcal{E}_{xx'} b^{T\alpha'} \times \pi^T_{\alpha'} dx'. \tag{5.26}$$

Introducing the decomposition (5.23) and the representation (5.26) in the Hamiltonian density yields

$$H = \int(\tfrac{1}{2}\pi^T_{\alpha i}\pi^{T\alpha i} - \tfrac{1}{4}B_{\alpha\beta i}B^{\alpha\beta i})dx + \tfrac{1}{2}\int\int \mathcal{E}_{xx'}{}^{i'}(b^T_\alpha \times \pi^{T\alpha})^i \times (b^T_\beta \times \pi^{T\beta'})_{i'} dx dx', \tag{5.27}$$

where, for clarity, we have reinstated some of the indices. The longitudinal part of the momentum has thereby been eliminated from the Hamiltonian.

We go on now to consider the static YM field. There can be some confusion in a local gauge theory as to the meaning of a static field since potentials which are time independent in one gauge may depend on time in another gauge. Loos[9] has clarified this problem by stating gauge invariant conditions satisfied by "static" fields. We take a result developed there and restate it in the context of the transverse-longitudinal decomposition. Let us adopt the following as a definition of a static YM field; the YM field is static if and only if the tranverse momentum vanishes. This is a gauge invariant statement since π^T_α is an isovector. For vanishing π^T_α, the constraint equation implies that

$$\nabla^\alpha \nabla_\alpha \varphi = 0. \tag{5.28}$$

It has been shown[9] that for $|x^\alpha|^2 |\varphi| \leq B$ as $x \to \infty$, and we have insisted on this asymptotic behavior here, that (5.27) has only the null solution. Thus, for static fields, as we have defined them, the Hamilton equations for the YM field become

$$\nabla^*_0 b^T_\alpha = 0, \qquad \nabla_\beta B_\alpha{}^\beta = 0, \tag{5.29}$$

and the constraint equation is

$$\varphi = 0. \tag{5.30}$$

6. A LORENTZ AND GAUGE-INVARIANT DECOMPOSITION OF THE GAUGE POTENTIALS

The transverse-longitudinal decomposition, being based on the transversality condition, is of course not Lorentz invariant. One can achieve a Lorentz invariant decomposition by proceeding in the same way as we did in Sec. 4, but with the Lorentz condition replacing the transversality condition.[7] The resulting decomposition yields in the special case of an Abelian field the invariant electromagnetic potentials obtained by Rohrlich and Strocchi[2] by their method of averaging over paths of the path-dependent potentials of DeWitt[3] and Mandelstam.[4]

Thus, we define an integrable part of b_κ and the solution of the equations

$$\partial_\lambda b^*_\kappa - \partial_\kappa b^*_\lambda - b^*_\lambda \times b^*_\kappa = 0, \tag{6.1}$$

$$\partial_\lambda b^{*\lambda} - b_\lambda \times b^{*\lambda} = \partial_\lambda b^\lambda. \tag{6.2}$$

In these equations the b_λ are independent. A variation of b_λ causes the change δb^*_λ which satisfies

$$\nabla^*_\lambda \delta b^*_\kappa - \nabla^*_\kappa \delta b^*_\lambda = 0, \tag{6.3}$$

$$\nabla_\lambda \delta b^{*\lambda} = \nabla^*_\lambda \delta b^\lambda. \tag{6.4}$$

Equation (6.3) implies that δb^*_κ is the gradient of an isovector, say χ. Then, for (6.4) we have

$$\nabla^\lambda \nabla^*_\lambda \chi = \nabla^{*\lambda} \delta b_\lambda. \tag{6.5}$$

We assume the existence of a fundamental solution $\mathcal{F}_{xx'}{}^{i'}{}_i$ which satisfies

$$\nabla^\lambda \nabla^*_\lambda \mathcal{F}_{xx'}{}^{i'}{}_i = \delta^{i'}_i(x - x'), \tag{6.6}$$

where x now refers to time-space components x^κ. Then δb^*_κ can be expressed formally

$$\delta b^*_\kappa = \nabla^*_\kappa \int \mathcal{F}_{xx'} \nabla^{*\lambda'} \delta b_{\lambda'} dx'. \tag{6.7}$$

For variations vanishing on the boundaries $|x^\alpha| = \infty$ and $x^0 \to \pm\infty$ we obtain

$$\delta b^*_\kappa = -\int \nabla^*_\kappa \nabla^{*\lambda'} \mathcal{F}_{xx'} \delta b_{\lambda'} dx'. \tag{6.8}$$

and consequently

$$\frac{\delta b^*_\kappa}{\delta b_{\lambda'}} = -\nabla^*_\kappa \nabla^{*\lambda'} \mathcal{F}_{xx'}. \tag{6.9}$$

An isovector b_λ is now defined as the difference between the full potential b_λ and b^*_λ or, equivalently,

$$b_\lambda = b_\lambda + b^*_\lambda. \tag{6.10}$$

If follows from (6.10) and (6.2) that

$$\nabla^{*\kappa} b_\kappa = 0. \tag{6.11}$$

From (6.10) and (6.9) we see that

$$\frac{\delta b_\kappa}{\delta b_{\lambda'}} = \delta^{\lambda'}_\kappa(x - x') + \nabla^*_\kappa \nabla^{*\lambda'} \mathcal{F}_{xx'}. \tag{6.12}$$

For infinitesimal local gauge transformations generated by η, where

$$|x^\alpha| |\eta| \leq B, \qquad |x^\alpha| \to \infty,$$

$$\eta = 0, \qquad , x^0 = \to \pm\infty,$$

the b_κ transform as isovectors, while b^*_κ undergo the change

$$\delta b^*_\kappa = -\nabla^*_\kappa \eta. \tag{6.13}$$

Using the decomposition (6.10) we express the gauge field in the form

$$B_{\kappa\lambda} = \nabla_\kappa b_\lambda - \nabla^*_\lambda b_\kappa. \tag{6.14}$$

Taking the divergence $\nabla^{*\kappa}$ of (6.14) and using the commutativity of $\nabla^{*\kappa}$ and $\nabla^{*\lambda}$, which is a consequence of the integrability of b^*_κ, yields

$$\nabla^{*\kappa} \nabla_\kappa b_\lambda = \nabla^{*\kappa} B_{\kappa\lambda}. \tag{6.15}$$

With the functional $\mathcal{F}_{xx'}$ we obtain

$$b_\lambda = -\int \nabla^{*\kappa} \mathcal{F}_{xx'} B_{\kappa\lambda} dx'. \tag{6.16}$$

This equation has the same form as the representation of the electromagnetic potentials given in Ref. 2. However, here the potentials occur explicitly in the integrand through the covariant derivative $\nabla^{*\kappa}$ and the functional $\mathcal{F}_{xx'}$.

For the electromagnetic field the gauge covariant derivative in (6.16) becomes the derivative ∂^κ, the functional $\mathcal{F}_{xx'}$ becomes the Green's function $|x - x'|^{-2}$ of the de Alembertian operator, and (6.16) acquires

the form

$$\mathcal{R}_\lambda = -\int \partial^{\kappa'} |x - x'|^{-2} F_{\kappa',\lambda} dx', \tag{6.17}$$

where

$$F_{\kappa\lambda} = \partial_\kappa A_\lambda - \partial_\lambda A_\kappa \tag{6.18}$$

is the electromagnetic field tensor and A_κ are the electromagnetic potentials. The $\mathcal{R}_\lambda$ are the gauge invariant potentials obtained by Rohrlich and Strocchi.[2] The same expression for an invariant potential has been obtained by Goldberg[17] by unitary transformation of electromagnetic potentials coupled to a spinor field.

Via (6.18) and

$$\partial_\kappa \partial^\kappa |x - x'|^{-2} = \delta(x - x') \tag{6.19}$$

in (6.17) yields

$$\mathcal{R}_\lambda = A_\lambda + \int \partial^{\kappa'} \partial_\lambda |x - x^1|^{-2} A_{\kappa'} dx'. \tag{6.20}$$

From our point of view the last term is the integrable part A^*_λ of the full potential A_λ, and (6.20) gives us the decomposition

$$A_\lambda = \mathcal{R}_\lambda + A^*_\lambda, \tag{6.21}$$

where the functional A^*_λ is defined by the integral in (6.20).

As we have seen, the decomposition of the form (6.20) admits a generalization to non-Abelian gauge fields. However, the fact that the functional $\mathcal{R}_\lambda$ can be expressed in terms of the field alone seems to be an accident due to the non-Abelian property of the electromagnetic field.

The generalization of the decomposition (6.20) to the non-Abelian case yields the homogeneously transforming part b_κ which is a functional of the full potential and cannot be expressed in terms of the field $B_{\kappa\lambda}$ alone.

7. NONLOCAL LAGRANGIANS

One can use the transverse-longitudinal decomposition or the decomposition based on the Lorentz condition to form functionals which are scalars under local gauge transformations. For example, the "massive" YM Lagrangian

$$\tfrac{1}{4} B_{\kappa\lambda} B^{\kappa\lambda} - \tfrac{1}{2}\mu(\varphi^2 + b^T_\alpha b^{T\alpha}). \tag{7.1}$$

is invariant under local gauge transformations; but it is not Lorentz invariant. Further, it is nonlocal since φ and b^T_α are functionals of b_α.

Using the expressions for the variations of φ and b^T_α which were obtained in the analysis of the transverse-longitudinal decomposition, the action principle

$$\delta \int [\tfrac{1}{4} B_{\kappa\lambda} B^{\kappa\lambda} - \tfrac{1}{2}\mu(\varphi^2 + b^T_\alpha b^{T\alpha})] dx dt = 0 \tag{7.2}$$

gives the nonlocal field equations

$$-\nabla_\kappa B^\kappa{}_\beta - \mu b^T_\beta + \int \nabla^*_\beta D_{xx'} \nabla^*_0 \varphi(x') dx' = 0, \tag{7.3}$$

$$-\nabla_\alpha B^\alpha_0 - \mu\varphi = 0. \tag{7.4}$$

Taking the divergence ∇^β of (7.3) and the time derivative ∇^0 of (7.4) and adding the resulting equations shows that

$$\nabla^\beta b^T_\beta = 0 \tag{7.5}$$

is an integrability condition for the equations of motion and constraint.

A gauge and Lorentz invariant mass term can be constructed with the aid of the decomposition

$$b_\kappa = b_\kappa + b^*_\kappa \tag{7.6}$$

We form the Lagrangian

$$\mathcal{L} = \tfrac{1}{4} B_{\kappa\lambda} B^{\kappa\lambda} - \tfrac{1}{2}\mu b_\kappa b^\kappa \tag{7.7}$$

which is nonlocal since b_κ is a functional of b_κ. Using the variation of b_κ given by (6.12), the action principle yields

$$-\nabla_\kappa B^\kappa{}_\lambda - \mu b_\lambda - \mu \int \nabla^{*\kappa'} \nabla^*_\lambda \mathfrak{F}_{xx'} b_{\kappa'} dx' dt = 0. \tag{7.8}$$

If the Lorentz condition (6.11) is acknowledged, we obtain the field equation

$$-\nabla_K B^K{}_\lambda - \mu b_\lambda = 0. \tag{7.9}$$

The Lorentz condition (6.11) is now an integrability condition for (7.9).

The invariant decompositions may also be used to couple the YM field invariantly to other fields. For example, consider the spinor–isospinor field $\psi_i{}^a$ where $i, j, = 1, 2, 3$ are algebra indices and $a, b, \cdots = 1, 2, 3, 4$ are Dirac spinor indices. We take the Lagrangian density

$$\mathcal{L} = \tfrac{1}{4} B_{\kappa\lambda i} B^{\kappa\lambda i} + i\bar{\psi}_{ia} \gamma^\kappa{}_b{}^a \nabla_\kappa \psi^{ib} - m\psi_{ia}\psi^{ia}, \tag{7.10}$$

for which the action principle gives the field equations

$$\nabla_\kappa B^\kappa{}_{\lambda i} = i c_{ijk} \psi^{\dagger j} \gamma_\lambda \psi^k, \tag{7.11}$$

$$i\gamma^\kappa \nabla_\kappa \psi_i - m\psi_i = 0, \tag{7.12}$$

where spinor indices are suppressed and $\psi^\dagger = \bar{\psi}_{ia}$ is the adjoint spinor. Introducing the Lorentz and gauge invariant decomposition (7.6) into Eqs. (7.11) and (7.12) gives for any basis i'

$$\nabla^*_\kappa B^\kappa{}_{\lambda i'} = C_{i'j'k'} b_\kappa{}^{j'} B^\kappa{}_\lambda{}^{k'} + i C_{i'j'k'} \psi^{\dagger j'} \gamma_\lambda \psi^{k'}, \tag{7.13}$$

$$i\gamma^\kappa \nabla^*_\kappa \psi_{i'} - m\psi_{i'} = C_{i'j'k'} \gamma^\kappa b_\kappa{}^{j'} \psi^{k'}. \tag{7.14}$$

The self- and external coupling terms in (7.13) and (7.14) are gauge scalars, but nonlocal because b_κ is a functional of b_κ. We do not attempt to attach any physical significance to the integrable part b^*_κ of the full potential. Rather, we regard the b^*_κ in somewhat the same way as one would view parameters of a linear connection for event vectors which arise in the flat Minkowski space as a consequence of using curvilinear coordinates.

In the theory of the classical field, which is pursued here, the functionals b^*_κ can be removed entirely from the theory by a process which in tensor analy-

sis[14] is called strangling of indices. Let us take basis vectors $\overset{i}{e}_j$ in the algebra space which are adapted to the Lorentz gauge. That is, in the gauge $b^*_\kappa = 0$, choose basis vectors

$$\overset{i}{e}_j \overset{*}{=} \delta^i_j, \tag{7.15}$$

where the $\overset{*}{=}$ means that (7.15) is asserted only in bases such that $b^*_\kappa = 0$. The index i in (7.15) is dead and serves only to label the independent basis vectors. A change in the local isobases represented by the local orthogonal transformation matrix $S^j_{j'}$ gives new components for the basis vectors

$$\overset{i}{e}_{j'} = S^j_{j'} \overset{i}{e}_j. \tag{7.16}$$

It follows by use of the nonhomogeneous transformation law (3.26) for Γ^*_κ that, in any basis j',

$$\nabla^*_\mu \overset{i}{e}_{j'} = 0. \tag{7.17}$$

Equation (7.17) shows that we should regard the basis vectors $\overset{i}{e}_{j'}$ as functionals of Γ^*_μ which in turn are functionals of the full potentials Γ_μ.

The free indices in Eqs. (7.13) and (7.14) are strangled by transvection of these equations with the basis vectors $\overset{i}{e}_{j'}$. Then, with the aid of the property (7.17), transvection of (7.13) and (7.14) gives

$$i\gamma^\kappa \partial_\kappa \underset{i}{\psi} - m \underset{i}{\psi} = C_{ijk} \gamma^\kappa \overset{j}{b}_\kappa \overset{k}{\psi}, \tag{7.18}$$

$$\partial_\kappa \underset{i}{B^\kappa}{}_\lambda = C_{ijk} \overset{j}{b}_\kappa \overset{k}{B^\kappa}{}_\lambda + iC_{ijk} \overset{j}{\psi}{}^\dagger \gamma_\lambda \overset{k}{\psi}, \tag{7.19}$$

where we have placed the isospin index above or below the kernels (except for the structure constants) to emphasize that all the isospin indices in (7.18) and (7.19) are dead and are not subject to gauge transformation. Equations (7.18) and (7.19) thus involve only isotopic scalars.

We could apply the same method of strangling indices to the Hamiltonian form of the YM equations. There, one would choose basis vectors $\overset{i}{e}_j$ adapted to the transverse gauge where, in the bases such that

$$\partial_\alpha b^{T\alpha} \overset{*}{=} 0, \quad b^*_\alpha \overset{*}{=} 0, \tag{7.20}$$

the basis vectors are given by

$$\overset{i}{e}_j \overset{*}{=} \delta^i_j, \tag{7.21}$$

and the vectors $\overset{i}{e}_{j'}$ in any local base j' satisfy

$$\nabla^*_\alpha \overset{i}{e}_{j'} = 0, \quad \nabla^*_0 \overset{i}{e}_{j'} = 0. \tag{7.22}$$

8. REMARKS

The arbitrariness of the YM potentials up to local gauge transformations is not a source of difficulty in the theory of the classical field. One removes the arbitrariness by fixing a gauge by the imposition of a gauge-variant condition on the potentials. However, when one attempts to fix a gauge in a quantized theory, just as in quantum electrodynamics, the problem of compatibility of the gauge conditions, field equations, and the quantum rules arise. One of the motivations for this study was that the gauge-invariant decompositions of the potentials might be useful in the quantum theory.

There are several apparently different quantized theories which have been shown to be gauge and Lorentz invariant.[6,19,20,11] Whether these theories are physically equivalent seems not to be known. If one follows the ideas of Schwinger,[6] the b^T_α of the work here would be identified as the true quantum variable while the b^*_α would play the role of Schwinger's group parameters. If it would be possible to develop a consistent quantum theory with this identification, it should be manifestly gauge invariant. In Loos' canonical quantization[11] the secondary constraint on the YM field is $\nabla^\alpha \pi_{\alpha} \Psi = 0$, where Ψ is a quantum state functional. This constraint is satisfied by any state $\Psi[b_\alpha]$ which is invariant under the change $b_\alpha \to b_\alpha - \nabla_\alpha \eta$, i.e., under a gauge transformation generated by η where $|x\alpha|^2 \eta \leq B$ as $|x^\alpha| \to \infty$. The secondary constraint is satisfied automatically be states which are scalar functionals of the homogeneously transforming quantities $b^T_\alpha[b], \nabla_\beta b^T_\alpha[b], \nabla_\gamma \nabla_\beta b^T_\alpha[b]$. For example

$$\Psi = \exp\left(-\int b^T_{\alpha i} b^{T\alpha i} dx\right), \quad |x^\alpha|^2 |b_{\beta i}| \leq B, \ |x^\alpha| \to \infty,$$

is such an invariant functional. If one adopts the canonical quantization rules of Ref. 11 for b_α and π_α, then possibly the quantum rules for b^T_α and π^T_α are to a large extent committed. But these rules have not been worked out.

After constructing the gauge scalar functional $\mathfrak{R}_\kappa$, Rohrlich and Strocchi[2] found quantization rules for the $\mathfrak{R}_\kappa$ which are manifestly gauge and Lorentz invariant. It may be possible to carry out a similar program for the homogeneously transforming functional b_κ. However, one anticipates complications which usually attend attempts to generalize an Abelian to a non-Abelian theory.

ACKNOWLEDGMENTS

I extend thanks to Professor H. G. Loos for enjoyable and illuminating discussions on Yang–Mills theory. Discussions with Dr. C. Uzes and Dr. J. Grodnik were much appreciated, and I thank the Aspen Institute for Physics, where part of this work was carried out, for their hospitality.

APPENDIX: A SERIES REPRESENTATION OF $\mathfrak{D}_{xx'}$

The functional $\mathfrak{D}_{xx'}$, which was encountered in the study of the transverse-longitudinal decomposition of the gauge potentials, was defined (Sec. 3) as the solution of the equations

$$\nabla^\alpha \nabla^*_\alpha \mathfrak{D}_{xx',i}{}^{i'} = \delta^{i'}_i (x - x'). \tag{A1}$$

Equation (A1) is the gauge invariant version of

$$\partial^\alpha \partial_\alpha \mathfrak{D}_{xx',}{}^{i'} - \Gamma^{T}_{\alpha i}{}^{j} \partial^\alpha \mathfrak{D}_{xx',j}{}^{i'} = \delta_i{}^{i'}(x - x'), \tag{A2}$$

where $\Gamma^{T}_{\alpha i}{}^{j}$ are the transverse potentials in the transverse gauge, i.e., $\partial_\alpha \Gamma^{T\alpha j}_i = 0$. The functional $\mathfrak{D}_{xx'}$ which satisfies (A2) arises also in the quantum theory of the YM field studied in Refs. 6 and 10. We derive in

this appendix a series representation of $\mathfrak{D}_{xx'}$. To generate the series, we first derive the functional derivative equation which is satisfied by $\mathfrak{D}_{xx'}$. Let $\mathfrak{D}_{xx'}$ denote the variation brought about by the variation of the coefficients in (A2). It is understood that the variation is taken within the transversality condition so that $\Gamma^T_\alpha + \delta\Gamma^T_\alpha$ also satisfies the transversality condition. Taking the variation of (A2) shows that $\delta\mathfrak{D}_{xx'}$ satisfies

$$\partial^\alpha\partial_\alpha\delta\mathfrak{D}_{xx'i}{}^{i'} - \Gamma^{Tj}_{\alpha i}\partial^\alpha\mathfrak{D}_{xx'j}{}^{i'} = \delta\Gamma^{Tj}_{\alpha i}\partial^\alpha\mathfrak{D}_{xx'i}{}^{i'}. \quad \text{(A3)}$$

If $\mathfrak{D}_{xx'}$ exists, then we can write the integral representation

$$\delta\mathfrak{D}_{xx'i}{}^{i_1} = \int\mathfrak{D}_{xx_1}\delta\Gamma_{\alpha_1 i_1}{}^{j_1}\partial^{\alpha_1}\mathfrak{D}_{x_1x'j_1}{}^{i'}dx_1. \quad \text{(A4)}$$

Then the functional derivative of $\mathfrak{D}_{xx'}$ is given by

$$\frac{\delta\mathfrak{D}_{xx'}{}^{i'}}{\delta\Gamma^T_{\alpha_1 i_1}{}^{j_1}} = \mathfrak{D}_{xx_1}{}^{i_1}\partial^{\alpha_1}\mathfrak{D}_{x_1x'j_1}{}^{i'}. \quad \text{(A5)}$$

For later use it is convenient to have (A5) in terms of b^T_α. The equation

$$\frac{\delta\mathfrak{D}_{xx_1i}{}^{i'}}{\delta b^T_{\alpha_1}{}^{k_1}} = \mathfrak{D}_{xx_1}{}^{i_1}C_{k_1i_1}{}^{j_1}\mathfrak{D}_{x_1x'j_1}{}^{i'}. \quad \text{(A6)}$$

has the same mathematical content as (A5) and is easily deduced by introducing the structure constants and b^T_α in the integrand of (A4).

Let us try to express the solution of (A5) as a power series of functionals[18] (a Volterra series) at the origin of Ω, i.e., at the point $\Gamma^T_\alpha = 0$, $\Gamma^*_\alpha = 0$. We write formally

$$\begin{aligned}\mathfrak{D}_{xx_1i}{}^{i'} &= \overline{\mathfrak{D}}_{xx_1i}{}^{i'} + \int_{x_1}\frac{\delta\overline{\mathfrak{D}}_{xx_1i}{}^{i'}}{\delta\Gamma^T_{\alpha_1i_1}{}^{j_1}}\Gamma^T_{\alpha_1t_1}{}^{j_1}dx_1 \\ &+ \tfrac{1}{2}\int_{x_1}\int_{x_2}\frac{\delta^2\overline{\mathfrak{D}}_{xx'i}{}^{i'}}{\delta\Gamma^T_{\alpha_2i_2}{}^{j_2}\delta\Gamma^T_{\alpha_1i_1}{}^{j_1}}\Gamma^T_{\alpha_1i_1}{}^{j_1}dx_1dx_2 + \cdots \\ &+ \frac{1}{n}\int_{x_1}\cdots\int_{x_2}\frac{\delta^n\overline{\mathfrak{D}}_{xx'i}{}^{i'}}{\delta\Gamma^T_{\alpha_ni_n}{}^{j_n}\cdots\delta\Gamma^T_{\alpha_1i_1}{}^{j_1}} \\ &\times\Gamma^T_{\alpha_1i_1}{}^{j_1}\cdots\Gamma^T_{\alpha_ni_n}{}^{j_n}dx_1\cdots dx_n. \end{aligned} \quad \text{(A7)}$$

The bar over the $\mathfrak{D}$ indicates that the functional derivatives are evaluated at $\Gamma_\alpha = 0$. The first functional derivative is obtained from the functional derivative equation itself. The higher order functional derivatives are obtained by repeated functional differentiation of (A5) and the subsequent use of (A5) to eliminate the derivatives from the right-hand side. For example, for the second functional derivative we obtain

$$\begin{aligned}\frac{\delta^2\mathfrak{D}_{xx'i}{}^{i'}}{\delta\Gamma^T_{\alpha_2i_2}{}^{j_2}\delta\Gamma^T_{\alpha_1i_1}{}^{j_1}} &= \mathfrak{D}_{xx_2}{}^{i_2}\partial^{\alpha_2}\mathfrak{D}_{x_2x_1j_2}{}^{i_1}\partial^{\alpha_1}\mathfrak{D}_{x_1x'j_1}{}^{i'} \\ &+ \mathfrak{D}_{xx_1}{}^{i_1}\partial^{\alpha_1}\mathfrak{D}_{x_1x_2j_1}{}^{i_2}\partial^{\alpha_2}\mathfrak{D}_{x_2x'j_2}{}^{i'}.\end{aligned}$$

For each term in the Volterra expansion (A7) we can collect under the nth integral the $n!$ terms of the nth functional derivative and, thereby, obtain

$$\begin{aligned}\mathfrak{D}_{xx'i}{}^{i'} &= \overline{\mathfrak{D}}_{xx'i}{}^{i'} + \int_{x_1}\overline{\mathfrak{D}}_{xx_1}{}^{i_1}\partial^{\alpha_1}\overline{\mathfrak{D}}_{x_1x'j_1}{}^{i'}\Gamma^T_{\alpha_1i_1}{}^{j_1}dx_1 \\ &+ \cdots + \int_{x_1}\cdots\int_{x_n}\overline{\mathfrak{D}}_{xx_1i}{}^{i_1}\partial^{\alpha_1}\overline{\mathfrak{D}}_{x_1x_2\alpha_1}{}^{i_2}\cdots \\ &\times\partial^{\alpha_n}\overline{\mathfrak{D}}_{x_nx'j_n}{}^{i'}\Gamma^T_{\alpha_ni_n}{}^{j_n}dx_1\cdots dx \end{aligned} \quad \text{(A8)}$$

Suppressing the isotopic spin indices and the space coordinates we write (A8) in the compact form

$$\begin{aligned}\mathfrak{D} = \overline{\mathfrak{D}} * [&\delta + \Gamma^{T\alpha_1}\partial_{\alpha_1}\overline{\mathfrak{D}} + \Gamma^{T\alpha_1}\partial_{\alpha_1}\overline{\mathfrak{D}} * \Gamma^{T\alpha_2}\partial_{\alpha_2}\overline{\mathfrak{D}} \\ &+ \cdots + (\Gamma^{T\alpha_1}\partial_{\alpha_1}\overline{\mathfrak{D}} * \cdots * \Gamma^{T\alpha_n}\partial_{\alpha_n}\overline{\mathfrak{D}}) + \cdots],\end{aligned} \quad \text{(A9)}$$

where the second factor on the right-hand side of (A9) is a geometric series in $\Gamma^{T\alpha}\partial_\alpha\overline{\mathfrak{D}}$. In (A9), δ is the Dirac-delta function and we have used the * symbol to indicate integration.

Schwinger[6] has noted that $\mathfrak{D}$ is self-adjoint. To prove this we will examine the series (A8) term by term. First we note from the defining differential equation (A2) that it follows that $\mathfrak{D}$ evaluated at $\Gamma_\alpha = 0$ is given by

$$\mathfrak{D}_{xx'i}{}^{i'} = \delta^{i'}_i/|x - x'|. \quad \text{(A10)}$$

Equation (A2) also implies that

$$\partial_\alpha\overline{\mathfrak{D}}_{xx'i}{}^{i'} = \delta^{i'}_i\partial_\alpha/|x - x'|. \quad \text{(A11)}$$

It is easy to show the self-adjointness of $\mathfrak{D}$ if we introduce (A10) and (A11) into the Volterra expansion (A8). Then (A8) becomes

$$\begin{aligned}\mathfrak{D}_{xx'i}{}^{i'} &= \frac{1}{|x - x'|}\delta^{i'}_i + \int_{x_1}\frac{1}{|x - x_1|}\Gamma^T_{\alpha_1i}{}^{i'} \\ &\times\partial^{\alpha_1}\frac{1}{|x_1 - x'|}dx_1 + \cdots + \int_{x_1}\cdots\int_{x_n}\frac{1}{|x - x_1|}\Gamma^T_{\alpha_1i}{}^{i_1} \\ &\times\partial^{\alpha_1}\frac{1}{|x_1 - x_2|}\cdots\Gamma_{\alpha_ni_n}{}^{i'}\partial^{\alpha_n} \\ &\times\frac{1}{|x_n - x_2|}dx_1\cdots dx_n \end{aligned} \quad \text{(A12)}$$

and the adjoint of $\mathfrak{D}_{xx'i}{}^{i'}$ is

$$\begin{aligned}\mathfrak{D}_{x'x}{}^{i'}{}_i &= \frac{1}{|x - x^1|}\delta^{i'}_i + \int_{x_1}\frac{1}{|x^1 - x_1|}\Gamma^T_{\alpha_1}{}^{i'}{}_i \\ &\times\partial^{\alpha_1}\frac{1}{|x_1 - x|}dx_1 + \cdots + \int_{x_1}\cdots\int_{x_n}\frac{1}{|x^1 - x_1|}\Gamma^{T\ i'}_{\alpha_1}{}_{i_1} \\ &\times\partial^{\alpha_1}\frac{1}{|x_1 - x_2|}\cdots\Gamma^T_{\alpha_n}{}^{i_n}{}_i\partial^{\alpha_n}\frac{1}{|x_n - x|}dx_1\cdots dx_n \end{aligned} \quad \text{(A13)}$$

Use of the divergenceless and anti-self-adjoint property of $\Gamma^{Ti'}_{\alpha i}$, parts integration, and the self-adjointness of $1/|x - x'|$ permits rearrangement of the terms in the integrands of (A13) so that it is evident that

$$\mathfrak{D}_{xx'i}{}^{i'} = \mathfrak{D}_{x'x}{}^{i'}{}_i. \quad \text{(A14)}$$

The convergence of the series for $\mathfrak{D}$ is not considered here except to note that if b^T_α is a member of the configuration space Ω then each multiple integral in the Volterra series exists for $x \neq x'$.

* Partially supported by the West Virginia University Senate Committee on Research.

1 F. J. Belinfante, Phys. Rev. **128**, 2832 (1961).
2 F. Rohrlich and F. Strocchi, Phys. Rev. **139**, B476 (1965).
3 B. DeWitt, Phys. Rev. **125**, 2189 (1965).
4 S. Mandelstam, Ann. Phys. (N.Y.)**19**, 1 (1962).
5 C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
6 J. Schwinger, Nuovo Cimento **30**, 278 (1963).
7 D. G. Boulware, Ann. Phys. (N.Y.) 56, 740 (1970), has studied the representation of any potential as the gauge transform of a transverse potential, which is equivalent to the decomposition based on the Lorentz condition used here.
8 J. Madore, Nuovo Cimento **4**, 599 (1971), uses a decomposition of the same form in a study of massive Yang–Mills fields.
9 H. G. Loos, Nuovo Cimento **53A**, 201 (1968).
10 H. G. Loos, J. Math. Phys. **11**, 3258 (1970).
11 H. G. Loos, Phys. Rev. **188**, 2342 (1969).
12 Actually $\Gamma_{\alpha i}{}^{j} = -b_{\alpha}^{k} C_{ki}{}^{j}$ and $\Gamma^{*}_{\alpha}{}^{j}{}_{i} = -b^{*}_{\alpha}{}^{k} C_{ki}{}^{j}$ transform as parameters of connection, and $\Gamma_{\alpha i}{}^{j}$ defines the equivalence of vectors belonging to the algebra space at neighboring points in the event space.
13 B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1964).
14 J. A. Schouten, *Ricci-Calculus* (Springer, New York, 1954), p. 14.
15 H. G. Loos (private communication).
16 R. L. Arnowitt and S. I. Fickler, Phys. Rev. **127**, 1821 (1962).
17 I. Goldberg, Phys. Rev. **139**, 1665 (1965).
18 V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover, New York, 1959), p. 35.
19 B. S. DeWitt, Phys. Rev. **162**, 1195 (1967).
20 S. Mandelstam, Phys. Rev. **175**, 1580 (1968).

Null Plane Restrictions of Current Commutators*

Richard A. Brandt† and Paul Otterson

Department of Physics, New York University, New York, New York 10012

(Received 19 April 1971)

We consider some mathematical aspects of the problems of defining restrictions of quantum fields and commutators of such fields to null planes. We give precise meanings to these restrictions and discuss how these lead to unambiguous derivations of the usual formal results. We discuss and relate various definitions of null plane charges and derive some of their properties such as vacuum annihilation. We define and exhibit finite null plane restrictions for causal solutions of the Klein–Gordon equation. Commutator functions defined by integral representations with various spectral functions are then considered. Light cone operator product expansions are used to calculate some null plane current commutators. In this way we can give precise derivations of Fubini sum rules and electroproduction structure function asymptotic behavior.

1. INTRODUCTION

There is growing evidence that many interesting physical processes can be understood in terms of the behavior of quantum field operators near the light cone (LC). In this paper we shall study some mathematical aspects of this LC physics. This will, on the one hand, put some previous work on a sounder mathematical footing and, on the other hand, will indicate ways to extend the domain of applicability of LC techniques.

One of the earliest encounters of the LC in particle physics was the observation that the infinite momentum techniques[1,2] of current algebra lead to statements about the LC commutator of the currents.[3] In order to describe this development, we consider the forward spin-averaged connected covariant retarded current-proton scattering amplitude

$$T_{\mu\nu}^{ab} = i\int d^4x\, e^{iqx}\theta(x_0)\langle p|[J_\mu^a(x), J_\nu^b(0)]|p\rangle + g_{\mu i} g_\nu^i d^{abc}\Sigma^c(0) = p_\mu p_\nu T_2^{ab}(\kappa,\nu) + \cdots . \tag{1.1}$$

Here $\nu = q\cdot p$, $\kappa = q^2$, $p^2 = 1$, and we have allowed for the presence of an operator Schwinger term of the form appearing in

$$\delta(x_0)[J_0^a(x), J_k^b(0)] = if^{abc}J_k^c(0) + id^{abc}\Sigma^c(0)\,\partial_k\delta(x). \tag{1.2}$$

In this paper we shall explicitly consider only the $SU(3)$ vector current $J_\mu^a(x)$ although all our equations can be immediately generalized to include the axial vector currents. The absorptive part of (1.1) is

$$W_{\mu\nu}^{ab} = (1/2\pi)\int d^4x\, e^{iqx}\langle p|[J_\mu^a(x), J_\nu^b(0)]|p\rangle = p_\mu p_\nu W_2^{ab}(\kappa,\nu) + \cdots \tag{1.3}$$

so that

$$W_2 = (1/\pi)\ \mathrm{Im}\ T_2. \tag{1.4}$$

We write

$$T_2^{(\pm)ab} = \tfrac{1}{2}(T_2^{ab} \pm T_2^{ba})$$

and similarly define $W_2^{(\pm)ab}$.

The currents are, of course, assumed to satisfy the local chiral algebra

$$\delta(x_0)[J_0^a(x), J_0^b(0)] = if^{abc}J_0^c\,\delta^4(x) \tag{1.5}$$

etc. When this relation is expressed in terms of a q_0 integral of (1.3) with $\mathbf{q}$ fixed, and an infinite momentum limit is taken inside of this integral, there results the Fubini–Dashen–Gell-Mann sum rule[2]

$$\int d\nu\, W_2^{(-)ab}(\kappa,\nu) = -f^{abc}F^c, \tag{1.6}$$

where we have defined the forward form factor by

$$ip_\mu F^c = \langle p|J_\mu^c(0)|p\rangle.$$

Equation (1.6) corresponds to the asymptotic behavior

$$T_2^{(-)ab}(\kappa,\nu) \underset{R}{\longrightarrow} (1/\nu) f^{abc}F^c$$

in the Regge limit (R limit: $\nu\to\infty$, κ fixed) characteristic of a fixed pole in the complex j plane at $j = 1$.[4]

Because of the need for the above infinite momentum assumption, (1.6) is not implied by the equal-time commutation relation (1.5), but rather by the LC commutation relation[3,5]

$$\tfrac{1}{2}\int dx_+\,\delta(x_-)[J_-^a(x), J_-^b(0)] = if^{abc}J_-^c(0)\,\delta(x_-, x_\perp). \tag{1.7}$$

Equation (1.6) follows immediately from (1.7) by integrating (1.3) over ν in the frame $q_- = 0, p = (1,\mathbf{0})$, in which

$$T_2^{ab}(\kappa,\nu) = i\int d^4x\, e^{iqx}\theta(x_0)\langle p|[J_-^a(x), J_-^b(0)]|p\rangle - d^{abc}\Sigma^c(0), \tag{1.8}$$

$$W_2^{ab}(\kappa,\nu) = \frac{1}{2\pi}\int d^4x\, e^{iqx}\langle p|[J_-^a(x), J_-^b(0)]|p\rangle. \tag{1.9}$$

The relation (1.7) is, however, a formal one since the indicated formal operation on the left side need not

* Partially supported by the West Virginia University Senate Committee on Research.

1 F. J. Belinfante, Phys. Rev. **128**, 2832 (1961).
2 F. Rohrlich and F. Strocchi, Phys. Rev. **139**, B476 (1965).
3 B. DeWitt, Phys. Rev. **125**, 2189 (1965).
4 S. Mandelstam, Ann. Phys. (N.Y.)**19**, 1 (1962).
5 C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).
6 J. Schwinger, Nuovo Cimento **30**, 278 (1963).
7 D. G. Boulware, Ann. Phys. (N.Y.) 56, 740 (1970), has studied the representation of any potential as the gauge transform of a transverse potential, which is equivalent to the decomposition based on the Lorentz condition used here.
8 J. Madore, Nuovo Cimento **4**, 599 (1971), uses a decomposition of the same form in a study of massive Yang–Mills fields.
9 H. G. Loos, Nuovo Cimento **53A**, 201 (1968).
10 H. G. Loos, J. Math. Phys. **11**, 3258 (1970).
11 H. G. Loos, Phys. Rev. **188**, 2342 (1969).
12 Actually $\Gamma_{\alpha i}{}^{j} = -b_{\alpha}^{k} C_{ki}{}^{j}$ and $\Gamma^{*}_{\alpha}{}^{j}{}_{i} = -b^{*}_{\alpha}{}^{k} C_{ki}{}^{j}$ transform as parameters of connection, and $\Gamma_{\alpha i}{}^{j}$ defines the equivalence of vectors belonging to the algebra space at neighboring points in the event space.
13 B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1964).
14 J. A. Schouten, *Ricci-Calculus* (Springer, New York, 1954), p. 14.
15 H. G. Loos (private communication).
16 R. L. Arnowitt and S. I. Fickler, Phys. Rev. **127**, 1821 (1962).
17 I. Goldberg, Phys. Rev. **139**, 1665 (1965).
18 V. Volterra, *Theory of Functionals and of Integral and Integro-Differential Equations* (Dover, New York, 1959), p. 35.
19 B. S. DeWitt, Phys. Rev. **162**, 1195 (1967).
20 S. Mandelstam, Phys. Rev. **175**, 1580 (1968).

Null Plane Restrictions of Current Commutators*

Richard A. Brandt† and Paul Otterson

Department of Physics, New York University, New York, New York 10012

(Received 19 April 1971)

We consider some mathematical aspects of the problems of defining restrictions of quantum fields and commutators of such fields to null planes. We give precise meanings to these restrictions and discuss how these lead to unambiguous derivations of the usual formal results. We discuss and relate various definitions of null plane charges and derive some of their properties such as vacuum annihilation. We define and exhibit finite null plane restrictions for causal solutions of the Klein–Gordon equation. Commutator functions defined by integral representations with various spectral functions are then considered. Light cone operator product expansions are used to calculate some null plane current commutators. In this way we can give precise derivations of Fubini sum rules and electroproduction structure function asymptotic behavior.

1. INTRODUCTION

There is growing evidence that many interesting physical processes can be understood in terms of the behavior of quantum field operators near the light cone (LC). In this paper we shall study some mathematical aspects of this LC physics. This will, on the one hand, put some previous work on a sounder mathematical footing and, on the other hand, will indicate ways to extend the domain of applicability of LC techniques.

One of the earliest encounters of the LC in particle physics was the observation that the infinite momentum techniques[1,2] of current algebra lead to statements about the LC commutator of the currents.[3] In order to describe this development, we consider the forward spin-averaged connected covariant retarded current-proton scattering amplitude

$$T_{\mu\nu}^{ab} = i\int d^4x\, e^{iqx}\theta(x_0)\langle p|[J_\mu^a(x), J_\nu^b(0)]|p\rangle + g_{\mu i} g_\nu^i d^{abc}\Sigma^c(0) = p_\mu p_\nu T_2^{ab}(\kappa,\nu) + \cdots . \tag{1.1}$$

Here $\nu = q\cdot p$, $\kappa = q^2$, $p^2 = 1$, and we have allowed for the presence of an operator Schwinger term of the form appearing in

$$\delta(x_0)[J_0^a(x), J_k^b(0)] = if^{abc}J_k^c(0) + id^{abc}\Sigma^c(0)\,\partial_k\delta(x). \tag{1.2}$$

In this paper we shall explicitly consider only the $SU(3)$ vector current $J_\mu^a(x)$ although all our equations can be immediately generalized to include the axial vector currents. The absorptive part of (1.1) is

$$W_{\mu\nu}^{ab} = (1/2\pi)\int d^4x\, e^{iqx}\langle p|[J_\mu^a(x), J_\nu^b(0)]|p\rangle = p_\mu p_\nu W_2^{ab}(\kappa,\nu) + \cdots \tag{1.3}$$

so that

$$W_2 = (1/\pi)\ \mathrm{Im}\ T_2. \tag{1.4}$$

We write

$$T_2^{(\pm)ab} = \tfrac{1}{2}(T_2^{ab} \pm T_2^{ba})$$

and similarly define $W_2^{(\pm)ab}$.

The currents are, of course, assumed to satisfy the local chiral algebra

$$\delta(x_0)[J_0^a(x), J_0^b(0)] = if^{abc}J_0^c\,\delta^4(x) \tag{1.5}$$

etc. When this relation is expressed in terms of a q_0 integral of (1.3) with $\mathbf{q}$ fixed, and an infinite momentum limit is taken inside of this integral, there results the Fubini–Dashen–Gell-Mann sum rule[2]

$$\int d\nu\, W_2^{(-)ab}(\kappa,\nu) = -f^{abc}F^c, \tag{1.6}$$

where we have defined the forward form factor by

$$ip_\mu F^c = \langle p|J_\mu^c(0)|p\rangle.$$

Equation (1.6) corresponds to the asymptotic behavior

$$T_2^{(-)ab}(\kappa,\nu) \underset{R}{\longrightarrow} (1/\nu) f^{abc}F^c$$

in the Regge limit (R limit: $\nu\to\infty$, κ fixed) characteristic of a fixed pole in the complex j plane at $j = 1$.[4]

Because of the need for the above infinite momentum assumption, (1.6) is not implied by the equal-time commutation relation (1.5), but rather by the LC commutation relation[3,5]

$$\tfrac{1}{2}\int dx_+\,\delta(x_-)[J_-^a(x), J_-^b(0)] = if^{abc}J_-^c(0)\,\delta(x_-, x_\perp). \tag{1.7}$$

Equation (1.6) follows immediately from (1.7) by integrating (1.3) over ν in the frame $q_- = 0, p = (1,\mathbf{0})$, in which

$$T_2^{ab}(\kappa,\nu) = i\int d^4x\, e^{iqx}\theta(x_0)\langle p|[J_-^a(x), J_-^b(0)]|p\rangle - d^{abc}\Sigma^c(0), \tag{1.8}$$

$$W_2^{ab}(\kappa,\nu) = \frac{1}{2\pi}\int d^4x\, e^{iqx}\langle p|[J_-^a(x), J_-^b(0)]|p\rangle. \tag{1.9}$$

The relation (1.7) is, however, a formal one since the indicated formal operation on the left side need not

be well defined. A major purpose of this paper will be to study such questions.

The simplicity of (1.7) is directly reflected in a simplified Feynman diagram structure in nice theories for good currents in the infinite momentum limit. It was, in fact, early observed that, for good–good free field current components, the three-particle intermediate state contribution to the equal time commutator vanishes in the IM limit. These IM techniques were then exploited by Weinberg,[6] who showed that, for simpler scalar field theories in the IM limit, many undesirable diagrams disappear and the remaining ones can be described by new Feynman rules involving integrals over two-dimensional transverse momenta. Similar results were afterwards obtained in other models.[7]

Weinberg's results were generalized and formalized by Susskind,[3] who emphasized the relevance of the two-dimensional Galilean group. Susskind, and independently Bardakci and Segre,[3] also notice that the IM limit of an ordinary formal charge

$$Q(x_0) = \int d^3x J_0(x) \tag{1.10}$$

is, apart from a scale factor, a null plane (NP) charge

$$Q(x_-) = \int dx_+ dx_\perp J_-(x). \tag{1.11}$$

The connection with (1.7) is clear: If the IM limit can be taken inside of the sum over intermediate states implicit in the ET commutation relation (1.5), the LC commutation relation (1.7) is obtained. Thus this interchange is formally equivalent to the interchange of the IM limit and the q_0-integration mentioned below in Eq. (1.5).

Bardakci and Halpern[8] carried further the exploitation of the Galilean subgroup and constructed Galilean-invariant interacting relativistic systems. Quantum electrodynamics has also been formulated from this point of view.[9]

An independent source of evidence for the significance of the LC appeared when it was recognized[10] that the behavior of scattering amplitudes in the scaling limit[11] is controlled by the behavior of relevant matrix elements of current operators near the LC. The assumption of exact scaling then determines the strength of the leading LC singularities.[12-14] It is emphasized in Ref. 13 that exact scaling is formally equivalent to the existence of LC restrictions of certain components of the current commutator.

These works, however, do not predict scaling but only show that it is equivalent to other assumptions. Somewhat more predictive statements became possible when it was shown[15] that the behavior of products of local field operators near the LC could be described in terms of operator product expansions. In this framework the scaling is equivalent to the effective canonical dimensionality of the fields in the expansions.[15] The existence of these canonical expansions gave rise to a number of further fruitful applications of LC physics.[16,17] It seems possible that even the ordinary Regge limit can be described in these terms.[18]

A connection based on a universality assumption between these two situations (the IM limit and the scaling limit), where the LC is relevant was proposed in Ref. 10. This proposal was further studied in Ref. 12, and an operator generalization was compared with the canonical operator product expansions in Ref. 19. The universality statement was

$$\tfrac{1}{2}\int dx_+ \delta(x_-)\epsilon(x_+)[J_-^a(x), J_-^b(0)] = -id^{abc}\tfrac{1}{2}\{M^{-1}P_-,\ S^c(0)\}\delta(x_-, x_\perp) \tag{1.12}$$

or, with (1.7),

$$\int dx_+ \delta(x_-)\theta(x_+)[J_-^a(x), J_-^b(0)] = \tfrac{1}{2} iM^{-1}[f^{abc}MJ_-^c(0) - d^{abc}P_-S^c_{(0)}]\delta(x_-, x_\perp). \tag{1.13}$$

Here P_μ is the total momentum operator, $M^2 \equiv P_\mu P_\mu$, and $S^c(x)$ is the local scalar density in $U(12)$ ($\frac{1}{2}\bar{\psi}\lambda^a\psi$ in the free quark model).

Precise definitions of expressions like (1.12) and (1.13) will be discussed in Sec. 2. Let us now, however, proceed heuristically and note that, via the methods of Refs. 10 and 12, (1.11) is seen to imply the existence of a finite Bjorken asymptotic limit (limit $\nu \to \infty$, $\omega \equiv -q^2/\nu$ fixed):

$$\nu W_2^{(+)ab}(\kappa, \nu) \underset{A}{\longrightarrow} F_2^{ab}(\omega). \tag{1.14}$$

The asymptotic structure function is related to the Fourier transform of the light cone restriction of

$$\int dx_\perp \langle p|[J_-^a(x), J_-^b(0)]|p\rangle \equiv \theta(x_+x_-)\mathfrak{M}^{ab}(x_+, x_-) \tag{1.15}$$

according to

$$F_2^{ab}(\omega) = (i/2\pi)\int d\lambda\ e^{-i\lambda\omega}\epsilon(\lambda)\mathfrak{M}^{ab}(2\lambda, 0). \tag{1.16}$$

The presence of the $\theta(x_+x_-)$ factor in (1.15) follows from causality.[12] The existence of (1.16) follows from (1.11), which implies the existence of the integral

$$\tfrac{1}{2}\int dx_+ \epsilon(x_+)\tfrac{1}{2}\mathfrak{M}^{ab}(x_+, 0) = -id^{abc}D^c. \tag{1.17}$$

Here we have defined the forward scalar form factor

$$D^c \equiv \langle p|S^c(0)|p\rangle. \tag{1.18}$$

Equations (1.16) and (1.17) now give

$$F_2^{ab}(0) = (1/\pi)d^{abc}D^c, \tag{1.19}$$

and so (1.11) is seen to imply a constant asymptotic behavior for $\omega \to 0$. Finally, (1.12) implies (when $\Sigma^c = 0$) that

$$T_2^{ab}(\kappa, \nu) \underset{R'}{\longrightarrow} (1/\nu)(f^{abc}F^c + id^{abc}D^c) \tag{1.20}$$

in the Regge limit with $|\kappa|$ large (R' limit: $\nu \gg |\kappa| \gg 1$).

Finally, let us mention some previous mathematical investigations of null plane restrictions. Klauder, Leutweyler, and Streit[20] have studied quantum field theory on lightlike slabs, and Neville and Rohrlich[21] have studied quantum fields and differential equations with initial data specified on lightlike planes. The present work extends portions of their results. Neville and Rohrlich have calculated the null-plane restrictions of first-order derivatives of $\Delta(x; m)$; we extend the calculation to include derivatives of all orders and provide a natural mechanism for remov-

ing infinite coefficients which occur in previous calculations. An analysis is given of the lightlike charges obtained by integrating current densities over diffuse cutoff null slabs, then removing the cutoff and shrinking the slabs to planes. When this is done carefully, the resulting charge annihilates vacuum states; however, this alone provides no control over other matrix elements of the charges. In order to control other matrix elements, we insert convergence factors into the sequentially defined charges. These convergence factors are later used to obtain finite sum rules from bad–good and bad–bad commutator matrix elements.

In Sec. 2 we first define NP charges and discuss some of their properties. We then discuss NP restrictions of commutators. We consider progressively more complicated situations by using DGS representations with different spectral functions. In particular, the connection with Regge asymptotic behavior in the energy is discussed. Methods for giving meaning to formally divergent NP restrictions are presented. In Sec. 3, LC operator product expansions for the general $SU(3)$ current commutator are given and discussed. Symmetry and equal-time restrictions are noted and the A and R' limits are formally discussed. Section 4 combines the results of Secs. 2 and 3. The good–good NP restriction is calculated and the restrictions from requiring (1.7) and (1.2) are noted. We conclude by mentioning how good–bad and bad–bad commutators can be treated.

2. NULL PLANE RESTRICTIONS AND CHARGES

Our purpose here is to give precise meanings to the formal expression

$$\int dx'_- \delta(x_- - x'_-)\, a(x) \equiv A(x_-; x_+, x_\perp) \tag{2.1}$$

for the sharp NP density of the field operator $a(x)$, the formal expression

$$\int dx_+ dx_\perp A(x_-; x_+, x_\perp) \equiv \mathfrak{A}(x_-) \tag{2.2}$$

for the NP charge, and the formal expressions

$$\delta(x_- - y_-)[A(x_-; x_+, x_\perp), B(y_-; y_+, y_\perp)] \tag{2.3}$$

and

$$\delta(x_- - y_-)[\mathfrak{A}(x_-), \mathfrak{B}(y_-)] \tag{2.4}$$

for the LC commutators of such objects.

Let us first consider the formal charge

$$Q(x'_-) = \int d^4x\, j(x)\, \delta(x'_- - x_-). \tag{2.5}$$

We always assume that $\langle 0|\, j(x)\,|0\rangle = 0$. We shall attempt to give this meaning as a limit of a sequence of well-defined expressions. Thus we consider

$$Q_h(x'_-) = \lim_{n_1, n_2 \to \infty} [j(x)]\, [h_{n_1, n_2}(x'_-; x)], \tag{2.6}$$

where

$$h_{n_1 n_2}(x'_-; x) = n_1 h\left(n_1(x_- - x'_-), \frac{x_+}{n_2}, \frac{x_\perp}{n_2}\right), \tag{2.7}$$

$$h(x_-, x_+, x_\perp) = w(x_-)\, v(x_+, x_\perp), \tag{2.8}$$

$$w \in \mathcal{S}(R), \quad \int w = 1, \quad v \in \mathcal{S}(R^3), \quad v(0) = 1. \tag{2.9}$$

Our treatment is thus analogous to the treatment of ordinary charges given, for example, in Ref. 22. We choose our testing functions to have compact support in momentum space rather than in position space. This does not contradict desired locality properties.[22]

The limit in (2.6) has not been fully specified; we note the possibilities

$$\begin{aligned} Q_{I,h} &= \lim_{n_1 \to \infty} [\lim_{n_2 \to \infty} [j]\, (h_{n_1, n_2})], \\ Q_{II,h} &= \lim_{n_2 \to \infty} [\lim_{n_1 \to \infty} [j]\, (h_{n_1, n_2})], \\ Q_{\beta,h} &= \lim_{n \to \infty} [j]\, (h_{n, n^\beta}). \end{aligned} \tag{2.10}$$

Thus, if the support $\hat{h}_{n_1 n_2}(k)$ is contained in a box surrounding the origin in momentum space, then:

I In the limit I the diameter shrinks to 0, and then the length increases;

II In the limit II the box stretches out along the k_+ axis keeping fixed diameter, and then its diameter shrinks to 0;

III in the limit III stretching and shrinking occur simultaneously, the parameter β indicating the relative rates.

We first note that if h has compact support in momentum space and there is a mass gap in the theory, then $Q_{\mathrm{II},h}$ is defined on the vacuum and annihilates it. The same is true for $Q_{\beta,h}$ if $\beta > 1$ or, if $\beta = 1$, provided $\hat{h}_{1,1}(k)$ vanishes for $k^2 > M^2$, where M is the minimal mass in the theory. This is so because under the above conditions, for appropriately large n_1 and n_2, the support of $\hat{h}_{n_1, n_2}(k)$ is disjoint from the mass spectrum and so, from the Kallen–Lehmann representation, $\|Q(x_-)|0\rangle\| = 0$.

Finally, if $\hat{v}(k_-, k_\perp) = 0$ for $k_- \geq 0$, then $Q_{1,h}$, $Q_{\mathrm{II},h}$, and $Q_{h,\beta}$ each annihilate the vacuum; however, since $h(x)$ is not real, they are not generally limits of sequences of Hermitian operators.

Since the preceding arguments depend only on support properties of $h_{n_1 n_2}$, they remain valid for new charges obtained by replacing $h_{n_1 n_2}$ with $n_1^a n_2^b h_{n_1 n_2}$—these new charges will be used to obtain finite results from "bad–bad" commutators.

The charges $Q_{\beta,h}$ can easily be related to the infinite momentum limits of ordinary charges. We define

$$\Lambda_n = \frac{1}{2n} \begin{bmatrix} 1+n^2 & 1-n^2 \\ 1-n^2 & 1+n^2 \end{bmatrix} = (\Lambda_{n^{-1}})^{-1} \tag{2.11}$$

so that

$$\begin{pmatrix} t' \\ z' \end{pmatrix} = \Lambda_n \begin{pmatrix} t \\ z \end{pmatrix}$$

satisfy $t' \pm z' = n^{\pm 1}(t \pm z)$. Thus

$$h_{n,n}(x) = n \tilde{h}_n(\Lambda_n x),$$

where

$$\tilde{h}_n(x_-, x_+, x_\perp) = h(x_-, x_+, x_\perp/n)$$

so that

$$\langle \alpha |\, Q_{1,h}(x_-)\,|\, \beta\rangle = \lim_{n \to \infty} \langle \Lambda_n \alpha |\, [j]\, (\tilde{h}_n)\,|\, \Lambda_n \beta\rangle. \tag{2.12}$$

One can similarly discuss currents with Lorentz indices and commutators of such currents.

Also of interest are the semicharges C^i formally obtained by integrating a current density over the null semiplane x_- constant, $(-1)^i x_+ \geq 0$. Indeed, if these are defined by sequences $[j](h_n^i(x))$, where $h_n^i(x) = 0$ unless $(-1)^i x_+ > -1$, $|x_-| < C$, then there exist nonempty open sets R^i which are spacelike with respect to the support of each h_n^i. Let P_i and Q_i be elements of the polynomial algebra associated with R^i—since they commute with all C_n^i, one has $C_n^i P_i|0\rangle = P_i C_n^i|0\rangle$ and $\langle Q_i\Omega, C^i P^i|0\rangle = (P_i^+ Q_i\Omega, C^i\Omega)$. The states $P_i|0\rangle$ are dense in H (by the Reeh–Schlieder theorem); thus, if $C_n^i|0\rangle$ is a Cauchy sequence, C^i is a densely defined sesquilinear form (and if $C_n^i|0\rangle$ converge to an element of D, the common domain of all P_i, then C^i is a densely defined operator).

Unfortunately, the previous simple support arguments cannot be used to control $C_n^i|0\rangle$. The use of convergence factors may provide a partial alternative. Let $u \in C^\infty(R)$ satisfy $u(x) + u(-x) = 1$, $u'(x) \geq 0$, $u(x) = 1$ for $x > 1$; and let $u^i(x) = u((-1)^i x)$. Define $C^i = \lim [j](h_n^i)$,

$$h_n^i(x) = f(n)\, u^i(x_+/n)\, w(nx_-)\, v(x_+/n, x_\perp/n),$$

with w, v as before with the added requirement that w is to have compact support. (The charge $C_n = C_n^0 + C_n^1$ still annihilates the vacuum provided v vanishes for $k \leq 0$.) Using the Kallen–Lehmann representation, one has

$$\|C_n^i|0\rangle\|^2 = [D(\kappa)]\,(|\hat{h}_n^i(\kappa)|^2)$$

and

$$\|(C_m^i - C_n^i)|0\rangle\|^2 = [D(\kappa)]\,(|\hat{h}_m^i - \hat{h}_n^i|^2)$$

where $D(\kappa)$ is a positive Lorentz-invariant distribution. For any finite pair r, s one may pick an inverse polynomial $f(n)$ such that $\|h_n(k)\|_{r,s}^2 \to 0$; since $D(\kappa)$ is bounded in some norm $\|\ \|_{r,s}$ one may always equate $f(n)$ to an inverse power of n such that $C_n^i|0\rangle$ is a Cauchy sequence converging to zero. However, it seems difficult to obtain a Cauchy sequence which does not converge to zero.

We now turn our attention to the NP restrictions of commutators of local fields. We explicitly consider the commutator

$$D(x) \equiv [j(x), j(0)] \tag{2.13}$$

of scalar "currents" $j(x)$. We shall treat (2.13) as a bilinear form, rather than as an operator, and further simplify matters by considering only diagonal bounded momentum matrix elements. Of course, we should really work with the eight-dimensional distribution $[j(x), j(y)]$, but this can be trivially related to (2.13) by translation invariance. Our final notational simplification will be to work with sharp momentum states. Momentum smearing is trivial and can be done afterwards if desired.

The above conventions can be conveniently summarized by the DGS representation

$$D(x) = \frac{-i}{2\pi} \int_0^\infty da \int_{-1}^1 db\ \sigma(a,b)\, e^{-ibp\cdot x}\Delta(x; a + b^2). \tag{2.14}$$

Our analysis will be based on properties of (2.14) for various choices of the spectral function $\sigma(a, b)$. Our results can be easily extended to other representations.

Precisely analogous to (2.10), we define, for distributions $d(x)$, the restrictions $D_{\mathrm{I},h}(x_-), D_{\mathrm{II},h}(x_-), D_{\beta,h}(x_-)$ corresponding to the formal expression (2.4). We shall proceed in three steps, first discussing $\Delta(x; m^2)$, then $e^{-ix\cdot p}\Delta(x; m^2)$, and then all of (2.14) for some interesting σ's.

We begin by noting that $[\mathfrak{D}\Delta(x; m^2)]_{\mathrm{I},h}$ and $[\mathfrak{D}\Delta(x; m^2)]_{\beta,h}$, $\beta > 1$, where

$$\mathfrak{D} = \partial_{x_+}^{\alpha_1}\partial_{x_-}^{\alpha_2}\partial_{x_1}^{\alpha_3}\partial_{x_2}^{\alpha_2}$$

is any differential operator, may be made to vanish by appropriately restricting the support of h as discussed above. [Thus when such restrictions are made, discrete singularities of $\sigma(a, b)$ at $b = 0$ do not contribute to the null plane charge obtained from Eq. (2.14).] At the other extreme we consider

$$[\mathfrak{D}\Delta(x; m^2)]_{\mathrm{II}\,h} = \lim_{n\to\infty} [\delta(x_-)\mathfrak{D}\Delta(x; m^2)]_w(v(x_+/n, x_\perp/n)) \tag{2.15}$$

where $[\delta(x_-)d(x)]_w$ is the null plane restriction of $d(x)$, defined below.

We define the restriction to the null plane $x_- = 0$ of the distribution $d(x) \in S'(R^4)$ taken with respect to $w \in S(R)$, $\int w = 1$, by

$$[\delta(x_-)d(x)]_w = \lim_{n\to\infty} [nw(nx_-)d(x)] \tag{2.16}$$

provided the limit exists in $S'(R^4)$. {We do not define $[\delta(x_-)d(x)]$ if the limit does not exist.} If (2.16) does exist and is the same for all w, we write it simply as $\delta(x_-)d(x)$. It is useful to place restrictions on the behavior of w at the origin. We therefore define the space S_k of functions w which satisfy

$$\left[\left(\frac{d}{d\tau}\right)^m w_k(\tau)\right]\Big|_{\tau=0} = 0, \qquad m = 0, \ldots, k = 1. \tag{2.17}$$

If $[\delta(x_-)d(x)]_{w_k}$ is defined and is independent of $w_k \in S_k$, we write it as $[\delta(x_-)d(x)]_k$. We note that the set of functions w_k is quite large and includes the simple functions

$$[2^{2p}\Gamma(l)/\sqrt{\pi}\ \Gamma(2l)]\, y^{2l} e^{-y^2}, \qquad 2l \geq k. \tag{2.18}$$

We list (in Table I) the results of the computations (2.15). The derivations and proofs of sequence independence are given in Appendix A. Additional combinations may be obtained by using

$$\frac{d}{dx_+}\frac{d}{dx_-} = (\Box + m^2) + \partial_\perp^2 - m^2. \tag{2.19}$$

Note that, by using an appropriate testing function w_k, all of the restrictions exist as distributions in $S(R^4)$.

As an application of the results in Table I, and for use below, we consider a homogeneous distribution $g_\lambda(x)$ of degree λ, $\lambda \in R$:

$$g_\lambda(\alpha x) = \alpha^\lambda g_\lambda(x), \qquad \alpha \in R,\ \alpha \neq 0,$$

where, of course,

$$[g(\alpha x)](u(x)) = \alpha^{-d}[g(x)](u(x/\alpha)), \tag{2.20}$$

and d is the dimension of space. If $d(x)$ is homogeneous of degree λ, then

TABLE I.

I. In general

$$[\delta(t+z)\cdot P(d_t-d_z,d_{x_1},d_{x_2})\,T(x)]_w = P(d_t-d_z,d_{x_1},d_{x_2})[\delta(t+z)\cdot T(x)]_w$$

$$[\delta^{(k)}(t+z)\cdot T(x)]_w = \frac{1}{2^k}\sum_{l=0}^{k}(-)^l\binom{k}{l}(d_t+d_z)^{k-l}[\delta(t+z)\cdot(d_t+d_z)^l T(x)]_w$$

II.

$$[\delta(t+z)\cdot\Delta(x;m^2)]_w = \tfrac{1}{4}\,\delta(t+z)\,\epsilon(t)\,\delta(x_\perp)$$

$$[\delta(t+z)\cdot(d_t-d_z)^k\Delta(x,m^2)]_w = \tfrac{1}{2}(d_t-d_z)^{k-1}\,\delta(x)$$

$$[\delta(t+z)\cdot\epsilon(x_0)\,\Delta(x;m^2)]_w = \tfrac{1}{4}\,\delta(t+z)\,\delta(x_\perp)$$

$$[\delta(t+z)\cdot(d_t-d_z)^k\,\epsilon(x_0)\,\Delta(x;m^2)]_w = 0, \quad k\geq 1$$

III.

$$[\delta(t+z)\cdot(d_t+d_z)^k\Delta(x;m^2)]_{w_k} = \frac{\epsilon(t-z)[(z-t)(\partial_\perp^2-m^2)]^k\,\delta(t+z;x_\perp)}{2^{k+2}\,k!}$$

IV.

$$[\delta(t+z)(d_t+d_z)^k\,\epsilon(x_0)\,\Delta(x;m^2)]_{w_k} = \frac{[(z-t)(\partial_\perp^2-m^2)]^k\,\delta(t+z;x_\perp)}{2^{k+2}\,k!}$$

V. Additional combinations may be obtained by using

$$(d_t-d_z)(d_t+d_z) = (\Box+m^2)+(\partial_\perp^2-m^2)$$

$$[\delta(t+z)\cdot(d_t+d_z)^k\,\delta(x)]_{w_b} = 0, \quad b>k$$

$$[d(x)](nw(nx_-)v(x_+/n^\beta, x_\perp/n^\beta)) = n^{A(\lambda,\beta)}[d(x)](nw(nx_-)v(x_+,x_\perp)), \tag{2.21}$$

where

$$A(x,\beta) = (\lambda+3)\beta/(1+\beta). \tag{2.22}$$

Thus, if $\lambda=-3$, then

$$D_{\beta,h} = [\delta(x_-)\,d(x)]_w(v). \tag{2.23}$$

If $\lambda\to -3$ and $[\delta(x_-)\,d(x)]_w$ is finite, then

$$D_{\beta,h} = 0, \tag{2.24}$$

and, if $\lambda\to -3$, then $D_{\beta,h}$ can only exist if $[\delta(x_-)\,d(x)]_w = 0$.

We next consider $d(x) = e^{ip\cdot x}\Delta(x;m^2)$ with $p_-\neq 0$. In this case, $D_{\mathrm{I}h}$, $D_{\mathrm{II}h}$, and $D_{\beta h}$ are all equal because, as $n_1\to\infty$, the tube representing the support of $\hat{h}_{n_1n_2}(k)$ stretches out at a 45° angle with respect to the energy axis and punctures every mass hyperboloid. The calculation is straightforward and gives

$$D_{\mathrm{I},h} = D_{\mathrm{II},h} = D_{\beta,h} = -2/p_-. \tag{2.25}$$

We are now ready to consider the general representation (2.14). We shall only consider spectral functions $\sigma(a,b)$ which are rapidly decreasing functions of a uniformly in b. We do, however, allow singularities at $b=0$. These spectral functions are of physical interest since they lead to scaling behavior in the A limit and can lead to Regge asymptotic behavior in the R limit.[12] For use below, we recall here some results from the Appendix of Ref. 12.

The Fourier transform [$p=(1,\mathbf{0})$]

$$W(\kappa,\nu) = \int_0^\infty da\int_{-1}^1 db\;\sigma(a,b)\,\delta(\kappa+2b\nu-a) \tag{2.26}$$

of (2.14) will satisfy scaling

$$\nu W \xrightarrow[A]{} F(\rho) < \infty \tag{2.27}$$

with

$$F(\rho) = \tfrac{1}{2}\int da\;\sigma(a,1/2\rho) \tag{2.28}$$

and will have Regge behavior

$$W \xrightarrow[R]{} w(\kappa)\nu^\alpha, \quad -1\leq\alpha\leq 1, \tag{2.29}$$

with

$$w(\kappa) = \tfrac{1}{2}\int da\;\sigma(a)[2/(a-\kappa)]^{\alpha+1} \tag{2.30}$$

provided

$$\sigma(a,b)\sim b^{-(1+\alpha)}\sigma(\alpha) \quad \text{for } b\sim 0. \tag{2.31}$$

It follows that

$$D(x) = \hat{W}(x^2;x\cdot p) \xrightarrow[x^2\to 0]{} \delta(x^2)\epsilon(x\cdot p)f(x\cdot p) \tag{2.32}$$

with

$$f(\lambda) \xrightarrow[\lambda\to\infty]{} \lambda^\alpha. \tag{2.33}$$

If (2.14) represents a true scalar commutator, then $f(\lambda)$ must be even. In order to include the affects of internal indices, however, we shall not so restrict $f(\lambda)$.

Since, according to (2.25), we have

$$\delta(x_-)\,e^{-ipx}\,\Delta(x;m^2) = -\tfrac{1}{8}\,\delta(x_-)\,\epsilon(x_+)\,\delta(x_\perp)\,e^{-2ip_-x_-}, \tag{2.34}$$

we formally obtain from (2.14)

$$\int dx_+\,\delta(x_-)D(x) = (-i/4p_-)\int da db\,[\sigma(a,b)/b]\times\delta(x_-)\,\delta(x_\perp). \tag{2.35}$$

If $\sigma(a,b)$ is even in b, then (2.35) vanishes. If it is odd in b, then, according to (2.31), (2.35) is only finite for $\alpha\leq -1$. For $\alpha<-1$, we expect the three D's to be equal and to give

$$(-i/4p_-)\int da db\;\sigma(a,b)/b \equiv \Sigma. \tag{2.36}$$

This is easily seen to be the case. The corresponding "sum rule" is

$$\int d\nu\,W(\kappa,\nu)\sim\Sigma. \tag{2.37}$$

Let us finally consider the case $-1<\alpha$. Then, according to (2.31), (2.36) does not exist and, according to (2.29), neither does (2.37). These difficulties amount to the sequence dependence of

$$\int d\lambda\;f(\lambda)\;v(\lambda/n). \tag{2.38}$$

A way to surmount the difficulty is suggested by the identity

$$\int d\lambda\;\lambda^\alpha\,v(\lambda/n) = n^{\alpha+1}\int d\lambda\;v(\lambda); \tag{2.39}$$

for the good situation $\alpha<-1$, this always vanishes for $n\to\infty$. In the bad situation $\alpha>-1$, however, it diverges for $n\to\infty$ unless the integral vanishes identically:

$$\int d\lambda\;\lambda^\alpha\;v(x) = 0. \tag{2.40}$$

This is quite consistent with the interpretation of the limit of (2.38) as a formal charge, which only requires that $v(0)=1$.

It seems natural to use these observations to define finite NP restrictions in the bad situation. Suppose, for example, that

$$f(\lambda) = \lambda^{\alpha} + g(\lambda), \quad g(\lambda) = \mathcal{O}(\lambda^{-1-\epsilon}), \quad \alpha > -1. \tag{2.41}$$

With $v(\lambda)$ chosen to satisfy (2.40), we have

$$\int_v dx_+ \delta(x_-) D(x) \sim \delta(x_-)\delta(x_\perp) \int d\lambda\, g(\lambda). \tag{2.42}$$

The corresponding sum rule will have the form

$$\int d\nu\, \tilde{w}(\kappa, \nu) \sim \int d\lambda\, g(\lambda) \tag{2.43}$$

where $\tilde{w}$ is obtained from w by subtracting out the leading large ν behavior (2.29).

3. OPERATOR PRODUCT EXPANSIONS

In this section we discuss the LCOPE's for the $SU(3)$ currents $J_\mu^a(x)$. These expansions have been shown to be valid in each order of any renormalizable field theory, in soluble models, and more generally.[15,23] As in previous cases,[15] we shall ignore the logarithmic factors ($\log x^2$) which occur in perturbation theory.

The methods of Ref. 15 show that the LC expansion for the commutator

$$C_{\mu\nu}^{ab}(x) = [J_\mu^a(x/2), J_\nu^b(-x/2)] \tag{3.1}$$

has the form

$$\begin{aligned} C_{\mu\nu}^{ab}(x) \xrightarrow[x^2\to 0]{} \sum_{\pm} d_{\pm}^{abc}[&O_1^{(\pm)c}(x)\partial_\mu\partial_\nu + O_2^{(\pm)c}(x) g_{\mu\nu}\partial/\partial x^2 \\ &+ iO_{3\mu}^{(\pm)c}(x)\partial_\nu + iO_{3\nu}^{(\pm)c}(x)\partial_\mu \\ &+ iO_{4\mu}^{(\pm)c}(x)\partial_\nu - O_{4\nu}^{(\pm)c}(x)\partial_\mu \\ &+ O_{5\beta}^{(\pm)c}(x)\epsilon_{\mu\nu}^{\alpha\beta}\partial_\alpha \\ &+ O_{6\mu\nu}^{(\pm)c}(x) + iO_{7\mu\nu}^{(\pm)c}(x)]\Delta(x), \end{aligned} \tag{3.2}$$

where

$$d_+^{abc} = id^{abc}, \quad d_-^{abc} = f^{abc} \tag{3.3}$$

and

$$\Delta(x) = (1/2\pi)\epsilon(x_0)\delta(x^2). \tag{3.4}$$

We have assumed that $SU(3)$ is only broken by mass terms so that the leading LC singularities have the above $SU(3)$ symmetric structure. O_6 is symmetric and O_7 is antisymmetric under $\mu \leftrightarrow \nu$. The $O(x)$'s in (3.2) have the expansions

$$O_K^{(\pm)c}(x) = \sum_{n=0}^{\infty} (i)^n x^{\alpha_1}\ldots x^{\alpha_n} P_{K\alpha_1\cdots\alpha_n}^{(\pm)c}(0) \tag{3.5}$$

in terms of a suitable basis $\{P_{K\alpha_1\cdots\alpha_n}^{(\pm)c}\}$ of local field operators. Here K takes on the values $1, 2, 3\mu, 4\mu, 5\beta, 6\mu\nu$, and $7\mu\nu$.

The symmetry property

$$C_{\mu\nu}^{ab}(x) = -C_{\nu\mu}^{ba}(-x) \tag{3.6}$$

implies that

$$O_K^{(\pm)c}(-x) = \pm O_K^{(\pm)c}(x), \quad K = 1, 2, 4\mu, 5\beta, 6\mu\nu, \tag{3.7a}$$

and

$$O_K^{(\pm)c}(-x) = \mp O_K^{(\pm)c}(x), \quad K = 3\mu, 7\mu\nu. \tag{3.7b}$$

The adjoint property

$$C_{\mu\nu}^{ab}(x)^* = -C_{\mu\nu}^{ab}(x) \tag{3.8}$$

implies that

$$O_K^{(\pm)c}(x)^* = \pm O_K^{(\pm)c}(x), \quad K = 1, 2, 4\mu, 5\beta, 6\mu\nu \tag{3.9a}$$

and

$$O_K^{(\pm)c}(x)^* = \mp O_K^{(\pm)c}(x), \quad K = 3\mu, 7\mu\nu. \tag{3.9b}$$

Equations (3.7) and (3.8) imply that the P_K in (3.5) are Hermitian. Equations (3.7) imply that

$$P_{K\alpha_1\cdots\alpha_{2n+1}}^{(+)c} = P_{K\alpha_1\cdots\alpha_{2n}}^{(-)} = 0, \quad K = 1, 2, 4\mu, 5\beta, 6\mu\nu \tag{3.10a}$$

and

$$P_{K\alpha_1\cdots\alpha_{2n+1}}^{(-)c} = P_{K\alpha_1\cdots\alpha_{2n}}^{(+)c} = 0, \quad K = 3\mu, 7\mu\nu. \tag{3.10b}$$

The equal-time commutation relations

$$C_{00}^{abc}(x)\delta(x_0) = if^{abc}J_0^c(0)\delta(x), \tag{3.11a}$$

$$C_{0k}^{ab}(x)\delta(x_0) = if^{abc}J_k^c(0)\delta(x) + id^{abc}\Sigma^c(0)\partial_k\delta(x) \tag{3.11b}$$

$$C_{ij}^{ab}(x)\delta(x_0) = if^{abc}\delta_{ij}J_0^c(0)\delta(x) + id^{abc}\epsilon_{ijk}A^{ck}(0)\delta(x) \tag{3.11c}$$

imply that

$$\begin{aligned} P_1^{(+)c} = \Sigma^c, \quad P_{1\alpha}^{(-)c} = 0, \quad P_{2\alpha}^{(-)c} = -2J_\alpha^c, \\ P_{3\mu}^{(-)c} = J_\mu^c, \quad P_{4\mu}^{(+)c} = 0, \quad P_{5\beta}^{(+)c} = A_\beta^c. \end{aligned} \tag{3.12}$$

The remaining low-indexed terms in (3.2) contribute to the time derivative commutator according to

$$\begin{aligned} \int d^4x\, \delta(x_0)[\dot{J}_i^a(x), J_j^b(0)] = 2id^{abc}[&-E_{ij}^{(+)c}(0) \\ &+ F_{ij}^{(+)c}(0) + \tfrac{1}{2}G_{ij}^{(+)c}(0)] \end{aligned} \tag{3.13}$$

$+ ij$ terms $+ (i \leftrightarrow j)$ antisymmetric terms, where we have used notation to be introduced below.

Especially interesting is the good–good commutator, which we write as

$$\begin{aligned} [J_-^a(x), J_-^b(0)] \xrightarrow[x^2\to 0]{} \sum_{\pm} d_{\pm}^{abc}[&\mathcal{E}^{(\pm)c}(x)\partial_-\partial_- \\ &+ 2i\mathcal{F}^{(\pm)c}(x)\partial_- + \mathcal{G}_{--}^{(\pm)c}(x)]\Delta(x), \end{aligned} \tag{3.14}$$

where

$$\begin{aligned} \mathcal{E}^{(\pm)c}(x) &= \sum_{n=0}^{\infty} (i)^n x^{\alpha_1}\ldots x^{\alpha_n} E_{\alpha_1\ldots\alpha_n}^{(\pm)c}(0), \\ \mathcal{F}_\mu^{(\pm)c}(x) &= \sum_{n=0}^{\infty} (i)^n x^{\alpha_1}\ldots x^{\alpha_n} F_{\mu\alpha_1\ldots\alpha_n}^{(\pm)c}(0), \\ \mathcal{G}_{\mu\nu}^{(\pm)c}(x) &= \sum_{n=0}^{\infty} (i)^n x^{\alpha_1}\ldots x^{\alpha_n} G_{\mu\nu\alpha_1\ldots\alpha_n}^{(\pm)c}(0) \end{aligned} \tag{3.15}$$

with the E's, F's, and G's Hermitian local field operators. Let us directly compute from (3.14) the behavior of the single-particle rest matrix elements. We define

$$\begin{aligned} \langle p \mid \mathcal{E}^{(\pm)c}(x) \mid p\rangle &= e^{(\pm)c}(x\cdot p) + \cdots, \\ \langle p \mid \mathcal{F}^{(\pm)c}(x) \mid p\rangle &= f^{(\pm)c}(x\cdot p) + \cdots, \\ \langle p \mid \mathcal{G}_{\mu\nu}^{(\pm)c}(x) \mid p\rangle &= g^{(\pm)c}(x\cdot p) + \cdots, \end{aligned} \tag{3.16}$$

where e, f, and g are scalar functions and the omitted terms do not contribute to the leading asymptotic behaviors. By the methods of Ref. 4, we find from (3.14) that (1.14) is valid with

$$F_2^{ab}(\omega) = (i/4\pi)\int d\lambda\, e^{-i\lambda\omega}h^{ab}(\lambda), \tag{3.17}$$

where

$$h^{ab}(\lambda) = \sum_{\pm} d_{\pm}^{abc}[g^{(\pm)c}(\lambda) - 2if^{(\pm)c\prime}(\lambda) + e^{(\pm)c\prime\prime}(\lambda)]$$
$$\equiv d_{+}^{abc}\, h_{+}^{c}(\lambda) + d_{-}^{abc} h_{-}^{c}(\lambda), \quad (3.18)$$

so that

$$F_2^{ab}(0) = (-1/4\pi) d^{abc} \int d\lambda\ h_{+}^{c}(\lambda). \quad (3.19)$$

Using (3.12), we find further that

$$T_2^{ab}(\kappa, \nu) \xrightarrow[R']{} (-i/4\nu)\, d^{abc} \int d\lambda h^{(\pm)c}(\lambda) + (f^{abc}/\nu) \times [F^c - \tfrac{1}{4} \int d\lambda \epsilon(\lambda) g^{(-)c}_{(\lambda)}]. \quad (3.20)$$

Equations (3.17) and (3.18) tell us that

$$\int d\omega\, F_2^{ab}(\omega) = (i/2)\, h^{ab}(0) = -\tfrac{1}{2}\, d^{abc} \times [g^{(+)c}(0) - 2if^{(+)c\prime}(0) + e^{(+)c\prime\prime}(0)]. \quad (3.21)$$

By virtue (3.13), this is just the Callan–Gross result. We have derived it without taking an infinite momentum limit because in (3.16) there occur no $g_{\mu\nu}$ terms since $g_{--} = 0$.

This concludes our discussion of the LCOPE's for the $SU(3)$ currents. In Sec. 4 we shall consider the LC restriction of (3.14) and briefly consider some of the other components of (3.2).

4. NULL PLANE CURRENT COMMUTATORS

In this section we combine some of the results of Secs. 2 and 3 in order to compute the null plane restrictions of some current commutators. We use the formal notations of Secs. 1 and 3 rather than the cumbersome, although precise, notations of Sec. 2. We will, of course, always have in mind the precise definitions of Sec. 2.

We consider first the expression (3.14) for the LC behavior of the good–good current commutator. Using Table I, we find

$$\delta(x_-)[J_-^a(x), J_-^b(0)] = \tfrac{1}{2} \sum_{\pm} d_{\pm}^{abc}[2\mathcal{E}^{(\pm)c}(x)\partial_-\delta(x) + 4i\mathcal{F}_-^{(\pm)c}(x) \times \delta(x) + \mathcal{G}_{--}^{(\pm)c}(x)\, \delta(x_-)\epsilon(x_+)\delta(x_\perp)]. \quad (4.1)$$

We note that this result is sequence independent.

We next want to integrate (4.1) over x_+. The $\mathcal{E}$ and $\mathcal{F}$ terms are clearly integrable and the integrability of the $\mathcal{G}$ term depends on the large-x_+ behavior of $G_{--}(x_+, 0, 0)$. Assuming this is integrable, we obtain

$$\tfrac{1}{2} \int dx_+\, \delta(x_-)[J_-^a(x) J_-^b(0)] = if^{abc} J_-^c(0)\, \delta(x_-, x_\perp) + \tfrac{1}{4} f^{abc} \int dx_+ \epsilon(x_+) \mathcal{G}_{--}^{(-)c}(x)\, \delta(x_-, x_\perp). \quad (4.2)$$

Thus, as in Ref. 19, we see that the LC commutation relation (1.7) or, equivalently, the sum rule (1.6) need not be satisfied in general even though the ET commutation relations (1.5) are satisfied. The condition for the validity of (1.6) and (1.7) is

$$\int dx_+ \epsilon(x_+)\, \mathcal{G}_{--}^{(-)c}(x_+, 0, 0_\perp) = 0. \quad (4.3)$$

Let us now consider the other null plane restriction (1.12). In order to avoid undefined expressions like $\delta(0)$, we shall make use of a different form of (3.14). We can write

$$[J_-^a(x), J_-^b(0)] \xrightarrow[x^2 \to 0]{} \sum_{\pm} d_{\pm}^{abc}\{\mathcal{K}_{--}^{(\pm)c}(x)\Delta(x) + \partial_-[\mathcal{G}_{--}^{(\pm)c}(x)\Delta(x)] + \partial_-\partial_-[\mathcal{J}^{(\pm)c}(x)\Delta(x)]\} \quad (4.4)$$

where $\mathcal{K}$, $\mathcal{G}$, and $\mathcal{J}$ are uniquely related to $\mathcal{E}$, $\mathcal{F}$, and $\mathcal{G}$. For example,

$$\mathcal{K}_{--}^{(\pm)c} = \partial_-\partial_-\mathcal{E}^{(\pm)c} - 2i\partial_-\mathcal{F}_-^{(\pm)c} + \mathcal{G}_{--}^{(\pm)c}. \quad (4.5)$$

We use the form (4.4) to uniquely define the "nonderivative part" $[J_-^a(x), J_-^b(0)]_N$ of the commutator so that

$$[J_-^a(x), J_-^b(0)]_N \xrightarrow[x^2 \to 0]{} \sum_{\pm} d_{\pm}^{abc} \mathcal{K}_{--}^{(\pm)c}(x)\Delta(x). \quad (4.6)$$

Using (4.6) and Table I, we formally obtain

$$\tfrac{1}{2} \int dx_+\, \delta(x_-)\epsilon(x_+)[J_-^a(x), J_-^b(0)]_N = \delta(x_-)\delta(x_\perp) i d^{abc} \int dx_+ \mathcal{K}_{--}^{(+)c}(x_+, 0, 0). \quad (4.7)$$

This result is correct provided $H_{--}^{(+)}(x_+, 0, 0)$ is integrable. Equation (1.12), interpreted in this way, thus requires the validity of the identity

$$\int dx_+ \mathcal{K}_{--}^{(+)}(x_+, 0, 0_\perp) = -\{M^{-1}P_-, S^c(0)\}. \quad (4.8)$$

Taking one-particle matrix elements of (4.7) and (4.8), using (3.16) and (3.19), we obtain the desired result (1.19). The same result (1.19) was obtained formally in Ref. 19, but there, because all x_+ integrals were assumed to be rapidly convergent, only the $g(\lambda)$ term in (3.18) was kept in (3.19) and only the G term in (4.5) was kept in (4.7). Here we allow for the possibility that, for example, $f(\lambda) \sim \epsilon(\lambda)$ so that $\int d\lambda f'(\lambda) \neq 0$.

So far we have only considered the good–good commutator (3.14). Working with the other components of (3.1) leads to two additional complications. The first is that the formal expression $\delta(x_-)C_{\mu\nu}^{ab}(x)$ is not well defined. This matter was resolved in Sec. 2, where it was shown that, by using suitable sequences w_k, each term in (3.2) could be restricted to the null plane $x_- = 0$. The second implication is that the large x_+ behaviors are not expected to be good. This follows from the result (2.34) and the fact that the Regge behavior for these components has $\alpha > -1$. An attack on this problem was also made in Sec. 2. There is was shown that with a suitable testing function $v(\lambda)$, the contributions of the bad large-x_+ pieces could be made to vanish so that a finite result is obtained. Possible forms for this result could thus be postulated and, via the corresponding finite sum rule, compared with experiment.

Note in manuscript: We have recently learned that C. H. Woo has derived related results about the existence and properties of NP charges.

APPENDIX: NULL PLANE RESTRICTIONS OF CAUSAL SOLUTIONS OF THE KLEIN–GORDON EQUATION

We define the restriction to the null plane $t + z = 0$ of the distribution $T(x) \in \mathcal{S}'(R)$ taken with respect to the function $w \in \mathcal{S}(R)$, $\int dy\, w(y) = 1, w(y) = w(-y)$ by

$$[\delta(t + z)\, T(x)]_w = \lim_{n \to \infty} [nw(n(t + z))\, T(x)]; \quad (A1)$$

this definition gives precise meaning to the formal expression $\delta(t+z)\,T(x)$. It is useful to place restrictions on the behavior of w at the origin; thus w_k will denote a function $w_k \in \mathcal{S}(R)$, $w_k(y) = w_k(-y)$, $\int dy\, w_k(y) = 1$, $[d_y{}^m w_k(y)]\,|_{y=0} = 0, m < k$. In the present section (the results of which are summarized in Table I) we calculate the null plane restrictions of the distributions $\Delta(x;m)$ and $\epsilon(x_0)\Delta(x;m)$ and derivatives of these distributions.

1. $\delta(x_0+x_3)\delta(x^2)$ and $\delta(x_0+x_3)\epsilon(x_0)\delta(x^2)$, and Derivatives

Recalling the definition of $\delta(x^2)$ and $\epsilon(x_0)\,\delta(x^2)$

$$\delta(x^2) = (1/2\,|t|)\,\delta(|t|-r),\ \epsilon(x_0)\,\delta(x^2) = \delta(|t|-r)/2t \tag{A2}$$

we have, for $u \in \mathcal{S}(R^4)$, $u(x) = u(t, r, \cos\theta, \phi)$,

$$\begin{aligned}[nw(n(t+z))\,\delta(x^2)]\,(u) &= \tfrac{1}{2}\int dt \int_{-1}^{1} d\cos\theta\,|t|\,\tilde{u}\,(t,|t|,\cos\theta)\\ &\quad\times nw(n(|t|\cos\theta + t))\\ &= \tfrac{1}{2}\int_{-\infty}^{\infty} dt \int_0^{\infty} dy\,\theta(2n|t|-y)\,\tilde{u}\\ &\quad\times (t,|t|,\epsilon|t|\,(y/n|t|-1))\,w(y),\end{aligned} \tag{A3}$$

where

$$\tilde{u}(t,|t|,\cos\theta) = \int_0^{2\pi} d\phi\ u(t,|t|,\cos\theta,\phi). \tag{A4}$$

Now the integrands $u(t,|t|,\pm(1-y/n|t|)w(y)$ are dominated by the integrable functions $|w(\pm y)| \times \sup_z |u(t,|t|,z)|$, so the Lebesque convergence theorem may be used to evaluate the limit

$$\lim_{n\to\infty}[nw(n(t+z))\,\delta(x^2)](u) = \tfrac{1}{4}\int_{-\infty}^{\infty} dt\ \tilde{u}(t,|t|,\epsilon(-t)). \tag{A5}$$

Since $\tilde{u}(t,|t|,-\epsilon(t)) = 2\pi u(x)|_{x_0=-x_3,\,x_1=x_2=0}$, we have

$$[\delta(t+z)\,\delta(x^2)]_w = (\pi/2)\,\delta(t+z)\,\delta(x_1,x_2) \tag{A6}$$

and similarly

$$[\delta(t+z)\,\epsilon(x_0)\,\delta(x^2)]_w = (\pi/2)\,\epsilon(t-z)\,\delta(t+z)\,\delta(x_1,x_2). \tag{A7}$$

Since $w(x_0+x_3)$ is independent of the variables $(x_0-x_3), x_1, x_2$, one may replace u by its appropriate derivative to obtain

$$[\delta(t+z)\cdot(d_t-d_z)^{a_1}d_{x_1}^{a_2}d_{x_2}^{a_3}\,\delta(x^2)]_w = (\pi/2)\,\delta_{a_1,0}\,d_{x_1}^{a_1}d_{x_2}^{a_2}\,\delta(t+z)\,\delta(x_1,x_2) \tag{A8}$$

and $(a_1 \geq 1)$

$$[\delta(t+z)\cdot(d_t-d_z)^{a_1}d_{x_1}^{a_2}d_{x_2}^{a_3}\epsilon(x_0)\,\delta(x^2)] = \pi(d_t-d_z)^{a_1-1}\,d_{x_1}^{a_2}\,d_{x_2}^{a_3}\,\delta(x). \tag{A9}$$

In order to calculate $\delta(x_0+x_3)(d_t+d_z)^n\delta(x^2)$, it is convenient to first guess the answer which may be done as follows. We recall [or verify directly from (A2)] that $\delta(x^2)$ and $\epsilon(x_0)\,\delta(x^2)$ satisfy the Klein–Gordon equations

$$\begin{aligned}&\Box\,\delta(x^2) = 2\,\delta(x),\\ &\Box\,\epsilon(x_0)\,\delta(x^2) = 0.\end{aligned} \tag{A10}$$

Thus, if $w_k \in \mathcal{S}(R)$ vanishes at the origin together with its first $(k-1)$ derivatives and $u(x)$ is such that $u = (d_t - d_z)^k\,h$ with $h \in \mathcal{S}(R^4)$,

$$h(x) = h(x_+, x_-, x_\perp) = \frac{1}{2^k}\int_0^{x_-} dx'_- \cdots \int_0^{x'_{-k-1}} dx_{-k}\,u(x_+, x_-^k, x_\perp), \tag{A11}$$

then (A10) implies

$$[\delta(x^2)](d_t+d_z)^k uw(t+z)) = [\delta(x^2)] \times ((d_t+d_z)^k(d_t-d_z)^k hw) = [\delta(x^2)](w\partial_\perp^2 h), \tag{A12}$$

and

$$[\delta(t+z)\cdot(d_t+d_z)^k\,\delta(x^2)]_{w_k}(u) = \int dx_-\left(\frac{(-x_-)^k}{2^{k+2}\,k!}\right)(\partial_\perp^2 u|_{\substack{x_+=0\\ x_-=0}}). \tag{A13}$$

If we conjecture that these relations hold in general, we have

$$[\delta(t+z)\cdot(d_t+d_z)^k\,\delta(x^2)]_{w_k} = \frac{\pi[(z-t)\,\partial_\perp^2]^k}{2^{k+1}\,k!}\,\delta(x_\perp)\,\delta(t+z) \tag{A14}$$

and similarly

$$[\delta(t+z)'(d_t+d_z^k)\,\epsilon(x_0)\,\delta(x^2)]_{w_k} = \frac{\pi\epsilon(t-z)}{2^{k+1}\,k!}[((z-t)\,\partial_\perp^2)^k\,\delta(x_\perp)\,\delta(t+z)]. \tag{A15}$$

It remains to verify (A14) and (A15) explicitly. This may be done by integrating the analogs of (A3) by parts, a fairly tedious task which we now carry out for the case $k=1$.

2. Verification of (A14) and (A15) for $k=1$

In order to verify the preceding result, we begin by evaluating $L_\pm(u) = \lim_{n\to\infty}[\theta(\pm t)\,\delta(x^2)]\,[(d_t+d_z)nw_1(n(t+z))]\,u(x)$. One has

$$\begin{aligned}I_{n\pm}(w,u) &= [\theta(\pm x_0)\,\delta(x^2)]\,[2n^2w_1'(n(t+z))u(x)]\\ &= \int_{-\infty}^{\infty} dt\ \theta(\pm t)\int_0^{2n|t|} dy\,\epsilon(t)nw_1'(y)\,\tilde{u}\\ &\quad\times (t,|t|,\epsilon(t)(y/n|t|-1)).\end{aligned} \tag{A16}$$

Integrating this by parts (recalling that w_1 vanishes at the origin), one obtains

$$I_{n\pm} = I_{n\pm}^1 + I_{n\pm}^2 \tag{A17}$$

where

$$I'_{n\pm} = \int_{-\infty}^{\infty} dt[\pm\,\theta(\pm t)nw_1(2n|t|)\,\tilde{u}(t,|t|,\epsilon(t))] \to \pm\,\tfrac{1}{4}\,\tilde{u}\,(0,0,1) = \pm\,(\pi/2)\,u(0)$$

and

$$I_{n\pm}^2 = -\int_{-\infty}^{\infty} dt\ \theta(\pm t)\int dy\ w(y)\,\tilde{\tilde{u}}(t,|t|,\epsilon(t)(y/n|t|-1)) \to \pm\,\tfrac{1}{2}\int dt\ \theta(\pm t)\,\tilde{\tilde{u}}\,(t,|t|,-\epsilon(t))$$

with

$$\begin{aligned}\tilde{\tilde{u}}(t,r,\cos\theta) &= r^{-1}\frac{d}{d\cos\theta}\,\tilde{u}(t,r,\cos\theta)\\ &= r^{-1}\frac{d}{d\cos\theta}\int_0^{2\pi} d\phi\ u(t,r\sin\theta\cos\phi,\\ &\quad\times r\sin\theta\sin\phi, r\cos\theta)\\ &= \int_0^{2\pi} d\phi(-\cot\theta(u_{,1}\cos\phi + u_{,2}\sin\phi) + u_3)\end{aligned}$$

$$= \int_0^{2\pi} d\phi[(-r\cos\theta)(\sin^2\phi\, u_{,11} + \cos^2\phi\, u_{,22} - 2\sin\phi\cos\phi\, u_{,2} + u_{,3}]$$

and so

$$\tilde{\tilde{u}}(t, |t|, -\epsilon(t)) = 2\pi\left[\left(\frac{d}{dz} - \frac{z}{2}\,\partial_\perp^2\right)u\right]\Big|_{z=-t,\,x_\perp=0}.$$

Thus we have

$$L_\pm(u) = \pm\frac{\pi}{2}u(0) - \pi\int dt\;\theta(\pm t) \times\left[\left(dz - \frac{z}{2}\,\partial_\perp^2\right)u\right]\Big|_{z=-t,\,x_\perp=0}. \quad \text{(A18)}$$

Now

$$[\delta(t+z)\cdot(d_t + d_z)\,\theta(\pm t)\,\delta(x^2)]_{w_1}(u) = -L_\pm(u) - [\delta(t+z)\cdot\theta(\pm t)\,\delta(x^2)](u) = -(\pi/2)\int dt\;\theta(\pm t)(z\partial_\perp^2 u), \quad \text{(A19)}$$

which is the result given before. It should be emphasized that the vanishing of w_1 at the origin is critical for the validity of (A14), although this is not evident in the heuristic derivation.

3. $\delta(t+z)\cdot\theta(x^2)$, $\delta(t+z)\cdot\Delta(x;m^2)$, and $\delta(t+z)\cdot\epsilon(x_0)\Delta(x;m^2)$

Expressions of the form $\delta(t+z)\cdot\theta(x^2)f(x^2)$ with f everywhere locally integrable may be seen to vanish as follows. One has

$$\begin{aligned}[\theta(x^2)f(x^2)]\,(nw(n(t+z))u(x)) &= \tfrac{1}{2}\int_{-\infty}^{\infty}dt\int_0^\infty dr\left(\frac{r^2}{|t|}\right)\int_{-1}^{1}d\cos\theta\; f(t^2-r^2)\tilde{u} \\ &\quad\times(t,r,\cos\theta)nw(n(t+r\cos\theta)) \\ &= \tfrac{1}{2}\int dt\,dr\,dy\left(\frac{r}{|t|}\right)f(t^2-r^2)w(y)\tilde{u}\left(t,r,\frac{y}{nr}-\frac{t}{r}\right) \\ &\quad\times\theta(|t|-r)\,\theta(y-n(t-r))\,\theta(n(t+r)-y). \end{aligned} \quad \text{(A20)}$$

Since the integrand of (A20) is dominated by the integrable function obtained by replacing $u(t,r,y/nr - t/r)$ by $\sup_y|u(t,r,y)|$ and setting $n = 0$ elsewhere, and since the integrand approaches zero almost everywhere, the Lebesgue convergence theorem implies that the limit vanishes.

Since $2\pi\Delta(x;m)$ differs from $\epsilon(x_0)\,\delta(x^2)$ by a locally integrable function, we obtain

$$[\delta(t+z)\cdot\Delta(x;m)]_w = \tfrac{1}{4}\,\delta(x_\perp)\,\delta(t+z),$$
$$[\delta(t+z)\cdot\epsilon(x_0)\,\Delta(x,m^2)]_w = \tfrac{1}{4}\,\epsilon(t)\,\delta(t+z)\,\delta(x_\perp). \quad \text{(A21)}$$

$(2\pi)^{-1}\epsilon(x_0)\Delta(x\cdot m)$ and $(2\pi)^{-1}\Delta(x;m)$ satisfy the Klein-Gordon equations obtained by replacing $\partial_\perp^2$ with $(\partial_\perp^2 - m^2)$ in the preceding calculation; the reasoning used in passing from (A10) to (A14) may be generalized by making the same substitutions.

4. $\delta^{(k)}(t+z)\,T(x)$

Distributions of the form $\delta^{(k)}(t+z)\cdot T(x)$ occur when one carries out the null-plane analog of the Bjorken expansion. Such distributions are incorporated in the foregoing analysis by means of the relation

$$(\partial^k w(x))\,T(x) = \sum_{l=0}^{k}\binom{k}{l}\partial_{x_i}^{k-l}(w(x)(-\partial_x)^l T(x)) \quad \text{(A22)}$$

valid for every c^∞ function w, distribution T, and first-order derivative ∂. Thus we define

$$[\delta^{(k)}(t+z)\cdot T(x)]_w = \lim_{n\to\infty}[n^{k+1}w^{(k)}(n(t+z))\,T(x)]$$

and obtain

$$[\delta^{(k)}(t+z)\cdot T(x)]_w = \frac{1}{2^k}\sum_{l=0}^{k}\binom{k}{l}(d_t+d_z)^{k-l}[\delta(t+z)\cdot(-d_t-d_z)^l T(x)]_w \quad \text{(A23)}$$

whenever each term in the sum is well defined.

---

* Supported in part by funds from the National Science Foundation Grant No. GP-13332.

† Alfred P. Sloan Foundation Fellow.

1 S. Fubini and G. Furlan, Physics **1**, 229 (1965).

2 S. Fubini, Nuovo Cimento **43A**, 475 (1966); R. F. Dashen and M. Gell-Mann, *Proceedings of the Third Coral Gables Conference on Symmetry Principles at High Energy*, edited by B. Kursunoglu *et al.* (Freeman, San Francisco, 1966).

3 K. Bardakci and G. Segre, Phys. Rev. **159**, 1263 (1967); L. Susskind, *ibid.* **165**, 1535 (1967); J. Jersak and J. Stern, Nuovo Cimento **59**, 315 (1969); H. Leutwyler, *Proceedings of the Summer School for Theoretical Physics, Karlsruhe*, Springer Tracts in Modern Physics (Springer-Verlag, Berlin, 1969), Vol. 50, p. 29.

4 J. B. Bronzan, I. S. Gerstein, B. W. Lee, and F. E. Low, Phys. Rev. Letters **18**, 32 (1966); Phys. Rev. **157**, 1448 (1967); V. Singh, Phys. Rev. Letters **18**, 36 (1967); **E18**, 300 (1967).

5 For a 4-vector A_μ we write $A_\pm = A_0 \pm A_3$, $A_\perp = (A_1, A_2)$. Also $\delta(x_\perp) = \delta(x_1)\,\delta(x_2)$.

6 S. Weinberg, Phys. Rev. **150**, 1313 (1966).

7 L. Susskind, Phys. Rev. **165**, 1535 (1967); D. Flory, *ibid.* **186**, 1701 (1969); *ibid.* D **1**, 2795 (1970); S. Cheng and S. Ma, Phys. Rev. **180**, 1506 (1969).

8 K. Bardakci and M. Halpern, Phys. Rev. **176**, 1686 (1968).

9 J. Kogut and D. Soper, Phys. Rev. D **1**, 2901 (1970).

10 R. A. Brandt, Phys. Rev. Letters **23**, 1260 (1969).

11 J. D. Bjorken, Phys. Rev. **179**, 1547 (1969).

12 R. A. Brandt, Phys. Rev. D **1**, 2808 (1970).

13 H. Leutwyler and J. Stern, Nucl. Phys. **B20**, 77 (1970).

14 R. Jackiw, R. Van Royen, and G. West, Phys. Rev. D **10**, 2473 (1970).

15 R. A. Brandt and G. Preparata, Nucl. Phys. **B27**, 541 (1971).

16 R. A. Brandt and G. Preparata, Phys. Rev. Letters **25**, 1530 (1970).

17 G. Altarelli, R. A. Brandt, and G. Preparata, Phys. Rev. Letters **26**, 42 (1971).

18 R. A. Brandt and C. Orzalesi, Phys. Letters **34B**, 641 (1971).

19 R. A. Brandt, Phys. Letters **33B**, 312 (1970).

20 R. Klauder, H. Leutwyler, and L. Streit, Nuovo Cimento **66**, 536 (1970).

21 D. Neville and F. Rohrlich, Phys. Rev. D **3**, 1692 (1971).

22 P. Otterson, NYU Preprint 7/70 (New York University, 1970).

23 W. Zimmermann, *Proceedings of the 1970 Brandeis Summer School*, to be published. G. F. Dell'Antonio and W. Zimmermann, in preparation.

Quantization of a General Dynamical System by Feynman's Path Integration Formulation*

K. S. Cheng
Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11790
(Received 7 July 1972)

The Schrödinger equation is obtained by Feynman's path integration method of quantization for a general dynamical system. The meaning of the results is discussed.

I. INTRODUCTION

We present a derivation of the "Schrödinger equation" for a general dynamical system using Feynman's path integral[1] method of quantization. The result differs from the "usual Schrödinger equation" in that there is an additional term proportional to the total curvature R of the coordinate space defined with a geometry given by the kinetic energy. This result had been given before by DeWitt.[2] In a curved space or in cases of constraints where $R \neq$ numerical constant, the presence of this additional term would change the energy spectrum of the whole system. In Sec. III we discuss the meaning of this additional term.

II. DERIVATION OF THE SCHRÖDINGER EQUATION

We will give a detailed derivation of the Schrödinger equation for a general mechanical system by using the path-integral method of Feynman. For a given mechanical system described by a set of coordinates $q\ (q^1, q^2, \cdots q^N)$, let the Lagrangian be

$$L(\dot{q}(t), q(t)) = \tfrac{1}{2} g_{ij}(q(t))\dot{q}^i\dot{q}^j. \tag{1}$$

Following Ref. 1, we can generalize Eqs. (1)–(18) to the above system, that is,

$$\psi(q(t+\epsilon), t+\epsilon) = (1/A)\int \exp[(i/\hbar)S(q(t+\epsilon), q(t))] \times \psi(q(t), t)\sqrt{g(q(t))}\,dq(t), \tag{2}$$

where $\psi(q(t+\epsilon), t+\epsilon)$ and $\psi(q(t), t)$ are, respectively, wavefunctions at time $t+\epsilon$ and t, $S(q(t+\epsilon), q(t))$ is the classical action, that is,

$$S(q(t+\epsilon), q(t)) = \text{minimum of } \int_t^{t+\epsilon} L(\dot{q}(t'), q(t'))dt' \tag{3}$$

with the boundary conditions

$$q(t')|_{t'=t} = q(t), \qquad q(t')|_{t'=t+\epsilon} = q(t+\epsilon). \tag{4}$$

A is a normalization factor to be determined later and g is the determinant of (g_{ij}). Taking the limit of Eq. (2) when $\epsilon \to 0$, we can derive the Schrödinger equation. Now as $\epsilon \to 0$, the factor $\exp[(i/\hbar)S(q(t+\epsilon), q(t))]$ oscillates very rapidly. Only the vicinity of the stationary point of $S(q(t+\epsilon), q(t))$ contributes to the integral in Eq. (2). The stationary point is

$$q(t) = q(t+\epsilon). \tag{5}$$

As we shall see the region which contributes to the integral in Eq. (2) is $|\Delta q| = |q(t) - q(t+\epsilon)| \lesssim \epsilon^{1/2}$. Thus we can expand $S(q(t+\epsilon), q(t))$ as a power series of Δq. This is done in Appendix A and gives

$$\begin{aligned} S(q(t+\epsilon), q(t)) = {} & \frac{1}{2\epsilon} g_{ij}(q(t+\epsilon)) \\ & \times \Big[\Delta q^i\Delta q^j - \left\{ {i \atop mn} \right\} \Delta q^j\Delta q^m\Delta q^n \\ & + \frac{1}{4}\left\{ {i \atop mn} \right\}\left\{ {j \atop \alpha\beta} \right\} \Delta q^m\Delta q^n\Delta q^\alpha\Delta q^\beta \\ & + \frac{1}{3}\left(\frac{\partial}{\partial q^l}\left\{ {i \atop mn} \right\} + \left\{ {i \atop \alpha l} \right\}\left\{ {\alpha \atop mn} \right\}\right)\Delta q^j\Delta q^m\Delta q^n\Delta q^l + \cdots\Big]. \end{aligned} \tag{6}$$

We also need the following expansions:

$$\sqrt{g(q(t))} = \sqrt{g(q(t+\epsilon))} - \Delta q^i \frac{\partial\sqrt{g}}{\partial q^i} + \tfrac{1}{2}\Delta q^i\Delta q^j \frac{\partial^2\sqrt{g}}{\partial q^i\partial q^j} + \cdots, \tag{7}$$

$$\psi(q(t), t) = \psi(q(t+\epsilon), t) - \Delta q^i \frac{\partial\psi}{\partial q^i} + \tfrac{1}{2}\Delta q^i\Delta^j \frac{\partial^2\psi}{\partial q^i\partial q^j} + \cdots. \tag{8}$$

In these equations

$$\Delta q = q(t+\epsilon) - q(t) \tag{9}$$

and $\left\{ {i \atop mn} \right\}$ is the Christoffel symbol,

$$\left\{ {i \atop mn} \right\} = g^{ik}[mn, k], \tag{10}$$

$$[mn, k] = \frac{1}{2}\left(\frac{\partial g_{mk}}{\partial q^n} + \frac{\partial g_{nk}}{\partial q^m} - \frac{\partial g_{mn}}{\partial q^k}\right), \tag{11}$$

and (g^{ik}) is the inverse matrix of (g_{ik}). Keeping the zero-order term $(1/2\epsilon)g_{ij}\Delta q^i\Delta q^j$ in the exponential and expanding higher-order terms into power series, we get from Eq. (2)

$$\begin{aligned} \psi(q(t+\epsilon), t+\epsilon) = {} & \frac{1}{A}\int \exp\left(\frac{i}{2\hbar\epsilon} g_{ij}\Delta q^i\Delta q^j\right) \\ & \times \Big[1 - \frac{i}{2\hbar\epsilon} g_{ij}\left\{ {i \atop mn} \right\}\Delta q^j\Delta q^m\Delta q^n \\ & + \frac{i}{8\hbar\epsilon} g_{ij}\left\{ {i \atop mn} \right\}\left\{ {j \atop \alpha\beta} \right\}\Delta q^m\Delta q^n\Delta q^\alpha\Delta q^\beta \\ & + \frac{i}{6\hbar\epsilon} g_{ij}\left(\frac{\partial}{\partial q^l}\left\{ {i \atop mn} \right\} + \left\{ {i \atop \alpha l} \right\}\left\{ {\alpha \atop mn} \right\}\right)\Delta q^j\Delta q^m\Delta q^n\Delta q^l \\ & - \frac{g_{ij}g_{st}}{8\hbar^2\epsilon^2}\left\{ {i \atop mn} \right\}\left\{ {s \atop \alpha\beta} \right\}\Delta q^j\Delta q^t\Delta q^m\Delta q^n\Delta q^\alpha\Delta q^\beta + \cdots\Big] \\ & \times \left(\sqrt{g(q(t+\epsilon))} - \Delta q^i\frac{\partial\sqrt{g}}{\partial q^i} + \tfrac{1}{2}\Delta q^i\Delta q^j\frac{\partial^2\sqrt{g}}{\partial q^i\partial q^j}\right) \\ & \times \left(\psi(q(t+\epsilon), t) - \Delta q^i\frac{\partial\psi}{\partial q^i} + \tfrac{1}{2}\Delta q^i\Delta q^j\frac{\partial^2\psi}{\partial q^i\partial q^j} + \cdots\right) d(\Delta q^1)\cdots d(\Delta q^N). \end{aligned} \tag{12}$$

The following are two useful identities:

$$\int\int_{-\infty}^{\infty}\cdots\int \exp\left(\frac{i}{2\hbar\epsilon} g_{ij}\Delta q^i\Delta q^j\right) d(\Delta q) = (i\pi\hbar\epsilon)^{N/2} g^{-1/2} \tag{13}$$

$$\begin{aligned} & \int\int_{-\infty}^{\infty}\cdots\int \exp\left(\frac{i}{2\hbar\epsilon} g_{ij}\Delta q^i\Delta q^j\right) \\ & \quad \times \Delta q^{\alpha_1}\Delta q^{\alpha_2}\cdots\Delta q^{\alpha_{2m}} d(\Delta q) \\ & = (i\pi\hbar\epsilon)^{N/2} g^{-1/2} (i\hbar\epsilon)^m \{g^{\alpha_1\alpha_2} g^{\alpha_3\alpha_4}\cdots g^{\alpha_{2m-1}\alpha_{2m}} \end{aligned}$$

+ terms with other possible permutations of

$$\alpha_1, \alpha_2, \cdots, \alpha_{2m}. \tag{14}$$

There are altogether $(2m-1)(2m-3)\cdots 5\cdot 3\cdot 1$ terms.

Using the two identities, we can easily find out the coefficients of $\psi(q(t+\epsilon),t), \partial\psi/\partial q^m, \partial^2\psi/\partial q^m \partial q^n$. The calculations are in Appendix B.

Here we just write down the results. Equation (12) becomes

$$\psi(q(t+\epsilon),t) + \epsilon \frac{\partial\psi}{\partial t} + \cdots = \frac{(i\pi\hbar\epsilon)^{N/2}}{A}\left\{\psi(q(t+\epsilon),t) + i\hbar\epsilon\left[\frac{1}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^m}\left(\sqrt{g}\, g^{mn}\frac{\partial\psi}{\partial q^n}\right) - \frac{R}{6}\psi\right]\right\}, \tag{15}$$

where

$$R = g^{ij}R_{ij}, \tag{16}$$

$$R_{ij} = R^{\alpha}_{ij\alpha}, \tag{17}$$

and

$$R^{l}_{ijk} = \frac{\partial}{\partial q^k}\left\{\begin{matrix} l \\ ij \end{matrix}\right\} - \frac{\partial}{\partial q^j}\left\{\begin{matrix} l \\ ik \end{matrix}\right\} + \left\{\begin{matrix} \alpha \\ ij \end{matrix}\right\}\left\{\begin{matrix} l \\ \alpha k \end{matrix}\right\} - \left\{\begin{matrix} \alpha \\ ik \end{matrix}\right\}\left\{\begin{matrix} l \\ \alpha j \end{matrix}\right\}. \tag{18}$$

Compare the coefficient up to order ϵ in Eq. (15). We get

$$A = (i\pi\hbar\epsilon)^{N/2} \tag{19}$$

and

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^m}\left(\sqrt{g}\, g^{mn}\frac{\partial\psi}{\partial q^n}\right) + \frac{\hbar^2 R}{6}\psi. \tag{20}$$

Equation (20) is the "Schrödinger equation" using Feynman's path integration formulation of quantum mechanics.

III. DISCUSSION

(a) Equation (20) above is different from the "usual Schrödinger equation" in which the term $\hbar^2 R/6$ is absent. Notice that *both* equations are convariant under any coordinate transformation $q^1 \cdots q^N \to Q^1 \cdots Q^N$.

(b) In case the curvature R vanishes, one does not have to discuss which of the two equations is to be preferred, since they are the same. Such is the case when the kinetic energy is that of a collection of nonrelativistic particles in Euclidean space where $N =$ (3 times the number of particles).

(c) If $R \neq 0$, it may seem at first sight that canonical quantization rules will yield the "usual Schrödinger equation." That is incorrect! In fact, only in the case $g_{ij} =$ constants are the canonical quantization rules

$$[P_i, q^j] = -i\hbar\delta_i{}^j$$

unambiguous and independent of coordinate transformations (if they maintain $g_{ij} =$ const). If $R \neq 0$, "canonical quantization rules" are ambiguous.

(d) The limit $\hbar \to 0$ of both equations give[2] the same results as classical mechanics, since the term $-\hbar^2 R/6$ is an equivalent potential energy and approaches 0 as $\hbar$ approaches 0.

(e) If $R \neq 0$, one can always embed the coordinate space $q^1 \cdots q^N$ as a curved subspace in a Euclidean space S of larger dimension. Does canonical quantization in the larger space S lead to a unique Schrödinger equation in the subspace? The answer to this question is no. To analyze this question, one would have to investigate first the constraint to be applied to the system in S so as to restrict the motion to the subspace. This constraint is to confine motion to a thin layer of "thickness" $\Delta(q^1 \cdots q^N)$ around the subspace and then to approach the limit $\Delta \to 0$. In classical mechanics *any* nondissipative constraint would yield the same result in the limit $\Delta \to 0$. The limiting trajectories would satisfy the Lagrangian equations for the q's, and one need not concern oneself with the larger space S. In particular the thickness Δ can depend on $q^1 \cdots q^N$. E.g., one could have

$$\Delta = A(q^1 \cdots q^N)\epsilon + O(\epsilon^2), \tag{21}$$

and take the limit $\epsilon \to 0$.

In quantum mechanics, however, the constraint produces a zero point energy. The limit for the Schrödinger equation would then depend on precisely how the limit $\Delta \to 0$ is taken. If one takes (21), and the fact that $A \neq$ const, the Schrödinger equation would acquire an infinite term $\alpha(A\epsilon)^{-2}$ which varies wildly over the q's. Consequently, the Schrödinger equation approaches no definite limit. If, on the other hand, one takes $A =$ const, then everything depends on the higher order terms in $O(\epsilon^2)$ in (21).

(f) To summarize, for a case $R \neq 0$, canonical quantization does not produce a unique Schrödinger equation, and embedding the system in a higher-dimensional Euclidean space would not help to produce a unique Schrödinger equation. The correspondence limit also does not uniquely determine a Schrödinger equation. Feynman's path integration formulation of quantization, however, does produce a unique equation, which is Eq. (20) above. The "usual Schrödinger equation" appears to be foundationless.

ACKNOWLEDGMENTS

The author wishes to thank Professor C. N. Yang for his encouragement and guidance and Dr. C. S. Hsue for many useful discussions.

APPENDIX A

In this appendix, we want to expand $S(q(t+\epsilon), q(t))$ as a power series of Δq. The equations of motion are

$$g_{mj}\ddot{q}^j = -\frac{1}{2}\left(\frac{\partial g_{mj}}{\partial q^\alpha} + \frac{\partial g_{m\alpha}}{\partial q^j} - \frac{\partial g_{\alpha j}}{\partial q^m}\right)\dot{q}^\alpha \dot{q}^j \tag{A1}$$

or

$$\ddot{q}^k = -\left\{\begin{matrix} k \\ \alpha\beta \end{matrix}\right\}\dot{q}^\alpha \dot{q}^\beta. \tag{A2}$$

Via Eq. (A2) it is not very difficult to prove

$$\frac{d}{dt}\left(\tfrac{1}{2} g_{ij}\dot{q}^i \dot{q}^j\right) = 0. \tag{A3}$$

That is,

$$S(q(t+\epsilon), q(t)) = \int_t^{t+\epsilon} L\,dt = \left[\tfrac{1}{2} g_{ij}(q(t+\epsilon))\,\dot{q}^i(t+\epsilon)\,\dot{q}^j(t+\epsilon)\right]\epsilon. \tag{A4}$$

Now if we know $\dot{q}^i(t+\epsilon)$ as a series of Δq, we know $S(q(t+\epsilon), q(t))$. In order to find out $\dot{q}^i(t+\epsilon)$, we need $\ddot{q}^k$. From Eq. (A2), we find

$$\dddot{q}^k = -\left(\frac{\partial}{\partial q^\gamma}\left\{\begin{matrix}k\\ \alpha\beta\end{matrix}\right\} - 2\left\{\begin{matrix}k\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \alpha\gamma\end{matrix}\right\}\right)\dot{q}^\alpha\dot{q}^\beta\dot{q}^\gamma. \tag{A5}$$

Now for small ϵ and $|\Delta q| \lesssim \epsilon^{1/2}$, we can write

$$q^i(t) = q^i(t+\epsilon) - \epsilon\dot{q}^i(t+\epsilon) + (\epsilon^2/2!)\ddot{q}^i(t+\epsilon) - (\epsilon^3/3!)\dddot{q}^i(t+\epsilon) + \cdots \tag{A6}$$

or

$$q^i(t) = q^i(t+\epsilon) - \epsilon\dot{q}^i(t+\epsilon) - \frac{\epsilon^2}{2!}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\dot{q}^\alpha\dot{q}^\beta + \frac{\epsilon^3}{3!}\left(\frac{\partial}{\partial q^\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\} - 2\left\{\begin{matrix}i\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \alpha\gamma\end{matrix}\right\}\right)\dot{q}^\alpha\dot{q}^\beta\dot{q}^\gamma + \cdots. \tag{A7}$$

From Eq. (A7) we get

$$\dot{q}^i(t+\epsilon) = \frac{\Delta q^i}{\epsilon} - \frac{1}{2\epsilon}\left\{\begin{matrix}i\\ mn\end{matrix}\right\}\Delta q^m\Delta q^n + \frac{1}{6\epsilon}\left(\frac{\partial}{\partial q^l}\left\{\begin{matrix}i\\ mn\end{matrix}\right\} + \left\{\begin{matrix}i\\ \alpha l\end{matrix}\right\}\left\{\begin{matrix}\alpha\\ mn\end{matrix}\right\}\right)\Delta q^m\Delta q^n\Delta q^l + \cdots. \tag{A8}$$

Substituting Eq. (A8) into Eq. (A4), we get

$$S(q(t+\epsilon), q(t)) = \frac{1}{2\epsilon}g_{ij}\left[\Delta q^i\Delta q^j - \left\{\begin{matrix}i\\ mn\end{matrix}\right\}\Delta q^j\Delta q^m\Delta q^n + \frac{1}{4}\left\{\begin{matrix}i\\ mn\end{matrix}\right\}\left\{\begin{matrix}j\\ \alpha\beta\end{matrix}\right\}\Delta q^m\Delta q^n\Delta q^\alpha\Delta q^\beta + \frac{1}{3}\left(\frac{\partial}{\partial q^l}\left\{\begin{matrix}i\\ mn\end{matrix}\right\} + \left\{\begin{matrix}i\\ \alpha l\end{matrix}\right\}\left\{\begin{matrix}\alpha\\ mn\end{matrix}\right\}\right)\Delta q^j\Delta q^m\Delta q^n\Delta q^l + \cdots\right]. \tag{A9}$$

This is Eq. (6).

APPENDIX B

In this appendix, we calculate the coefficients of ψ, $\partial\psi/\partial q$, $\partial^2\psi/\partial q\partial q$ by using Eqs. (12)–(14).

(a) $$\frac{\partial^2\psi}{\partial q^m\partial q^n} : \frac{1}{2A}\int \exp\left(\frac{i}{2\hbar\epsilon}g_{ij}\Delta q^i\Delta q^j\right)\Delta q^m\Delta q^n\sqrt{g}d(\Delta q) = \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(i\hbar\epsilon)\frac{g^{mn}}{2} \tag{B1}$$

(b) $$\frac{\partial\psi}{\partial q^n} : \frac{1}{A}\int \exp\left(\frac{1}{2\hbar\epsilon}g_{ij}\Delta q^i\Delta q^j\right)\sqrt{g}\left(\frac{i}{2\hbar\epsilon}\right) \times g_{ij}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\Delta q^n\Delta q^j\Delta q^\alpha\Delta q^\beta d(\Delta q) + \frac{1}{A}\int \exp\left(\frac{i}{2\hbar\epsilon}g_{ij}\Delta q^i\Delta q^j\right)\frac{\partial\sqrt{g}}{\partial q^i}\Delta q^i\Delta q^n$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(-i\hbar\epsilon)\frac{1}{2}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}g_{ij}\{g^{nj}g^{\alpha\beta} + g^{n\alpha}g^{j\beta} + g^{n\beta}g^{j\alpha}\} + \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(i\hbar\epsilon)g^{n\alpha}\left\{\begin{matrix}\beta\\ \alpha\beta\end{matrix}\right\}$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(-i\hbar\epsilon)\left(\tfrac{1}{2}g^{\alpha\beta}\left\{\begin{matrix}n\\ \alpha\beta\end{matrix}\right\} + g^{n\alpha}\left\{\begin{matrix}\beta\\ \alpha\beta\end{matrix}\right\}\right) + \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(i\hbar\epsilon)g^{n\alpha}\left\{\begin{matrix}\beta\\ \alpha\beta\end{matrix}\right\}$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(-i\hbar\epsilon)\tfrac{1}{2}g^{\alpha\beta}\left\{\begin{matrix}n\\ \alpha\beta\end{matrix}\right\}$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(i\hbar\epsilon)\frac{1}{2}\frac{1}{\sqrt{g}}\frac{\partial}{\partial q^m}(\sqrt{g}g^{mn}). \tag{B2}$$

In obtaining Eq. (B2) we use the identities

$$\frac{1}{\sqrt{g}}\frac{\partial\sqrt{g}}{\partial q^i} = \left\{\begin{matrix}\beta\\ i\beta\end{matrix}\right\} \tag{B3}$$

and

$$\frac{\partial g^{mn}}{\partial q^k} = -g^{m\alpha}g^{n\beta}\frac{\partial g_{\alpha\beta}}{\partial q^k}. \tag{B4}$$

Equation (B4) can be derived from

$$\frac{\partial}{\partial q^k}(g_{m\alpha}g^{n\alpha}) = \frac{\partial}{\partial q^k}(\delta^n_m) = 0. \tag{B5}$$

(c) ψ.

(1) $$\frac{1}{A}\int \exp\left(\frac{i}{2\hbar\epsilon}g_{ij}\Delta q^i\Delta q^j\right)\sqrt{g}\,d(\Delta q) = \frac{(i\pi\hbar\epsilon)^{N/2}}{A} \tag{B6}$$

(2) $$\frac{1}{A}\int \exp\left(\frac{i}{2\hbar\epsilon}g_{ij}\Delta q^i\Delta q^j\right)\times\sqrt{g} \times\left[\frac{i}{2\hbar\epsilon}g_{i\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \delta m\end{matrix}\right\}\Delta q^\alpha\Delta q^\beta\Delta q^\gamma\Delta q^\delta + \frac{1}{2}\left(\left\{\begin{matrix}\beta\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}\alpha\\ n\alpha\end{matrix}\right\} + \frac{\partial}{\partial q^n}\left\{\begin{matrix}\beta\\ m\beta\end{matrix}\right\}\right)\Delta q^m\Delta q^n + \left(\frac{i}{8\hbar\epsilon}g_{ij}\left\{\begin{matrix}j\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ \gamma\delta\end{matrix}\right\} + \frac{i}{6\hbar\epsilon}g_{i\delta}\times\left(\frac{\partial}{\partial q^\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\} + \left\{\begin{matrix}i\\ m\alpha\end{matrix}\right\}\left\{\begin{matrix}m\\ \beta\gamma\end{matrix}\right\}\right)\right)\Delta q^\alpha\Delta q^\beta\Delta q^\gamma\Delta q^\delta - \frac{1}{8\hbar^2\epsilon^2}g_{i\gamma}g_{j\delta}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ mn\end{matrix}\right\} \times\Delta q^\alpha\Delta q^\beta\Delta q^\gamma\Delta q^\delta\Delta q^m\Delta q^n\Big]d(\Delta q)$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}(i\hbar\epsilon)\times\Big\{-\tfrac{1}{2}g_{i\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \delta m\end{matrix}\right\}(\alpha\beta\gamma\delta) + \frac{1}{2}\left(\left\{\begin{matrix}\beta\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}\alpha\\ n\alpha\end{matrix}\right\} + \frac{\partial}{\partial q^n}\left\{\begin{matrix}\beta\\ m\beta\end{matrix}\right\}\right)(mn) - \left[\frac{1}{8}g_{ij}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ \partial\delta\end{matrix}\right\} + \frac{1}{6}g_{i\delta}\left(\frac{\partial}{\partial q^\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\} + \left\{\begin{matrix}i\\ m\alpha\end{matrix}\right\}\left\{\begin{matrix}m\\ \beta\gamma\end{matrix}\right\}\right)\right](\alpha\beta\gamma\delta) + \frac{1}{8}g_{i\gamma}g_{j\delta}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ mn\end{matrix}\right\}(\alpha\beta\gamma\delta mn)\Big\}. \tag{B7}$$

In these equations, $(\alpha_1, \alpha_2 \cdots \alpha_{2m})$ stands for $(g^{\alpha_1\alpha_2}\cdots g^{\alpha_{2m-1}\alpha_{2m}})$ + terms with other possible permutation of $(\alpha_1\alpha_2\cdots\alpha_{2m})$.

The term

$$\tfrac{1}{8}g_{i\gamma}g_{j\delta}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ mn\end{matrix}\right\}(\alpha\beta\gamma\delta mn) = \tfrac{1}{8}g_{i\gamma}g_{j\delta}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ mn\end{matrix}\right\}[g^{m\delta}(\alpha\beta\gamma n) + g^{n\delta}(\beta\gamma\delta m) + g^{\alpha\delta}(mn\gamma\beta) + g^{\beta\delta}(mn\alpha\gamma) + g^{\gamma\delta}(mn\alpha\beta)]$$

$$= \tfrac{1}{8}g_{i\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ m\delta\end{matrix}\right\}(\alpha\beta\gamma\delta) + \tfrac{1}{8}g_{i\gamma}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ m\delta\end{matrix}\right\}(\alpha\beta\gamma\delta) + \tfrac{1}{8}g_{i\gamma}\left\{\begin{matrix}i\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \alpha\delta\end{matrix}\right\}(\alpha\beta\gamma\delta) + \tfrac{1}{8}g_{i\gamma}\left\{\begin{matrix}i\\ m\beta\end{matrix}\right\}\left\{\begin{matrix}m\\ \alpha\delta\end{matrix}\right\}(\alpha\beta\gamma\delta) + \tfrac{1}{8}g_{ij}\left\{\begin{matrix}i\\ \alpha\beta\end{matrix}\right\}\left\{\begin{matrix}j\\ \gamma\delta\end{matrix}\right\}(\alpha\beta\gamma\delta).$$

Thus Eq. (B7) becomes

$$(\text{B7}) = \frac{(i\pi\hbar\epsilon)^{N/2}}{A}\,(i\hbar\epsilon)\times\Big[\Big(-\frac{1}{4}\,g_{i\gamma}\begin{Bmatrix} i \\ \alpha\beta\end{Bmatrix}\begin{Bmatrix} m \\ \delta m\end{Bmatrix} - \tfrac{1}{6}\,g_{i\delta}\,\frac{\partial}{\partial q^{\gamma}}\begin{Bmatrix} i \\ \alpha\beta\end{Bmatrix} + \frac{1}{12}\,g_{i\delta}\begin{Bmatrix} i \\ m\alpha\end{Bmatrix}\begin{Bmatrix} m \\ \beta\gamma\end{Bmatrix}\Big)\times(\alpha\beta\gamma\delta)$$

$$+\frac{1}{2}\Big(\begin{Bmatrix}\beta\\ m\beta\end{Bmatrix}\begin{Bmatrix}\alpha\\ n\alpha\end{Bmatrix} + \frac{\partial}{\partial q^{n}}\begin{Bmatrix}\beta\\ m\beta\end{Bmatrix}\Big)(mn)\Big]$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}\Big(\frac{-ih\epsilon}{6}\Big)\Big(\frac{\partial}{\partial q^{m}}\begin{Bmatrix} m \\ \alpha\beta\end{Bmatrix} - \frac{\partial}{\partial q^{\alpha}}\begin{Bmatrix} m \\ \beta m\end{Bmatrix} + \begin{Bmatrix} m \\ mn\end{Bmatrix}\begin{Bmatrix} n \\ \alpha\beta\end{Bmatrix} - \begin{Bmatrix} m \\ m\alpha\end{Bmatrix}\begin{Bmatrix} n \\ n\beta\end{Bmatrix}\Big)g^{\alpha\beta}. \tag{B8}$$

---

* Supported in part by NSF Grant GP32998X.

[1] R. P. Feynman, Rev. Mod. Phys. 20, 327 (1948).

[2] B. S. DeWitt, Rev. Mod. Phys. 29, 377 (1957).

An "*H*-Theorem" for Multiplicative Stochastic Processes

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 7 June 1972)

In a recent paper the author showed how multiplicative stochastic processes lead to a potentially comprehensive theory for nonequilibrium phenomena. In this paper an "*H* theorem" is proved from results obtained using multiplicative stochastic processes.

INTRODUCTION

In another paper,[1] the theory of multiplicative stochastic processes was explained, and it was shown how such a mathematical theory leads to a formalism for nonequilibrium thermodynamics: In this paper the thermodynamical "*H* function" will be introduced, and a proof of an "*H* theorem" will be presented.

RECAPITULATION

The Schrödinger equation for nonrelativistic quantum mechanics may be written in matrix form as

$$i\frac{d}{dt}C_{\alpha}(t) = \sum_{\alpha'} M_{\alpha\alpha'}C_{\alpha'}(t), \tag{1}$$

where $M_{\alpha\alpha'} = M^{*}_{\alpha\alpha}$, which is the condition of Hermiticity, and $\sum_{\alpha} C^{*}_{\alpha}(t)C_{\alpha}(t) = 1$, which is the condition of conservation of total probability. The Hermiticity of $M_{\alpha\alpha'}$ in (1) is necessary and sufficient for the conservation of total probability. Suppose that a fluctuating contribution to the Hamiltonian is considered. Then (1) becomes

$$i\frac{d}{dt}C_{\alpha}(t) = \sum_{\alpha'}M_{\alpha\alpha'}C_{\alpha'}(t) + \sum_{\alpha'}\tilde{M}_{\alpha\alpha'}(t)C_{\alpha'}(t), \tag{2}$$

where $\tilde{M}_{\alpha\alpha'}(t) = \tilde{M}^{*}_{\alpha'\alpha}(t)$, and the following properties hold for the averaged moments of $\tilde{M}_{\alpha\alpha'}(t)$[1]:

$$\langle\tilde{M}_{\alpha\alpha'}(t)\rangle = 0, \tag{3}$$

$$\langle\tilde{M}_{\alpha\alpha'}(t)\tilde{M}_{\beta\beta'}(s)\rangle = 2Q_{\alpha\alpha'\beta\beta'}\delta(t-s), \tag{4}$$

$$\langle\tilde{M}_{\mu_1\nu_1}(t_1)\cdots\tilde{M}_{\mu_{2n-1}\nu_{2n-1}}(t_{2n-1})\rangle = 0 \quad \text{for } n = 1, 2, \ldots, \tag{5}$$

$$\begin{aligned}\langle\tilde{M}_{\mu_1\nu_1}(t_1)\cdots\tilde{M}_{\mu_{2n}\nu_{2n}}(t_{2n})\rangle &= \frac{1}{2^n n!}\sum_{\rho\in S_{2n}}\prod_{j=1}^{n}\langle\tilde{M}_{\mu_{\rho(2j-1)}\nu_{\rho(2j-1)}} \\ &\quad\times(t_{\rho(2j-1)})\tilde{M}_{\mu_{\rho(2j)}\nu_{\rho(2j)}}(t_{\rho(2j)})\rangle \\ &= \frac{1}{2^n n!}\sum_{\rho\in S_{2n}}\prod_{j=1}^{n}2Q_{\mu_{\rho(2j-1)}\nu_{\rho(2j-1)}\mu_{\rho(2j)}\nu_{\rho(2j)}} \\ &\quad\times\delta(t_{\rho(2j-1)} - t_{\rho(2j)}),\end{aligned} \tag{6}$$

where S_{2n} is the symmetric group of order $(2n)!$ The properties given by (3)–(6) are those appropriate for a purely random, Gaussian, stochastic matrix.

A density matrix representation for the Schrödinger equation is obtained in terms of the density matrix $\rho_{\alpha\beta}$, which is defined by

$$\rho_{\alpha\beta}(t) \equiv C^{*}_{\alpha}(t)C_{\beta}(t). \tag{7}$$

If $L_{\alpha\beta\alpha'\beta'}$ and $\tilde{L}_{\alpha\beta\alpha'\beta'}(t)$ are defined by

$$\begin{aligned}L_{\alpha\beta\alpha'\beta} &\equiv \delta_{\alpha\alpha'}M_{\beta\beta'} - \delta_{\beta\beta'}M^{*}_{\alpha\alpha'}, \\ \tilde{L}_{\alpha\beta\alpha'\beta'}(t) &\equiv \delta_{\alpha\alpha'}\tilde{M}_{\beta\beta'}(t) - \delta_{\beta\beta'}\tilde{M}^{*}_{\alpha\alpha'}(t),\end{aligned} \tag{8}$$

then Eq. (2) may be used to directly verify

$$i\frac{d}{dt}\rho_{\alpha\beta}(t) = \sum_{\alpha'}\sum_{\beta'}[L_{\alpha\beta\alpha'\beta'} + \tilde{L}_{\alpha\beta\alpha'\beta'}(t)]\rho_{\alpha'\beta'}(t). \tag{9}$$

This is the density matrix equation. By averaging over the stochastic contribution by means of properties (3)–(6), an equation for the averaged density matrix, $\langle\rho_{\alpha\beta}(t)\rangle$, may be obtained, although only after significant computation[1]:

$$\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle = -i\sum_{\alpha'}\sum_{\beta'}L_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle - \sum_{\alpha'}\sum_{\beta'}R_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle. \tag{10}$$

The matrix $R_{\alpha\beta\alpha'\beta'}$ which appears in (10) is defined by[1]

$$R_{\alpha\beta\alpha'\beta'} \equiv \delta_{\alpha\alpha'}\sum_{\theta}Q_{\beta\theta\theta\beta'} + \delta_{\beta\beta'}\sum_{\theta}Q_{\theta\alpha\alpha'\theta} - Q_{\beta\beta'\alpha'\alpha} - Q_{\alpha'\alpha\beta\beta'}. \tag{11}$$

It is also provable that for arbitrary complex matrices $X_{\alpha\beta}$,

$$\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}X^{*}_{\alpha\beta}R_{\alpha\beta\alpha'\beta'}X_{\alpha'\beta'} \geq 0 \tag{12}$$

and

$$\sum_{\alpha}R_{\alpha\alpha\mu\nu} = 0. \tag{13}$$

Conditions (12) and (13) lead to irreversible behavior in (10) with the equilibrium state being proportional

Thus Eq. (B7) becomes

$$(\text{B7}) = \frac{(i\pi\hbar\epsilon)^{N/2}}{A}\,(i\hbar\epsilon)\times\Big[\Big(-\frac{1}{4}\,g_{i\gamma}\begin{Bmatrix} i \\ \alpha\beta\end{Bmatrix}\begin{Bmatrix} m \\ \delta m\end{Bmatrix} - \tfrac{1}{6}\,g_{i\delta}\,\frac{\partial}{\partial q^{\gamma}}\begin{Bmatrix} i \\ \alpha\beta\end{Bmatrix} + \frac{1}{12}\,g_{i\delta}\begin{Bmatrix} i \\ m\alpha\end{Bmatrix}\begin{Bmatrix} m \\ \beta\gamma\end{Bmatrix}\Big)\times(\alpha\beta\gamma\delta)$$

$$+\frac{1}{2}\Big(\begin{Bmatrix}\beta\\ m\beta\end{Bmatrix}\begin{Bmatrix}\alpha\\ n\alpha\end{Bmatrix} + \frac{\partial}{\partial q^{n}}\begin{Bmatrix}\beta\\ m\beta\end{Bmatrix}\Big)(mn)\Big]$$

$$= \frac{(i\pi\hbar\epsilon)^{N/2}}{A}\Big(\frac{-ih\epsilon}{6}\Big)\Big(\frac{\partial}{\partial q^{m}}\begin{Bmatrix} m \\ \alpha\beta\end{Bmatrix} - \frac{\partial}{\partial q^{\alpha}}\begin{Bmatrix} m \\ \beta m\end{Bmatrix} + \begin{Bmatrix} m \\ mn\end{Bmatrix}\begin{Bmatrix} n \\ \alpha\beta\end{Bmatrix} - \begin{Bmatrix} m \\ m\alpha\end{Bmatrix}\begin{Bmatrix} n \\ n\beta\end{Bmatrix}\Big)g^{\alpha\beta}. \tag{B8}$$

---

* Supported in part by NSF Grant GP32998X.

[1] R. P. Feynman, Rev. Mod. Phys. 20, 327 (1948).

[2] B. S. DeWitt, Rev. Mod. Phys. 29, 377 (1957).

An "*H*-Theorem" for Multiplicative Stochastic Processes

Ronald Forrest Fox

School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 7 June 1972)

In a recent paper the author showed how multiplicative stochastic processes lead to a potentially comprehensive theory for nonequilibrium phenomena. In this paper an "*H* theorem" is proved from results obtained using multiplicative stochastic processes.

INTRODUCTION

In another paper,[1] the theory of multiplicative stochastic processes was explained, and it was shown how such a mathematical theory leads to a formalism for nonequilibrium thermodynamics: In this paper the thermodynamical "*H* function" will be introduced, and a proof of an "*H* theorem" will be presented.

RECAPITULATION

The Schrödinger equation for nonrelativistic quantum mechanics may be written in matrix form as

$$i\frac{d}{dt}C_{\alpha}(t) = \sum_{\alpha'} M_{\alpha\alpha'}C_{\alpha'}(t), \tag{1}$$

where $M_{\alpha\alpha'} = M^{*}_{\alpha\alpha}$, which is the condition of Hermiticity, and $\sum_{\alpha} C^{*}_{\alpha}(t)C_{\alpha}(t) = 1$, which is the condition of conservation of total probability. The Hermiticity of $M_{\alpha\alpha'}$ in (1) is necessary and sufficient for the conservation of total probability. Suppose that a fluctuating contribution to the Hamiltonian is considered. Then (1) becomes

$$i\frac{d}{dt}C_{\alpha}(t) = \sum_{\alpha'}M_{\alpha\alpha'}C_{\alpha'}(t) + \sum_{\alpha'}\tilde{M}_{\alpha\alpha'}(t)C_{\alpha'}(t), \tag{2}$$

where $\tilde{M}_{\alpha\alpha'}(t) = \tilde{M}^{*}_{\alpha'\alpha}(t)$, and the following properties hold for the averaged moments of $\tilde{M}_{\alpha\alpha'}(t)$[1]:

$$\langle\tilde{M}_{\alpha\alpha'}(t)\rangle = 0, \tag{3}$$

$$\langle\tilde{M}_{\alpha\alpha'}(t)\tilde{M}_{\beta\beta'}(s)\rangle = 2Q_{\alpha\alpha'\beta\beta'}\delta(t-s), \tag{4}$$

$$\langle\tilde{M}_{\mu_1\nu_1}(t_1)\cdots\tilde{M}_{\mu_{2n-1}\nu_{2n-1}}(t_{2n-1})\rangle = 0 \quad \text{for } n = 1, 2, \ldots, \tag{5}$$

$$\begin{aligned}\langle\tilde{M}_{\mu_1\nu_1}(t_1)\cdots\tilde{M}_{\mu_{2n}\nu_{2n}}(t_{2n})\rangle &= \frac{1}{2^n n!}\sum_{\rho\in S_{2n}}\prod_{j=1}^{n}\langle\tilde{M}_{\mu_{\rho(2j-1)}\nu_{\rho(2j-1)}} \\ &\quad\times(t_{\rho(2j-1)})\tilde{M}_{\mu_{\rho(2j)}\nu_{\rho(2j)}}(t_{\rho(2j)})\rangle \\ &= \frac{1}{2^n n!}\sum_{\rho\in S_{2n}}\prod_{j=1}^{n}2Q_{\mu_{\rho(2j-1)}\nu_{\rho(2j-1)}\mu_{\rho(2j)}\nu_{\rho(2j)}} \\ &\quad\times\delta(t_{\rho(2j-1)} - t_{\rho(2j)}),\end{aligned} \tag{6}$$

where S_{2n} is the symmetric group of order $(2n)!$ The properties given by (3)–(6) are those appropriate for a purely random, Gaussian, stochastic matrix.

A density matrix representation for the Schrödinger equation is obtained in terms of the density matrix $\rho_{\alpha\beta}$, which is defined by

$$\rho_{\alpha\beta}(t) \equiv C^{*}_{\alpha}(t)C_{\beta}(t). \tag{7}$$

If $L_{\alpha\beta\alpha'\beta'}$ and $\tilde{L}_{\alpha\beta\alpha'\beta'}(t)$ are defined by

$$\begin{aligned}L_{\alpha\beta\alpha'\beta} &\equiv \delta_{\alpha\alpha'}M_{\beta\beta'} - \delta_{\beta\beta'}M^{*}_{\alpha\alpha'}, \\ \tilde{L}_{\alpha\beta\alpha'\beta'}(t) &\equiv \delta_{\alpha\alpha'}\tilde{M}_{\beta\beta'}(t) - \delta_{\beta\beta'}\tilde{M}^{*}_{\alpha\alpha'}(t),\end{aligned} \tag{8}$$

then Eq. (2) may be used to directly verify

$$i\frac{d}{dt}\rho_{\alpha\beta}(t) = \sum_{\alpha'}\sum_{\beta'}[L_{\alpha\beta\alpha'\beta'} + \tilde{L}_{\alpha\beta\alpha'\beta'}(t)]\rho_{\alpha'\beta'}(t). \tag{9}$$

This is the density matrix equation. By averaging over the stochastic contribution by means of properties (3)–(6), an equation for the averaged density matrix, $\langle\rho_{\alpha\beta}(t)\rangle$, may be obtained, although only after significant computation[1]:

$$\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle = -i\sum_{\alpha'}\sum_{\beta'}L_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle - \sum_{\alpha'}\sum_{\beta'}R_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle. \tag{10}$$

The matrix $R_{\alpha\beta\alpha'\beta'}$ which appears in (10) is defined by[1]

$$R_{\alpha\beta\alpha'\beta'} \equiv \delta_{\alpha\alpha'}\sum_{\theta}Q_{\beta\theta\theta\beta'} + \delta_{\beta\beta'}\sum_{\theta}Q_{\theta\alpha\alpha'\theta} - Q_{\beta\beta'\alpha'\alpha} - Q_{\alpha'\alpha\beta\beta'}. \tag{11}$$

It is also provable that for arbitrary complex matrices $X_{\alpha\beta}$,

$$\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}X^{*}_{\alpha\beta}R_{\alpha\beta\alpha'\beta'}X_{\alpha'\beta'} \geq 0 \tag{12}$$

and

$$\sum_{\alpha}R_{\alpha\alpha\mu\nu} = 0. \tag{13}$$

Conditions (12) and (13) lead to irreversible behavior in (10) with the equilibrium state being proportional

to the identity matrix.[1] In the following these properties will be proved within the context of the "H theorem."

PROOF OF THE "H THEOREM"

By virtue of its definition, $\rho_{\alpha\beta}(t)$, and also $\langle\rho_{\alpha\beta}(t)\rangle$, is a positive definite Hermitian matrix. This implies that there exists a unitary matrix $U_{\alpha\beta}(t)$, such that

$$U_{\alpha\alpha'}(t)\langle\rho_{\alpha'\beta'}(t)\rangle U^{-1}_{\beta'\beta}(t) = D_{\alpha\beta}(t), \tag{14}$$

where $D_{\alpha\beta}(t)$ is diagonal

$$D_{\alpha\beta}(t) = D_{\alpha}(t)\delta_{\alpha\beta} \tag{15}$$

with positive diagonal elements $D_{\alpha}(t)$. Note that at each instant of time, there will correspond a particular, time-dependent unitary matrix $U_{\alpha\beta}(t)$.

The logarithm of the averaged density matrix is defined by

$$\begin{aligned}[\log_e(\langle\boldsymbol{\rho}(t)\rangle)]_{\alpha\beta} &\equiv \sum_{\alpha'}\sum_{\beta'} U^{-1}_{\alpha\alpha'}(t)\,\log_e(D_{\alpha'}(t))\delta_{\alpha'\beta'}U_{\beta'\beta}(t)\\ &= \sum_{\theta} U^{-1}_{\alpha\theta}\,\log_e(D_{\theta}(t))U_{\theta\beta}(t).\end{aligned} \tag{16}$$

This procedure is required in order that the logarithm of a density matrix is well defined.

The H function is given in terms of the averaged density matrix by[2,3]

$$\begin{aligned}H(t) &= \mathrm{Tr}[\langle\boldsymbol{\rho}(t)\rangle\,\log_e(\langle\boldsymbol{\rho}(t)\rangle)]\\ &= \sum_{\alpha}\sum_{\beta}\sum_{\theta}\sum_{\theta'} U^{-1}_{\alpha\theta}D_{\theta}(t)U_{\theta\beta}\,U^{-1}_{\beta\theta'}\,\log_e(D_{\theta'}(t))\,U_{\theta'\alpha}\\ &= \sum_{\theta} D_{\theta}(t)\,\log_e(D_{\theta}(t)),\end{aligned} \tag{17}$$

where the last two equalities follow from (14)–(16).

Consider the time derivative of $H(t)$:

$$\frac{d}{dt}H(t) = \sum_{\theta}\frac{dD_{\theta}(t)}{dt}\,\log_e(D_{\theta}(t)) + \sum_{\theta} D_{\theta}(t)\frac{1}{D_{\theta}(t)}\frac{dD_{\theta}(t)}{dt}. \tag{18}$$

The second sum in (18) is simply $\sum_{\theta}(d/dt)D_{\theta}(t)$ which vanishes by virtue of conservation of total probability. Therefore, the time change of $H(t)$ becomes

$$\frac{d}{dt}H(t) = \sum_{\theta}\frac{dD_{\theta}(t)}{dt}\log_e(D_{\theta}(t)). \tag{19}$$

In order to compute this remaining sum it is necessary to compute $dD_{\theta}(t)/dt$.

Using (14) gives

$$\begin{aligned}\frac{d}{dt}D_{\theta}(t) &= \frac{d}{dt}\sum_{\alpha}\sum_{\beta}U_{\theta\alpha}(t)\langle\rho_{\alpha\beta}(t)\rangle U^{-1}_{\beta\theta}(t)\\ &= \sum_{\alpha}\sum_{\beta}\Big[\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)\langle\rho_{\alpha\beta}(t)\rangle U^{-1}_{\beta\theta}(t)\\ &\quad + U_{\theta\alpha}(t)\Big(\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle\Big)U^{-1}_{\beta\theta}(t)\\ &\quad + U_{\theta\alpha}(t)\langle\rho_{\alpha\beta}(t)\rangle\Big(\frac{d}{dt}U^{-1}_{\beta\theta}(t)\Big)\Big].\end{aligned} \tag{20}$$

Consider the first and third terms of the last multiple sum. Again using (14) gives

$$\begin{aligned}&\sum_{\alpha}\sum_{\beta}\Big[\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)\langle\rho_{\alpha\beta}(t)\rangle U^{-1}_{\beta\theta}(t)\\ &\qquad + U_{\theta\alpha}(t)\langle\rho_{\alpha\beta}(t)\rangle\Big(\frac{d}{dt}U^{-1}_{\beta\theta}(t)\Big)\Big]\\ &= \sum_{\alpha}\sum_{\beta}\sum_{\mu}\Big[\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)U^{-1}_{\alpha\mu}(t)D_{\mu}(t)U_{\mu\beta}(t)U^{-1}_{\beta\theta}(t)\\ &\qquad + U_{\theta\alpha}(t)U^{-1}_{\alpha\mu}(t)D_{\mu}(t)U_{\mu\beta}(t)\Big(\frac{d}{dt}U^{-1}_{\beta\theta}(t)\Big)\Big]\\ &= \sum_{\alpha}\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)U^{-1}_{\alpha\theta}(t)D_{\theta}(t)\\ &\qquad + \sum_{\beta}D_{\theta}(t)U_{\theta\beta}(t)\Big(\frac{d}{dt}U^{-1}_{\beta\theta}(t)\Big)\\ &= D_{\theta}(t)\Big\{\sum_{\alpha}\Big[\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)U^{-1}_{\alpha\theta}(t) + U_{\theta\alpha}(t)\Big(\frac{d}{dt}U^{-1}_{\alpha\theta}(t)\Big)\Big]\Big\}.\end{aligned} \tag{21}$$

However, because $\sum_{\alpha}U_{\theta\alpha}(t)U^{-1}_{\alpha\theta}(t) = 1$ since $U_{\alpha\beta}(t)$ is unitary for each time t, then it follows that

$$\begin{aligned}&\frac{d}{dt}\sum_{\alpha}U_{\theta\alpha}(t)U^{-1}_{\alpha\theta}(t)\\ &\quad = \sum_{\alpha}\Big[\Big(\frac{d}{dt}U_{\theta\alpha}(t)\Big)U^{-1}_{\alpha\theta}(t) + U_{\theta\alpha}(t)\Big(\frac{d}{dt}U^{-1}_{\alpha\theta}(t)\Big)\Big] = 0.\end{aligned} \tag{22}$$

Therefore, only the middle term in the last multiple sum of (20) yields a possibly nonzero result. Therefore, (20) becomes

$$\begin{aligned}\frac{d}{dt}D_{\theta}(t) &= \sum_{\alpha}\sum_{\beta}U_{\theta\alpha}(t)\Big(\frac{d}{dt}\langle\rho_{\alpha\beta}(t)\rangle\Big)U^{-1}_{\beta\theta}(t)\\ &= \sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)(-iL_{\alpha\beta\alpha'\beta'} - R_{\alpha\beta\alpha'\beta'})\\ &\quad\times\langle\rho_{\alpha'\beta'}(t)\rangle U^{-1}_{\beta\theta}(t),\end{aligned} \tag{23}$$

where the second equality follows from (10).

Consider the term in (23) which contains $L_{\alpha\beta\alpha'\beta'}$. Using (14) and (15) gives

$$\begin{aligned}&-i\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)L_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle U^{-1}_{\beta\theta}(t)\\ &\quad = -i\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}\sum_{\mu}U_{\theta\alpha}(t)L_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\mu}(t)\\ &\qquad\times D_{\mu}(t)U_{\mu\beta'}(t)U^{-1}_{\beta\theta}(t)\\ &\quad = -i\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}\sum_{\mu}U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)L_{\alpha\beta\alpha'\beta'}\\ &\qquad\times U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)D_{\mu}(t).\end{aligned} \tag{24}$$

By returning to (8) it is readily seen that

$$\begin{aligned}&\sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)L_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &\quad = \sum_{\alpha}\sum_{\beta}\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)(\delta_{\alpha\alpha'}M_{\beta\beta'}\\ &\qquad - \delta_{\beta\beta'}M^{*}_{\alpha\alpha'})U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &\quad = \delta_{\theta\mu}\Big(\sum_{\beta}\sum_{\beta'}U^{-1}_{\beta\theta}(t)M_{\beta\beta'}U_{\mu\beta'}(t)\\ &\qquad - \sum_{\alpha}\sum_{\alpha'}U_{\theta\alpha}(t)M^{*}_{\alpha\alpha'}U^{-1}_{\alpha'\mu}(t)\Big)\\ &\quad = 0\end{aligned} \tag{25}$$

The last equality follows directly from the Hermiticity of $M_{\alpha\alpha'}$ and the unitarity of $U_{\alpha\beta}(t)$ which lead to

$$\sum_{\beta}\sum_{\beta'}U^{-1}_{\beta\theta}(t)M_{\beta\beta'}U_{\theta\beta'}(t) = \sum_{\alpha}\sum_{\alpha'}U_{\theta\alpha}(t)M^{*}_{\alpha\alpha'}U^{-1}_{\alpha'\theta}(t). \tag{26}$$

Note that the $\delta_{\theta\mu}$ in (25) has been used to replace μ by θ in (26). Returning to (23), it has been proved that

$$\frac{d}{dt}D_\theta(t) = -\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)R_{\alpha\beta\alpha'\beta'}\langle\rho_{\alpha'\beta'}(t)\rangle U^{-1}_{\beta\theta}(t). \quad (27)$$

Using (14) and (15) converts (27) into

$$\begin{aligned}\frac{d}{dt}D_\theta(t) &= -\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'}\sum_\mu U_{\theta\alpha}(t)R_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\mu}(t)\\ &\quad\times D_\mu(t)U_{\mu\beta'}(t)U^{-1}_{\beta\theta}(t)\\ &= -\sum_\mu\Bigg(\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)R_{\alpha\beta\alpha'\beta'}\\ &\quad\times U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\Bigg)D_\mu(t)\\ &= -\sum_\mu W_{\theta\mu}(t)D_\mu(t), \end{aligned}\quad (28)$$

where $W_{\theta\mu}(t)$ is defined by

$$W_{\theta\mu}(t) \equiv \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)R_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t). \quad (29)$$

Using (28) in (19) gives

$$\frac{d}{dt}H(t) = -\sum_\theta\sum_\mu W_{\theta\mu}(t)D_\mu(t)\log_e(D_\theta(t)). \quad (30)$$

From (13), (29), and the unitarity of $U_{\alpha\beta}(t)$, it follows that

$$\sum_\mu W_{\mu\theta}(t) = 0. \quad (31)$$

This may be expressed by

$$W_{\theta\theta}(t) = -\sum_{\substack{\mu\\ \mu\neq\theta}} W_{\mu\theta}(t). \quad (32)$$

Combining (30) with (32) results in

$$\frac{d}{dt}H(t) = -\sum_\theta\sum_\mu[W_{\theta\mu}(t)D_\mu(t) - W_{\mu\theta}(t)D_\theta(t)]\log_e(D_\theta(t)). \quad (33)$$

By interchanging μ and θ, (33) also gives

$$\begin{aligned}\frac{d}{dt}H(t) &= -\sum_\theta\sum_\mu[W_{\mu\theta}(t)D_\theta(t) - W_{\theta\mu}(t)D_\mu(t)]\log_e(D_\mu(t))\\ &= \sum_\theta\sum_\mu[W_{\theta\mu}(t)D_\mu(t) - W_{\mu\theta}(t)D_\theta(t)]\log_e(D_\mu(t)). \end{aligned}\quad (34)$$

Combining (33) and (34) gives

$$\frac{d}{dt}H(t) = -\tfrac{1}{2}\sum_\theta\sum_\mu[W_{\theta\mu}(t)D_\mu(t) - W_{\mu\theta}(t)D_\theta(t)]\log_e\left(\frac{D_\theta(t)}{D_\mu(t)}\right). \quad (35)$$

Before consideration of (35) may be completed it is necessary to prove two additional properties of $W_{\theta\mu}(t)$.

The two properties of $W_{\theta\mu}(t)$ to be proved below are

$$W_{\theta\mu}(t) = W_{\mu\theta}(t) \quad (36)$$

and

$$W_{\theta\mu}(t) \leq 0 \quad \text{for every} \quad \theta \neq \mu. \quad (37)$$

The proof of (36) starts with (29) and uses (11):

$$\begin{aligned}W_{\mu\theta}(t) &= \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\mu\alpha}(t)U^{-1}_{\beta\mu}(t)R_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\theta}(t)U_{\theta\beta'}(t)\\ &= \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)R_{\beta'\alpha'\beta\alpha}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t), \end{aligned}\quad (38)$$

where the last equality follows upon the index exchanges $\alpha \leftrightarrow \beta'$ and $\beta \leftrightarrow \alpha'$. From (11) it follows that

$$R_{\beta'\alpha'\beta\alpha} = \delta_{\beta\beta'}\sum_\theta Q_{\alpha'\theta\theta\alpha} + \delta_{\alpha\alpha'}\sum_\theta Q_{\theta\beta'\beta\theta} - Q_{\alpha'\alpha\beta\beta'} - Q_{\beta\beta'\alpha'\alpha}. \quad (39)$$

From (4) it follows that

$$Q_{\alpha\beta\mu\nu} = Q_{\mu\nu\alpha\beta}. \quad (40)$$

Using (40) in (39) and looking at (11) gives

$$R_{\beta'\alpha'\beta\alpha} = R_{\alpha\beta\alpha'\beta'}. \quad (41)$$

Therefore, (38) becomes

$$W_{\mu\theta}(t) = \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)R_{\alpha\beta\alpha'\beta}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t) = W_{\theta\mu}(t). \quad (42)$$

This completes the proof of (36).

The proof of (37) requires use of (11) and (4) also. Taking (29) and using (11) for $\mu \neq \theta$ gives

$$\begin{aligned}W_{\theta\mu}(t) &= \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'} U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)R_{\alpha\beta\alpha'\beta'}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &= \sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'}[U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)\delta_{\alpha\alpha'}\sum_\theta Q_{\beta\theta\theta\beta'}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &\quad + U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)\delta_{\beta\beta'}\sum_\theta Q_{\theta\alpha\alpha'\theta}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &\quad - U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)Q_{\beta\beta'\alpha'\alpha}\ U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &\quad - U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)Q_{\alpha'\alpha\beta\beta'}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)]\\ &= \delta_{\theta\mu}\sum_\beta\sum_{\beta'}U^{-1}_{\beta\theta}(t)\sum_\theta Q_{\beta\theta\theta\beta'}\,U_{\mu\beta'}(t)\\ &\quad + \delta_{\theta\mu}\sum_\alpha\sum_{\alpha'}U_{\theta\alpha}(t)\sum_\theta Q_{\theta\alpha\alpha'\theta}U^{-1}_{\alpha'\mu}(t)\\ &\quad - 2\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)Q_{\beta\beta'\alpha'\alpha}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t)\\ &= -2\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'}U_{\theta\alpha}(t)U^{-1}_{\beta\theta}(t)Q_{\beta\beta'\alpha'\alpha}U^{-1}_{\alpha'\mu}(t)U_{\mu\beta'}(t). \end{aligned}\quad (43)$$

The terms with the $\delta_{\theta\mu}$ factor give zero for $\mu \neq \theta$, and (40) has been used to combine the remaining terms. With (4), $Q_{\beta\beta'\alpha'\alpha}$ may be written as

$$Q_{\beta\beta'\alpha'\alpha} = \int_0^t \langle\tilde{M}_{\beta\beta'}(t)M_{\alpha'\alpha}(s)\rangle\,ds. \quad (44)$$

Using (44) in the last equality of (43) gives

$$\begin{aligned}W_{\theta\mu}(t) &= -2\sum_\alpha\sum_\beta\sum_{\alpha'}\sum_{\beta'}\int_0^t\langle[U^{-1}_{\beta\theta}(t)\tilde{M}_{\beta\beta'}(t)U_{\mu\beta'}(t)]\\ &\quad\times[U_{\theta\alpha}(t)\tilde{M}_{\alpha'\alpha}(s)U^{-1}_{\alpha'\mu}(t)]\rangle\,ds\\ &= -2\int_0^t\langle[\sum_\beta\sum_{\beta'}U_{\mu\beta'}(t)\tilde{M}^*_{\beta'\beta}(t)U^{-1}_{\beta\theta}(t)]\\ &\quad\times[\sum_\alpha\sum_{\alpha'}U^*_{\mu\alpha'}(t)\tilde{M}_{\alpha'\alpha}(s)U^{-1\,*}_{\alpha\theta}(t)]\rangle\,ds\\ &= -2\int_0^t\langle[\sum_\alpha\sum_{\alpha'}U^*_{\mu\alpha}(t)\tilde{M}_{\alpha\alpha'}(t)U^{-1\,*}_{\alpha'\theta}(t)]^*\\ &\quad\times[\sum_\alpha\sum_{\alpha'}U^*_{\mu\alpha}(s)\tilde{M}_{\alpha\alpha'}(s)U^{-1\,*}_{\alpha'\theta}(s)]\rangle\,ds. \end{aligned}\quad (45)$$

This last equality involves renaming indices in the first factor and setting all time variables equal to s in the second factor. The time variable change is permissible because in (4) a delta function $\delta(t-s)$

occurs. The expression given by the last equality of (45) has an integrand which is positive since the integrand is the product of a function with its own complex conjugate, and the two-time variables result in a delta function when the stochastic average is performed. These properties correspond to the positive definiteness of $Q_{\beta\beta'\alpha'\alpha}$ in (44). The final result is that $W_{\theta\mu}(t) \leq 0$, which proves (37).

Having now proved (36) and (37), it is possible to complete consideration of (35). Using (36) immediately gives

$$\frac{d}{dt}H(t) = -\tfrac{1}{2}\sum_{\theta}\sum_{\mu} W_{\theta\mu}(t)[D_{\mu}(t) - D_{\theta}(t)] \log_e \frac{D_{\theta}(t)}{D_{\mu}(t)}. \quad (46)$$

Noting (37) and the inequality

$$(X - Y)\log_e(Y/X) \leq 0 \quad \text{for all} \quad X \text{ and } Y \quad (47)$$

gives

$$\frac{d}{dt}H(t) \leq 0. \quad (48)$$

This completes the proof of an "H theorem" for multiplicative stochastic processes.

CONCLUDING REMARKS

The approach to the mathematical formulation of nonequilibrium phenomena based upon multiplicative stochastic processes has resulted, in this paper, in a derivation of an "H theorem." Once the assumption of a fluctuating contribution to the Hamiltonian of a system has been made, all consequences quoted or derived in this paper follow rigorously from that single assumption. How far one can go toward a rigorously established, comprehensive theory for nonequilibrium phenomena remains to be investigated.

ACKNOWLEDGMENT

The author is very grateful to Joel Keizer for many provocative conversations related to this work.

[1] R. F. Fox, J. Math. Phys. **8**, 1196 (1972).
[2] L. Landau and E. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958), Chap. 1.
[3] G. Uhlenbeck and G. Ford, *Lectures in Statistical Mechanics* (Amer. Math. Soc., Providence, R.I., 1963), Chaps. 1 and 4.

Algebraic Equations for Bethe-Salpeter and Coulomb Green's Functions

E. Kyriakopoulos
Nuclear Research Center "Democritos," Aghia Paraskevi, Attikis, Athens, Greece
(Received 7 February 1972)

The equation for the Bethe-Salpeter Green's function in the case of two scalar quarks interacting via the exchange of a scalar particle of zero mass is transformed to an algebraic equation with the help of the dynamical group $SO(5,2)$. From this equation a one-parameter integral representation of the Green's function is obtained in the case of maximal binding, and from this representation the Green's function is calculated in terms of a hypergeometric function. The equation for the f-dimensional nonrelativistic Coulomb Green's function is also transformed to an algebraic equation.

1. INTRODUCTION

It is well known that the symmetry group of the Bethe-Salpeter (BS) equation[1-4] for two scalar quarks interacting via the exchange of a scalar particle of zero mass to form a bound state with vacuumlike 4-momentum is the group $SO(5)$.[5] The eigenfunctions of the problem, which are the five-dimensional spherical harmonics, can be accommodated into a single irreducible representation of the group $SO(5,2)$, which is the noninvariance or dynamical group of the above BS equation.[6] If the energy-momentum 4-vector of the bound state is timelike, spacelike, or lightlike, the representation $SO(5,2)$ splits into a direct sum of $SO(4,2)$ representations.[6,7] These group theoretical aspects of the above BS equation were used to transform this equation to an infinite component wave equation. It seems not to have been recognized, however, that taking into account these group theoretical properties one can construct explicitly the Green's function of the problem.

In Sec. 2 the BS equation for the Green's function in the general case is transformed to an algebraic equation, i.e., an equation involving tensors and generators of the group $SO(5,2)$. If we drop the δ-function term we obtain the infinite component wave equation for the BS wavefunction, generalizing in this way the corresponding result of Ref. 6. In the case of maximal binding, the equation for the Green's function is solved, and a one-parameter integral representation is obtained. The integration is performed, and the Green's function is expressed in terms of a hypergeometric function.

It has been found that the group $SO(f+1)$ is the symmetry group[8,9] of the f-dimensional analog of the nonrelativistic H atom for $f = 2, 3, \ldots$, and that the group $SO(f+1,2)$ is the dynamical group of this problem.[6] Schwinger[10] used the $SO(4)$ symmetry of the H atom to obtain a one-parameter integral representation of the Green's function. The f-dimensional Coulomb Green's function in momentum-space has been calculated recently by Hostler.[11] In Sec. 3 the equation for the Green's function of the f-dimensional H atom is transformed to a simple algebraic equation. The Coulomb problem is treated in Appendix B.

2. BETHE-SALPETER GREEN'S FUNCTION

In a previous paper[6] the BS equation of two scalar quarks of equal mass, interacting via the exchange of a scalar particle of zero mass, was transformed to an algebraic equation with the help of the dynamical group $SO(5,2)$. We shall derive in this section the algebraic equation in the general case of quarks with unequal masses, arbitrary energy-momentum 4-vector of the bound state, and also the algebraic equation for the Green's function of this model.

occurs. The expression given by the last equality of (45) has an integrand which is positive since the integrand is the product of a function with its own complex conjugate, and the two-time variables result in a delta function when the stochastic average is performed. These properties correspond to the positive definiteness of $Q_{\beta\beta'\alpha'\alpha}$ in (44). The final result is that $W_{\theta\mu}(t) \leq 0$, which proves (37).

Having now proved (36) and (37), it is possible to complete consideration of (35). Using (36) immediately gives

$$\frac{d}{dt}H(t) = -\tfrac{1}{2}\sum_{\theta}\sum_{\mu} W_{\theta\mu}(t)[D_{\mu}(t) - D_{\theta}(t)] \log_e \frac{D_{\theta}(t)}{D_{\mu}(t)}. \quad (46)$$

Noting (37) and the inequality

$$(X - Y)\log_e(Y/X) \leq 0 \quad \text{for all} \quad X \text{ and } Y \quad (47)$$

gives

$$\frac{d}{dt}H(t) \leq 0. \quad (48)$$

This completes the proof of an "H theorem" for multiplicative stochastic processes.

CONCLUDING REMARKS

The approach to the mathematical formulation of nonequilibrium phenomena based upon multiplicative stochastic processes has resulted, in this paper, in a derivation of an "H theorem." Once the assumption of a fluctuating contribution to the Hamiltonian of a system has been made, all consequences quoted or derived in this paper follow rigorously from that single assumption. How far one can go toward a rigorously established, comprehensive theory for nonequilibrium phenomena remains to be investigated.

ACKNOWLEDGMENT

The author is very grateful to Joel Keizer for many provocative conversations related to this work.

[1] R. F. Fox, J. Math. Phys. **8**, 1196 (1972).
[2] L. Landau and E. Lifshitz, *Statistical Physics* (Pergamon, New York, 1958), Chap. 1.
[3] G. Uhlenbeck and G. Ford, *Lectures in Statistical Mechanics* (Amer. Math. Soc., Providence, R.I., 1963), Chaps. 1 and 4.

Algebraic Equations for Bethe-Salpeter and Coulomb Green's Functions

E. Kyriakopoulos
Nuclear Research Center "Democritos," Aghia Paraskevi, Attikis, Athens, Greece
(Received 7 February 1972)

The equation for the Bethe-Salpeter Green's function in the case of two scalar quarks interacting via the exchange of a scalar particle of zero mass is transformed to an algebraic equation with the help of the dynamical group $SO(5,2)$. From this equation a one-parameter integral representation of the Green's function is obtained in the case of maximal binding, and from this representation the Green's function is calculated in terms of a hypergeometric function. The equation for the f-dimensional nonrelativistic Coulomb Green's function is also transformed to an algebraic equation.

1. INTRODUCTION

It is well known that the symmetry group of the Bethe-Salpeter (BS) equation[1-4] for two scalar quarks interacting via the exchange of a scalar particle of zero mass to form a bound state with vacuumlike 4-momentum is the group $SO(5)$.[5] The eigenfunctions of the problem, which are the five-dimensional spherical harmonics, can be accommodated into a single irreducible representation of the group $SO(5,2)$, which is the noninvariance or dynamical group of the above BS equation.[6] If the energy-momentum 4-vector of the bound state is timelike, spacelike, or lightlike, the representation $SO(5,2)$ splits into a direct sum of $SO(4,2)$ representations.[6,7] These group theoretical aspects of the above BS equation were used to transform this equation to an infinite component wave equation. It seems not to have been recognized, however, that taking into account these group theoretical properties one can construct explicitly the Green's function of the problem.

In Sec. 2 the BS equation for the Green's function in the general case is transformed to an algebraic equation, i.e., an equation involving tensors and generators of the group $SO(5,2)$. If we drop the δ-function term we obtain the infinite component wave equation for the BS wavefunction, generalizing in this way the corresponding result of Ref. 6. In the case of maximal binding, the equation for the Green's function is solved, and a one-parameter integral representation is obtained. The integration is performed, and the Green's function is expressed in terms of a hypergeometric function.

It has been found that the group $SO(f+1)$ is the symmetry group[8,9] of the f-dimensional analog of the nonrelativistic H atom for $f = 2, 3, \ldots$, and that the group $SO(f+1,2)$ is the dynamical group of this problem.[6] Schwinger[10] used the $SO(4)$ symmetry of the H atom to obtain a one-parameter integral representation of the Green's function. The f-dimensional Coulomb Green's function in momentum-space has been calculated recently by Hostler.[11] In Sec. 3 the equation for the Green's function of the f-dimensional H atom is transformed to a simple algebraic equation. The Coulomb problem is treated in Appendix B.

2. BETHE-SALPETER GREEN'S FUNCTION

In a previous paper[6] the BS equation of two scalar quarks of equal mass, interacting via the exchange of a scalar particle of zero mass, was transformed to an algebraic equation with the help of the dynamical group $SO(5,2)$. We shall derive in this section the algebraic equation in the general case of quarks with unequal masses, arbitrary energy-momentum 4-vector of the bound state, and also the algebraic equation for the Green's function of this model.

Let m_j, p_j, $j = 1, 2$, be the masses and the 4-momenta of the quarks 1 and 2. We write

$$m_1 = m(1 + \Delta), \quad m_2 = m(1 - \Delta), \tag{2.1}$$

and we introduce the total momentum q of the bound state and the relative momentum p of the quarks by

$$p_1 = p + \tfrac{1}{2}(1 + \Delta)q, \quad p_2 = -p + \tfrac{1}{2}(1 - \Delta)q. \tag{2.2}$$

Then the equation for the Green's function is the following:

$$[(p + \tfrac{1}{2}(1 + \Delta)q)^2 + (1 + \Delta)^2 m^2] \times [(p - \tfrac{1}{2}(1 - \Delta)q)^2 + (1 - \Delta)^2 m^2] G(p, p') - \frac{\lambda}{\pi^2} \int \frac{d^4 p''}{(p - p'')^2} G(p'', p') = \delta^4(p - p'). \tag{2.3}$$

In the above equation the Wick rotation[12] has been performed.

Equation (2.3) can be easily transformed to an infinite component equation by a straightforward application of the method of Ref. 6. We shall present in this section an easier method of performing this transformation. We make a stereographic projection of the four-dimensional space into a sphere in five-dimensional space by introducing the variables η_i:

$$\eta_\alpha = 2p_5 p_\alpha / p_i^2, \quad \alpha = 1, \ldots, 4,$$
$$\eta_5 = (p_5^2 - p_\alpha^2)/p_i^2, \quad i = 1, \ldots, 5, \tag{2.4}$$

where

$$p_5 \equiv [(m^2 + \tfrac{1}{4} q^2)(1 - \Delta^2)]^{1/2}. \tag{2.5}$$

An immediate calculation shows that

$$d^4 p = [(p_\alpha^2 + p_5^2)/2p_5]^4 \, d^5\Omega = [p_5/(1 + \eta_5)]^4 \, d^5\Omega, \tag{2.6}$$

where $d^5\Omega$ is the element of area of a five-dimensional sphere. Another form of the above relation is

$$\delta^4(p - p') = [(1 + \eta_5)/p_5]^4 \, \delta^5(\Omega - \Omega'). \tag{2.7}$$

Also we have

$$(p - p'')^2 = \{2p_5^2/[(1 + \eta_5)(1 + \eta_5'')]\}(1 - \eta_i \eta_i'') \tag{2.8}$$

and

$$\{[p + \tfrac{1}{2}(1 + \Delta)q]^2 + (1 + \Delta)^2 m^2\} \times \{[p - \tfrac{1}{2}(1 - \Delta)q]^2 + (1 - \Delta)^2 m^2\} = \{4p_5^2/[(1 - \Delta^2)(1 + \eta_5)^2]\}[p_5^2 - \tfrac{1}{4}(q_i' \eta_i)^2], \tag{2.9}$$

where the five-dimensional vector q_i' is given by

$$q_i' = \{(1 - \Delta^2) q_\alpha, 2\Delta p_5\}. \tag{2.10}$$

Using Eqs. (2.6)–(2.9), we can write Eq. (2.3) in the form

$$4(1 - \Delta^2)^{-1}[p_5^2 - \tfrac{1}{4}(q_i' \eta_i)^2] \Gamma'(\Omega, \Omega') - \frac{\lambda}{2\pi^2} \int d^5\Omega'' \frac{\Gamma'(\Omega'', \Omega')}{1 - \eta_i \eta_i''} = \delta^5(\Omega - \Omega'), \tag{2.11}$$

where

$$\Gamma'(\Omega, \Omega') = \left(\frac{p_5}{1 + \eta_5}\right)^3 G(p, p') \left(\frac{p_5}{1 + \eta_5'}\right)^3. \tag{2.12}$$

It follows from the Introduction that in our treatment of the BS equation we shall make use of the representation of the group $SO(5, 2)$ whose basic functions are the five-dimensional spherical harmonics. We introduce a spherical polar coordinate system in the five-dimensional sphere as follows:

$$\eta_4 = \cos\chi, \quad \eta_5 = \sin\chi \cos\psi, \quad \eta_3 = \sin\chi \sin\psi \cos\theta,$$
$$\eta_2 = \sin\chi \sin\psi \sin\theta \sin\vartheta, \quad \eta_1 = \sin\chi \sin\psi \sin\theta \cos\vartheta. \tag{2.13}$$

We shall take the following representation of the five-dimensional spherical harmonics

$$Y_{N,n,l,m}(\chi, \psi, \theta, \vartheta) = (2\pi)^{-1/2} (\sin\chi)^{n-1} \mathcal{C}_{N-n}^{n+1/2} \times (\cos\chi)(\sin\psi)^l \mathcal{C}_{n-l-1}^{l+1}(\cos\psi) \times (\sin\theta)^m \mathcal{C}_{l-m}^{m+(1/2)}(\cos\theta) e^{im\vartheta}, \tag{2.14}$$

where

$$N - 1 \geq n - 1 \geq l \geq |m|. \tag{2.15}$$

The $\mathcal{C}_n^\nu(x)$ are normalized Gegenbauer polynomials given by

$$\mathcal{C}_n^\nu(x) = \left(\frac{(\nu + n)\Gamma(n + 1)[\Gamma(\nu)]^2}{2^{1-2\nu}\pi\Gamma(n + 2\nu)}\right)^{1/2} C_n^\nu(x) \tag{2.16}$$

with $C_n^\nu(x)$ the usual Gegenbauer polynomials.[13] The above-defined spherical harmonics are orthonormal.

To transform Eq. (2.11) to algebraic form, we consider the relation

$$\frac{1}{1 - \eta_i \eta_i''} = 8\pi^2 \sum_{N,n,l,m} \frac{Y_{N,n,l,m}(\eta) Y_{N,n,l,m}^*(\eta'')}{N(N + 1)}, \tag{2.17}$$

which is a consequence of the Funck–Hecke theorem.[14] Let us assume that the function $\Gamma'(\Omega'', \Omega')$ has an expansion of the form

$$\Gamma'(\Omega'', \Omega') = \sum_{N',n',l',m'} h_{N',n',l',m'}(\eta') Y_{N',n',l',m'}(\eta''). \tag{2.18}$$

Using Eqs (2.17), (2.18), and the orthogonality properties of the spherical harmonics, we get

$$\int d^5\Omega'' \frac{\Gamma'(\Omega'', \Omega')}{1 - \eta_i' \eta_i''} = 8\pi^2 \sum_{N,n,l,m} \frac{h_{N,n,l,m}(\eta') Y_{N,n,l,m}(\eta)}{N(N + 1)}. \tag{2.19}$$

Let $L'_{\mu,\nu}$ $\mu, \nu = 1, 2, \ldots, 7$, be the generators of the $SO(5, 2)$ representation whose basic functions are the five-dimensional spherical harmonics. The diagonal generator L'_{67} satisfies the relation[15]

$$L'_{67} Y_{N,n,l,m} = (N + \tfrac{1}{2}) Y_{N,n,l,m}. \tag{2.20}$$

Therefore, we get from Eqs (2.19) and (2.20)

$$\int d^5\Omega'' \frac{\Gamma'(\Omega'', \Omega')}{1 - \eta_i \eta_i''} = 8\pi^2 (L'^2_{67} - \tfrac{1}{4})^{-1} \Gamma'(\Omega, \Omega'), \tag{2.21}$$

and Eq. (2.11) can be written in the form

$$4(1 - \Delta^2)^{-1}[p_5^2 - \tfrac{1}{4}(q_i' \eta_i)^2] \Gamma'(\Omega, \Omega') - 4\lambda(L'^2_{67} - \tfrac{1}{4})^{-1} \Gamma'(\Omega, \Omega') = \delta^5(\Omega - \Omega') \tag{2.22}$$

or

$$4(1 - \Delta^2)^{-1}(L'^2_{67} - \tfrac{1}{4})[p_5^2 - \tfrac{1}{4}(q_i' \eta_i)^2] \Gamma'(\Omega, \Omega') - 4\lambda\Gamma'(\Omega, \Omega') = (L'^2_{67} - \tfrac{1}{4}) \delta^5(\Omega - \Omega'). \tag{2.23}$$

Defining $\Gamma(\Omega,\Omega')$ by

$$\Gamma(\Omega,\Omega') = 4(1-\Delta^2)^{-1}[p_5^2 - \tfrac{1}{4}(q_i'\eta_i)^2]\,\Gamma'(\Omega,\Omega'), \quad (2.24)$$

Eq. (2.23) becomes

$$\left\{\left[1-\left(\frac{q_i'\eta_i}{2p_5}\right)^2\right](L_{67}'^2 - \tfrac{1}{4}) - g\right\}\Gamma(\Omega,\Omega') = \left[1-\left(\frac{q_i'\eta_i}{2p_5}\right)^2\right](L_{67}'^2 - \tfrac{1}{4})\,\delta^5(\Omega-\Omega'), \quad (2.25)$$

where

$$g = \lambda(1-\Delta^2)/p_5^2. \quad (2.26)$$

It has been shown in Ref. 6 that [see equation preceding Eq. (4.26) of Ref. 6]

$$\eta_i Y_{N,n,l,m} = (z_i/z)Y_{N,n,l,m} = L_{i7}'L_{67}'^{-1}Y_{N,n,l,m}. \quad (2.27)$$

Therefore, Eq. (2.25) becomes

$$\left\{\left[1-\left(\frac{q_i'L_{i7}'L_{67}'^{-1}}{2p_5}\right)^2\right](L_{67}'^2 - \tfrac{1}{4}) - g\right\}\Gamma(\Omega,\Omega') = \left[1-\left(\frac{q_i'L_{i7}'L_{67}'^{-1}}{2p_5}\right)^2\right](L_{67}'^2 - \tfrac{1}{4})\,\delta^5(\Omega-\Omega') \quad (2.28)$$

The above expression is the infinite component equation of the BS Green's function in the case of two scalar quarks of unequal masses interacting through the exchange of a scalar boson of zero mass to form a bound state with an arbitrary total momentum 4-vector (timelike, spacelike, lightlike, or vacuumlike), in an arbitrary Lorentz frame. If we omit the term which contains the δ function, we get the infinite component wave equation of the BS equation in the above specified case. If the quarks have equal masses, the total momentum 4-vector is timelike, and, if we go in the Lorentz frame in which the bound state is at rest, we find the expression of Ref. 6. We can get rid of the inverse generator $L_{67}'^{-1}$ by using the relation

$$(L_{67}'^2 - 1)L_{i7}'L_{67}'^{-1} = 2L_{67}'L_{i7}' - L_{i7}'L_{67}'. \quad (2.29)$$

We shall calculate the function $\Gamma(\Omega,\Omega')$ for $q_\alpha = \Delta = 0$. We get

$$\Gamma(\Omega,\Omega') = [L_{67}'^2 - \tfrac{1}{4} - g]^{-1}(L_{67}'^2 - \tfrac{1}{4})\delta^5(\Omega-\Omega'). \quad (2.30)$$

From Eqs (2.20), (2.30), and the expression

$$\delta^5(\Omega-\Omega') = \sum_{N,n,l,m} Y_{N,n,l,m}(\xi)Y^*_{N,n,l,m}(\xi'), \quad (2.31)$$

where ξ and ξ' are two unit vectors in the five-dimensional Euclidean space specified by the set of angles Ω and Ω', respectively, we get

$$\Gamma(\Omega,\Omega') = \sum_{N,n,l,m} \frac{N(N+1)}{N(N+1)-g} \times Y_{N,n,l,m}(\xi)\; Y^*_{N,n,l,m}(\xi'). \quad (2.32)$$

The above expression has poles at

$$\lambda = N(N+1)m^2, \quad (2.33)$$

i.e., at the physical eigenvalue spectrum as expected.

We want to find an integral representation for the function $\Gamma(\Omega,\Omega')$ of Eq. (2.32). We have[16]

$$\sum_{n,l,m} Y_{N,n,l,m}(\xi)Y^*_{N,n,l,m}(\xi') = [(2N+1)/8\pi^2]\,C_{N-1}^{3/2}(\xi_i\xi_i'). \quad (2.34)$$

Then Eq. (2.32) becomes

$$\Gamma(\Omega,\Omega') = \sum_{N=1}^{\infty} \frac{N(N+1)}{N(N+1)-g}\,\frac{2N+1}{8\pi^2}\,C_{N-1}^{3/2}(\xi_i\xi_i') = \delta(\Omega-\Omega') + \frac{g}{8\pi^2}\sum_{N=1}^{\infty}\left(\frac{1}{N+\frac{1}{2}+a} + \frac{1}{N+\frac{1}{2}-a}\right) \times C_{N-1}^{3/2}(\xi_i\xi_i'), \quad (2.35)$$

where

$$a \equiv (g+\tfrac{1}{4})^{1/2}. \quad (2.36)$$

Equations (2.35) give the expansion of the BS Green's function in a series of Gegenbauer polynomials. We have[13]

$$\frac{1}{[(1-\rho)^2 + \rho(\xi-\xi')^2]^{3/2}} = \sum_{N=1}^{\infty}\rho^{N-1}\,C_{N-1}^{3/2}(\xi_i\xi_i'). \quad (2.37)$$

Therefore, the second term on the right-hand side of Eq. (2.35) can be written in integral form. We find

$$\Gamma(\Omega,\Omega') = \delta(\Omega-\Omega') + \frac{g}{8\pi^2}\int_0^1 d\rho\,\frac{\rho^{(1/2)+a} + \rho^{(1/2)-a}}{[(1-\rho)^2 + \rho(\xi-\xi')^2]^{3/2}} \quad (2.38)$$

with the restriction (for a real and positive)

$$a < \tfrac{3}{2}. \quad (2.39)$$

This restriction comes from the fact that after expansion of the denominator of the integrand of Eq. (2.38) and integration term by term, all exponents of ρ must be positive in order to give a finite result at $\rho = 0$.

The function $\Gamma(\Omega,\Omega')$ of Eq. (2.38) can be expressed in terms of a hypergeometric function. Introducing the variable ρ' by $\rho' = 1/\rho$ we get

$$\int_0^1 d\rho\,\frac{\rho^{(1/2)+a}}{[(1-\rho)^2 + \rho(\xi-\xi')^2]^{3/2}} = \int_1^\infty d\rho'\,\frac{\rho'^{(1/2)-a}}{[(1-\rho')^2 + \rho'(\xi-\xi')^2]^{3/2}}. \quad (2.40)$$

Therefore, we get[17]

$$\int_0^1 d\rho\,\frac{\rho^{(1/2)+a} + \rho^{(1/2)-a}}{[(1-\rho)^2 + \rho(\xi-\xi')^2]^{3/2}} = \int_0^\infty d\rho\,\rho^{(1/2)-a}(\rho-y_+)^{-3/2}(\rho-y_-)^{-3/2} = (-y_+)^{-3/2}\,(-y_-)^{-a}\,B(\tfrac{3}{2}-a,\tfrac{3}{2}+a) \times {}_2F_1(\tfrac{3}{2},\tfrac{3}{2}-a;3;1-(y_-/y_+)), \quad \tfrac{3}{2} > \mathrm{Re}a > -\tfrac{3}{2}, \quad (2.41)$$

where $y_\pm \equiv \xi_i\xi_i' \pm [(\xi_i\xi_i')^2 - 1]^{1/2}$ and B is a beta function. For

$$Z = [(\xi_i\xi_i')^2 - 1]/(\xi_i\xi_i')^2, \quad (2.42)$$

we get[18]

$${}_2F_1(\tfrac{3}{2}-a,\tfrac{3}{2};3;1-(y_-/y_+)) = {}_2F_1(\tfrac{3}{2}-a,\tfrac{3}{2};3;2Z^{1/2}/(1+Z)^{1/2}) = (1+Z^{1/2})^{(3/2)-a}\,{}_2F_1(\tfrac{3}{4}-\tfrac{1}{2}a,\tfrac{5}{4}-\tfrac{1}{2}a;2;Z). \quad (2.43)$$

The above hypergeometric series can be further simplified. Using Eq. (2.42), we get[19]

$$ {}_2F_1(\tfrac{3}{4}-\tfrac{1}{2}a,\tfrac{5}{4}-\tfrac{1}{2};2;[(\xi_i\xi_i')^2-1]/(\xi_i\xi_i')^2) = (-\xi_i\xi_i')^{(3/2)-a}\,{}_2F_1(\tfrac{3}{2}-a,\tfrac{3}{2}+a;2;\tfrac{1}{2}(1+\xi_i\xi_i')). \tag{2.44} $$

From Eqs. (2.38) and (2.41)–(2.44), we get

$$ \Gamma(\Omega,\Omega') = \delta(\Omega-\Omega') + (g/8\pi^2)B(\tfrac{3}{2}-a,\tfrac{3}{2}+a) \times {}_2F_1(\tfrac{3}{2}-a,\tfrac{3}{2}+a;2;1-\tfrac{1}{4}(\xi-\xi')^2). \tag{2.45} $$

Another one-parameter integral representation of $\Gamma(\Omega,\Omega')$ is the following:

$$ \Gamma(\Omega,\Omega') = \frac{1}{8\pi^2}\int_0^1 d\rho(\rho^{(1/2)+a}+\rho^{(1/2)-a}) \times \frac{d^2}{d\rho^2}\frac{\rho^2}{[(1-\rho)^2+\rho(\xi-\xi')^2]^{3/2}} \tag{2.46} $$

with $a<\frac{3}{2}$. To prove that Eq. (2.38) follows from Eq. (2.46), we integrate Eq. (2.46) by parts twice and get

$$ \Gamma(\Omega,\Omega') = \lim_{\sigma\to 0}\ \frac{3}{4\pi^2}\,\frac{\sigma}{[\sigma^2+(\xi-\xi')^2]^{5/2}} + \frac{g}{8\pi^2}\int_0^1 d\rho \times \frac{\rho^{(1/2)+a}+\rho^{(1/2)-a}}{[(1-\rho)^2+\rho(\xi-\xi')^2]^{3/2}}. \tag{2.47} $$

It is shown in Appendix A that

$$ \lim_{\sigma\to 0}\ \frac{\Gamma[\frac{1}{2}(f+1)]}{\pi^{(f+1)/2}}\,\frac{\sigma}{[\sigma^2+(\zeta-\zeta')^2]^{(f+1)/2}} = \delta^{f+1}(\Omega-\Omega'), \tag{2.48} $$

where ζ and ζ' are two unit vectors in an $(f+1)$-dimensional space, specified by the set of angles Ω and Ω', respectively. Using Eq. (2.48) for $f=4$ we easily find that the expressions (2.38) and (2.46) are identical.

The integral representations for the Green's function $G(p,p')$ can be easily found from Eqs (2.4), (2.7), (2.8), (2.12), (2.24), (2.38), and (2.46). We get

$$ G(p,p') = \frac{1}{p_i^2p_j'^2}\,\delta^4(p-p') + \frac{2m^4g}{\pi^2}\int_0^1 d\rho \times \frac{\rho^{(1/2)+a}+\rho^{(1/2)-a}}{[p_i^4p_j'^4(1-\rho)^2+4m^2p_i^2p_j'^2\rho(p-p')^2]^{3/2}} \tag{2.49} $$

and

$$ G(p,p') = \frac{2m^4}{\pi^2}\int_0^1 d\rho(\rho^{(1/2)+a}+\rho^{(1/2)-a})\,\frac{d^2}{d\rho^2} \times \frac{\rho^2}{[p_i^4p_j'^4(1-\rho)^2+4m^2p_i^2p_j'^2\rho(p-p')^2]^{3/2}}. \tag{2.50} $$

Also Eq. (2.45) gives

$$ G(p,p') = \frac{1}{p_i^2p_j'^2}\,\delta^4(p-p') + \frac{2m^4g}{\pi^2}\,B(\tfrac{3}{2}-a,\tfrac{3}{2}+a) \times {}_2F_1\left(\tfrac{3}{2}-a,\tfrac{3}{2}+a;2;1-\frac{m^4}{p_i^2p_j'^2}(p-p')^2\right). \tag{2.51} $$

3. ALGEBRAIC EQUATION FOR THE COULOMB GREEN'S FUNCTION

Making use of the $O(f+1)$ symmetry[9] of the f-dimensional analog of the Schrödinger equation for the nonrelativistic H atom, one can calculate the Green's function of this problem for $f=2,3,\cdots$. The case $f=3$ has been investigated by Schwinger.[10]

The f-dimensional equation for the Green's function $G_C(\mathbf{p},\mathbf{p}')$ in momentum-space is

$$ (\mathbf{p}^2+p_0^2)G_C(\mathbf{p},\mathbf{p}')\,\frac{Ze^2\mu\Gamma[\frac{1}{2}(f-1)]}{\pi^{(f+1)/2}\hbar}\int\frac{d^f\mathbf{p}''G_C(\mathbf{p}'',\mathbf{p}')}{(\mathbf{p}-\mathbf{p}'')^{f-1}} = -2\mu\delta^f(\mathbf{p}-\mathbf{p}'), \tag{3.1} $$

where $p_0^2=-2\mu E$. Consider the new coordinate system

$$ \zeta_\beta = \frac{2p_0p_\beta}{\mathbf{p}^2+p_0^2},\quad \zeta_{f+1} = \frac{\mathbf{p}^2-p_0^2}{\mathbf{p}^2+p_0^2},\quad \beta=1,2,\ldots,f. \tag{3.2} $$

We get then[20]

$$ d^f\mathbf{p}'' = [(\mathbf{p}''^2+p_0^2)/2p_0]^f d^{f+1}\Omega, \tag{3.3} $$

$$ (\mathbf{p}-\mathbf{p}'')^{f-1} = \left(\frac{(\mathbf{p}^2+p_0^2)(\mathbf{p}''^2+p_0^2)}{(2p_0)^2}\right)^{(f-1)/2}(\zeta-\zeta'')^{f-1}, \tag{3.4} $$

and Eq. (3.1) becomes

$$ (\mathbf{p}^2+p_0^2)^{(f+1)/2}G_C(\mathbf{p},\mathbf{p}') - \frac{Ze^2\mu\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}\hbar p_0} \times\int d^{f+1}\Omega''\,\frac{(\mathbf{p}''^2+p_0^2)^{(f+1)/2}G_C(\mathbf{p}''\mathbf{p}')}{(\zeta-\zeta'')^{f-1}} = -2\mu(\mathbf{p}^2+p_0^2)^{(f-1)/2}\,\delta^f(\mathbf{p}-\mathbf{p}'). \tag{3.5} $$

Via the relation

$$ \delta^f(\mathbf{p}-\mathbf{p}') = \left(\frac{2p_0}{\mathbf{p}'^2+p_0^2}\right)^f\delta^{f+1}(\Omega-\Omega') \tag{3.6} $$

and the definitions

$$ H_f(\Omega,\Omega') = -\frac{1}{2\mu(2p_0)^f}\,(\mathbf{p}^2+p_0^2)^{(f+1)/2} \times G_C(\mathbf{p},\mathbf{p}')(\mathbf{p}'^2+p_0^2)^{(f+1)/2}, \tag{3.7} $$

$$ \kappa_f = Ze^2\mu/p_0\hbar, \tag{3.8} $$

Eq. (4.5) becomes

$$ H_f(\Omega,\Omega') - \frac{\kappa_f\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}}\int d^{f+1}\Omega''\,\frac{H_f(\Omega'',\Omega')}{(\zeta-\zeta'')^{f-1}} = \delta^{f+1}(\Omega-\Omega'). \tag{3.9} $$

The above equation can be easily transformed to an algebraic equation by the method which was used in the case of BS equation. We have[13,16] ($j=1,2,\ldots,f+1$)

$$ \frac{1}{(1-2\rho\zeta_j\zeta_j''+\rho^2)^{(f-1)/2}} = \sum_{N=1}^{\infty}\rho^{N-1}\,C_{N-1}^{(f-1)/2}(\zeta_j\zeta_j'') = \frac{4\pi^{(f+1)/2}}{\Gamma[\frac{1}{2}(f-1)]}\sum_{N,n,\ldots}\rho^{N-1} \times\frac{1}{2N+f-3}\,Y_{N,n\ldots}(\zeta)Y^*_{N,n,\ldots}(\zeta''). \tag{3.10} $$

For $\rho=1$ the above equation gives

$$ \frac{1}{(\zeta-\zeta'')^{f-1}} = \frac{4\pi^{f+1/2}}{\Gamma[\frac{1}{2}(f-1)]} $$

$$\times \sum_{N,n,\cdots} \frac{1}{2N+f-3}\, Y_{N,n,\cdots}(\zeta) Y^*_{N,n,\cdots}(\zeta''). \tag{3.11}$$

If $H_f(\Omega'', \Omega')$ has an expansion of the form

$$H_f(\Omega'', \Omega') = \sum_{N,n,\cdots} h_{N,n,\cdots}(\zeta') Y_{N,n,\cdots}(\zeta''), \tag{3.12}$$

we get, using Eq. (3.11),

$$\int d^{f+1}\Omega'' \frac{H_f(\Omega'', \Omega')}{(\zeta - \zeta'')^{f-1}} = \frac{4\pi^{(f+1)/2}}{\Gamma[\frac{1}{2}(f-1)]} \times \sum_{N,n,\cdots} \frac{1}{2N+f-3} h_{N,n,\cdots}(\zeta') Y_{N,n,\cdots}(\zeta) = \frac{2\pi^{(f+1)/2}}{\Gamma[\frac{1}{2}(f-1)]} L'^{-1}_{f+2,f+3} H_f(\Omega, \Omega'), \tag{3.13}$$

where the diagonal generator $L'_{f+2,f+3}$ of the group $SO(f+1,2)$ satisfies the relation[6]

$$L'_{f+2,f+3} Y_{N,n,\cdots} = [N + \tfrac{1}{2}(f-3)] Y_{N,n,\cdots}. \tag{3.14}$$

Therefore, the algebraic form of Eq. (3.9) is

$$H_f(\Omega, \Omega') - \kappa_f L'^{-1}_{f+2,f+3} H_f(\Omega, \Omega') = \delta^{f+1}(\Omega - \Omega'). \tag{3.15}$$

The above equation can be easily solved. We get

$$H_f(\Omega, \Omega') = (L'_{f+2,f+3} - \kappa_f)^{-1} L'_{f+2,f+3} \delta^{f+1}(\Omega - \Omega') = \sum_{N,n,\cdots} \frac{2N+f-3}{2N+f-3-2\kappa_f} \times Y_{N,n,\cdots}(\zeta) Y^*_{N,n,\cdots}(\zeta'). \tag{3.16}$$

The Coulomb Green's function is examined in Appendix B.

ACKNOWLEDGMENT

I would like to thank Professor Abdus Salam and Professor P. Budini, as well as the International Atomic Energy Agency and UNESCO for hospitality at the International Center for Theoretical Physics, where this work was partially done.

APPENDIX A: PROOF OF EQ. (2.48)

We want to prove Eq. (2.48), which becomes if we use Eqs (3.4) and (3.6),

$$\lim_{\omega\to 0} \frac{\Gamma[\frac{1}{2}(f+1)]}{\pi^{(f+1)/2}} \frac{\omega}{[\omega^2 + (\mathbf{p}-\mathbf{p}')^2]^{(f+1)/2}} = \delta^f(\mathbf{p}-\mathbf{p}'), \tag{A1}$$

where

$$\omega = \{[(\mathbf{p}^2 + p_0^2)(\mathbf{p}'^2 + p_0^2)]/(2p_0)^2\}^{1/2}\sigma. \tag{A2}$$

The f-dimensional Dirac δ-function is represented by ($p = |\mathbf{p}|, r = |\mathbf{x}|$)

$$\delta^f(\mathbf{p}) = \frac{1}{(2\pi)^f} \int d^f x e^{i\mathbf{p}\mathbf{x}} = \frac{1}{(2\pi)^f} \int_0^\infty dr\, r^{f-1} \times \int_0^\pi d\theta_1 (\sin\theta_1)^{f-2} e^{ipr\cos\theta_1} \int d^{f-1}\Omega. \tag{A3}$$

Since[21]

$$\int d^{f-1}\Omega = \frac{2\pi^{(f-1)/2}}{\Gamma[\frac{1}{2}(f-1)]} \tag{A4}$$

and[22]

$$\int_0^\pi d\theta_1 (\sin\theta_1)^{2\nu} e^{ipr\cos\theta_1} = \pi^{1/2}\left(\frac{2}{pr}\right)^\nu \Gamma(\nu+\tfrac{1}{2}) J_\nu(pr), \quad \mathrm{Re}\,\nu > -\tfrac{1}{2}, \tag{A5}$$

where J_ν is a Bessel function, we get for $f = 2\mu$

$$\delta^f(\mathbf{p}) = \frac{1}{(2\pi)^\mu p^{\mu-1}} \int_0^\infty dr\, r^\mu J_{\mu-1}(pr). \tag{A6}$$

We write[23] ($\omega > 0$)

$$\delta^f(\mathbf{p}) = \lim_{\omega\to 0} \frac{1}{(2\pi)^\mu p^{\mu-1}} \int_0^\infty dr\, r^\mu J_{\mu-1}(pr) e^{-\omega r} = \lim_{\omega\to 0} \frac{\Gamma(2\mu)}{(2\pi)^\mu} \frac{(\omega^2+p^2)^{-(\mu+1)/2}}{p^{\mu-1}} \times P_\mu^{-\mu+1}\left(\frac{\omega}{(\omega^2+p^2)^{1/2}}\right), \tag{A7}$$

where $P_\mu^{-\mu+1}$ is a Legendre function of the first kind. The relation[24]

$$(\nu' - \mu' + 1) P^{\mu'}_{\nu'+1}(Z) = (2\nu'+1) Z P^{\mu'}_{\nu'}(Z) - (\nu' + \mu') P^{\mu'}_{\nu'-1}(Z) \tag{A8}$$

gives for $\mu' = -\mu + 1$ and $\nu' = \mu - 1$

$$P_\mu^{-\mu+1}(Z) = Z P_{\mu-1}^{-\mu+1}(Z). \tag{A9}$$

For $Z = \cos\theta$, $-1 \le Z \le 1$, we get[24]

$$P_{\mu-1}^{-\mu+1}(\cos\theta) = [2^{-\mu+1}(\sin\theta)^{\mu-1}]/\Gamma(\mu). \tag{A10}$$

We have in our case

$$Z = \cos\theta = \omega/(\omega^2+p^2)^{1/2}. \tag{A11}$$

Equation (A7) becomes, if we use Eqs (A9), (A10) and (A11),

$$\delta^f(\mathbf{p}) = \lim_{\omega\to 0} \frac{\Gamma(2\mu)}{2^{2\mu-1}\pi^\mu \Gamma(\mu)} \frac{\omega}{(\omega^2+p^2)^{(2\mu+1)/2}} \tag{A12}$$

or since $f = 2\mu$

$$\delta^f(\mathbf{p}) = \lim_{\omega\to 0} \frac{\Gamma[\frac{1}{2}(f+1)]}{\pi^{(f+1)/2}} \frac{\omega}{(\omega^2+p^2)^{(f+1)/2}}, \tag{A13}$$

which is identical to Eq. (A1) if we make the substitution $p^2 \to (\mathbf{p}-\mathbf{p}')^2$. Therefore we have proven Eq. (A1).

APPENDIX B: f-DIMENSIONAL COULOMB GREEN'S FUNCTION

In this appendix we shall give integral representations of the f-dimensional Coulomb Green's function, an expression in terms of a generalized hypergeometric function of two variables, series expansions in terms of Gegenbauer polynomials, and an approximate expression for finite angle scattering. Some of these results have also been obtained by Hostler.[11] By a straightforward generalization of the approach of Schwinger[10] one gets from Eq. (3.16) the following integral representation of $H(\Omega, \Omega')$.

$$H_f(\Omega, \Omega') = \delta^{f+1}(\Omega - \Omega') + \frac{[\nu + \frac{1}{2}(f-3)]\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}} \times \left[\frac{1}{(\zeta-\zeta)^{f-1}} + \left(\nu + \frac{f-3}{2}\right)\right.$$

$$\times \int_0^1 d\rho \, \frac{\rho^{-\nu}}{[(1-\rho)^2 + \rho(\zeta - \zeta')^2]^{(f-1)/2}} \tag{B1}$$

valid for $\nu < 1$, where the parameter ν is defined by $\kappa_f = \nu + \frac{1}{2}(f-3)$. Integration by parts gives

$$H_f(\Omega, \Omega') = \delta^{f+1}(\Omega - \Omega') + \frac{[\nu + \frac{1}{2}(f-3)]\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}} \times \int_0^1 d\rho \rho^{-\nu-[(f-3)/2]} \times \frac{d}{d\rho} \frac{\rho^{(f-1)/2}}{[(1-\rho)^2 + \rho(\zeta - \zeta')^2]^{(f-1)/2}}. \tag{B2}$$

Integrating by parts again and using Eq. (2.48), we can write the above equation in the form

$$H_f(\Omega, \Omega') = \frac{\Gamma[\frac{1}{2}(f+1)]}{2\pi^{(f+1)/2}} \int_0^1 d\rho \, \rho^{-\nu-[(f-3)/2]} \times \frac{d}{d\rho} \frac{\rho^{(f-1)/2}(1-\rho^2)}{[(1-\rho)^2 + \rho(\zeta - \zeta')^2]^{(f+1)/2}}. \tag{B3}$$

The function $H_f(\Omega, \Omega')$ of Eq. (B1) can be expressed in terms of the generalized hypergeometric function in two variables[25] F_1. We have[26]

$$\int_0^1 d\rho \, \rho^{-\nu}[(1-\rho)^2 + \rho(\zeta - \zeta')^2]^{-(f-1)/2} = \int_0^1 d\rho \, \rho^{-\nu}(\rho - Z_+)^{-(f-1)/2}(\rho - Z_-)^{-(f-1)/2} = -\nu F_1(1-\nu, \tfrac{1}{2}(f-1), \tfrac{1}{2}(f-1), 2-\nu; Z_+, Z_-), \quad \mathrm{Re}\,\nu < 1, \tag{B4}$$

where

$$Z_\pm = \zeta_j \zeta_j' \pm [(\zeta_j \zeta_j')^2 - 1]^{1/2}. \tag{B5}$$

From Eqs. (B1) and (B4) we get

$$H_f(\Omega, \Omega') = \delta^{f+1}(\Omega, \Omega') + \frac{[\nu + \frac{1}{2}(f-3)]\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}} \times \left[\frac{1}{(\zeta - \zeta')^{f-1}} - \nu\left(\nu + \frac{f-3}{2}\right) \times F_1\left(1-\nu, \frac{f-1}{2}, \frac{f-1}{2}, 2-\nu; Z_+, Z_-\right)\right]. \tag{B6}$$

Expanding the denominators of the integrands of Eqs. (B1), (B2), and (B3) in series of Gegenbauer polynomials and integrating term by term, we get, respectively,

$$H_f(\Omega, \Omega') = \delta^{f+1}(\Omega - \Omega') + \frac{[\nu + \frac{1}{2}(f-3)]\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{(f+1)/2}} \times \left[\frac{1}{(\zeta - \zeta')^{f-1}} + \left(\nu + \frac{f-3}{2}\right) \times \sum_{N=1}^{\infty} \frac{C_{N-1}^{(f-1)/2}(\zeta_j \zeta_j')}{N - \nu}\right], \tag{B7}$$

$$H_f(\Omega, \Omega') = \delta^{f+1}(\Omega - \Omega') + \frac{[\nu + \frac{1}{2}(f-3)]\,\Gamma[\frac{1}{2}(f-1)]}{2\pi^{f+1/2}} \times \sum_{N=1}^{\infty} \frac{N + \frac{1}{2}(f-3)}{N - \nu} C_{N-1}^{(f-1)/2}(\zeta_j \zeta_j'), \tag{B8}$$

$$H_f(\Omega, \Omega') = \frac{\Gamma[\frac{1}{2}(f+1)]\,[\nu + \frac{1}{2}(f-3)]}{\pi^{f+1/2}} \times \sum_{N=1}^{\infty} \frac{1}{(N-\nu)(N-\nu+2)} C_{N-1}^{(f+1)/2}(\zeta_j \zeta_j'). \tag{B9}$$

Also Eq. (3.16) gives

$$H_f(\Omega, \Omega') = \frac{\Gamma[\frac{1}{2}(f-1)]}{2\pi^{f+1/2}} \sum_{N=1}^{\infty} \frac{[N + \frac{1}{2}(f-3)]^2}{N - \nu} \times C_{N-1}^{(f-1)/2}(\zeta_j \zeta_j'). \tag{B10}$$

Equations (B7) and (B8) can be easily derived from Eq. (B10).

The expressions (B1)–(B3), (B6), and (B7)–(B10) can be easily written in terms of the Green's function $G_C(\mathbf{p}, \mathbf{p}')$. For example, Eq. (B1) gives

$$G_C(\mathbf{p}, \mathbf{p}') = \frac{\delta^f(\mathbf{p} - \mathbf{p}')}{E - T} - C\left[1 + \left(\nu + \frac{f-3}{2}\right) \times \int_0^1 d\rho \, \frac{\rho^{-\nu}}{[\rho - d(1-\rho)^2]^{(f-1)/2}}\right], \tag{B11}$$

where

$$C = \frac{(2\nu + f - 3)\,\Gamma[\frac{1}{2}(f-1)]}{4\pi^{(f+1)/2}} \times \frac{p_0}{\mu(E-T)(E-T')(\mathbf{p} - \mathbf{p}')^{f-1}}, \tag{B12}$$

$$d = \frac{\mu(T-E)(T'-E)}{2E(\mathbf{p} - \mathbf{p}')^2}, \tag{B13}$$

$$T = \frac{\mathbf{p}^2}{2\mu}, \quad T' = \frac{\mathbf{p}'^2}{2\mu} \tag{B14}$$

For finite angle scattering we have

$$E - T \sim 0, \quad E - T' \sim 0, \quad (\mathbf{p} - \mathbf{p}')^2 > 0. \tag{B15}$$

In this case d is small and a simple approximate expression of the function $G_C(\mathbf{p}, \mathbf{p}')$ can be obtained from Eq. (B11). Since the main contribution to the integral comes from small ρ, we get

$$G_C(\mathbf{p}, \mathbf{p}') \sim -C\left[1 + \left(\nu + \frac{f-3}{2}\right) \int_0^1 d\rho \, \rho^{-\nu}(\rho - d)^{-(f-1)/2}\right] = -C\left[1 + \left(\nu + \frac{f-3}{2}\right)(-d)^{-\nu-(f-3)/2} \times \int_0^{1/d} dt \, t^{-\nu}(t+1)^{-(f-1)/2}\right]. \tag{B16}$$

For d small, the above expression gives

$$G_C(\mathbf{p}, \mathbf{p}') \sim -C\{1 + [\nu + \tfrac{1}{2}(f-3)](-d)^{-\nu-[(f-3)/2]} \times B(1-\nu, \tfrac{1}{2}(f-3) + \nu)\}, \tag{B17}$$

where $B(1-\nu, \frac{1}{2}(f-3) + \nu)$ is a beta function. For $f = 3$ the second term of the above expression reduces to the Green's function found by Schwinger.[10]

1 Y. Nambu, Progr. Theoret. Phys. (Kyoto) 5, 614 (1950).
2 E. E. Salpeter and H. A. Bethe, Phys. Rev. 84, 1232 (1951).
3 J. Schwinger, Proc. Natl. Acad. Sci. (U.S.) 37, 455 (1951).
4 M. Gell-Mann and F. Low, Phys. Rev. 84, 350 (1951).
5 R. E. Cutkosky, Phys. Rev. 96, 1135 (1954).
6 E. Kyriakopoulos, Phys. Rev. 174, 1846 (1968).
7 F. T. Hadjioannou and E. Kyriakopoulos, Intern. J. Theoret. Phys. 5, 199 (1972).

[8] V. Fock, Z. Physik **98**, 145 (1935).
[9] S. P. Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys. JETP **6**, 156 (1958)].
[10] J. Schwinger J. Math. Phys. **5**, 1606 (1964).
[11] L. Hostler, J. Math. Phys. **12**, 2311 (1971).
[12] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[13] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1954), Vol. 1, Sec. 3-15-1.
[14] Reference 13, Vol. 2, pp. 247 and 248.
[15] Reference 6, Eq. (2. 14b) for $n = 5$ and $t = N - 1$.
[16] Reference 13, Vol. 2, Sec. 11. 4, Theorem 4.
[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products* (Academic, New York, 1965), expression 3. 197. 1.
[18] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), expression 15-3-20.
[19] Reference 18, expression 15-3-29 for $z = \frac{1}{2}(1 + \xi_i \xi_i')$.
[20] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[21] Reference 13, Vol. 2, p. 234.
[22] Reference 13, Vol. 2, p. 81.
[23] Reference 17, expression 6. 621. 1.
[24] Reference 18, p. 334.
[25] Reference 13, Chap. V.
[26] Reference 13, p. 582.

Absence of Long-Range Order in Thin Films*

John C. Garrison, Jack Wong, and Harry L. Morrison†
Lawrence Livermore Laboratory, University of California, Livermore, California 94550
(Received 8 October 1971; Revised Manuscript Received 2 June 1972)

Thin films are described as idealized systems having finite extent in one direction but infinite extent in the other two. For systems of particles interacting through smooth potentials (e.g., no hard cores), it is shown that an equilibrium state for a homogeneous thin film is necessarily invariant under any continuous internal symmetry group generated by a conserved density. For short-range interactions it is also shown that equilibrium states are necessarily translation invariant. The absence of long-range order follows from its relation to broken symmetry. The only properties of the state required for the proof are local normality, spatial translation invariance, and the Kubo-Martin-Schwinger boundary condition. The argument employs the Bogoliubov inequality and the techniques of the algebraic approach to statistical mechanics. For inhomogeneous systems, the same argument shows that all order parameters defined by anomalous averages must vanish. Identical results can be obtained for systems with infinite extent in one direction only.

1. INTRODUCTION

In the past few years, a number of authors[1-6] have used the Bogoliubov inequalities to show that specific long-range order parameters must vanish for one- and two-dimensional systems. Since their conclusions apparently contradict the experimental observations of thin-film superconductivity[7] and superfluidity,[8] it is natural to be skeptical of these arguments. The purpose of this paper is to reinforce the previous conclusions by means of a more general proof, which uses the rigorous techniques and results of the algebraic approach to statistical mechanics.

The first result we need is the existence of a general relation between long-range order and broken symmetry. This relation allows us to avoid any questions about the validity of the Bogoliubov quasi-averaging method[9] used in previous proofs, and it also allows us to carry out the argument without making a specific choice of order parameter. With the assumption that the Hamiltonian can be written as the sum of a kinetic energy and an interaction term, we prove directly that a state describing a homogeneous thin film cannot exhibit any broken internal symmetry. This conclusion holds equally for long- and short-range potentials; therefore, the conventional descriptions of superconductivity and superfluidity in terms of broken gauge invariance will not work for thin films. We also show that a thin film state cannot exhibit broken translation invariance provided only that the potential satisfies a range condition previously found by Mermin.[3] Thus the usual descriptions of crystalline and magnetic ordering are not applicable to thin films. The assumption made above about the Hamiltonian excludes the class of singular potentials for which there is no clear separation of kinetic and potential energy; consequently, we cannot draw any conclusions for hard-core systems.

For inhomogeneous thin films, the argument can be adapted to prove that all order parameters defined by anomalous averages must vanish, but one cannot conclude that the state has no broken symmetry. Since no rigorous connection between long-range order (defined by ordinary averages) and the existence of anomalous averages has been established in the case, the absence of the latter does not preclude the existence of the former.

It will become obvious that our argument also serves to exclude long-range order for one-dimensional systems. Therefore, this case will not be explicitly considered.

In Sec. 2 we sketch the necessary theoretical background, and in Sec. 3 we present the argument that establishes the connection between long-range order and broken symmetry. A proof of the Bogoliubov inequality is outlined in Sec. 4; in Sec. 5, the inequality is applied to prove the impossibility of broken symmetries for homogeneous thin films. The case of inhomogeneous films is treated in Sec. 6 and is followed in Sec. 7 by a discussion of the results.

2. THEORETICAL BACKGROUND

The appropriate setting for our discussion is given by the algebraic formulation of statistical mechanics as described for example, by Haag, Hugenholtz, and Winnink[10] (HHW). We begin with the Fock space $\mathfrak{H}_F(\Gamma)$, where Γ is the configuration space relevant to the problem. One usually takes $\Gamma = \mathbf{R}^\nu$; but we will consider instead the space $\Gamma = \mathbf{R}^2 \times I$, where I is an interval of length L. In other words, we assume that the physical system involved lies between two infinitely extended parallel planes with separation L. For each bounded region $\Lambda \subset \Gamma$ there is a subspace $\mathfrak{H}_F(\Lambda)$ and a von Neumann algebra $\mathfrak{A}(\Lambda)$ consisting of bounded operators on $\mathfrak{H}_F(\Lambda)$; the C^*-algebra $\mathfrak{A}$ of quasilocal observables is then defined as the norm-closure of $\mathfrak{A}_L \equiv \cup_\Lambda \mathfrak{A}(\Lambda)$. A *state* ω is a positive, continuous, normalized linear functional on $\mathfrak{A}$. A state is said to be *locally normal* if its restriction to each $\mathfrak{A}(\Lambda)$ is given by a density matrix, and ω is a Gibbs state when the following limit exists for each $A \in \mathfrak{A}_L$:

$$\omega(A) = \lim_{n\to\infty} \omega_n(A), \quad \omega_n(A) = \mathcal{Q}_n^{-1} \operatorname{tr}(e^{-\beta(H_n - \mu N_n)} A),$$

[8] V. Fock, Z. Physik **98**, 145 (1935).
[9] S. P. Alliluev, Zh. Eksp. Teor. Fiz. **33**, 200 (1957) [Sov. Phys. JETP **6**, 156 (1958)].
[10] J. Schwinger J. Math. Phys. **5**, 1606 (1964).
[11] L. Hostler, J. Math. Phys. **12**, 2311 (1971).
[12] G. C. Wick, Phys. Rev. **96**, 1124 (1954).
[13] Bateman Manuscript Project, *Higher Transcendental Functions* (McGraw-Hill, New York, 1954), Vol. 1, Sec. 3-15-1.
[14] Reference 13, Vol. 2, pp. 247 and 248.
[15] Reference 6, Eq. (2. 14b) for $n = 5$ and $t = N - 1$.
[16] Reference 13, Vol. 2, Sec. 11. 4, Theorem 4.
[17] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals Series and Products* (Academic, New York, 1965), expression 3. 197. 1.
[18] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1965), expression 15-3-20.
[19] Reference 18, expression 15-3-29 for $z = \frac{1}{2}(1 + \xi_i \xi_i')$.
[20] M. Bander and C. Itzykson, Rev. Mod. Phys. **38**, 330 (1966).
[21] Reference 13, Vol. 2, p. 234.
[22] Reference 13, Vol. 2, p. 81.
[23] Reference 17, expression 6. 621. 1.
[24] Reference 18, p. 334.
[25] Reference 13, Chap. V.
[26] Reference 13, p. 582.

Absence of Long-Range Order in Thin Films*

John C. Garrison, Jack Wong, and Harry L. Morrison†
Lawrence Livermore Laboratory, University of California, Livermore, California 94550
(Received 8 October 1971; Revised Manuscript Received 2 June 1972)

Thin films are described as idealized systems having finite extent in one direction but infinite extent in the other two. For systems of particles interacting through smooth potentials (e.g., no hard cores), it is shown that an equilibrium state for a homogeneous thin film is necessarily invariant under any continuous internal symmetry group generated by a conserved density. For short-range interactions it is also shown that equilibrium states are necessarily translation invariant. The absence of long-range order follows from its relation to broken symmetry. The only properties of the state required for the proof are local normality, spatial translation invariance, and the Kubo-Martin-Schwinger boundary condition. The argument employs the Bogoliubov inequality and the techniques of the algebraic approach to statistical mechanics. For inhomogeneous systems, the same argument shows that all order parameters defined by anomalous averages must vanish. Identical results can be obtained for systems with infinite extent in one direction only.

1. INTRODUCTION

In the past few years, a number of authors[1-6] have used the Bogoliubov inequalities to show that specific long-range order parameters must vanish for one- and two-dimensional systems. Since their conclusions apparently contradict the experimental observations of thin-film superconductivity[7] and superfluidity,[8] it is natural to be skeptical of these arguments. The purpose of this paper is to reinforce the previous conclusions by means of a more general proof, which uses the rigorous techniques and results of the algebraic approach to statistical mechanics.

The first result we need is the existence of a general relation between long-range order and broken symmetry. This relation allows us to avoid any questions about the validity of the Bogoliubov quasi-averaging method[9] used in previous proofs, and it also allows us to carry out the argument without making a specific choice of order parameter. With the assumption that the Hamiltonian can be written as the sum of a kinetic energy and an interaction term, we prove directly that a state describing a homogeneous thin film cannot exhibit any broken internal symmetry. This conclusion holds equally for long- and short-range potentials; therefore, the conventional descriptions of superconductivity and superfluidity in terms of broken gauge invariance will not work for thin films. We also show that a thin film state cannot exhibit broken translation invariance provided only that the potential satisfies a range condition previously found by Mermin.[3] Thus the usual descriptions of crystalline and magnetic ordering are not applicable to thin films. The assumption made above about the Hamiltonian excludes the class of singular potentials for which there is no clear separation of kinetic and potential energy; consequently, we cannot draw any conclusions for hard-core systems.

For inhomogeneous thin films, the argument can be adapted to prove that all order parameters defined by anomalous averages must vanish, but one cannot conclude that the state has no broken symmetry. Since no rigorous connection between long-range order (defined by ordinary averages) and the existence of anomalous averages has been established in the case, the absence of the latter does not preclude the existence of the former.

It will become obvious that our argument also serves to exclude long-range order for one-dimensional systems. Therefore, this case will not be explicitly considered.

In Sec. 2 we sketch the necessary theoretical background, and in Sec. 3 we present the argument that establishes the connection between long-range order and broken symmetry. A proof of the Bogoliubov inequality is outlined in Sec. 4; in Sec. 5, the inequality is applied to prove the impossibility of broken symmetries for homogeneous thin films. The case of inhomogeneous films is treated in Sec. 6 and is followed in Sec. 7 by a discussion of the results.

2. THEORETICAL BACKGROUND

The appropriate setting for our discussion is given by the algebraic formulation of statistical mechanics as described for example, by Haag, Hugenholtz, and Winnink[10] (HHW). We begin with the Fock space $\mathfrak{H}_F(\Gamma)$, where Γ is the configuration space relevant to the problem. One usually takes $\Gamma = \mathbf{R}^\nu$; but we will consider instead the space $\Gamma = \mathbf{R}^2 \times I$, where I is an interval of length L. In other words, we assume that the physical system involved lies between two infinitely extended parallel planes with separation L. For each bounded region $\Lambda \subset \Gamma$ there is a subspace $\mathfrak{H}_F(\Lambda)$ and a von Neumann algebra $\mathfrak{A}(\Lambda)$ consisting of bounded operators on $\mathfrak{H}_F(\Lambda)$; the C^*-algebra $\mathfrak{A}$ of quasilocal observables is then defined as the norm-closure of $\mathfrak{A}_L \equiv \cup_\Lambda \mathfrak{A}(\Lambda)$. A *state* ω is a positive, continuous, normalized linear functional on $\mathfrak{A}$. A state is said to be *locally normal* if its restriction to each $\mathfrak{A}(\Lambda)$ is given by a density matrix, and ω is a Gibbs state when the following limit exists for each $A \in \mathfrak{A}_L$:

$$\omega(A) = \lim_{n\to\infty} \omega_n(A), \quad \omega_n(A) = \mathcal{Q}_n^{-1} \operatorname{tr}(e^{-\beta(H_n - \mu N_n)} A),$$

where H_n and N_n are respectively the Hamiltonian and number operator for the region Λ_n; the trace is taken over $\mathfrak{H}(\Lambda_n)$; $\mathcal{Q}_n$ is the normalization constant; and the regions Λ_n satisfy $\Lambda_{n+1} \supset \Lambda_n, \bigcup_n \Lambda_n = \Gamma$. Since $\mathfrak{A}(\Lambda)$ is a von Neumann algebra, any Gibbs state is locally normal.

The Gel'fand–Naimark–Segal[10] (GNS) construction for any state ω yields a Hilbert space $\mathfrak{H}$, a cyclic vector Ω, and a representation map $\pi: \mathfrak{A} \to \mathcal{L}(\mathfrak{H})$ (the bounded operators on $\mathfrak{H}$) with the property

$$\omega(A) = (\Omega, \pi(A)\Omega) \quad \forall A \in \mathfrak{A}.$$

For later use we need to add certain unbounded observables to $\mathfrak{A}$; this can be done as follows: We say that a closed, densely defined operator Q is affiliated[11] to $\mathfrak{A}(\Lambda)$ if the spectral projections of $(Q^\dagger Q)^{1/2}$ all belong to $\mathfrak{A}(\Lambda)$. Sewell[12] has shown that if ω is locally normal, the map π can be extended to the closed densely defined operators affiliated to $\mathfrak{A}(\Lambda)$ for some Λ. For each such operator Q, $\pi(Q)$ is a densely defined, closed operator on $\mathfrak{H}$. We will say that Q is ω-*affiliated* to $\mathfrak{A}$ if it is affiliated to some $\mathfrak{A}(\Lambda)$ and if $\Omega \in \mathrm{Dom}[\pi(Q)]$.

In the algebraic approach to statistical mechanics, the description of time translation is based on the sequence of maps $\{\alpha_n(t)\}$ defined by

$$\alpha_n(t)A = U_n(t)AU_n(t)^{-1}, \quad A \in \mathfrak{A}_L,$$
$$U_n(t) = \exp[i(H_n - \mu N_n)t].$$

The algebra $\mathfrak{A}(\Lambda_n)$ is to be constructed so that α_n is an automorphism on $\mathfrak{A}(\Lambda_n)$. In HHW it was assumed that for each $A \in \mathfrak{A}_L$ the sequence $\{\alpha_n(t)A\}$ is norm-convergent; this is sufficient to guarantee the existence of an automorphism $\alpha(t)$ on $\mathfrak{A}$ representing time translation. This limit has been shown to exist for certain lattice systems; but continuous systems present greater difficulties, as evidenced by the example of the ideal Bose gas for which it can be shown that $\alpha_n(t)$ cannot converge in the required sense. In view of these difficulties, it is important to note that our argument will not depend directly on the existence of $\alpha(t)$, but only on the following consequences of its existence:

(a) Time translations are realized by a group of automorphisms $\{\gamma(t)\}$ acting on the bicommutant $\pi(\mathfrak{A})''$; the group is implemented by a strongly continuous group of unitary operators $\{U(t)\}$ acting on $\mathfrak{H}$ and leaving the cyclic vector Ω invariant.

(b) The Kubo–Martin–Schwinger (KMS) boundary condition is satisfied in the form

$$\int dt\, f(t - i\beta)(\Omega, \pi(B)\gamma(t)\pi(A)\Omega) = \int dt\, f(t)(\Omega, \gamma(t)\pi(A)\pi(B)\Omega),$$

$\forall A, B \in \mathfrak{A}$ and $f \in \mathfrak{D}$ (the space of C^∞ functions with compact support in R).

The fact that our argument involves only (a) and (b) is of particular interest because of the alternative treatment of time translations proposed by Dubin and Sewell.[13] In their theory, the strong convergence condition on $\{\alpha_n(t)\}$ is replaced by a weaker condition on the behavior of time-dependent correlation functions for a Gibbs state; in this way, the problem of the ideal Bose gas is resolved, while the consequences (a) and (b) are retained. Since the necessary properties (a) and (b) follow either from the strong theory of time translations given in HHW or from the weaker assumptions of Dubin and Sewell, we will take (a) and (b) as the starting point for our discussion.

The other groups of physical symmetry operations (for example, space translations, gauge transformations, etc.) can always be represented by automorphisms on $\mathfrak{A}$, and they also have a local structure[14] that will be of interest later on. A one-parameter group $F = \{\alpha_\lambda\}$ of automorphisms is said to be *locally generated* if there is an operator-valued distribution $q(\mathbf{x})$ acting on $\mathfrak{H}_F(\Gamma)$ that satisfies

(1) For a real test function $f \in \mathfrak{D}(\Gamma)$, $q(f)$ is a self-adjoint operator ω-affiliated to $\mathfrak{A}$.

(2) For a suitable sequence $\{f_n\} \subset \mathfrak{D}(\Gamma)$, and any $A \in \mathfrak{A}, \alpha_\lambda \in F$,

$$\alpha_\lambda A = \operatorname*{norm-lim}_{n\to\infty} \exp[i\lambda q(f_n)]A \exp[-i\lambda q(f_n)]. \tag{2.1}$$

The choice of $\{f_n\}$ depends on Γ; we will exhibit in Sec. 5 the sequence appropriate to our problem. It is reasonable to suppose that all groups of physical interest are locally generated. Finally, we remark that the useful automorphism groups are symmetries of the Hamiltonian; this implies that the local generator $q(\mathbf{x})$ must satisfy a continuity equation. In the present context, this equation takes the form[12]

$$\left(\Phi, \left[\frac{\partial}{\partial t}\pi(q(\mathbf{x}))_t + \nabla \cdot \pi(\mathbf{I}(\mathbf{x}))\right]\Omega\right) = 0 \quad \forall \Phi \in \mathfrak{H}, \tag{2.2}$$

where $\pi(\cdot)_t \equiv \gamma(t)\pi(\cdot)$, and $\mathbf{I}$ is an operator of the same type as q.

We will always assume that the Hamiltonian H_n for a finite region Λ_n can be written as the sum of a kinetic energy term H_{n0} and an interaction term H_{n1} and that the commutators $[q(f), H_{n0}]$ and $[q(f), H_{n1}]$ have a common dense domain in $\mathfrak{H}(\Lambda_n)$. Under these conditions the current $\mathbf{I}$ can be written as the sum of contributions from the kinetic and interaction terms in the Hamiltonian. This assumption is essentially a regularity condition on the interaction potential; therefore, it is unlikely that the arguments to follow can be applied to singular potentials such as those with hard cores.

3. LONG-RANGE ORDER AND BROKEN SYMMETRY

Let G be a group of physical symmetry operations represented by an automorphism group $\{\alpha_g : g \in G\}$. Then a state ω is said to be G-invariant if $\omega(\alpha_g A) = \omega(A)\ \forall g \in G, A \in \mathfrak{A}$—and to be G-*ergodic* if it is an extremal point of the convex set of G-invariant states. Ruelle[15] has argued that pure thermodynamic phases should be described by G-ergodic states, where G is the invariance group for the Hamiltonian; therefore, we may suppose that all states of interest are G-ergodic.

For homogeneous systems, G will include the spatial translations T as a subgroup; we introduce the following notation for the action of T on $\mathfrak{A}$:

$$A(\mathbf{x}) \equiv \alpha_{\mathbf{x}} A \quad \forall \mathbf{x} \in T,\ A \in \mathfrak{A},$$

and also define the *correlation function* C_A by

$$C_A(\mathbf{x}) = \omega(\{\tilde{A}(\mathbf{x}), \tilde{A}^\dagger\}),$$

where $\{\cdot,\cdot\}$ denotes the anticommutator and $\tilde{A} \equiv A - \omega(A)$. A state is said to be *strongly clustering* if

$$\lim_{|x|\to\infty} C_A(\mathbf{x}) = 0, \quad \forall A \in \mathfrak{A}_L,$$

and *weakly clustering* if

$$\lim_{\Lambda\to\infty} [1/V(\Lambda)] \int_\Lambda d^3x\ C_A(\mathbf{x}) = 0, \quad \forall A \in \mathfrak{A}_L, \qquad (3.1)$$

where, for example, $\Lambda \to \infty$ in the sense of van Hove.[16] We will say that a state ω exhibits *long-range order* if, for some $A \in \mathfrak{A}_L$, the weak-clustering condition [Eq. (3.1)] is violated. Note that this definition is stronger than simply requiring $C_A(x) \neq 0$ as $|x| \to \infty$, which would be a violation of the strong cluster property. There are two reasons for choosing this definition: (1) It agrees with the conventional definitions used in the study of Bose condensation, superconductivity, crystalline order, etc.; and (2) there are rigorous general theorems relating the weak cluster property to the symmetry properties of the state. The first such result is Theorem 3.1.

Theorem 3.1: A T-invariant state ω is T-ergodic if and only if it is weakly clustering.[17] In other words, ω exhibits long-range order if and only if it fails to be T-ergodic. The connection between long-range order and broken symmetry will be provided by the following theorem.

Theorem 3.2: Every locally normal T-invariant, KMS state ω is the resultant of a unique probability measure μ_E carried by the locally normal, T-ergodic, KMS states,

$$\omega(A) = \int d\mu_E(\sigma)\sigma(A), \quad \forall A \in \mathfrak{A}.$$

The measure μ_E defines the *ergodic decomposition* of ω. This theorem, without the KMS condition, is given by Ruelle.[18] To include the KMS condition, we need two other facts about integral decompositions on C^*-algebras. First in addition to the ergodic decomposition, there is a *central decomposition*[19] defined by a probability measure μ_c whose support contains the support of μ_E. Second, the central decomposition preserves the KMS condition; i.e., the measure μ_c for a KMS state is carried by the KMS states.[19] The combination of these two remarks with the theorem as given by Ruelle yields Theorem 3.2.

The states which appear in the ergodic decomposition of ω are labeled by parameters that are unimportant for the measurement of local observables, e.g., the phase of a condensate wavefunction, the orientation and location of crystal axes, etc. Since a state ω exhibiting long-range order cannot be T-ergodic states, and ω can, for all physical purposes, be replaced by any one of the states in the decomposition. Note that these states cannot be fully G-invariant, since the original state was assumed to be G-ergodic. This remark provides the connection between long-range order and broken symmetry. In other words, a state with long-range order, that is, one that violates weak clustering, can always be replaced by a state that is weakly clustering but has less symmetry than the Hamiltonian for the system.

4. BOGOLIUBOV INEQUALITIES

The principal result needed for the remainder of this paper is the well-known Bogoliubov inequality.[9] In the context of infinite volume systems, the inequality is based on the following theorem.

Theorem 4.1: If ω is a [locally normal] KMS state, then the bilinear form (X, Y) defined by

$$(X,Y) = \frac{1}{\beta}\int_0^\beta d\tau\ (\Omega, X^\dagger_{-i\tau} Y \Omega), \quad X, Y \in \pi(\mathfrak{A})'', \quad (4.1)$$

is a norm-continuous inner product on $\pi(\mathfrak{A})''$.

In another publication,[20] a version of this theorem was proved under the assumption that time translations are given by automorphisms on $\mathfrak{A}$; however, the proof is easily adapted to cover the case of a state satisfying (a) and (b) of Sec. 2. One simply replaces $\mathfrak{A}$ by $\pi(\mathfrak{A})''$ throughout the argument.

We should also note that Theorem 4.1 remains valid if we drop the assumption of local normality. However, this assumption is essential for applications to unbounded operators.

The Bogoliubov inequalities follow from [Eq. (4.1)] by choosing $X = \pi(K)$ and $Y = i(\partial/\partial t)\pi(M^\dagger)_t$. We obtain

$$\langle\{K, K^\dagger\}\rangle \langle[M, (i\dot{M})^\dagger]\rangle \geq (2/\beta)\,|\langle[K, M^\dagger]\rangle|^2, \qquad (4.2)$$

where $K, M \in \mathfrak{A}$, $\langle\cdot\rangle \equiv (\Omega, \pi(\cdot)\Omega)$; and, by abuse of notation, $\dot{M}$ stands for $(\partial/\partial t)\pi(M)_t|_{t=0}$.

5. ABSENCE OF BROKEN SYMMETRIES IN HOMOGENEOUS THIN FILMS

A *thin film* system is one which has the configuration space $\Gamma = \mathbf{R}^2 \times I$ discussed in Sec. 2. We take the z axis along the interval I and apply hard-wall boundary conditions at $z = 0, L$. With these conventions, the Hamiltonian is invariant under translations and rotations in the x-y plane; therefore, we take G to be the product of the two-dimensional Euclidean group with the relevant internal symmetry group (for example, gauge transformation, spin rotations, etc.). We also assume that the interaction potential is invariant under the internal symmetries. This assumption is always satisfied for gauge transformations, but for spin rotations it means that the potential is spin dependent.

Let ω_0 be a G-ergodic state describing a pure phase of the thin film. We want to know if this state can exhibit long-range order as we have defined it. According to the discussion in Sec. 3, this is the same as asking whether the state ω_0 has a nontrivial ergodic decomposition into non-G-invariant states. We recall that if ω_0 is a locally normal, KMS state, the states in the decomposition will have the same properties. A negative answer to the last question is then provided by showing that there are no broken symmetries. We begin by considering the internal symmetries.

Theorem 5.1: Every T-invariant, locally normal, KMS state ω for a thin-film system is necessarily invariant under any one-parameter group $F \subset G$ of

internal symmetries locally generated by a conserved density $q(x)$. Hence, there can be no long-range order associated with a broken internal symmetry.

For the proof, we first choose operators K and M to be substituted into the Bogoliubov inequality [Eq. (4.2)],

$$K = \int_{S_n} d^2x\ e^{-i\mathbf{k}\cdot\mathbf{x}}\,\tilde{A}(\mathbf{x}),$$
$$M = \int d^3r\, f_n(\mathbf{r})\, e^{-i\mathbf{k}\cdot\mathbf{r}}\, q(\mathbf{r}), \tag{5.1}$$

where $A \in \mathfrak{A}_L$ and $q(\mathbf{r})$ is the local generator for F. We have adopted the convention that $\mathbf{x}, \mathbf{x}'$, etc., are vectors with vanishing z component, while $\mathbf{r}, \mathbf{r}'$ denote general vectors; also, the momentum vector $\mathbf{k}$ has no z component. The area S_n is related to the sequence $\{f_n\}$ of test functions that we must now specify more precisely. Let Λ_n be a right-circular cylinder with axis parallel to the z axis, radius R_n, and height $L_n < L$; we denote by Λ_n^* the coaxial cylinder with radius $R_n + a$ and height L. The function $f_n \in \mathfrak{D}(\Gamma)$ is chosen to satisfy

$$f_n(\mathbf{r}) = \begin{cases} 1, & \mathbf{r} \in \Lambda_n, \\ 0, & \mathbf{r} \in \Gamma \backslash \Lambda_n^*. \end{cases}$$

We take $S_n = \pi R_n^2$ to be the area of the base Λ_n. We substitute the chosen K and M into Eq. (4.2), divide through by $V(\Lambda_n)^2$, and take the limit $\Lambda_n \to \infty$, which means that we first let $L_n \to L$ and then let $R_n \to \infty$ as $n \to \infty$. We briefly sketch the calculation of the various terms involved:

$$\lim_{n\to\infty} (1/S_n)\langle\{K, K^\dagger\}\rangle = \int d^2x\ e^{-i\mathbf{k}\cdot\mathbf{x}}\langle\{\tilde{A}(\mathbf{x}), \tilde{A}^\dagger\}\rangle \equiv \hat{C}_A(\mathbf{k}), \tag{5.2}$$

where we have used translation invariance and the notation $\hat{}$ for the Fourier transform. In a similar way we find

$$\lim_{n\to\infty} (1/S_n)\langle[K, M^\dagger]\rangle = \lim_{n\to\infty} \int d^3r\ f_n(\mathbf{r}) e^{+i\mathbf{k}\cdot\mathbf{r}}\langle[A, q(\mathbf{r})]\rangle. \tag{5.3}$$

Since $A \in \mathfrak{A}_L$, that is, $A \in \mathfrak{A}(\Lambda_0)$ for some finite region Λ_0, the right-hand side of Eq. (5.3) will become independent of Λ_n as soon as $\Lambda_n \supset \Lambda_0$; consequently, the limit $k \to 0$ will commute with the limit $n \to \infty$. Thus we have

$$\lim_{k\to 0}\lim_{n\to\infty} \frac{1}{S_n}\langle[K, M]\rangle = \lim_{n\to\infty}\langle[A, q(f_n)]\rangle = i\frac{\partial}{\partial\lambda}\langle\alpha_\lambda A\rangle_{\lambda=0}, \tag{5.4}$$

where the last line follows from Eq. (2.1). The remaining calculation involves the continuity equation (2.2) for q. We have

$$\lim_{n\to\infty} \frac{1}{S_n L_n^2}\langle[M, (i\dot{M})^\dagger]\rangle = \lim_{n\to\infty} \frac{(-i)}{S_n L_n^2} \int d^3r'\, f_n(\mathbf{r}') e^{i\mathbf{k}\cdot\mathbf{r}'} \times \left(\Omega, \left[\pi(M), \frac{\partial}{\partial t}\,\pi[q(\mathbf{r}')]_t\right]\Omega\right)_{t=0}$$

We have assumed that M is ω-affiliated to $\mathfrak{A}$; consequently, $\Omega \in \mathrm{Dom}(\pi(M))$, and the weak continuity equation (2.2) gives, after some integrations by parts,

$$\lim_{n\to\infty} \frac{1}{S_n L_n^2}\langle[M, (i\dot{M})^\dagger]\rangle = k_\alpha \int d^2x\ e^{-i\mathbf{k}\cdot\mathbf{x}}[\bar{q}(\mathbf{x}), \bar{I}_\alpha(0)]\rangle, \tag{5.5}$$

where $\bar{q}(\mathbf{x}) \equiv L^{-1}\int_0^L dz\ q(\mathbf{x}, z)$, etc., and repeated vector indices are summed. In obtaining Eq. (5.5), we have dropped terms containing factors of $\nabla' f_n(\mathbf{r}')$, which vanish everywhere except in the "skin" region between Λ_n and Λ_n^*. The contributions from the "skin" on the sides of Λ_n are eliminated by the factor S_n^{-1}; and the contributions from the regions above and below Λ_n vanish by virtue of the hard-wall boundary condition, which we impose in the form

$$\lim_{m\to\infty} \int d^3r\, h_m(\mathbf{r})\langle I_z(\mathbf{r})\rangle = 0,$$

whenever h_m is a delta sequence in the z coordinate concentrated at either boundary plane. The terms $\nabla_z f_m$ evidently form such a sequence.

At this point we first make use of the assumption that the internal symmetry generator $q(\mathbf{r})$ commutes with the interaction term in the Hamiltonian. One typically has

$$q(\mathbf{r}) = \psi^\dagger(\mathbf{r})\,\Gamma\psi(\mathbf{r}),$$

where Γ is a Hermitian matrix acting on the internal degrees of freedom only. The corresponding current is

$$I = (2im)^{-1}(\psi^\dagger\Gamma\nabla\psi - \nabla\psi^\dagger\Gamma\psi),$$

and by using the equal-time commutation relations we can compute the integral in Eq. (5.5) explicitly to obtain

$$\lim_{n\to\infty} (1/S_n L_n^2)\langle[M, (i\dot{M})^\dagger]\rangle = Wk^2, \tag{5.6}$$

where

$$W = (mL)^{-1}\langle\overline{\psi^\dagger\Gamma^2\psi}\rangle < \infty.$$

Substitution of Eqs. (5.2), (5.4), and (5.6) into Eq. (4.2) yields

$$k^2\,\hat{C}_A(\mathbf{k}) \geq \frac{2}{\beta W}\left|\frac{\partial}{\partial\lambda}\langle\alpha_\lambda A\rangle_{\lambda=0}\right|^2. \tag{5.7}$$

Thus the Bogoliubov inequalities give us information about the small-k behavior of $\hat{C}_A(\mathbf{k})$ for any A. We obtain additional information by noting that the correlation function for a bounded observable is a continuous, positive-definite function;[21] i.e., the matrix $C_A(x_j - x_k)$ is positive definite for any N distinct points $\{x_1, \ldots, x_N\}$. For functions of this type we have the following results.[22]

Theorem 5.2 (Bochner): Every continuous, positive-definite function is the Fourier transform of a finite positive measure. Thus the correlation function for a bounded observable can be written as

$$C_A(\mathbf{x}) = \int d\mu(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}},$$

where the finite, positive measure μ is related to the (formal) Fourier transform $\hat{C}_A$ by (we now consider a general ν-dimensional configuration space)

$$d\mu(k) = \hat{C}_A(k)\, d^\nu k/(2\pi)^\nu.$$

Strictly speaking, the steps leading from Eq. (5.1) to (5.7) which involve $\hat{C}_A$ are only formally valid; however, the derivation is easily made rigorous by integrating both sides of the Bogoliubov inequality over an

arbitrary bounded region in k-space. We can now employ an argument similar to Hohenberg's;[1] first integrate Eq. (5.7) over a small sphere of radius k_0 centered at the origin to get

$$\int_{k<k_0} d\mu(k)\, k^2 \geq \frac{2}{\beta W}\left|\frac{\partial}{\partial\lambda}\langle\alpha_\lambda A\rangle_{\lambda=0}\right|^2 g(\nu)k_0^\nu, \qquad (5.8)$$

where $g(\nu)k_0^\nu$ is the volume of the sphere. In general $d\mu$ may have a contribution of positive mass concentrated on the origin; we split this term off explicitly and define $d\mu_1$ by

$$d\mu_1(k) = d\mu(k) - b\,\delta_\nu(\mathbf{k})\,d^\nu k,$$

where $b > 0$ is the mass at the origin. The measure μ_1 evidently satisfies

$$k^2 d\mu_1(k) = k^2 d\mu(k), \quad \lim_{k_0\to 0}\int_{k<k_0} d\mu_1(k) = 0.$$

The first condition allows us to estimate the left-hand side of Eq. (5.8) by

$$\int_{k<k_0} d\mu(k)k^2 = \int_{k<k_0} d\mu_1(k)k^2 \leq k_0^2 \int_{k<k_0} d\mu_1(k).$$

Using this estimate in Eq. (5.8), we find

$$\int_{k<k_0} d\mu_1(k) \geq \frac{2}{\beta W}\left|\frac{\partial}{\partial\lambda}\langle\alpha_\lambda A\rangle_{\lambda=0}\right|^2 g(\nu)k_0^{\nu-2}.$$

The left-hand side of this inequality vanishes as k_0 approaches 0; therefore, if $\nu \leq 2$, we must have

$$\frac{\partial}{\partial\lambda}\langle\alpha_\lambda A\rangle_{\lambda=0} = 0 \quad \forall A \in \mathfrak{A}_L.$$

Replace A by $\alpha_\mu A$; then we have

$$\frac{\partial}{\partial\mu}\langle\alpha_\mu A\rangle = 0 \quad \forall A \in \mathfrak{A}_L,\ \alpha_\mu \in F;$$

this yields

$$\omega(\alpha_\mu A) = \omega(A) \quad \forall A \in \mathfrak{A}_L,\ \alpha_\mu \in F;$$

that is, ω is F-invariant. This completes the proof of Theorem 5.1.

The simplest but most important application of this theorem is to forbid broken gauge invariance for thin films. The conventional descriptions of Bose condensation in helium and Cooper-pairing of conduction electrons in metals require broken gauge invariance; therefore, the phenomena of thin film superfluidity and superconductivity cannot be described by the conventional theory. Note that the treatment of Cooper pairing of electrons interacting through the Coulomb potential presents no special difficulties since the proof of Theorem 5.1 is independent of particle statistics and the potential range. Furthermore, the argument excludes all higher order mechanisms such as the formation of quartets, etc.

We now turn our attention to the problem of crystal formation (including magnetic crystals). It follows from the results already obtained that we may describe the system by a state which is T-ergodic and invariant under internal symmetries. By analogy with the general definition of long-range order we will say that a T-ergodic state ω_1 has crystalline *long-range order* if there is a discrete subgroup (lattice group) $T_L \subset T$ such that ω_1 is not T_L-ergodic. The decomposition theorem (Theorem 3.2) remains valid if T is replaced by T_L; consequently, any state having crystalline long-range order can be decomposed into states that are T_L-ergodic but not T-invariant. This corresponds physically to fixing the location of the crystal lattice in space. The following theorem shows that this situation cannot arise for thin films; that is, every T-ergodic state is also T_L-ergodic so that it cannot have crystalline long-range order.

Theorem 5.3: If the average kinetic and potential energy densities are finite, and the potential $v(r)$ satisfies

$$\lim_{r\to\infty} r^{4+\epsilon}\,\nabla^2 v(r) = 0,$$

for some $\epsilon > 0$, then every T_L-invariant, thin-film state ω is also T-invariant; consequently, every T-ergodic state ω_1 is also T_L-ergodic. The second part of the conclusion follows easily from the first which is established by a suitable modification of the proof of Theorem 5.1. Let $\{\mathbf{x}_j : j \in T_L\}$ be the lattice generated by T_L and choose a positive function $\phi \in D(\mathbf{R}^2)$ normalized by

$$\int d^2x\ \phi(\mathbf{x}) = 1$$

with support containing the origin but no other lattice point. Next define $A_\phi(\mathbf{x})$ by

$$A_\phi(\mathbf{x}) = \sum_{j\in T_L}\phi(\mathbf{x}-\mathbf{x}_j)A(\mathbf{x}_j);$$

the sum exists and defines a bounded operator since at most one term is nonvanishing for any given $\mathbf{x}$. With the notation used in the proof of Theorem 5.1 we put

$$K = \int_{S_n} d^2x\ e^{-i\mathbf{k}\cdot\mathbf{x}}\tilde{A}_\phi(\mathbf{x}),$$

and

$$q(\mathbf{r}) \equiv \hat{\mathbf{e}}\cdot\mathbf{J}(\mathbf{r}),$$

where $\hat{\mathbf{e}}$ is a unit vector in the x-y plane and the momentum $\mathbf{J}$ is the local generator for spatial translations. We have explicitly[2,3]

$$\mathbf{J} = (2i)^{-1}(\psi^\dagger\nabla\psi - \nabla\psi^\dagger\psi)$$

and

$$\dot{J}_\beta(\mathbf{r}) = -\nabla_\lambda T^{(0)}_{\lambda\beta} - \int d^3r'\ \nabla_\beta v(\mathbf{r}-\mathbf{r}') \times \psi^\dagger(\mathbf{r})\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}),$$

with

$$T^{(0)}_{\lambda\beta} = \tfrac{1}{2}(\nabla_\lambda\psi^\dagger\nabla_\beta\psi + \nabla_\beta\psi^\dagger\nabla_\lambda\psi) - \tfrac{1}{4}\delta_{\lambda\beta}(\nabla^2\psi^\dagger\psi + \psi^\dagger\nabla^2\psi + 2\nabla\psi^\dagger\cdot\nabla\psi).$$

We next compute the various quantities which appear in the Bogoliubov inequality. First consider the correlation function

$$\hat{C}(\mathbf{k}) = \lim_{n\to\infty}(1/S_n)\langle\{K,K^\dagger\}\rangle = \lim_{n\to\infty}\hat{C}_n(\mathbf{k}),$$

$$\hat{C}_n(\mathbf{k}) = \int d^2x\ e^{-i\mathbf{k}\cdot\mathbf{x}}\,C_n(\mathbf{x}),$$

$$C_n(\mathbf{x}) = (1/S_n)\int d^2x'\ \chi_n(\mathbf{x}+\mathbf{x}')\chi_n(\mathbf{x}') \times\langle\{\tilde{A}_\phi(\mathbf{x}+\mathbf{x}'),\tilde{A}^\dagger_\phi(\mathbf{x}')\}\rangle,$$

where χ_n is the characteristic function for S_n. The function C_n is bounded by $\|A\|^2$ for all $\mathbf{x}$ and n; therefore, $\hat{C}(\mathbf{k})$ is the Fourier transform of the bounded

function $C(\mathbf{x}) = \lim C_n(\mathbf{x})$. Next we calculate

$$\Delta = \lim_{n\to\infty} (1/S_nL_n)\langle[K, M^\dagger]\rangle = \lim_{n\to\infty} \Delta_n,$$

where we set $\mathbf{k} = 0$ to begin with. Inserting the definitions of K and M, we have

$$\begin{aligned}\Delta_n &= (1/S_nL_n)\int d^2x \int d^3r'\ \chi_n(\mathbf{x}) f_n(\mathbf{r}')\langle[A_\phi(\mathbf{x}), q(\mathbf{r}')]\rangle \\ &= (1/S_nL_n)\int d^2x \int d^3r'\ \chi_n(\mathbf{x}) f_n(\mathbf{r}') \sum_{j\in T_L} \phi(\mathbf{x}-\mathbf{x}_j) \\ &\quad \times \langle[A, q(\mathbf{r}'-\mathbf{x}_j)]\rangle.\end{aligned}$$

In the last step we used the T_L invariance of the state. This result may be rewritten as

$$\Delta_n = (1/L_nS_n)\sum_{j\in T_L} \int d^3r' \left[\int d^2x\ \chi_n(\mathbf{x})\,\phi(\mathbf{x}-\mathbf{x}_j)\right] \times f_n(\mathbf{r}'+\mathbf{x}_j)\langle[A, q(\mathbf{r}')]\rangle.$$

The interchange of summation over T_L and the integrals is made legitimate by the fact that the overlap integral in square brackets vanishes for $\mathbf{x}_j$ sufficiently far from S_n. Thus we have, apart from surface terms which will not contribute in the limit,

$$\Delta_n = \int d^3r' \left(\frac{1}{L_nS_n}\sum_{j\in T_L} \chi_n(\mathbf{x}_j) f_n(\mathbf{r}'+\mathbf{x}_j)\right)\langle[A, q(\mathbf{r}')]\rangle.$$

The commutator vanishes for large $\mathbf{r}'$; consequently, we may take the limit inside the integral to obtain

$$\begin{aligned}\Delta &= n_L \int d^3r'\ \langle[A, q(\mathbf{r}')]\rangle \\ &= -in_L\, \mathbf{e}\cdot\nabla\langle A(\mathbf{x})\rangle|_{x=0},\end{aligned}$$

where n_L is the density of lattice sites.

Finally we need to calculate

$$\begin{aligned}Y &= \lim_{n\to\infty} (1/S_nL_n^2)\langle[M, (i\dot{M})^\dagger]\rangle = \lim_{n\to\infty} Y_n, \\ Y_n &= -(i/S_nL_n^2)\hat{e}_\alpha\hat{e}_\beta \int d^3r \int d^3r'\ f_n(\mathbf{r}) f_n(\mathbf{r}') e^{-i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \\ &\quad \times \langle[J_\alpha(\mathbf{r}), \dot{J}_\beta(\mathbf{r}')]\rangle.\end{aligned}$$

The contribution to Y from the $T^{(0)}_{\lambda\beta}$ term in the continuity equation for $\mathbf{J}$ can be obtained explicitly from the commutation relations; thus

$$Y^{(0)} = w_0 k^2 + \tfrac{1}{4}(\bar{n}/L)\, k^2(\mathbf{k}\cdot\hat{\mathbf{e}})^2,$$

where $\bar{n}$ is the average particle density and

$$w_0 = \hat{e}_\alpha\hat{e}_\beta \lim_{n\to\infty} (1/S_nL_n^2)\int d^3r\ f_n(\mathbf{r})\langle\nabla_\alpha\psi^\dagger(\mathbf{r})\nabla_\beta\psi(\mathbf{r})\rangle.$$

The coefficient w_0 is finite since the integrand can be bounded by the kinetic energy density. If we now impose the T_L invariance of the state, the large-volume average can be replaced by an average over the unit cell s_0 of the lattice

$$w_0 = (1/L)\hat{e}_\alpha\hat{e}_\beta(1/s_0L)\int_{s_0\times I} d^3r\,\langle\nabla_\alpha\psi^\dagger(\mathbf{r})\nabla_\beta\psi(\mathbf{r})\rangle.$$

The contribution from the potential term requires a somewhat more complicated but still straightforward treatment which leads to

$$\begin{aligned}Y^{(1)} &= -\hat{e}_\alpha\hat{e}_\beta \int d^3r\ (e^{-i\mathbf{k}\cdot\mathbf{r}}-1)\nabla_\alpha\nabla_\beta v(\mathbf{r})\,\theta(z(L-z)) \\ &\quad \times (1/s_0L^2)\int_{s_0\times I} d^3r'\ \langle\psi^\dagger(\mathbf{r}+\mathbf{r}') \\ &\quad \times \psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}+\mathbf{r}')\rangle.\end{aligned}$$

Since the potential $v(r)$ will usually have a singularity at $r = 0$, we first examine the convergence of the r integral for small r by splitting off a cylinder Λ_0 of height L and radius R_0 centered on the z axis. After two integrations by parts the dominant term in the integral over Λ_0 is found to be

$$(\mathbf{k}\cdot\mathbf{e})^2\,(1/s_0L^2)\int_{s_0\times I} d^3r' \int_{\Lambda_0} d^3r\ v(r) \times \langle\psi^\dagger(\mathbf{r}+\mathbf{r}')\psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}+\mathbf{r}')\rangle.$$

The integral over Λ_0 is finite by virtue of the assumption that the potential energy density is finite; therefore, the integral defining $Y^{(1)}$ converges at small r. Turning to the large r behavior, one sees that the integral converges uniformly in $\mathbf{k}$ if $r^{2+\epsilon}\nabla^2 v(r) \to 0$ as $r\to\infty$; however, we require in addition that $k^{-2}Y^{(1)}$ remains finite as $k\to 0$ so we must impose $r^{4+\epsilon}\nabla^2 v(r)\to 0$. If the latter condition is satisfied, we find

$$\begin{aligned}\lim_{k\to 0} k^{-2}Y^{(1)} &= \hat{e}_\alpha\hat{e}_\beta \int d^3r\ \theta(z(L-z))(\hat{\mathbf{k}}\cdot\mathbf{r})^2\nabla_\alpha\nabla_\beta v(r) \\ &\quad \times (1/s_0L^2)\int_{s_0\times I} d^3r'\ \psi^\dagger(\mathbf{r}+\mathbf{r}') \\ &\quad \times \psi^\dagger(\mathbf{r}')\psi(\mathbf{r}')\psi(\mathbf{r}+\mathbf{r}'),\end{aligned}$$

where $\hat{\mathbf{k}} = \mathbf{k}/k$.

Combining the results for $Y^{(0)}$ and $Y^{(1)}$, we see that there is a constant $w < \infty$ such that

$$Y \le wk^2$$

for sufficiently small $\mathbf{k}$. From this point on the argument is identical to that in the proof of Theorem (5.1), and the conclusion is that the state is T-invariant.

We have now shown that thin films cannot possess crystalline long-range order. The customary description of magnetically ordered system presupposes an underlying structure having crystalline long-range order; therefore, magnetic systems are also forbidden by the last result.

6. LONG-RANGE ORDER IN INHOMOGENEOUS THIN FILMS

In the preceding sections, the property of translation invariance was crucial in establishing the connection between long-range order and broken symmetry via the ergodic decomposition (Theorem 3.2). Since inhomogeneous systems are by definition not translation-invariant, this general relation between the two properties is lost. An alternative approach can be based on the idea of anomalous averages. In this context is is pointless to discuss crystalline states; therefore, we restrict our attention to a group F of internal symmetries. Let $\mathfrak{A}^0$ be the subalgebra of $\mathfrak{A}$ formed by F-invariant operators; then $\mathfrak{A}$ (as a vector space) has the decomposition $\mathfrak{A} = \mathfrak{A}^0 \oplus \mathfrak{A}^\perp$. Here $\mathfrak{A}^\perp$ is the complementary subspace defined by the projection operation $\mathcal{P}$:

$$\mathcal{P}(A) = A - \lim_\alpha \int_F d\lambda\ \chi_\alpha(\lambda)\alpha_\lambda A,$$

where $\{\chi_\alpha\}$ is an M-net on F, that is, an increasing sequence of normalized characteristic functions on F. A state ω is said to have an *anomalous average* (with respect to F) if $\omega(A) \neq 0$ for some $A \in \mathfrak{A}^\perp$. The corresponding *order parameter* Ψ_A is conventionally defined by

$$\Psi_A = \lim_{\Lambda\to\infty} [1/V(\Lambda)] \int_\Lambda d^\nu y\ \omega(A(\mathbf{y})),$$

that is, as the volume mean of an anomalous average. One possible definition of long-range order is that $\Psi_A \neq 0$ for some $A \in \mathfrak{A}$. In the present case the corresponding correlation function depends on both $\mathbf{x}$ and $\mathbf{x}'$,

$$C_A(\mathbf{x},\mathbf{x}') \equiv \langle\{\tilde{A}(\mathbf{x}),\tilde{A}^\dagger(\mathbf{x}')\}\rangle;$$

but it is more convenient to regard it as a function of $\mathbf{x}-\mathbf{x}'$ and $\mathbf{X} \equiv \frac{1}{2}(\mathbf{x}+\mathbf{x}')$. We define a state ω to be *mildly inhomogeneous* if the following "center-of-mass" average exists for each A:

$$D_A(\mathbf{x}-\mathbf{x}') \equiv \lim_{s\to\infty} (1/s) \int_s d^2X\ C_A(\mathbf{x},\mathbf{x}').$$

We can now adapt the argument of Sec. 5 to prove the following.

Theorem 6.1: Let ω be a locally normal, mildly inhomogeneous KMS state for a thin film; then $\Psi_A = 0$ for all $A \in \mathfrak{A}^\perp$. We again define K and M by Eqs. (5.1), and we find

$$\lim_{n\to\infty} (1/S_n)\langle\{K,K^\dagger\}\rangle = \hat{D}_A(\mathbf{k}). \tag{6.1}$$

This result is established by first integrating the left-hand side of Eq. (6.1) (for finite n) with a test function and then taking the limit $n \to \infty$. The other factors appearing in the Bogoliubov inequality can be treated similarly, with the results as follows:

$$\lim_{k\to 0}\lim_{n\to\infty} \frac{1}{S_n}\langle[K,M^\dagger]\rangle = \lim_{n\to\infty} \frac{1}{S_n}\int_{S_n} d^2x\ i\frac{\partial}{\partial\lambda} \times \langle\alpha_\lambda A(\mathbf{x})\rangle_{\lambda=0}, \tag{6.2}$$

$$\lim_{n\to\infty} \frac{1}{S_n L_n^2}\langle[M,(i\dot{M})^\dagger]\rangle = k_\alpha \int d^2(x-x')\ e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} \times \lim_{n\to\infty} \frac{1}{S_n}\int_{S_n} d^2X\ \langle[\bar{q}(\mathbf{x}'),\bar{I}_\alpha(\mathbf{x})]\rangle. \tag{6.3}$$

Just as before, Eq. (6.3) has the small-k form

$$\lim_{n\to\infty} (1/S_n L_n^2)\langle[M,(i\dot{M})^\dagger]\rangle \leq Wk^2.$$

The Bogoliubov inequality now reads

$$k^2\hat{D}_A(\mathbf{k}) \geq \frac{2}{\beta W}\left|\lim_{n\to\infty}\frac{1}{S_n}\int_{S_n} d^2x\ \frac{\partial}{\partial\lambda}\langle\alpha_\lambda A(\mathbf{x})\rangle_{\lambda=0}\right|^2.$$

The function $D_A(\mathbf{x}-\mathbf{x}')$ is a continuous, positive-definite function; therefore, the previous argument requires that

$$\lim_{n\to\infty}\frac{1}{S_n}\int_{S_n} d^2x\ \frac{\partial}{\partial\lambda}\langle\alpha_\lambda A(\mathbf{x})\rangle_{\lambda=0} = 0. \tag{6.4}$$

The map $\alpha_\lambda : \mathfrak{A}^\perp \to \mathfrak{A}^\perp$ is bijective; consequently, the algebra generated by $\{((\partial/\partial\lambda)\alpha_\lambda A)_{\lambda=0} : A \in \mathfrak{A}\}$ is dense in $\mathfrak{A}^\perp$. Equation (6.4) implies $\Psi_A = 0\ \forall A \in \mathfrak{A}^\perp$. This completes the proof of Theorem 6.1.

This result is not sufficient to settle the question of long-range order in inhomogeneous films, since there is an alternative idea of order parameter that does not involve anomalous averages. In this approach, which is a generalization of Yang's concept of off-diagonal long-range order, one defines Φ_A for $A \in \mathfrak{A}^\perp$ by

$$\Phi_A^2 = \lim_{\Lambda\to\infty} [1/V(\Lambda)^2] \int_\Lambda d^\nu y \int_\Lambda d^\nu y'\ \langle A(\mathbf{y})A^\dagger(\mathbf{y}')\rangle.$$

If $\Phi_A \neq 0$, the "density matrix" $\langle A(\mathbf{y})A^\dagger(\mathbf{y}')\rangle$ is said to have a *macroscopic eigenvalue*. Long-range order may then be defined by a nonvanishing Φ_A for some $A \in \mathfrak{A}^\perp$. If the state ω satisfies the inhomogeneous form of the weak cluster property, that is,

$$\lim_{\Lambda\to\infty} [1/V(\Lambda)^2] \int_\Lambda d^\nu y \int_\Lambda d^\nu y'\ C_A(\mathbf{y},\mathbf{y}') = 0,$$

then the two order parameters are related by $\Phi_A^2 = |\Psi_A|^2$. Unfortunately, in this case there is no general result[24] like Theorem 3.2 that would assure us that we can always replace the original state by one that is weakly clustering.

The application of Theorem 6.1 to Bose condensation shows that anomalous (non-gauge-invariant) order parameters Ψ_A are excluded. Thus, Bose condensation as usually described in terms of Ψ_A is forbidden for inhomogeneous thin films, but the argument does not exclude long-range order of the Φ_A type.

7. DISCUSSION

The intuitive idea of long-range order is that some correlation function $C_A(\mathbf{x})$ fails to satisfy the strong cluster property; that is, $C_A(\mathbf{x}) \neq 0$ as $|\mathbf{x}| \to \infty$; however, the conventional definitions used for Bose condensation, crystalline ordering, etc., impose the stronger condition that C_A violate the weak cluster property [Eq. (3.1)]. We have adopted the latter, more stringent definition for long-range order. One of the principal advantages of this definition is that the ergodic decomposition theorem for homogeneous systems provides us with a general connection between long-range order and broken symmetry. On the other hand, we have shown that physically acceptable (locally normal, T-invariant, KMS) states for a thin film cannot exhibit any broken internal[25] symmetries. Therefore, the phenomenon of Bose condensation which is associated with broken gauge invariance is forbidden. Since the arguments used are independent of the statistics of the particles and the range of the potential, we conclude that the condensation of Cooper pairs usually associated with superconductivity is also forbidden. The formation of a crystal lattice involves the violation of translation invariance, which we have shown to be impossible as long as the potential satisfies the range condition $r^{4+\epsilon}\nabla^2 v(r) \to 0$. Since magnetically ordered systems possess an underlying crystal structure, they are also forbidden.

In the case of inhomogeneous thin films, the strong results discussed above are not available, because of the crucial role played by translation invariance. However, we have shown that, even for inhomogeneous films, the order parameters defined by anomalous averages must vanish.

In view of the very general nature of the assumptions required for the proof of Theorem 5.1, it seems unlikely to us that the ordered states observed in thin films (for example, superconductivity, superfluid helium films, etc.) can be described in the same way as the analogous phenomena in bulk systems. Thus, one is forced to look for alternative descriptions. One possibility is provided by the idea of "weak" long-range order,[26] which is that the generalized susceptibility for some observable A diverges. This is consistent with weak and even strong clustering.

ACKNOWLEDGMENTS

A preliminary version of this paper was presented at the 1971 Boulder Summer Institute for Theoretical Physics, and the authors would like to express their appreciation to Professor W. Britten for his hospitality. We also want to acknowledge our indebtedness to the participants, especially Professor R. Haag, Professor N. M. Hugenholtz, and Professor D. Ruelle, for stimulating and enlightening conversations.

APPENDIX: THE CASE OF UNBOUNDED OBSERVABLES

Let ω be a locally normal, KMS state on the algebra $\mathfrak{A}$ and suppose that the unbounded operator A is ω-affiliated to $\mathfrak{A}$. The correlation function $C_A(x)$ is then a positive-definite tempered distribution and we have the following result.[22]

Theorem (Bochner-Schwartz): Every positive-definite tempered distribution is the Fourier transform of a positive tempered measure. Thus there is a positive measure $d\mu(k)$ related to $\hat{C}_A(k)$ by

$$d\mu(k) = C_A(k)\, d^{\nu}k,$$

which satisfies

$$\int \frac{d\mu(k)}{1 + |k|^n} < \infty$$

for some $n > 0$; furthermore, the integral of $d\mu(k)$ over any compact set is finite.

The derivation of Eq. (5.8) from the Bogoliubov inequality would be technically more difficult for the case of an unbounded operator, but if one is willing to assume Eq. (5.8), then it is clear that the argument given in the text will go through equally well for unbounded observables. This observation could be used to avoid the rather complicated estimates used in Refs. 1–6 which involve assumptions about the finiteness of certain physical quantities (e.g., the compressibility).

---

* Work performed under the auspices of the U.S. Atomic Energy Commission.
† Department of Physics, University of California, Berkeley, California 94720.

1 P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).
2 N. D. Mermin and H. Wagner, Phys. Rev. Letters **17**, 1133 (1966).
3 N. D. Mermin, Phys. Rev. **176**, 250 (1968).
4 D. Jasnow and M. E. Fisher, Phys. Rev. Letters **23**, 286 (1969); Phys. Rev. B **3**, 895, 907 (1971).
5 J. F. Fernández, Phys. Rev. A **3**, 1104 (1971).
6 M. D. Girardeau, J. Math. Phys. **10**, 993 (1968).
7 K. Maki, "Gapless Superconductivity," in *Superconductivity*, edited by R. D. Parks (Dekker, New York, 1969), Vol. 2, p. 1035.
8 D. L. Goodstein and R. L. Elgin, Phys. Rev. Letters **22**, 383 (1969).
9 N. N. Bogoliubov, Phys. Abhandl. Sowjetunion **6**, 1 (1962). For English translation, see N. N. Bogoliubov, *Lectures on Quantum Statistics* (Gordon and Breach, New York, 1971), Vol. 2.
10 R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).
11 J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1969), 2nd ed., p. 15, Exercise 1.1.10.
12 G. L. Sewell, J. Math. Phys. **11**, 1868 (1970).
13 D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970).
14 The definition of a symmetry group by a local current is discussed in D. Kastler, D. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108 (1966).
15 D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), pp. 161ff.
16 Reference 15, p. 14.
17 Reference 15, p. 155, Prop. 6.3.5.
18 Reference 15, p. 159, Prop. 6.4.5.
19 D. Ruelle, lecture notes, in *Cargèse Lectures in Physics* (Gordon and Breach, New York, 1970), Vol. 4.
20 J. C. Garrison and J. Wong, Commun. Math. Phys. **26**, 1 (1972).
21 We are indebted to Professor O. E. Lanford for pointing out the importance of the positive-definiteness condition and suggesting the argument based on Bochner's theorem. We should also point out that a similar argument exists for unbounded observables if we assume the validity of Eq. (5.8). See the Appendix for details.
22 I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4, pp. 155ff.
23 J. A. Swieca, Commun. Math. Phys. **4**, 1 (1967).
24 There are results of this type for some special systems; see O. E. Lanford and D. Ruelle, Commun. Math. Phys. **13**, 194 (1969), and O. E. Lanford, *Colloques Internationaux No. 181, May 1969* (Centre National de la Recherche Scientifique, Paris, 1970).
25 It is worth remarking that the Ising model, the anisotropic Heisenberg model, etc., do not constitute counterexamples to this general result, since in each of these models no continuous symmetry of the Hamiltonian is broken in the ordered phase.
26 G. V. Chester, M. E. Fisher, and N. D. Mermin, Phys. Rev. **185**, 760 (1969).

Exponential Fourier Transforms for Coupled Harmonic Oscillator Chains

K. H. Lee
Department of Physics, University of Missouri, Columbia, Missouri 65201
(Received 19 June 1972)

The exponential Fourier transform is used to study the dynamics of semi-infinite and infinite chains of interacting harmonic oscillators. In addition to the harmonic coupling between nearest neighbors, each oscillator is subjected to frictional and other external time-dependent forces. In contrast with previous studies on such systems, the initial conditions (at $t = 0$) are not specified, and the motion of all the oscillators is expressed in terms of the given applied forces only. The analytic structure of the transforms as well as some properties of the propagators are studied for all possible values of physical constants including the limiting values for uncoupled oscillators. The inverse transforms not readily available from tables are obtained by carrying out the integrations explicitly.

The simple infinite chain of masses connected by ideal springs has been extensively studied as one of the very few many-body systems in which exact calculations are possible.[1] A more complex system, namely the infinite chain of harmonic oscillators, has also been studied.[2] However, very little has been done on the exact treatment of a semi-infinite chain. In a recent study,[3] the exact dynamics of semi-infinite and infinite chains of harmonic oscillators with frictional and other external forces was studied. The motion of each of the oscillators was expressed in terms of the given applied forces and initial conditions.

The present work uses the exponential Fourier transform method to study the exact dynamics of the above systems when the initial conditions (at $t = 0$) are not given. The motion of all the oscillators is

ACKNOWLEDGMENTS

A preliminary version of this paper was presented at the 1971 Boulder Summer Institute for Theoretical Physics, and the authors would like to express their appreciation to Professor W. Britten for his hospitality. We also want to acknowledge our indebtedness to the participants, especially Professor R. Haag, Professor N. M. Hugenholtz, and Professor D. Ruelle, for stimulating and enlightening conversations.

APPENDIX: THE CASE OF UNBOUNDED OBSERVABLES

Let ω be a locally normal, KMS state on the algebra $\mathfrak{A}$ and suppose that the unbounded operator A is ω-affiliated to $\mathfrak{A}$. The correlation function $C_A(x)$ is then a positive-definite tempered distribution and we have the following result.[22]

Theorem (Bochner-Schwartz): Every positive-definite tempered distribution is the Fourier transform of a positive tempered measure. Thus there is a positive measure $d\mu(k)$ related to $\hat{C}_A(k)$ by

$$d\mu(k) = C_A(k)\, d^{\nu}k,$$

which satisfies

$$\int \frac{d\mu(k)}{1 + |k|^n} < \infty$$

for some $n > 0$; furthermore, the integral of $d\mu(k)$ over any compact set is finite.

The derivation of Eq. (5.8) from the Bogoliubov inequality would be technically more difficult for the case of an unbounded operator, but if one is willing to assume Eq. (5.8), then it is clear that the argument given in the text will go through equally well for unbounded observables. This observation could be used to avoid the rather complicated estimates used in Refs. 1–6 which involve assumptions about the finiteness of certain physical quantities (e.g., the compressibility).

---

* Work performed under the auspices of the U.S. Atomic Energy Commission.
† Department of Physics, University of California, Berkeley, California 94720.

1 P. C. Hohenberg, Phys. Rev. **158**, 383 (1967).
2 N. D. Mermin and H. Wagner, Phys. Rev. Letters **17**, 1133 (1966).
3 N. D. Mermin, Phys. Rev. **176**, 250 (1968).
4 D. Jasnow and M. E. Fisher, Phys. Rev. Letters **23**, 286 (1969); Phys. Rev. B **3**, 895, 907 (1971).
5 J. F. Fernández, Phys. Rev. A **3**, 1104 (1971).
6 M. D. Girardeau, J. Math. Phys. **10**, 993 (1968).
7 K. Maki, "Gapless Superconductivity," in *Superconductivity*, edited by R. D. Parks (Dekker, New York, 1969), Vol. 2, p. 1035.
8 D. L. Goodstein and R. L. Elgin, Phys. Rev. Letters **22**, 383 (1969).
9 N. N. Bogoliubov, Phys. Abhandl. Sowjetunion **6**, 1 (1962). For English translation, see N. N. Bogoliubov, *Lectures on Quantum Statistics* (Gordon and Breach, New York, 1971), Vol. 2.
10 R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967).
11 J. Dixmier, *Les algèbres d'opérateurs dans l'espace hilbertien* (Gauthier-Villars, Paris, 1969), 2nd ed., p. 15, Exercise 1.1.10.
12 G. L. Sewell, J. Math. Phys. **11**, 1868 (1970).
13 D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970).
14 The definition of a symmetry group by a local current is discussed in D. Kastler, D. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108 (1966).
15 D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969), pp. 161ff.
16 Reference 15, p. 14.
17 Reference 15, p. 155, Prop. 6.3.5.
18 Reference 15, p. 159, Prop. 6.4.5.
19 D. Ruelle, lecture notes, in *Cargèse Lectures in Physics* (Gordon and Breach, New York, 1970), Vol. 4.
20 J. C. Garrison and J. Wong, Commun. Math. Phys. **26**, 1 (1972).
21 We are indebted to Professor O. E. Lanford for pointing out the importance of the positive-definiteness condition and suggesting the argument based on Bochner's theorem. We should also point out that a similar argument exists for unbounded observables if we assume the validity of Eq. (5.8). See the Appendix for details.
22 I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1964), Vol. 4, pp. 155ff.
23 J. A. Swieca, Commun. Math. Phys. **4**, 1 (1967).
24 There are results of this type for some special systems; see O. E. Lanford and D. Ruelle, Commun. Math. Phys. **13**, 194 (1969), and O. E. Lanford, *Colloques Internationaux No. 181, May 1969* (Centre National de la Recherche Scientifique, Paris, 1970).
25 It is worth remarking that the Ising model, the anisotropic Heisenberg model, etc., do not constitute counterexamples to this general result, since in each of these models no continuous symmetry of the Hamiltonian is broken in the ordered phase.
26 G. V. Chester, M. E. Fisher, and N. D. Mermin, Phys. Rev. **185**, 760 (1969).

Exponential Fourier Transforms for Coupled Harmonic Oscillator Chains

K. H. Lee
Department of Physics, University of Missouri, Columbia, Missouri 65201
(Received 19 June 1972)

The exponential Fourier transform is used to study the dynamics of semi-infinite and infinite chains of interacting harmonic oscillators. In addition to the harmonic coupling between nearest neighbors, each oscillator is subjected to frictional and other external time-dependent forces. In contrast with previous studies on such systems, the initial conditions (at $t = 0$) are not specified, and the motion of all the oscillators is expressed in terms of the given applied forces only. The analytic structure of the transforms as well as some properties of the propagators are studied for all possible values of physical constants including the limiting values for uncoupled oscillators. The inverse transforms not readily available from tables are obtained by carrying out the integrations explicitly.

The simple infinite chain of masses connected by ideal springs has been extensively studied as one of the very few many-body systems in which exact calculations are possible.[1] A more complex system, namely the infinite chain of harmonic oscillators, has also been studied.[2] However, very little has been done on the exact treatment of a semi-infinite chain. In a recent study,[3] the exact dynamics of semi-infinite and infinite chains of harmonic oscillators with frictional and other external forces was studied. The motion of each of the oscillators was expressed in terms of the given applied forces and initial conditions.

The present work uses the exponential Fourier transform method to study the exact dynamics of the above systems when the initial conditions (at $t = 0$) are not given. The motion of all the oscillators is

expressed in terms of the given time-dependent externally applied forces only. The analytic structure of the transforms as well as some properties of the propagators are studied for all possible values of physical constants including the limiting values for uncoupled oscillators. The inverse transforms not readily available from tables[4] are obtained by carrying out the integrations explicitly.

The coupled equations for the systems are

$$m\ddot{x}_n = -Kx_n - \beta\dot{x}_n + \phi_n - k(x_n - x_{n+1}) - k(x_n - x_{n-1})(1 - \delta_{n0}^{\text{semi}}), \quad (1)$$

where m is the particle mass, x_n represents the displacement of the nth particle ($n \geqq 0$ if semi-infinite) measured from its equilibrium position, K and k are the force constants, β is the friction coefficient, δ is the Kronecker delta (to be inserted if semi-infinite), and ϕ_n represents the external force applied to the nth particle and is assumed to be a known function of time. When x_0 is also specified in addition to all the ϕ's, Eq. (1) for $n = 0$ provides the consistency condition between x_0 and ϕ_0.

If one assumes that x_n and ϕ_n have the Fourier transforms

$$X_n(\omega) = F[x_n(t)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} dt x_n(t) \exp(i\omega t), \quad (2)$$

$$\Phi_n(\omega) = F[\phi_n(t)/k], \quad (3)$$

then Eq. (1) leads to

$$X_{n+1} + 4(\nu^2 + 2i\mu\nu - \alpha^2 + \delta_{n0}/4)X_n + (1 - \delta_{n0})X_{n-1} = -\Phi_n, \quad (4)$$

where[5]

$$\nu = \omega/(2\omega_0), \quad \omega_0 = |k/m|^{1/2}, \quad (5)$$

$$\mu = \beta/|16km|^{1/2}, \quad (6)$$

$$\alpha = |(\Omega_0/\omega_0)^2 + 2|^{1/2}/2, \quad \text{and} \quad \Omega_0 = |K/m|^{1/2}. \quad (7)$$

The solutions[6] for Eq. (4) are

$$X_n^{\text{inf}} = -\frac{1}{D}\sum_{-\infty}^{\infty} \xi^{|n-r|}\Phi_r \quad (8a)$$

$$= \xi^{|n|}X_0 - \frac{1}{D}\left(\sum_{1\,(n\geq 0)}^{\infty} \text{or} \sum_{-1\,(n\leq 0)}^{-\infty}\right)[\xi^{|n-r|} - \xi^{|n+r|}]\Phi_r, \quad (8b)$$

$$X_n^{\text{semi}} = -\frac{1}{D}\sum_{0}^{\infty}[\xi^{|n-r|} + \xi^{n+r+1}]\Phi_r \quad (9a)$$

$$= \xi^n X_0 - \frac{1}{D}\sum_{1}^{\infty}[\xi^{|n-r|} - \xi^{n+r}]\Phi_r, \quad (9b)$$

where

$$D = 4[(\lambda^2 - b^2)(\lambda^2 - b^2 - 1)]^{1/2}, \quad (10)$$

$$\xi = (D/2) - 2(\lambda^2 - b^2) + 1, \quad (11)$$

$$\lambda = \nu + i\mu, \quad (12)$$

$$b = (\alpha^2 - \mu^2 - 1/2)^{1/2} \quad (13)$$

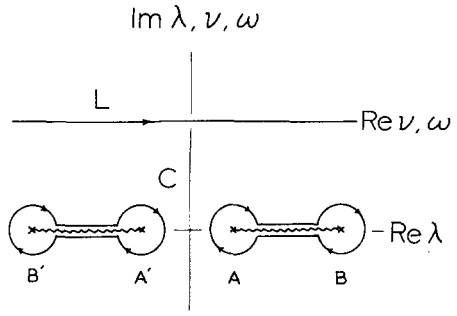


FIG. 1. Contour for underdamped chains ($b^2 > 0$). L is the original path for the inversion integral.

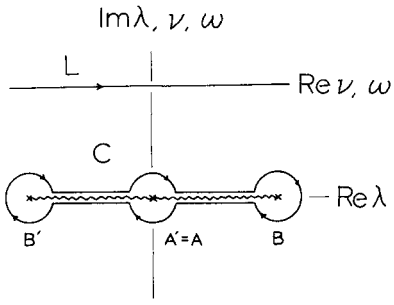


FIG. 2. Contour for critically damped chains ($b^2 = 0 \neq \beta$).

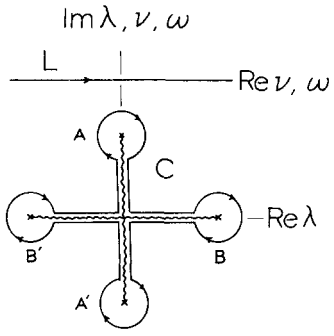


FIG. 3. Contour for overdamped chains ($-1 < b^2 < 0$).

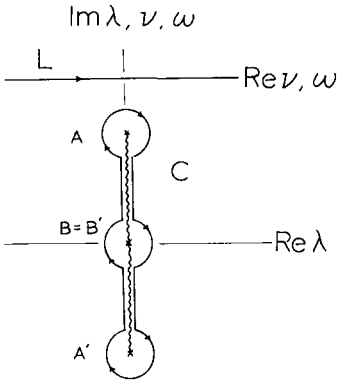


FIG. 4. Contour for overdamped chains ($b^2 = -1$).

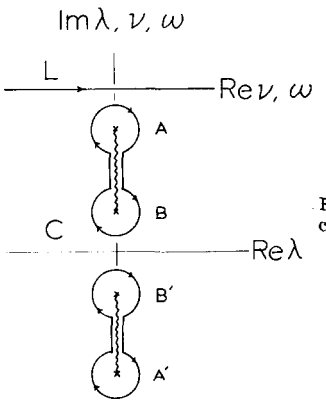


FIG. 5. Contour for overdamped chains ($b^2 < -1$).

$$= [(K - \beta^2/4m)/4k]^{1/2} \tag{14}$$

$$= [(\Omega_0^2 - \Gamma^2)/(2\omega_0)^2]^{1/2}, \tag{15}$$

and

$$\Gamma = 2\mu\omega_0 = \beta/2m. \tag{16}$$

The inverse transforms of X_n given in Eqs. (8a)–(9b) are not readily available from tables, and, hence, the inversion has to be carried out via explicit integration. To this end, one studies the analytic properties of X_n. At first, one observes from (14) that b^2 can take any value in the range $(-\infty, +\infty)$ depending on K, β, m, and k. Also $b^2 > 0, b^2 = 0 \neq \beta$ (and hence $K \neq 0$), and $b^2 < 0$ correspond to the under-, critically-, and over-damped oscillator chains, respectively. The systems with $b^2 = 0 = \beta$ (and hence $K = 0$) are the undamped simple chains, and the noninteracting oscillators are represented by the limiting values $b^2 = \pm\infty$. The branch cuts for X_n are shown in Figs. 1–5 for several values of b^2 in decreasing order. One observes that the cuts are in the lower halves of the ν and ω planes. Also,

$$\mathrm{Im}(\omega_{A,A',B,B'}) \to 0^- \quad \text{as} \quad \beta \to 0^+ \text{ for Figs. 1 and 2} \tag{17}$$

and

$$\mathrm{Im}(\omega_A) \to 0^- \quad \text{as} \quad K/k \to 0^+ \text{ for Figs. 3–5.} \tag{18}$$

By defining an integral

$$G_n(t) = \frac{\exp(-\Gamma t)}{\pi} \int_C d\lambda \, \frac{\xi^{|n|}\exp(-i2\omega_0 t\lambda)}{\omega_0 D}, \qquad t > 0, \tag{19}$$

along any clockwise closed contour C enclosing the cuts, one can write the inverses of (8a)–(9b) as

$$x_n(t) = \sum_{-\infty}^{\infty} G_{n-r}(t)^*[\phi_r(t)/m] \tag{20a}$$

$$= \delta_{n0}x_0(t) + [G_{|n|-1}(t) - G_{|n|+1}(t)]^*[kx_0(t)/m] + \left(\sum_{1 \atop (n\geq 0)}^{\infty} \text{or} \sum_{-1 \atop (n\leq 0)}^{-\infty}\right)[G_{n-r}(t) - G_{n+r}(t)]^*[\phi_r(t)/m], \tag{20b}$$

$$x_n(t) = \sum_0^{\infty}[G_{n-r}(t) + G_{n+r+1}(t)]^*[\phi_r(t)/m] \tag{21a}$$

$$= \delta_{n0}x_0(t) + [G_{n-1}(t) - G_{n+1}(t)]^*[kx_0(t)/m] + \sum_1^{\infty}[G_{n-r}(t) - G_{n+r}(t)]^*[\phi_r(t)/m], \tag{21b}$$

where

$$A(t)^*B(t) \equiv \int_{-\infty}^{t} dt' A(t-t')B(t'). \tag{22}$$

One can give a physical interpretation of Eqs. (20a)–(21b) similar to that given in Ref. 3. For example, the G's are Green functions (or propagators) such that $G_{n-r}(t)^*\phi_r(t)/m$ represents the displacement component in $x_n(t)$ due to ϕ_r during the prior time interval $(-\infty, t)$. The second propagators in [ ] of Eqs. (20b)–(21b) represent the following reflections: (21a) the reflections without phase reversal at the stiff-to-soft boundary located between the particle $n = 0$ and the missing particle $n = -1$, (20b) and (21b) the reflections with phase reversal at the soft-to-stiff boundary located at the particle $n = 0$. The second terms of (20b) and (21b) imply that the effect of specifying x_0 is equivalent to introducing a soft-to-stiff boundary at the particle $n = 0$ and an effective force kx_0 applied to the particle $|n| = 1$. It can be shown (see Appendix) that

$$G_n(t) = \frac{\exp(-\Gamma t)}{\pi} \int_0^{\pi} d\theta \cos(n\theta) \times \left(\frac{\sin\{2\omega_0 t[b^2 + \sin^2(\theta/2)]^{1/2}\}}{2\omega_0[b^2 + \sin^2(\theta/2)]^{1/2}}\right) \tag{23}$$

$$= \exp(-\Gamma t) \int_0^t dt' J_0(2b\omega_0|t^2 - t'^2|^{1/2}) J_{2n}(2\omega_0 t') \tag{24}$$

$$= \exp(-\Gamma t) \int_0^t dt' J_0\{[(\Omega_0^2 - \Gamma^2)(t^2 - t'^2)]^{1/2}\} J_{2n}(2\omega_0 t'). \tag{25}$$

Therefore, one observes that the present solutions in (20a)–(21b) are identical in form to those for the systems with specified initial conditions (at $t = 0$) except for two expected points: (i) The terms containing $x_r(0)$ and $\dot{x}_r(0)$ are missing and (ii) the lower limit of the * operation given in (22) is now $t = -\infty$ instead of $t = 0$.

The properties of the displacements x_n given in (20a)–(21b) can be studied by examining the properties of G. For example, one can show that[7]

$$\dot{G}_n = \exp(-\Gamma t)J_{2n}(2\omega_0 t) - \Gamma G_n - (t/2)(\Omega_0^2 - \Gamma^2)(G_n + G_n^1), \tag{26}$$

$$\dot{G}_n = -\Gamma G_n + (t\omega_0^2/2n)(G_{n-1} - G_{n+1}), \qquad n \neq 0, \tag{27}$$

$$\ddot{G}_n = -2\Gamma \exp(-\Gamma t)J_{2n}(2\omega_0 t) + (2\Gamma^2 - \Omega_0^2 - 2\omega_0^2)G_n + \omega_0^2(G_{n-1} + G_{n+1}) + (\Gamma t)(\Omega_0^2 - \Gamma^2)(G_n + G_n^1), \tag{28}$$

$$\ddot{G}_n = (2\Gamma^2 - \Omega_0^2 - 2\omega_0^2)G_n + \omega_0^2(G_{n-1} + G_{n+1}) - 2\Gamma(t\omega_0^2/2n)(G_{n-1} - G_{n+1}), \qquad n \neq 0, \tag{29}$$

$$G_n(0) = 0, \dot{G}_n(0) = \delta_{n0}, \ddot{G}_n(0) = -2\Gamma\delta_{n0}, \tag{30}$$

$$G_n(\infty) = \delta_{b0}\delta_{\Gamma 0}/(2\omega_0), \quad \text{and} \quad \dot{G}_n(\infty) = \ddot{G}_n(\infty) = 0, \tag{31}$$

where G_n^1 is an integral of the type G_n given in (24) and (25) with J_0 replaced by J_2. One sees from (30) that an immediate effect of the applied forces is to change the velocity and acceleration, but not the displacement, of the oscillator on which the force is applied. The first expression of (31) implies that only the undamped simple chains (i.e., $K = \beta = 0$) remember the effect of forces applied in the infinitely remote past. One can show from (25) that

$$\lim_{\omega_0 \to 0} G_n(t) = \delta_{n0}\exp(-\Gamma t)|\Omega_0^2 - \Gamma^2|^{-1/2} \times \left(\sin_{(\Omega_0 > \Gamma)} \text{or} \sinh_{(\Omega_0 < \Gamma)}\right)(t|\Omega_0^2 - \Gamma^2|^{1/2}). \tag{32}$$

This is the expected result because the limit taken is equivalent to the uncoupling limit $k \to 0$ and the right-hand side without δ_{n0} is the well-known Green function for the equation of motion for a forced harmonic oscillator.

ACKNOWLEDGMENTS

The author expresses his appreciation to Justin C. Huang and Roland A. Hultsch for many valuable discussions.

APPENDIX

Choosing the contour C for the integral (19) as shown in Figs. 1–5, one can show that the contribution to the integral from the circular arcs around A, A', B, and B' is zero (via vanishing or cancellation in pairs) in the limit of vanishing arc radii. For the remaining integral along the four edges of the branch cut, it is convenient to make a change of variable from λ to θ given by

$$\lambda = [b^2 + \sin^2(\theta/2)]^{1/2}, \quad \theta \in (0, \pi).$$

The sum of the four integrals then reduces to Eq. (23). For the equivalence of (23) and (24), see Appendix C of Ref. 3.

---

[1] See, for example, E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966); J. Hori, *Spectral Properties of Disordered Chains and Lattices* (Pergamon, Oxford, 1968); and A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York, 1971), 2nd ed.

[2] M. A. Huetra, H. S. Robertson, and J. C. Nearing, J. Math. Phys. **12**, 2305 (1971) and the papers quoted therein.

[3] K. H. Lee, J. Math. Phys. **13**, 1312 (1972).

[4] A. Erdélyi *et al.*, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 1, and F. Oberhettinger, *Tabellen zur Fourier Transformation* (Springer-Verlag, Berlin, 1957).

[5] To avoid possible confustion, $|z|^{1/2}$ will be used for the positive root of $|z|$ and $z^{1/2}$ for the double-valued function.

[6] When x_0 is specified, (8b) and (9b) are to be used. The consistency condition between x_0 and ϕ_0, i.e., Eq. (1) for $n = 0$, makes them equivalent to (8a) and (9a). The algebra leading to (8a)–(9b) is very similar to that of Ref. 3. Although $\xi^{\pm|n|} = [\pm D/2 - 2(\lambda^2 - b^2) + 1]^{|n|}$ are two solutions of the homogeneous part of Eq. (4), the terms in X_n of the types $\xi^{-|n|}$ and $\xi^{-|n|}/D$ for $\omega \to \pm\infty$ are inconsistent with conditions for the existence of the inversion integrals.

[7] Some properties of $g_n(t) \equiv \exp(\Gamma t) G_n(t)$ are studied in Ref. 3.

Some Asymptotic Expressions for Prolate Spheroidal Functions and for the Eigenvalues of Differential and Integral Equations of Which They Are Solutions

J. des Cloizeaux and M. L. Mehta

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP. n° 2–91–Gif-sur-Yvette, France

(Received 24 January 1972; Revised Manuscript Received 18 April 1972)

The asymptotic behavior of the prolate spheroidal functions of order zero $f_n(x, c)$, where n is the number of zeros of the function in the interval $-1 \le x \le 1$, is studied for large values of the parameter c and all values of n. The method used involves solving the differential equation which defines the functions by using a classical approximation. The corresponding eigenvalues χ_n are given by an implicit equation and the norm of the functions is calculated. The functions $f_n(x, c)$ are also solutions of an integral equation and associated with eigenvalues $\lambda_n(c)$. Asymptotic expressions of $[1 - \lambda_n(c)]$ are derived by using the values obtained for the norm of $f_n(x, c)$. All these results generalize and interpolate partial results obtained by Slepian and others in two special cases, namely, n finite and $n \simeq c$.

1. INTRODUCTION

For a discrete set of values χ_n of the parameter χ, the differential equation

$$\frac{d}{dx}(1 - x^2)\frac{df}{dx} + (\chi - c^2x^2)f = 0, \tag{1.1}$$

where c is a real nonnegative parameter, has a real continuous solution $f(x)$ which is finite for every x and unique except for a constant factor. It is convenient to fix its amplitude by assuming that $f(1) = 1$, a condition which is assumed throughout this article. For reasons of simplicity, we shall usually omit the explicit dependence on c of $f(x)$ and of other quantities. One may order the χ_n so that $\chi_0 < \chi_1 < \chi_2 < \cdots$, then the corresponding solution $f_n(x)$ has exactly n zeros, in the interval $-1 \le x \le 1$. The function $f_n(x)$ is even or odd according as n is even or odd;

$$f_n(-x) = (-1)^n f_n(x). \tag{1.2}$$

They are called prolate spheroidal functions of zero order.

Some physical applications of these functions arise from the fact that they are also solutions of the integral equation

$$\alpha_n f_n(x) = \int_{-1}^{1} e^{icxy} f_n(y)\,dy \tag{1.3}$$

and of its first iterate

$$\lambda_n f_n(x) = \int_{-1}^{1} \frac{\sin[c(x-y)]}{\pi(x-y)} f_n(y)\,dy, \tag{1.4}$$

where

$$\lambda_n = (c/2\pi)|\alpha_n|^2. \tag{1.5}$$

Many details of their properties can be found in the literature.[1] Sometimes one needs the behavior of $f_n(x)$ and of λ_n for large values of c. Partial results have been obtained by several people in this domain. In particular, a few years ago Slepian[2] gave asymptotic expansions of $f_n(x)$ and λ_n, for large values of c, in two different cases:

(1) when n is finite,

(2) when n is large and of the same order of magnitude as c $(n/c \simeq 1)$.

However, for practical applications, the behavior of $f_n(x)$ and λ_n for intermediate values of n is often needed. In this article, we aim at giving expressions of $f_n(x)$ and λ_n for all values of n $[n < c + O(\log c)]$, interpolating cases (1) and (2) treated by Slepian.

APPENDIX

Choosing the contour C for the integral (19) as shown in Figs. 1–5, one can show that the contribution to the integral from the circular arcs around A, A', B, and B' is zero (via vanishing or cancellation in pairs) in the limit of vanishing arc radii. For the remaining integral along the four edges of the branch cut, it is convenient to make a change of variable from λ to θ given by

$$\lambda = [b^2 + \sin^2(\theta/2)]^{1/2}, \quad \theta \in (0, \pi).$$

The sum of the four integrals then reduces to Eq. (23). For the equivalence of (23) and (24), see Appendix C of Ref. 3.

---

[1] See, for example, E. H. Lieb and D. C. Mattis, *Mathematical Physics in One Dimension* (Academic, New York, 1966); J. Hori, *Spectral Properties of Disordered Chains and Lattices* (Pergamon, Oxford, 1968); and A. A. Maradudin, E. W. Montroll, G. H. Weiss, and I. P. Ipatova, *Theory of Lattice Dynamics in the Harmonic Approximation* (Academic, New York, 1971), 2nd ed.

[2] M. A. Huetra, H. S. Robertson, and J. C. Nearing, J. Math. Phys. **12**, 2305 (1971) and the papers quoted therein.

[3] K. H. Lee, J. Math. Phys. **13**, 1312 (1972).

[4] A. Erdélyi *et al.*, *Tables of Integral Transforms* (McGraw-Hill, New York, 1954), Vol. 1, and F. Oberhettinger, *Tabellen zur Fourier Transformation* (Springer-Verlag, Berlin, 1957).

[5] To avoid possible confustion, $|z|^{1/2}$ will be used for the positive root of $|z|$ and $z^{1/2}$ for the double-valued function.

[6] When x_0 is specified, (8b) and (9b) are to be used. The consistency condition between x_0 and ϕ_0, i.e., Eq. (1) for $n = 0$, makes them equivalent to (8a) and (9a). The algebra leading to (8a)–(9b) is very similar to that of Ref. 3. Although $\xi^{\pm|n|} = [\pm D/2 - 2(\lambda^2 - b^2) + 1]^{|n|}$ are two solutions of the homogeneous part of Eq. (4), the terms in X_n of the types $\xi^{-|n|}$ and $\xi^{-|n|}/D$ for $\omega \to \pm\infty$ are inconsistent with conditions for the existence of the inversion integrals.

[7] Some properties of $g_n(t) \equiv \exp(\Gamma t) G_n(t)$ are studied in Ref. 3.

Some Asymptotic Expressions for Prolate Spheroidal Functions and for the Eigenvalues of Differential and Integral Equations of Which They Are Solutions

J. des Cloizeaux and M. L. Mehta

Service de Physique Théorique, Centre d'Etudes Nucléaires de Saclay, BP. n° 2–91–Gif-sur-Yvette, France

(Received 24 January 1972; Revised Manuscript Received 18 April 1972)

The asymptotic behavior of the prolate spheroidal functions of order zero $f_n(x, c)$, where n is the number of zeros of the function in the interval $-1 \le x \le 1$, is studied for large values of the parameter c and all values of n. The method used involves solving the differential equation which defines the functions by using a classical approximation. The corresponding eigenvalues χ_n are given by an implicit equation and the norm of the functions is calculated. The functions $f_n(x, c)$ are also solutions of an integral equation and associated with eigenvalues $\lambda_n(c)$. Asymptotic expressions of $[1 - \lambda_n(c)]$ are derived by using the values obtained for the norm of $f_n(x, c)$. All these results generalize and interpolate partial results obtained by Slepian and others in two special cases, namely, n finite and $n \simeq c$.

1. INTRODUCTION

For a discrete set of values χ_n of the parameter χ, the differential equation

$$\frac{d}{dx}(1 - x^2)\frac{df}{dx} + (\chi - c^2x^2)f = 0, \tag{1.1}$$

where c is a real nonnegative parameter, has a real continuous solution $f(x)$ which is finite for every x and unique except for a constant factor. It is convenient to fix its amplitude by assuming that $f(1) = 1$, a condition which is assumed throughout this article. For reasons of simplicity, we shall usually omit the explicit dependence on c of $f(x)$ and of other quantities. One may order the χ_n so that $\chi_0 < \chi_1 < \chi_2 < \cdots$, then the corresponding solution $f_n(x)$ has exactly n zeros, in the interval $-1 \le x \le 1$. The function $f_n(x)$ is even or odd according as n is even or odd;

$$f_n(-x) = (-1)^n f_n(x). \tag{1.2}$$

They are called prolate spheroidal functions of zero order.

Some physical applications of these functions arise from the fact that they are also solutions of the integral equation

$$\alpha_n f_n(x) = \int_{-1}^{1} e^{icxy} f_n(y)\,dy \tag{1.3}$$

and of its first iterate

$$\lambda_n f_n(x) = \int_{-1}^{1} \frac{\sin[c(x-y)]}{\pi(x-y)} f_n(y)\,dy, \tag{1.4}$$

where

$$\lambda_n = (c/2\pi)|\alpha_n|^2. \tag{1.5}$$

Many details of their properties can be found in the literature.[1] Sometimes one needs the behavior of $f_n(x)$ and of λ_n for large values of c. Partial results have been obtained by several people in this domain. In particular, a few years ago Slepian[2] gave asymptotic expansions of $f_n(x)$ and λ_n, for large values of c, in two different cases:

(1) when n is finite,

(2) when n is large and of the same order of magnitude as c $(n/c \simeq 1)$.

However, for practical applications, the behavior of $f_n(x)$ and λ_n for intermediate values of n is often needed. In this article, we aim at giving expressions of $f_n(x)$ and λ_n for all values of n $[n < c + O(\log c)]$, interpolating cases (1) and (2) treated by Slepian.

Our method is simple. Since c is large, we use a classical approximation (JBKW method or Langer's method) in all the regions of space which are not singular (i.e., far from the values $x = \pm 1$ or $x = \chi c^{-2}$). In the singular regions around $x = \pm 1$ (and sometimes $x = 0$), the spheroidal functions can be approximately represented by confluent hypergeometric functions. Solutions in two different regions must coincide in the domain where they overlap each other, and thus we obtain compatibility requirements which determine the solutions and the corresponding eigenvalues.

In Sec. 2, we give the notation used throughout this article and in Sec. 3 the results obtained. We derive, in Sec. 4, asymptotic expressions for the spheroidal function $f_n(x)$ in various regions, in Sec. 5 an implicit relation for χ_n, in Sec. 6 the normalization integral, and in Sec. 7 asymptotic expressions for $(1 - \lambda_n)$.

2. NOTATIONS

Throughout this paper we shall use the following symbols:

$$\chi = cu, \quad \epsilon = u/c,$$
$$b = \tfrac{1}{2}(c - u), \quad \beta = 2b/c = 1 - \epsilon. \tag{2.1}$$

The ranges of variations of the various quantities are as follows:

$$0 \le \chi < \infty, \quad 0 \le \epsilon < 1 + V/c + O(c^{-2}),$$
$$0 \le n < c + O(\ln c), \quad -\infty < b \le \tfrac{1}{2}c, \tag{2.2}$$

where V is a constant independent of c. The functions $\varphi(b)$ and $\eta(b)$ are defined by

$$\Gamma(\tfrac{1}{2} + i\tfrac{1}{2}b) = [\pi/\cosh(\tfrac{1}{2}\pi b)]^{1/2} e^{i\varphi(b)}, \tag{2.3}$$
$$\eta(b) = \varphi(b) - \tfrac{1}{2}b(\ln|\tfrac{1}{2}b| - 1). \tag{2.4}$$

We note that $\varphi(b)$ is real when b is real, a consequence of the identity

$$\Gamma(\tfrac{1}{2} + z)\Gamma(\tfrac{1}{2} - z) = \pi/\cos(\pi z). \tag{2.5}$$

Also, we note that, when $b \to \infty$,

$$\eta(b) = (1/12b) + O(1/b^2) \tag{2.6}$$

a consequence of Stirling's formula.

We need the function

$$\delta(\epsilon) = 0, \quad \text{if } \epsilon \ge 1 \tag{2.7}$$

$$\delta(\epsilon) = (1 - \epsilon)\left(\int_0^{\pi/2} \frac{\cos^2\alpha\, d\alpha}{(\cos^2\alpha + \epsilon \sin^2\alpha)^{1/2}} - \frac{\pi}{4}\right)$$
$$\equiv E[(1 - \epsilon)^{1/2}] - \epsilon K[(1 - \epsilon)^{1/2}] - \tfrac{1}{4}(1 - \epsilon), \quad \text{if } \epsilon \le 1, \tag{2.8}$$

where E and K are the elliptic integrals

$$E(k) = \int_0^{\pi/2} (1 - k^2 \sin^2\alpha)^{1/2} d\alpha, \tag{2.9}$$

$$K(k) = \int_0^{\pi/2} (1 - k^2 \sin^2\alpha)^{-1/2} d\alpha. \tag{2.10}$$

Finally the function $\mathfrak{D}(\epsilon)$ is defined as

$$\mathfrak{D}(\epsilon) = 2[(1 + e^{\pi b})/2\pi]^{1/2} \exp[c\delta(\epsilon)]. \tag{2.11}$$

Our results can be compared with those of Slepian;[2] by using the correspondence

$$b \leftrightarrow -\delta_{\text{Slepian}}, \quad \varphi(b) \leftrightarrow \varphi_0(\delta)_{\text{Slepian}}. \tag{2.12}$$

3. RESULTS

All the results given below are asymptotic when $c \to \infty$.

A. Spheroidal Function $f_n(x)$

The function $f_n(x)$ has different asymptotic expressions depending on the relative magnitudes of n, x, and c or, equivalently, of ϵ, x, and c. To achieve clarity, we draw up a map in the (ϵ, x)—plane and divide it into various regions. These regions have, in general, no sharp boundaries, they overlap over large areas, and they cover the whole plane. In each of these regions, $f_n(x)$ has an asymptotic expression. On the overlap of two or more regions, the corresponding expressions of $f_n(x)$ coincide approximately, as they should.

As $f_n(x)$ is either an even or an odd function, we need consider only $x \ge 0$. Moreover, $\epsilon \ge 0$. The (ϵ, x)—plane is divided into seven regions, called A, B, C, D^+, D^-, E, and F. (See Fig. 1.) In drawing the boundaries of the regions, the European rules of the road have been adopted.[3] In other words, starting from an interior point of a region one may cross at will a double line and still remain in the region if one crosses it from the dotted side. Starting from an interior point, one may not cross a double line from the solid side, nor a single solid line without leaving the region.

Equivalently, we may define the various regions as follows. Let V be any positive real constant $V \gg 1$;

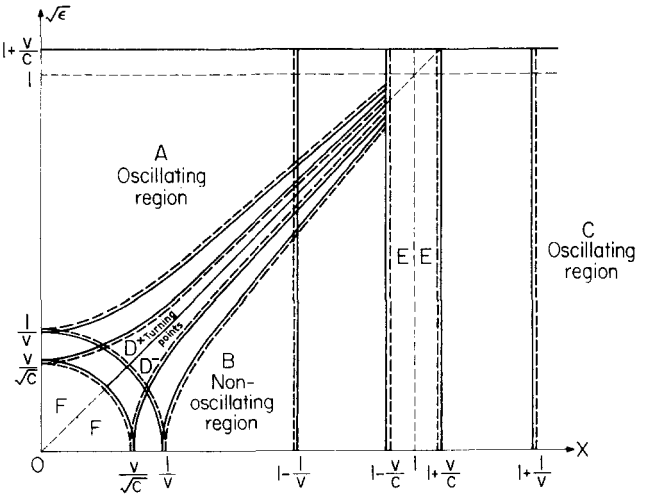


FIG. 1. Map of the regions A, B, C, D⁺, D⁻, E, and F. Starting from an internal point of a region one may freely cross a dotted line, one may also cross a double line and still remain in the region, if one crosses the dotted line first. Crossing a single line or a double line from the solid side is not permitted if one does not want to go outside the region. Note that many regions overlap each other. As V is a number $\gg 1$ and independent of c, $V \ll c$, the boundaries of the regions are not sharp, except for that between D⁺ and D⁻.

in particular, V does not increase with c so that quantities like $1/V, V/c, V/\sqrt{c}, V/\ln c$, etc. are all $\ll 1$. Then

Region A:

$$(1-x)c \geq V, \quad (\epsilon - x^2)\sqrt{c} \geq V. \tag{3.1}$$

Let us note that in Eq. (3.1) the lower limits for $(1-x)c$ and $(\epsilon - x^2)\sqrt{c}$ may be different constants V_1 and V_2. The same would be understood in other similar equations to follow.

Region B:

$$(1-x)c \geq V, \quad (x^2 - \epsilon)\sqrt{c} \geq V. \tag{3.2}$$

Region C:

$$(x-1)c \geq V. \tag{3.3}$$

Region D^+:

$$(1-x)c \geq V, \quad (\epsilon - x^2) \leq 1/V. \tag{3.4}$$

Region D^-:

$$(1-x)c \geq V, \quad (x^2 - \epsilon) \leq 1/V. \tag{3.5}$$

Region E:

$$|x-1| \leq 1/V. \tag{3.6}$$

Region F:

$$x^2 + \epsilon \leq 1/V. \tag{3.7}$$

In each of the regions, $f_n(x)$ has the asymptotic expressions written below. They are compatible with one another and represent approximate solutions of Eq. (1.1).

Region A:

$$f(x) = \mathfrak{D}(\epsilon)[c^2(1-x^2)(\epsilon - x^2)]^{-1/4} \times \cos[\alpha(x) + \eta(b) - \tfrac{1}{4}\pi], \tag{3.8}$$

where

$$\alpha(x) = c\int_x^{\min(1,\sqrt{\epsilon})} \left(\frac{\epsilon - y^2}{1-y^2}\right)^{1/2} dy. \tag{3.9}$$

Region B:

$$f(x) = \tfrac{1}{2}\mathfrak{D}(\epsilon)[c^2(1-x^2)(x^2-\epsilon)]^{-1/4}(\cos\eta(b)\, e^{-\beta(x)} + 2\sin\eta(b)\, e^{\beta(x)-2\delta(\epsilon)}), \tag{3.10}$$

where

$$\beta(x) = c\int_{\sqrt{\epsilon}}^x \left(\frac{y^2 - \epsilon}{1-y^2}\right)^{1/2} dy. \tag{3.11}$$

Region C:

$$f(x) = [(2/\pi)(1 + e^{-\pi b})]^{1/2}[c^2(x^2-1)(x^2-\epsilon)]^{-1/4} \times \cos[\gamma(x) - \eta(b) - \tfrac{1}{4}\pi], \tag{3.12}$$

where

$$\gamma(x) = c\int_{\max(1,\sqrt{\epsilon})}^x \left(\frac{y^2 - \epsilon}{y^2 - 1}\right)^{1/2} dy. \tag{3.13}$$

Regions D^+ and A:

$$f(x) = \mathfrak{D}(\epsilon)[c^2(1-x^2)(\epsilon - x^2)]^{-1/4}[\tfrac{2}{3}\pi\alpha(x)]^{1/2} \times \{\cos[\tfrac{1}{3}\pi + \eta(b)]J_{1/3}(\alpha(x)) + \cos[\tfrac{1}{3}\pi - \eta(b)]J_{-1/3}(\alpha(x))\}, \tag{3.14}$$

where $J_{\pm 1/3}$ are the usual Bessel functions; $\alpha(x)$ is given by (3.9) and in D^+

$$\alpha(x) \approx 2^{3/2}3^{-1}\epsilon^{1/4}(1-\epsilon)^{-1/2}(\epsilon^{1/2} - x)^{3/2}. \tag{3.15}$$

Regions D^- and B:

$$f(x) = \mathfrak{D}(\epsilon)[c^2(1-x^2)(x^2-\epsilon)]^{-1/4}[(2/\pi)\beta(x)]^{1/2} \times \{\pi e^{-2c\delta(\epsilon)}\sin\eta(b) I_{1/3}(\beta(x)) + \cos[\tfrac{1}{3}\pi - \eta(b)]K_{1/3}(\beta(x))\}, \tag{3.16}$$

where $I_{1/3}$ and $K_{1/3}$ are the usual Bessel functions; $\beta(x)$ is given by (3.11) and in region D^-

$$\beta(x) \approx 2^{3/2}3^{-1}\epsilon^{1/4}(1-\epsilon)^{-1/2}(x - \epsilon^{1/2})^{3/2}. \tag{3.17}$$

Region E:

$$f(x) = e^{-icz/2}\Phi(\tfrac{1}{2} + i(b/2), 1; icz) = e^{icz/2}\Phi(\tfrac{1}{2} - i(b/2), 1; -icz), \tag{3.18}$$

where $z = x^2 - 1$ and $\Phi(a, b; z)$ is the confluent hypergeometric function.

Region F:

$$f(x) = \pi^{-1/2}2^{-3n/2-1}c^{-n/2-1/2}\, e^c D_n(x\sqrt{2c}), \tag{3.19}$$

where $D_n(x)$ is the Weber–Hermite function.

B. Eigenvalues

The value of χ_n is implicitly determined by the quantization relation

$$(n + \tfrac{1}{2})\tfrac{1}{2}\pi = c\int_0^{\min(1,\sqrt{\epsilon})} \left(\frac{\epsilon - y^2}{1-y^2}\right)^{1/2} dy + \eta(b). \tag{3.20}$$

Once χ_n is known and, hence, the values of u_n, ϵ_n, and b_n, the eigenvalue λ_n is given by

$$1 - \lambda_n = \frac{\sqrt{2\pi}(u/2e)^u}{\Gamma(\frac{1}{2}(u+1))}\frac{\exp[-2c\delta(\epsilon)]}{1 + \exp(\pi b)}. \tag{3.21}$$

These relations are simplified:

(i) If n is finite, $n \ll c$, then

$$0 \lesssim \epsilon \ll 1, \quad u \cong 2n + 1, \tag{3.22}$$

$$\delta(\epsilon) \cong (1 - \tfrac{1}{4}\pi) + \tfrac{1}{4}\epsilon\ln\epsilon + \epsilon(\tfrac{1}{4}\pi - \ln 2 - \tfrac{1}{4}) \tag{3.23}$$

(see Appendix A),

$$1 - \lambda_n \cong \sqrt{\pi}\, 2^{3n+2}\, e^{-2c} c^{n+1/2}(n!)^{-1}. \tag{3.24}$$

(ii) If n is large, $n \sim c$, then

$$|1 - \epsilon| \ll 1, \tag{3.25}$$

$$\int_0^{\min(1,\sqrt{\epsilon})} \left(\frac{\epsilon - y^2}{1-y^2}\right)^{1/2} dy \cong 1 + \frac{1-\epsilon}{4}\{\ln|1-\epsilon| - 1 - 4\ln 2\} \tag{3.26}$$

(see Appendix B),

$$(n + \tfrac{1}{2})\tfrac{1}{2}\pi \cong c + \varphi(b) - \tfrac{1}{2}b \ln(4c), \tag{3.27}$$

$$1 - \lambda_n \cong (1 + e^{\pi b})^{-1}. \tag{3.28}$$

Relations (3.24), (3.27), and (3.28) agree with Slepian's results.

C. Normalization

The functions $f_n(x)$ are not normalized since the amplitude of $f_n(x)$ has been fixed by requiring that

$$f_n(1) = 1. \tag{3.29}$$

Let the norm then be defined by

$$N_n = \int_{-1}^{1} f_n^2(x)\,dx. \tag{3.30}$$

For N_n, we have

$$N_n = [(1 + e^{\pi b})/\pi c]e^{2c\delta(\epsilon)}\Gamma[\tfrac{1}{2}(u+1)](u/2e)^{-u/2}(2\pi)^{-1/2} \times [2K((1 - |\beta|)^{1/2}) + \ln|\tfrac{1}{2}b| - 2\varphi'(b)]. \tag{3.31}$$

This expression simplifies:

(i) if $n \ll c$; then $\epsilon \ll 1$, and

$$N_n \cong \pi^{-1/2}2^{-3n-2}c^{-n-3/2}e^{2c}n! \tag{3.32}$$

(ii) if $n \sim c$; then $|1 - \epsilon| \ll 1$, and

$$N_n \cong [(1 + e^{\pi b})/\pi c][\ln(4c) - 2\varphi'(b)]. \tag{3.33}$$

4. SPHEROIDAL FUNCTIONS: DERIVATION OF ASYMPTOTIC EXPRESSIONS

A. Classical Approximation in the Regions A, B, C, D⁺, and D⁻

By setting

$$f(x) = |x^2 - 1|^{-1/2}\Theta(x), \tag{4.1}$$

Eq. (1.1) is transformed into

$$\Theta''(x) + \mu(x)\Theta(x) = 0, \tag{4.2}$$

where

$$\mu(x) = c^2(x^2 - \epsilon)/(x^2 - 1) + (x^2 - 1)^{-2}. \tag{4.3}$$

One may neglect the second term on the right-hand side of this equation whenever $|x^2 - 1|c \gg 1$, i.e., in the regions A, B, C, D⁺, and D⁻. When c is large, $\mu(x)$ is large, and the BKW method can be used for Eq. (4.2) provided that[4]

$$|\mu'(x)|/|\mu(x)|^{3/2} \ll 1 \tag{4.4}$$

or approximately

$$|x(1 - \epsilon)| \ll c|x^2 - 1|^{1/2}|x^2 - \epsilon|^{3/2}. \tag{4.5}$$

Thus the approximation breaks down in the vicinity of the points $x = 1$ and $x = \sqrt{\epsilon}$ (turning point) and also when $x = 0$ and $\epsilon \ll 1$. In other words the BKW method is valid in the regions A, B, and C. Near the turning point (i.e. $x = \sqrt{\epsilon}$), connected solutions valid on both sides of it, i.e., in the regions D⁺ and D⁻, can be obtained by the prescriptions of Langer.[5]

In regions A and C, $\mu(x) > 0$, while in B, $\mu(x) < 0$. We write

$$a(x) = [\mu(x)]^{1/2} = c[(\epsilon - x^2)/(1 - x^2)]^{1/2}, \tag{4.6}$$

$$b(x) = [-\mu(x)]^{1/2} = c[(x^2 - \epsilon)/(1 - x^2)]^{1/2}, \tag{4.7}$$

so that in A

$$\Theta(x) = \mathfrak{A}[a(x)]^{-1/2}\cos[\alpha(x) + \eta_A - \tfrac{1}{4}\pi], \tag{4.8}$$

in region B

$$\Theta(x) = \mathfrak{B}[b(x)]^{-1/2}(2\sin\eta_B\, e^{\beta(x)} + \cos\eta_B\, e^{-\beta(x)}), \tag{4.9}$$

and in region C

$$\Theta(x) = \mathfrak{C}[a(x)]^{-1/2}\cos[\gamma(x) + \eta_C - \tfrac{1}{4}\pi], \tag{4.10}$$

where $\alpha(x)$, $\beta(x)$, and $\gamma(x)$ are given by Eq. (3.9), (3.11), and (3.13).

Langer's prescription gives in region D⁺

$$\Theta(x) = \mathfrak{D}[2\pi\alpha(x)/a(x)]^{1/2}[\cos(\tfrac{1}{3}\pi + \eta)J_{1/3}(\alpha(x)) + \cos(\tfrac{1}{3}\pi - \eta)J_{-1/3}(\alpha(x))] \tag{4.11}$$

and in D⁻

$$\Theta(x) = \mathfrak{D}[2\beta(x)/\pi b(x)]^{1/2}[\pi\sin\eta I_{1/3}(\beta(x)) + \cos(\tfrac{1}{3}\pi - \eta)K_{1/3}(\beta(x))]. \tag{4.12}$$

Equations (4.11) and (4.12) coincide approximately with Eqs. (3.14) and (3.16) if we choose

$$\mathfrak{D} = \mathfrak{D}(\epsilon), \qquad \eta = \eta(b)e^{-2c\delta(\epsilon)}. \tag{4.13}$$

This becomes clear if one remarks that for $|b|$ finite, $|1 - \epsilon| \ll 1$, $\delta(\epsilon) \approx \frac{1}{32}\pi(1 - \epsilon)^2 = \frac{1}{8}\pi(b^2/c^2)$ and thus $\exp(c\delta(\epsilon)) \approx 1$, while for $|b| \gg 1$, $\eta(b) \approx 1/12b^2 \ll 1$; so that for any ϵ, $0 < \epsilon < 1$, we have

$$\sin\eta \approx e^{-2c\delta(\epsilon)}\sin\eta(b), \tag{4.14}$$

$$\cos(\tfrac{1}{3}\pi \pm \eta) \approx \cos[\tfrac{1}{3}\pi \pm \eta(b)]. \tag{4.15}$$

For large values of $\alpha(x)$ and $\beta(x)$, we may replace the Bessel functions by their asymptotic expressions. Thus in the overlap region of A and D^+

$$\Theta(x) \approx \mathfrak{D}[a(x)]^{-1/2}\cos[\alpha(x) + \eta - \tfrac{1}{4}\pi], \tag{4.16}$$

while in B and D⁻

$$\Theta(x) \approx \tfrac{1}{2}\mathfrak{D}\{b(x)\}^{-1/2}\{2\sin\eta\, e^{\beta(x)} + \cos\eta\, e^{-\beta(x)}\}. \tag{4.17}$$

Choosing $\mathfrak{D}$ and η as in Eq. (4.13) and using (4.14) and (4.15), we get Eqs. (3.8) and (3.10). The constants in Eq. (4.10) as well as the expressions for $\mathfrak{D}(\epsilon)$, $\eta(b)$, and $\delta(\epsilon)$ will be derived later by comparing the solutions in the overlap regions of E and A, E and B, and E and C.

B. Region E

In E, $x - 1$ is small. So, by introducing the notation

$$z = x^2 - 1, \qquad f(x) \equiv \bar{f}(z), \tag{4.18}$$

Eq. (1.1) can be written as

$$(1 + z)z\frac{d^2\bar{f}}{dz^2} + (1 + \tfrac{3}{2}z)\frac{d\bar{f}}{dz} + \frac{c}{4}(2b + cz)\bar{f} = 0. \tag{4.19}$$

For $|z| \ll 1$, the solution of

$$z\frac{d^2\bar{f}_E}{dz^2} + \frac{d\bar{f}_E}{dz} + \frac{c}{4}(2b + cz)\bar{f}_E = 0 \tag{4.20}$$

is approximately the same as $\bar{f}(z)$. If we set

$$\xi = icz, \quad \bar{f}_E(z) = e^{-icz/2}g(\xi), \tag{4.21}$$

Eq. (4.20) takes the form

$$\xi\frac{d^2g}{d\xi^2} + (1-\xi)\frac{dg}{d\xi} + (-\tfrac{1}{2} - \tfrac{1}{2}ib)g = 0. \tag{4.22}$$

The solution of (4.22) which is regular at $z = 0$ is the confluent hypergeometric function[6]

$$g(\xi) = (\text{const})\Phi(\tfrac{1}{2} + ib/2, 1; \xi). \tag{4.23}$$

The constant is fixed by the condition

$$f(1) = \bar{f}(0) = \bar{f}_E(0) = g(0) = 1, \tag{4.24}$$

so that finally

$$f(x) \cong \bar{f}_E(z) = e^{-icz/2}\Phi(\tfrac{1}{2} + ib/2, 1; icz), \tag{4.25}$$

i.e., Eq. (3.18).

To connect the function (4.25) with the approximate solutions of Eq. (1.1) valid in the regions A, B, and C, we need an asymptotic expression of $\bar{f}_E(z)$ valid for $V/c < |z| < 1/V$ and any value of b. We show in Appendix C that for $|cz| \gg 1$ and $cz + b > 0$,

$$\bar{f}_E(z) = 2\exp[-\tfrac{1}{4}\pi b\epsilon(z)](\cosh(\tfrac{1}{2}\pi b))^{1/2}(\pi cz)^{-1/2} \times \cos[\tfrac{1}{2}cz - \tfrac{1}{4}\pi\epsilon(z) - \tfrac{1}{2}b\ln|cz| - \varphi(b)], \tag{4.26}$$

where

$$\epsilon(z) = \operatorname{sgn}(z). \tag{4.27}$$

On the other hand, in the interval $V/c < |z| < 1/V$, Eq. (4.20) can be solved by the BKW method.

Setting

$$\bar{f}_E(z) = |z|^{-1/2}\Theta_E(z) \tag{4.28}$$

in Eq. (4.20), we get

$$\Theta_E''(z) + \mu_E(z)\Theta_E(z) = 0 \tag{4.29}$$

with

$$\mu_E(z) = \tfrac{1}{4}[c^2 + (2bc/z) + z^{-2}]. \tag{4.30}$$

For $|cz| \gg 1$, $\mu_E(z) \gg 1$ and the term z^{-2} may be neglected. The BKW solution is

$$\bar{f}_E(z) = z^{-1/2}\Theta_E(z) \sim [za_E(z)]^{-1/2}\cos[\alpha_E^+(z) + \eta_E^+], \tag{4.31}$$

for $z > 0, z + \beta > 0$; and

$$\bar{f}_E(z) = (-z)^{-1/2}\Theta_E(z) \sim [|z|a_E(z)]^{-1/2}\cos[\alpha_E^-(z) + \eta_E^-] \tag{4.32}$$

for $z < 0, z + \beta < 0$. The functions $a_E(z)$ and $\alpha_E^{\pm}(z)$ used above are defined by

$$a_E(z) = \tfrac{1}{2}c[(z+\beta)/z]^{1/2}, \tag{4.33}$$

$$\alpha_E^+(z) = \frac{c}{2}\int_{\max(0,-\beta)}^{z}\left(\frac{t+\beta}{t}\right)^{1/2}dt = \tfrac{1}{2}c\{[z(z+\beta)]^{1/2} + \beta\ln[z^{1/2} + (z+\beta)^{1/2}] - \tfrac{1}{2}\beta\ln|\beta|\} \tag{4.34}$$

and

$$\alpha_E^-(z) = \frac{c}{2}\int_{z}^{\min(0,-\beta)}\left(\frac{t+\beta}{t}\right)^{1/2}dt = \tfrac{1}{2}c\{[z(z+\beta)]^{1/2} - \beta\ln[(-z)^{1/2} + (-z-\beta)^{1/2}] + \tfrac{1}{2}\beta\ln|\beta|\}. \tag{4.35}$$

For $|z| \gg |\beta|$, Eqs. (4.34) and (4.35) give

$$\alpha_E^{\pm}(z) \approx \pm\{\tfrac{1}{2}cz + \tfrac{1}{2}b\ln|cz| - \tfrac{1}{2}b(\ln|\tfrac{1}{2}b| - 1)\}. \tag{4.36}$$

Comparing Eq. (4.26) with (4.31), (4.32), and (4.36) we get the phases $\eta_E^{\pm}$ and the proportionality constant. Thus (4.31) and (4.32) may be written as

$$\bar{f}_E(z) = 2\left(\frac{1+e^{-\pi b}}{2\pi}\right)^{1/2}[c^2z(z+\beta)]^{-1/4} \times \cos(\alpha_E^+(z) - \eta(b) - \tfrac{1}{4}\pi) \tag{4.37}$$

for $z > 0, z + \beta > 0$ and

$$\bar{f}_E(z) = 2\left(\frac{1+e^{\pi b}}{2\pi}\right)^{1/2}[c^2z(z+\beta)]^{-1/4} \times \cos[\alpha_E^-(z) + \eta(b) - \tfrac{1}{4}\pi) \tag{4.38}$$

for $z < 0, z + \beta < 0$.

Replacing z by $x^2 - 1$ in the above equations, we obtain

$$f_E(x) = 2\left(\frac{1+e^{-\pi b}}{2\pi}\right)^{1/2}[c^2(x^2-1)(x^2-\epsilon)]^{-1/4} \times \cos(\alpha_E^+(x) - \eta(b) - \tfrac{1}{4}\pi) \tag{4.39}$$

when $x > 1, x^2 > \epsilon$, and

$$f_E(x) = 2\left(\frac{1+e^{\pi b}}{2\pi}\right)^{1/2}[c^2(1-x^2)(\epsilon-x^2)]^{-1/4} \times \cos(\alpha_E^-(x) + \eta(b) - \tfrac{1}{4}\pi), \tag{4.40}$$

when $x < 1, x^2 < \epsilon$. The functions $\alpha_E^{\pm}(x)$ from Eqs. (4.34) and (4.35) can now be written as

$$\alpha_E^+(x) = c\int_{\max(1,\sqrt{\epsilon})}^{x}\left(\frac{y^2-\epsilon}{y^2-1}\right)^{1/2}ydy \tag{4.41}$$

and

$$\alpha_E^-(x) = c\int_{x}^{\min(1,\sqrt{\epsilon})}\left(\frac{\epsilon-y^2}{1-y^2}\right)^{1/2}ydy. \tag{4.42}$$

To get an asymptotic expression for $f_E(x)$ when $\sqrt{\epsilon} < x < 1$, $1 - \epsilon < V/c$, one may use the connecting formulas of Langer[4] for Eqs. (4.28) and (4.29) near the turning point $x = \sqrt{\epsilon}$. We get for $\sqrt{\epsilon} < x < 1$,

$$f_E(x) = [(1+e^{\pi b})/2\pi]^{1/2}[c^2(1-x^2)(x^2-\epsilon)]^{-1/4} \times \{2\sin\eta(b)\exp[\beta_E(x)] + \cos\eta(b)\exp[-\beta_E(x)]\}, \tag{4.43}$$

where

$$\beta_E(x) = c\int_{\sqrt{\epsilon}}^{x}\left(\frac{y^2-\epsilon}{1-y^2}\right)^{1/2}ydy. \tag{4.44}$$

Thus in the overlap regions of E with A, B, and C, $f_E(x)$ is given by the approximate expressions (4. 40), (4. 43), and (4. 39), and we must show that they are nearly equal to the values of $f(x)$ in A, B, and C given, respectively, by Eqs. (3. 8), (3. 10), and (3. 12).

Thus, comparing Eqs. (4. 40) and (4. 41) with Eqs. (3. 9) and (3. 13), we see immediately that for $|x-1| \ll 1$,

$$\alpha_E^+(x) \simeq \gamma(x), \qquad \alpha_E^-(x) \simeq \alpha(x). \tag{4.45}$$

Thus, in the overlap region of E and C, Eq. (4. 39) coincides approximately with Eq. (3. 12).

Now, we remark that in the overlap region of E and A, we have $|1-\epsilon| < V/c$, $\delta(\epsilon) \simeq \frac{1}{32}\pi(1-\epsilon)^2$ [see Eq. (A3) in Appendix A], consequently $c\delta(\epsilon) \simeq 0$ and, therefore,

$$\mathfrak{D}(\epsilon) \simeq (2/\pi)^{1/2}(1+e^{\pi b})^{1/2}. \tag{4.46}$$

Thus, in this region, Eq. (4. 40) coincides approximately with Eq. (3. 8).

In the overlap region of E and B the difference $\beta(x) - \beta_E(x)$ must almost be a constant. We write (4. 43) as

$$\begin{aligned} f(x) \cong\ & [(1+e^{\pi b})/2\pi]^{1/2}[c^2(1-x^2)(x^2-\epsilon)]^{-1/4} \\ & \times \{2\sin\eta(b)\exp[\beta(x) - c\delta(\epsilon)] \\ & + \cos\eta(b)\exp[-\beta(x) + c\delta(\epsilon)]\}, \end{aligned} \tag{4.47}$$

where $\delta(\epsilon)$ is defined by

$$c\delta(\epsilon) = \beta(1) - \beta_E(1) = c\int_{\sqrt{\epsilon}}^{1}\left(\frac{y^2-\epsilon}{1-y^2}\right)^{1/2}(1-y)dy. \tag{4.48}$$

Making the transformation $y^2 = \cos^2\alpha + \epsilon\sin^2\alpha$, we get

$$\delta(\epsilon) = (1-\epsilon)\left(\int_0^{\pi/2}\frac{\cos^2\alpha d\alpha}{(\cos^2\alpha + \epsilon\sin^2\alpha)^{1/2}} - \frac{\pi}{4}\right), \tag{4.49}$$

i.e., Eq. (2. 8), for $\epsilon \leq 1$. On the other hand, we see that Eq. (4. 47) coincides with Eq. (3. 10).

Thus, concerning regions A, B, C, D^+, D^-, and E we have established all the expressions $f_n(x)$ given in Sec. 3A, by showing that in each region those expressions are approximate solutions of Eq. (1. 1) and that whenever two regions overlap, the corresponding expressions coincide approximately in the common domain.

C. Region F

For small values of ϵ, the classical approximation breaks down for $x \ll 1$. This case has been well studied,[1,2] and we will not repeat the discussion here. The asymptotic solution is

$$f_F(x) = \mathfrak{F}D_m(x\sqrt{2c}), \tag{4.50}$$

where m, the number of zeros of the Weber–Hermite function[7] D_m, is related to u by

$$u = 2m + 1. \tag{4.51}$$

To find the constant $\mathfrak{F}$ one may compare (4. 50) with (3. 10) in the overlap region of B and F. Inserting the asymptotic expression of D_m for $x\sqrt{2c} \gg 1$, we have

$$f_F(x) \approx \mathfrak{F}(2cx^2)^{m/2}e^{-cx^2/2}. \tag{4.52}$$

And from Eq. (3. 10) for $V/c < x < 1/V$,

$$f(x) \approx \tfrac{1}{2}\mathfrak{D}(\epsilon)(cx)^{-1/2}\exp[-\beta(x)], \tag{4.53}$$

where

$$\begin{aligned} \beta(x) &= c\int_{\sqrt{\epsilon}}^{x}\left(\frac{y^2-\epsilon}{1-y^2}\right)^{1/2}dy \\ &\approx \tfrac{1}{2}cx^2 - \tfrac{1}{2}u\ln(2x) + \tfrac{1}{4}u[\ln(u/c) - 1]. \end{aligned} \tag{4.54}$$

For $0 \leq \epsilon \ll 1$ (see Appendix A),

$$e^{c\delta(\epsilon)} \approx e^{-(\pi/4)(c-u)}2^{-u}c^{-u/4}u^{u/4}e^{c-u/4}. \tag{4.55}$$

Equation (2. 11) gives then $\mathfrak{D}(\epsilon)$.

Substituting these asymptotic values in Eq. (4. 53) we find

$$f(x) \cong (\pi c)^{-1/2}2^{-u/2-1/2}e^{c}x^{u/2-1/2}e^{-cx^2/2}, \tag{4.56}$$

and comparing with Eq. (4. 52) we get (since $u = 2m + 1$),

$$\mathfrak{F} = (2/\pi)^{1/2}(8c)^{-(m+1)/2}e^{c}. \tag{4.57}$$

In the next section we shall prove that $m = n$, where n is the number of zeros of $f_n(x)$ in the interval $-1 \leq x \leq 1$. This fact together with Eqs. (4. 52) and (4. 57) lead to Eq. (3. 19).

5. RELATION BETWEEN u AND n

We want to derive a relation between u, the eigenvalue of Eq. (1. 1), and n, the number of zeros of $f(x)$ in the range $-1 \leq x \leq 1$. Let us remark that $f_n(x)$ is even or odd according as n is even or odd,

$$f_n(-x) = (-1)^n f_n(x). \tag{5.1}$$

Also it never vanishes in the region B; this is clear from Eq. (3. 10) and from the knowledge that for $b > 0$, $0 < \eta(b) < \frac{1}{2}\pi$ (see Appendix D).

For $\epsilon < V/c$, $x \approx 0$, $f_n(x)$ is approximately proportional to $D_m(x\sqrt{2c})$. Thus $n = m$ and from Eq. (4. 53)

$$u = 2n + 1. \tag{5.2}$$

For $\epsilon > V/c$, a relation between u and n is obtained by remarking that $f_n(x)$ must be symmetric (antisymmetric) with respect to the origin if n is even (odd). As the function $f(x)$ given by Eq. (3. 8) must be consistent with this requirement, we may write

$$\alpha(0) + \eta(b) - \tfrac{1}{4}\pi = (n + 2\nu)\tfrac{1}{2}\pi. \tag{5.3}$$

Moreover, we remark that, for $\epsilon < 1$, the quantity $[\alpha(x) + \eta(b) - \frac{1}{4}\pi]$ can be used to count the number of zeros belonging to the interval $(x, \sqrt{\epsilon})$ contained in region A. As $|\eta(b) - \frac{1}{4}\pi| < \frac{1}{4}\pi$, $\alpha(\sqrt{\epsilon}) = 0$ and since region B does not contain any zero, this means that ν in Eq. (5. 3) must vanish. Thus, the relation between u and n is

$$\begin{aligned} \tfrac{1}{2}n\pi &= \alpha(0) + \eta(b) - \tfrac{1}{4}\pi \\ &\equiv c\int_0^{\min(1,\sqrt{\epsilon})}\left(\frac{\epsilon-y^2}{1-y^2}\right)^{1/2}dy + \eta(b) - \tfrac{1}{4}\pi, \end{aligned} \tag{5.4}$$

i.e., Eq. (3.20).

When $\epsilon \ll 1$, $\eta(b) \simeq 0$ and Eq. (5.4) reduces to Eq. (5.2). Therefore, Eq. (5.4) is valid for all values of ϵ, $0 \leq \epsilon < 1 + V/c$.

6. NORMALIZATION

The norm of $f_n(x)$ is defined by

$$N = \int_{-1}^{1} f_n^2(x)dx. \tag{6.1}$$

This quantity is evaluated in a different way in the regions $\epsilon < 1/V$, $V/c < \epsilon < 1 - V/c$, and $1 - V/c < \epsilon < 1 + V/c$.

(a) For small values of $\epsilon(\epsilon < 1/V)$, the main contribution to N comes from region F. Thus, as (6.2)

$$\int_{-\infty}^{+\infty} D_n^2(X)dX = (2\pi)^{1/2} n!. \tag{6.3}$$

Equation (3.19) gives immediately the result

$$N = \pi^{-1/2} 2^{-(3n+2)} c^{-n-3/2} e^{2c} n! \tag{6.4}$$

or in terms of $u(u = 2n + 1$ for these values of $\epsilon)$:

$$N = \pi^{-1/2} 2^{-(3u+1)/2} c^{-u/2-1} e^{2c} (\tfrac{1}{2}u - \tfrac{1}{2})!. \tag{6.5}$$

(b) In the intermediate range $(V/c < \epsilon < 1 - V/c)$, the main contribution to N comes from the integration of $f_n(x)$ in the oscillating region A where $f_n(x)$ is given by Eq. (3.8). Since the variations of $\alpha(x)$ are large the factor $\cos^2[\alpha(x) + \eta(b) - \frac{1}{4}\pi]$ can be replaced by its mean value, i.e., $\frac{1}{2}$. Thus, we may write

$$\begin{aligned} N &= \mathfrak{D}^2(\epsilon)c^{-1} \int_0^{\sqrt{\epsilon}} [(1-x^2)(\epsilon - x^2)]^{-1/2} dx \\ &= \mathfrak{D}^2(\epsilon)c^{-1} \int_0^{\pi/2} [1 - \epsilon \sin^2\alpha]^{-1/2} d\alpha \\ &= \mathfrak{D}^2(\epsilon)c^{-1}K(\sqrt{\epsilon}). \end{aligned} \tag{6.6}$$

This appears as a rather crude approximation; in particular, it seems that the regions (in x) around the turning point ($x = \sqrt{\epsilon}$) should give important additional corrections. However, a calculation shows that in first approximation this contribution vanishes.[8] Thus, the previous result turns out to be more exact than it may *a priori* look.

(c) When ϵ is of the order of one, $|\epsilon - 1| < V/c$, important contributions to N come from the vicinity of the point $x = 1$. Choosing a value μ such that $\mu = c^{-1/2}O(1)$, we may write

$$N = N_1 + N_2$$

with

$$N_1 = 2\int_0^{(1-\mu)^{1/2}} f^2(x)dx, \quad N_2 = 2\int_{(1-\mu)^{1/2}}^{1} f^2(x)dx \tag{6.7}$$

and calculate separately each integral.

The integral N_1 corresponds to an integration of $f^2(x)$ in region A. The calculation is similar to that of N in the preceding section:

$$N_1 = \mathfrak{D}^2(\epsilon)c^{-1} \int_0^{(1-\mu)^{1/2}} [(1-x^2)(\epsilon - x^2)]^{-1/2} dx. \tag{6.8}$$

Since $\mu \gg |\beta|$, we may replace ϵ by 1, then $\delta(\epsilon) \approx 1$, Eq. (2.11) gives $\mathfrak{D}(\epsilon)$, and we get the final expression

$$N_1 \simeq -\left(\frac{1 + e^{b\pi}}{\pi c}\right) \ln\left(\frac{\mu}{4}\right). \tag{6.9}$$

The integral N_2 corresponds to the integration of $f^2(x)$ in region E [see Eq. (3.18) and note that $z = x^2 - 1$]

$$N_2 = \int_0^{c\mu} \left|\Phi\left(\frac{1}{2} + \frac{ib}{2}, 1; -icz\right)\right|^2 dz. \tag{6.10}$$

Here $c\mu \gg 1$, but the integral diverges when $c\mu \to \infty$.

The calculation of N_2 requires more delicate manipulations. One studies the asymptotic behavior, when $\sigma \to 0$, of the classical formula[9]

$$\begin{aligned} &\int_0^{\infty} e^{-\sigma z} \Phi\left(\frac{1}{2} + \frac{ib}{2}, 1; -iz\right) \Phi\left(\frac{1}{2} - \frac{ib}{2}, 1; iz\right) dz \\ &\quad = e^{b\pi/2}(1 + i\sigma)^{-1/2-ib/2}(1 - i\sigma)^{-1/2+ib/2} \\ &\qquad \times F\left(\frac{1}{2} + \frac{ib}{2}, \frac{1}{2} - \frac{ib}{2}; 1; \frac{1}{1+\sigma^2}\right). \end{aligned} \tag{6.11}$$

Using an integral representation[10] of the function $F(a, a-1; 1; /(1 + \sigma^2))$ for $a = \frac{1}{2} + ib/2$, the final result is

$$N_2 = [(1 + e^{b\pi})/\pi c][\ln(c\mu) - 2\varphi'(b)]. \tag{6.12}$$

From Eqs. (6.9) and (6.12), we deduce the value of N and the arbitrary cutoff μ disappears as expected. Therefore, using Eq. (2.4), we may write

$$N = N_1 + N_2 = [(1 + e^{b\pi})/\pi c][\ln(4c) - 2\varphi'(b)]. \tag{6.13}$$

Thus in cases (a)–(c), the norm N is, respectively, given by Eq. (6.5), (6.6), and (6.13). However, these three different formulas can be replaced by a unique expression [which coincides with Eq. (3.31)], namely

$$\begin{aligned} N = c^{-1}\mathfrak{D}^2(\epsilon)(\tfrac{1}{2}u - \tfrac{1}{2})! \, [(u/2e)^{-u/2}/\sqrt{2\pi}] \\ \times [K(\sqrt{1 - |\beta|}) + \tfrac{1}{2} \ln|\tfrac{1}{2}b| - \varphi'(b)] \end{aligned} \tag{6.14}$$

which is valid everywhere $(\epsilon < 1 + V/c)$.

As will now be shown

(a) for small values of ϵ $(\epsilon < V/c)$ (see definition 2.11)

$$\mathfrak{D}^2(\epsilon) \simeq (2/\pi)e^{(c-u)\pi/2+2c\delta(\epsilon)}, \tag{6.15}$$

using Eq. (A6), we get

$$\mathfrak{D}^2(\epsilon) = \pi^{-1}e^{2c-u/2}2^{-2u+1}c^{-u/2}u^{u/2}. \tag{6.16}$$

On the other hand, according to Eqs. (2.4) and (2.7) $(b \gg 1)$,

$$\varphi'(b) - \log|\tfrac{1}{2}b| \simeq 0. \tag{6.17}$$

And, we have also

$$K(\sqrt{1 - |\beta|}) = K(\sqrt{\epsilon}) \simeq \tfrac{1}{2}\pi. \tag{6.18}$$

Bringing all these values in Eq. (6.14), we find Eq. (6.5).

(b) In the intermediate range $(V/c < \epsilon < 1 - V/c)$, the values of b and u are both large; Eq. (6.17) remains valid and we have

$$(\tfrac{1}{2}u - \tfrac{1}{2})! = (2\pi)^{1/2}(u/2e)^{u/2}. \tag{6.19}$$

Thus we see that in this case, Eqs. (6.14) and (6.6) coincide.

(c) When $\epsilon \simeq 1 (\epsilon < 1)$, u is also large, Eq. (6.19) remains valid, β is small and positive, and we have

$$\mathfrak{D}^2(\epsilon) \simeq 2(1 + e^{b\pi})/\pi, \tag{6.20}$$

$$\begin{aligned} K[(1-\beta)^{1/2}] &= \int_0^{\pi/2} (\sin^2\alpha + \beta\cos^2\alpha)^{-1/2} d\alpha \\ &\simeq \log 4 - \tfrac{1}{2}\log\beta \\ &= -\tfrac{1}{2}\ln(\tfrac{1}{2}b) + \tfrac{1}{2}\ln(4c). \end{aligned} \tag{6.21}$$

Bringing these values in Eq. (6.14), we find Eq. (6.13). Thus the validity of Eq. (6.14) is established in every case.

7. EIGENVALUES λ_n

As $c \to \infty$, the kernel in Eq. (1.4) tends to a delta function, and for any fixed n, $\lambda_n \to 1$. To know how $1 - \lambda_n$ decreases, we use the differential equation[11]

$$\frac{1}{\lambda_n}\frac{d\lambda_n}{dc} = \frac{2}{c}\,|f_n(1)|^2 \Big/ \int_{-1}^{1} f_n^2(x)dx. \tag{7.1}$$

Setting

$$N_n = \int_{-1}^{1} f_n^2(x)dx \tag{7.2}$$

and using our normalization convention

$$f_n(1) = 1, \tag{7.3}$$

we can write Eq. (7.1) as

$$\frac{1}{\lambda_n}\frac{d\lambda_n}{dc} = \frac{2}{c}\frac{1}{N_n}. \tag{7.4}$$

Our proof of Eq. (3.21),

$$\begin{aligned} 1 - \lambda_n = (2\pi)^{1/2}(u/2e)^{u/2}[\Gamma(\tfrac{1}{2}(u+1))]^{-1} \\ \times \exp[-2c\delta(\epsilon)](1 + e^{\pi b})^{-1} \end{aligned} \tag{7.5}$$

consists in verifying that

(i) $\lambda_n \to 1$ as $c \to \infty$, and

(ii) λ_n satisfies the differential equation (7.4).

When n is fixed and $c \to \infty$, we see from Eq. (3.20) that $\epsilon \to 0$, and hence $b \to \infty$, $\eta(b) \to 0$, and $u = 2n + 1$.

From Eqs. (A6) and (2.1),

$$\begin{aligned} \exp[-2c\delta(\epsilon)](1 + e^{\pi b})^{-1} &\approx e^{-2c}(c/u)^{u/2}e^{-u/2}2^{2u} \\ &\to 0, \quad \text{as } c \to \infty \end{aligned} \tag{7.6}$$

and, therefore, $\lambda_n \to 1$.

To see that λ_n given by (7.5) satisfies (7.4), we divide our discussion into two parts, $0 < \epsilon \le 1 - c^{-1/2}$ and $|1 - \epsilon| \le c^{-1/2}$.

First of all let us remark that we may regard the factor

$$F(u) \equiv (2\pi)^{1/2}(u/2e)^{u/2}[\Gamma(\tfrac{1}{2}(u+1))]^{-1} \tag{7.7}$$

as a constant, since either u is finite, $u = 2n + 1$, and the said factor is a constant, or u is large and the said factor is unity.

(i) $\epsilon < 1 - c^{-1/2}$.

In this case, $b > \frac{1}{2}c^{1/2} \gg 1$, $\eta(b) \approx 0$, and Eq. (7.5) gives

$$1 - \lambda_n \approx F(u)e^{-2c\delta(\epsilon) - \pi b}. \tag{7.8}$$

As $\lambda_n \approx 1$, we get

$$\begin{aligned} \frac{1}{\lambda_n}\frac{d\lambda_n}{dc} \approx F(u)e^{-2c\delta(\epsilon)-\pi b} 2\Big[\delta(\epsilon) + \frac{\pi}{4}(1-\epsilon) \\ + \Big(\delta'(\epsilon) - \frac{\pi}{4}\Big)c\frac{d\epsilon}{dc}\Big]. \end{aligned} \tag{7.9}$$

Now, we have

$$\begin{aligned} &\delta(\epsilon) + \tfrac{1}{4}\pi(1-\epsilon) \\ &\quad = \int_0^{(1-\epsilon)^{1/2}} \left(\frac{1-\epsilon-y^2}{1-y^2}\right)^{1/2} dy \\ &\quad = E[(1-\epsilon)^{1/2}] - \epsilon K[(1-\epsilon)^{1/2}], \end{aligned} \tag{7.10}$$

$$\begin{aligned} &\delta'(\epsilon) - \tfrac{1}{4}\pi \\ &\quad = -\tfrac{1}{2}\int_0^{(1-\epsilon)^{1/2}} \frac{dy}{[(1-\epsilon-y^2)(1-y^2)]^{1/2}} \\ &\quad = -\frac{K(\sqrt{1-\epsilon})}{2} \end{aligned} \tag{7.11}$$

and from Eq. (3.20), neglecting $\eta(b)$,

$$\begin{aligned} 0 &= \frac{c}{2}\frac{d\epsilon}{dc}\int_0^{\sqrt{\epsilon}} \frac{dy}{[(\epsilon-y^2)(1-y^2)]^{1/2}} + \int_0^{\sqrt{\epsilon}} \left(\frac{\epsilon-y^2}{1-y^2}\right)^{1/2} dy \\ &\equiv \frac{c}{2}\frac{d\epsilon}{dc}K(\sqrt{\epsilon}) + E(\sqrt{\epsilon}) - (1-\epsilon)K(\sqrt{\epsilon}). \end{aligned} \tag{7.12}$$

So that

$$\begin{aligned} &\delta(\epsilon) + \frac{\pi}{4}(1-\epsilon) + \Big(\delta'(\epsilon) - \frac{\pi}{4}\Big)c\frac{d\epsilon}{dc} \\ &\quad = E[(1-\epsilon)^{1/2}] - \epsilon K[(1-\epsilon)^{1/2}] \\ &\qquad + K[(1-\epsilon)^{1/2}]\frac{[E(\sqrt{\epsilon}) - (1-\epsilon)K(\sqrt{\epsilon})]}{K(\sqrt{\epsilon})} \\ &\quad = \tfrac{1}{2}\pi(K(\sqrt{\epsilon}))^{-1}, \end{aligned} \tag{7.13}$$

where in the last step we have used Legendre's relation[12]

$$\begin{aligned} E[(1-\epsilon)^{1/2}]K(\sqrt{\epsilon}) + E(\sqrt{\epsilon})K[(1-\epsilon)^{1/2}] \\ - K[(1-\epsilon)^{1/2}]K(\sqrt{\epsilon}) = \tfrac{1}{2}\pi. \end{aligned} \tag{7.14}$$

From Eqs. (7.9) and (7.13) we get

$$\frac{c}{2\lambda_n}\frac{d\lambda_n}{dc} = F(u)e^{-2c\delta(\epsilon)-\pi b}\tfrac{1}{2}c\pi(K(\sqrt{\epsilon}))^{-1}. \tag{7.15}$$

This is just the value of $1/N_n$ given by Eq. (6.20) for large values of b. [See Eqs. (2.4) and (2.7).] Equation (7.4) is, therefore, valid in this region:

(ii) $|\epsilon - 1| < c^{-1/2}$

From Eq. (A3), $\delta(\epsilon) \approx \frac{1}{32}\pi(1-\epsilon)^2$, $c\delta(\epsilon) \ll 1$, and Eq. (7.5) becomes

$$1 - \lambda_n \approx (1 + e^{\pi b})^{-1}. \tag{7.16}$$

Hence

$$\frac{1}{\lambda_n}\frac{d\lambda_n}{dc} = \frac{\pi}{1+e^{\pi b}}\frac{db}{dc}. \tag{7.17}$$

Also from Eq. (3.27)

$$(n+\tfrac{1}{2})\tfrac{1}{2}\pi \cong c + \varphi(b) - \tfrac{1}{2}b\ \ln(4c), \tag{7.18}$$

we get by differentiation

$$\frac{db}{dc} \approx [\tfrac{1}{2}\ \ln(4c) - \varphi'(b)]^{-1}. \tag{7.19}$$

In this region, according to Eq. (3.33),

$$N_n = [(1+e^{\pi b})/\pi c][\ln(4c) - 2\varphi'(b)]. \tag{7.20}$$

Equations (7.17), (7.19), and (7.20) are equivalent to Eq. (7.4).

APPENDIX A: PROPERTIES OF $\delta(\epsilon)$

The function $\delta(\epsilon)$ is defined as follows:

for $\epsilon \geq 1$ $\quad \delta(\epsilon) = 0$

for $0 \leq \epsilon \leq 1$

$$\delta(\epsilon) = (1-\epsilon)\left(\int_0^{\pi/2} \frac{\sin^2\alpha d\alpha}{[\sin^2\alpha + \epsilon\cos^2\alpha]^{1/2}} - \frac{\pi}{4}\right) \tag{A1}$$

and both formulas give $\delta(1) = 0$: (A2)

(a) For $0 < 1-\epsilon \ll 1$, we have

$$\delta(\epsilon) \simeq (\pi/32)(1-\epsilon)^2. \tag{A3}$$

This result is established by remarking that

$$\left(\frac{\delta(\epsilon)}{(1-\epsilon)}\right)_{\epsilon=1} = 0, \tag{A4}$$

$$\left(\frac{d}{d\epsilon}\ \frac{\delta(\epsilon)}{(1-\epsilon)}\right)_{\epsilon=1} = -\frac{\pi}{32}. \tag{A5}$$

(b) For $0 \leq \epsilon \ll 1$

$$\delta(\epsilon) \simeq (1-\tfrac{1}{4}\pi) + \tfrac{1}{4}\epsilon\ \ln\epsilon + \epsilon(\tfrac{1}{4}\pi - \ln 2 - \tfrac{1}{4}). \tag{A6}$$

The first term is an immediate consequence of Eq. (A1):

$$\delta(0) = 1 - \tfrac{1}{4}\pi. \tag{A7}$$

On the other hand, for small values of ϵ,

$$\delta'(\epsilon) \simeq -(1-\tfrac{1}{4}\pi) - \tfrac{1}{2}I(\epsilon), \tag{A8}$$

where

$$I(\epsilon) \equiv \int_0^{\pi/2} \frac{\sin^2\alpha\cos^2\alpha}{[\sin^2\alpha + \epsilon\cos^2\alpha]^{3/2}}\, d\alpha. \tag{A9}$$

For $\epsilon \ll 1$ it is easy to show that

$$I(\epsilon) \cong 2\ln 2 - 2 - \tfrac{1}{2}\ \ln\epsilon. \tag{A10}$$

Substituting this value in Eq. (A9) we find

$$\delta'(\epsilon) \cong \tfrac{1}{4}\ \ln\epsilon + \tfrac{1}{4}\pi - \ln 2, \tag{A11}$$

which by integration gives Eq. (A6).

APPENDIX B: ASYMPTOTIC PROPERTIES OF AN INTEGRAL

Let us show that the integral

$$J(\epsilon) = \int_0^{\mathrm{Min}(1,\sqrt{\epsilon})} \left(\frac{\epsilon - y^2}{1-y^2}\right)^{1/2} dy \tag{B1}$$

can be expanded, near $\epsilon = 1$, as

$$J(\epsilon) \simeq 1 + \tfrac{1}{4}(1-\epsilon)[\ln|1-\epsilon| - 1 - 4\ln 2]. \tag{B2}$$

We see immediately, that

$$J(1) = 1. \tag{B3}$$

On the other hand, we have

$$J'(\epsilon) = \frac{1}{2}\int_0^{\mathrm{Min}(1,\sqrt{\epsilon})} [(1-y^2)(\epsilon - y^2)]^{-1/2}\, dy \tag{B4}$$

$$\simeq \frac{1}{2}\int_0^{\pi/2} [|\epsilon - 1| + \sin^2\alpha]^{-1/2}\, d\alpha, \qquad |\epsilon - 1| \ll 1; \tag{B5}$$

we may write $(0 < \eta \ll 1, \eta^2/|\epsilon - 1| \gg 1)$

$$\begin{aligned} J'(\epsilon) &\simeq \tfrac{1}{2}\int_0^{\eta} [|\epsilon - 1| + \alpha^2]^{-1/2}\, d\alpha + \tfrac{1}{2}\int_{\eta}^{\pi/2} [\sin\alpha]^{-1}\, d\alpha \\ &= \tfrac{1}{2}\ln[\eta + (\eta^2 + |\epsilon - 1|)^{1/2}] \\ &\quad - \tfrac{1}{4}\ln|\epsilon - 1| - \tfrac{1}{2}\ \ln tg(\tfrac{1}{2}\eta) \\ &\simeq \ln 2 - \tfrac{1}{4}\ln|\epsilon - 1|. \end{aligned} \tag{B6}$$

Equation (B2) is a direct consequence of Eqs. (B3) and (B7).

APPENDIX C: ASYMPTOTIC VALUE OF $\bar{f}_E(z)$

The function $\bar{f}_E(z)$ is defined by

$$\bar{f}_E(z) = \exp(-icz/2)\Phi(\tfrac{1}{2} + \tfrac{1}{2}ib, 1; icz). \tag{C1}$$

We want to show that for $|z| \gg 1$

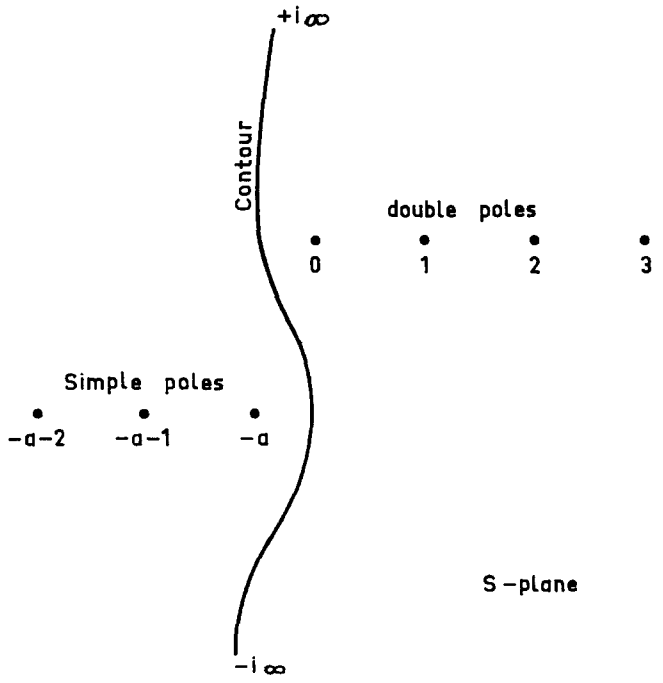


FIG. 2. Contour of integration in the complex plane of S for the calculation of $\psi(a, 1; x)$.

$$\bar{f}_E(z) \simeq 2\exp[-\tfrac{1}{4}b\pi\epsilon(z)][\cosh(\pi b/2)]^{1/2}(\pi cz)^{-1/2} \times \cos[\tfrac{1}{2}cz - \tfrac{1}{4}\pi\epsilon(z) + \tfrac{1}{2}b\ln|cz| - \varphi(b)], \quad \text{(C2)}$$

where $\varphi(b)$ is defined in Sec. 2 and $\epsilon(z) = \mathrm{sgn}(z)$.

The function $\Phi(a, c; x)$ can be expressed in terms of $\psi(a, c; x)$ and $\psi(a, c; -x)$ as[13]

$$\Phi(a, c; x) = \Gamma(c)\left(\frac{e^{i\epsilon a\pi}}{\Gamma(c-a)}\,\psi(a, c; x) + \frac{e^{i\epsilon\pi(a-c)}}{\Gamma(a)}\,e^{x}\psi(c-a, c; -x)\right) \quad \text{(C3)}$$

with $\epsilon = \mathrm{sgn}(\mathrm{Im}x)$

Thus, we may write (C1) as follows:

$$\bar{f}_E(z) = \exp\left(-\frac{b\pi}{2}\,\epsilon(z)\right)\left(\frac{\exp[i[\frac{1}{2}cz - \frac{1}{2}\pi\epsilon(z)]\}}{(-\frac{1}{2} + \frac{1}{2}ib)!} \times \psi(\tfrac{1}{2} - \tfrac{1}{2}ib, 1; -icz) + \text{c.c.}\right). \quad \text{(C4)}$$

The $\psi(a, 1; x)$ can be defined by[14]

$$\psi(a, 1; x) = \frac{1}{2\pi i}\int_{\mathcal{C}} \Gamma(S + a)\left(\frac{\Gamma(-S)}{\Gamma(a)}\right)^2 x^S dS. \quad \text{(C5)}$$

The contour $\mathcal{C}$ goes from $-i\infty$ to $+i\infty$ and passes between the poles of $\Gamma(S + a)$ and $\Gamma(-S)$ (see Fig. 2). When $|x| \to +\infty$, the asymptotic value is obtained by displacing the contour to the left; thus $\psi(a, 1; x)$ is equivalent to the contribution of the first pole which is met by the contour (i.e., $S = -a$)

$$\psi(a, 1; x) \simeq x^{-a}. \quad \text{(C6)}$$

Thus

$$\psi(\tfrac{1}{2} - \tfrac{1}{2}ib, 1; -icz) \simeq (-icz)^{-(\frac{1}{2}-\frac{1}{2}ib)} = |cz|^{-(\frac{1}{2}-\frac{1}{2}ib)}e^{\pi\epsilon(z)(b+i)/4}, \quad \epsilon(z) = \mathrm{sgn}(z). \quad \text{(C7)}$$

In Eq. (C4) let us replace $\psi(\frac{1}{2} + (ib/2), 1; icz)$ by the preceding expression, and let us express $[-\frac{1}{2} - (ib/2)]!$ in terms of $\varphi(b)$ by using Eq. (2. 4). We obtain the result (C2).

APPENDIX D: PROOF THAT $0 < \eta(b) < \frac{1}{2}\pi$ FOR $b > 0$

The phase shift $\eta(b)$ is defined by

$$\eta(b) = \varphi(b) - \tfrac{1}{2}b[\ln|\tfrac{1}{2}b| - 1], \quad \text{(D1)}$$

where $\varphi(b)$ is the phase of $\Gamma(\frac{1}{2} + \frac{1}{2}ib)$ [see Eq. (2. 4)]. According to a well-known formula[15]:

$$\frac{d}{dz}\ln\Gamma(z) = \int_0^\infty \left(\frac{e^{-t}}{t} - \frac{e^{-zt}}{1-e^{-t}}\right)dt \quad \text{(D2)}$$

which can be transformed into

$$\frac{d}{dx}\ln\Gamma\left(\frac{1}{2} + x\right) = \ln x + \int_0^\infty \left(\frac{1}{t} - \frac{e^{-t/2}}{1-e^{-t}}\right)e^{-xt}\,dt. \quad \text{(D3)}$$

By integration from 0 to x, we obtain

$$\ln\Gamma(\tfrac{1}{2} + x) = x(\ln x - 1) + \int_0^\infty \left(\frac{1}{t} - \frac{e^{-t/2}}{1-e^{-t}}\right)\left(\frac{1-e^{-xt}}{t}\right)dt, \quad \text{(D4)}$$

$$\mathrm{Im}\ln\Gamma(\tfrac{1}{2} + \tfrac{1}{2}ib) = \frac{b}{2}\left(\ln\left|\frac{b}{2}\right| - 1\right) + \frac{1}{2}\int_0^\infty \left(\frac{1}{u} - \frac{1}{\sinh u}\right)\frac{\sin bu}{u}\,du. \quad \text{(D5)}$$

Thus

$$\eta(b) = \frac{1}{2}\int_0^\infty \left(\frac{1}{u} - \frac{1}{\sinh u}\right)\frac{\sin bu}{u}\,du. \quad \text{(D6)}$$

To find bounds for $\eta(b)$, we shall use the fact that the function

$$h(u) \equiv \frac{1}{u}\left(\frac{1}{u} - \frac{1}{\sinh u}\right) \quad \text{(D7)}$$

is a decreasing function of u. This last property can be verified by differentiating $h(u)$, reducing to a common denominator, and expanding the numerator in powers of u.

As $h(u)$ is a positive decreasing function of u, we can derive from Eq. (D6) the following inequalities:

$$0 < \eta(b) < \frac{1}{2}\int_0^\infty \left(\frac{1}{u} - \frac{1}{\sinh u}\right)\frac{du}{u} = \tfrac{1}{2}\ln 2 \quad \text{(D8)}$$

and, therefore, the result

$$0 < \eta(b) < \tfrac{1}{2}\pi. \quad \text{(D9)}$$

---

[1] J. Meixner and F. W. Schafke, *Mathieusche Funktionen und Spheroidfunktionen* (Springer-Verlag, Berlin, 1954); C. Flammer, *Spheroidal Wave Functions* (Standford U.P., Palo Alto, Calif., 1957); J. A. Stratton, P. M. Morse, L. J. Chu, J. D. C. Little, and P. J. Corbatto, *Spheroidal Wave Functions* (Wiley, New York, 1956); L. Robin, *Fonctions Sphériques de Legendre et Fonctions Sphéroidales* (Gauthier-Villars, Paris, 1959).

[2] D. Slepian, J. Math. & Phys. **44**, 99 (1965).

[3] Code de la Route, Editions Egé, Paris 1970, Article R.5.

[4] See, for example, A. Messiah, *Mécanique quantique* (Dunod, Paris, 1959), Tome 1, pp. 194–97.

[5] R. E. Langer, Phys. Rev. **51**, 669 (1937). In Eq. (11. b) of this article, the factor 2 in front of the cosine should be dropped assuming as usual that $K_\nu(z) = \int_0^\infty \exp(-z\cosh t)\cosh(\nu t)dt$.

[6] Bateman manuscript project, *Higher Transcendental Functions*, edited by A. Erdelyi (McGraw-Hill, New York, 1953), Vol. 1, Chap. 6.

[7] Reference 6, Vol. 2, Chap. 8.

[8] This contribution can be calculated under the assumption that the potential is linear near the turning point. The result is proportional to the sum of the integrals $\int_0^\infty \{\frac{1}{3}\pi t^{1/3}[J_{1/3}(t) + J_{-1/3}(t)] - t^{2/3}\}dt \equiv -\pi 2^{-1/3}[\Gamma(\frac{1}{3})]^{-2}$ and $\int_0^\infty (1/\pi)[K_{1/3}(t)]^2 t^{1/3}\,dt \equiv \pi 2^{-1/3}[\Gamma(\frac{1}{3})]^{-2}$ and, therefore, vanishes.

[9] Reference 6, Vol. 1, p. 287, Eq. (6. 15. 22).

[10] Reference 6, Vol. 1, p. 114, Eq. (2. 12. 1).

[11] Reference 2, Eq. (4. 3), W. H. J. Fuchs, J. Math. Anal. Appl. **9**, 317 (1964).

[12] Reference 6, Vol. 2, p. 320, Eq. (13. 8. 15).

[13] Reference 6, Vol. 1, p. 259, Eq. (6. 7. 7).

[14] Reference 6, Vol. 1, p. 256, Eq. (6. 5. 5).

[15] Reference 6, Vol. 1, p. 16, Eq. (1. 7. 17).

Generalized Semidirect Product in Group Extensions

Dj. Šijački, M. Vujičić, and F. Herbut
Institute "Boris Kidrich"-Vinča, Belgrade, Yugoslavia
(Received 6 April 1972)

The case when there exists a homomorphism σ of a group G into Aut(K) of a non-Abelian group K [σ having at most one image in every coset of Aut(K) with respect to $I(K)$] is investigated. It is shown that any extension $E \in \mathrm{ext}_\sigma(G,K)$ can be obtained as a generalized semidirect product (GSP): $E = (K \textcircled{T} H)/C'$, where H belongs to $\mathrm{ext}_\sigma(G,C)$ (the group C being the center of K), the semidirect product of K and H is based on τ which equals $\sigma \circ n$ (n being the homomorphism of H onto G), and C' is the antidiagonal of $C \otimes C$. The GSP is a natural generalization of the central extensions, it is applicable to most groups in theoretical physics, and it has a suitable form for the derivation of the irreducible representations of E.

1. INTRODUCTION

A group E is called an extension of a group G by a group K, and denoted by (i, E, s), if there is an invariant subgroup $i(K)$ in E which is isomorphic to K and if $E/i(K) \cong G$.[1]

Group extensions play an important role in physics because they enable us to find more complete symmetry groups of physical systems when their partial symmetries are known.

By diagram the extension E for a given g is expressed as follows:

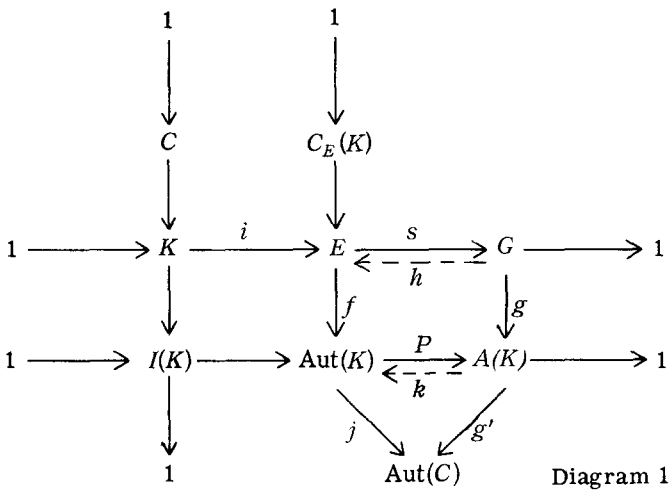


Diagram 1

where C is the center of K, $C_E(K)$ is the centralizer of K in E, $I(K)$ and Aut(K) are the groups of all inner automorphisms and of all automorphisms of K respectively, and $A(K) \equiv \mathrm{Aut}(K)/I(K)$. The mappings s and p are homomorphisms, and the maps h and k satisfy

$$s \circ h = I_G, \tag{1a}$$

$$p \circ k = I_{A(K)} \tag{1b}$$

(e.g., I_G denotes the identity transformation in G). Besides, h and k are normalized (i.e., they take the unit element into the unit one), and the following commutation relation is valid:

$$k \circ g = f \circ h. \tag{2}$$

The homomorphism j is due to the restriction of the domain of each automorphism in K to C.

As it is known, two extensions (i, E, s) and (i', E', s') of G by K with the same g are equivalent if there exists an isomorphism μ: $E \to E'$ such that

$$\mu \circ i = i', \tag{3a}$$

$$s' \circ \mu = s. \tag{3b}$$

The set of all extensions with the same g we denote by $\mathrm{ext}_g(G,K)$.

The problem of finding all inequivalent extensions was solved in the mathematical literature some time ago,[2] and Michel introduced this theory into theoretical physics.[1]

In the case of a non-Abelian K, one reduces the extensions of G by K to those of G by C.[2] It is desirable to do this in the most practical way. Michel has shown[3] that each central extension E (characterized by g being trivial) can be written in the simple form

$$E = (K \otimes H)/C', \tag{4}$$

where H is a certain central extension of G by C and

$$C' = \{(\gamma, \gamma^{-1}) | \gamma \in C\}. \tag{5}$$

The aim of this paper is to generalize this result of Michel to nontrivial g, but such that $k \circ g$ is a homomorphism. We are going to show that in this case all extensions from $\mathrm{ext}_g(G,K)$ can be written, within equivalence, in the following way:

$$E = (K \textcircled{T} H)/C', \tag{6}$$

where τ, H, and C' are defined below (see Sec. 4).

Expression (6) we call the *generalized semidirect product* (GSP). Relation (4) is obviously a special case of GSP, which therefore might be called the generalized direct product.

The scope of applicability of GSP as well as its significance for physics is discussed in Sec. 5.

2. SOME REMARKS ON EQUIVALENCE

In standard extension theory[1,4] it is known that an arbitrary element from $\mathrm{ext}_g(G,K)$ can be written, within equivalence, as the set of ordered pairs

$$E = \{(\alpha, a) | \alpha \in K, a \in G\}, \tag{7a}$$

with the composition law

$$(\alpha, a)(\beta, b) = (\alpha k \circ g[a](\beta)\omega(a,b), ab) \tag{7b}$$

where $k \circ g[a]$ (cf. Diag. 1) is an automorphism in K which is the image of $a \in G$, and $\omega(a,b) \in K, \forall a, b \in G$, is the normalized factor system. The homomorphisms i and s are given by

$$i(\alpha) = (\alpha, e), \tag{7c}$$

$$s(\alpha, a) = a, \tag{7d}$$

where e is the unit element of G.

The necessary and sufficient conditions for E given by (7) to be an element of $\mathrm{ext}_g(G,K)$ are

$$k\circ g[a](k\circ g[b](\alpha)) = \omega(a,b)k\circ g[ab](\alpha)\omega(a,b)^{-1}, \quad (8a)$$

$$\omega(a,b)\omega(ab,c) = k\circ g[a](\omega(b,c))\omega(a,bc) \quad (8b)$$

Lemma 1: Two extensions E and E' of the form (7) are equivalent if and only if there exists an isomorphism $\mu: E \to E'$ such that

$$\mu(\alpha,a) = (\alpha\phi(a)^{-1},a), \quad \forall\alpha\in K, \quad \forall a\in G, \quad (9)$$

where $\phi(a)\in K$ and $\phi(e)=\epsilon$, ϵ being the unit element of K.

Proof: Let E and E' be equivalent through the isomorphism μ. It follows from (7c) that, $\forall\alpha\in K$, $i(\alpha)=i'(\alpha)=(\alpha,e)$, and (3a) then gives $\mu(\alpha,e)=(\alpha,e)$. In the same way (7d) and (3b) imply $\mu(\alpha,a)=(\alpha',a)$, α' some element of K. Let us denote

$$\mu(\epsilon,a) = (\phi(a)^{-1},a), \quad \phi(a)\in K; \quad (10)$$

then $\phi(e)=\epsilon$. Finally, for the general element of E one has $\mu(\alpha,a)=\mu((\alpha,e)(\epsilon,a))=(\alpha,e)(\phi(a)^{-1},a)=(\alpha\phi(a)^{-1},a)$.

The sufficiency of (9) for (3a), (3b) follows immediately from its form. QED

Lemma 2: A set of elements from K, $\{\phi(a)|a\in G\}$, defines an isomorphism $\mu: E\to E'$ via (9) if and only if $\forall\alpha\in K$, $\forall a,b\in G$

$$k'\circ g[a](\alpha) = \phi(a)k\circ g[a](\alpha)\phi(a)^{-1}, \quad (11a)$$

$$\omega'(a,b) = \phi(a)k\circ g[a](\phi(b))\omega(a,b)\phi(ab)^{-1}. \quad (11b)$$

Proof: Let μ be an isomorphism. Then

$$\mu(\alpha,a)\mu(\beta,b) = \mu(\alpha k\circ g[a](\beta)\omega(a,b),ab). \quad (12)$$

Making use of (9), one obtains

$$(\alpha\phi(a)^{-1},a)(\beta\phi(b)^{-1},b) = (\alpha k\circ g[a](\beta)\omega(a,b)\phi(ab)^{-1},ab) \quad (13)$$

lhs $=(\alpha\phi(a)^{-1}k'\circ g[a](\beta\phi(b)^{-1})\omega'(a,b),ab)$. Putting $\beta\phi(b)^{-1}=\beta'$ and equating lhs and rhs in (13), one gets

$$k'\circ g[a](\beta')\omega'(a,b) = \phi(a)k\circ g[a](\beta')k\circ g[a](\phi(b))\omega(a,b)\phi(ab)^{-1}. \quad (14)$$

For $\beta'=\epsilon$, (11b) follows immediately. Replacing (11b) in (14), one finally arrives at (11a).

On the other hand, one can easily see that (11a), (11b) imply (12). QED

Corollary 1: Two equivalent extensions E and E' have $k'=k$ if and only if $\phi(a)\in C$, $\forall a\in G$ [cf. (9)].

Proof: If $k'=k$, then from (11a) it follows that $f(\phi(a))=I_K$ (see Diag. 1), which means $\phi(a)\in C_E(K)$. Since $\phi(a)\in K$ and $C_E(K)\cap K = C$, one has $\phi(a)\in C$, $\forall a\in G$. The sufficiency is obvious from (11a). QED

Corollary 2: Let k' be an arbitrary given normalized mapping which satisfies (1b) and (2). In every class of equivalent extensions there exists at least one of the form (7) with this k' in the composition law.

Proof: In any class of equivalent extensions there is one, say E, of the form (7). From the definition of the mappings k and k' one concludes that their corresponding images differ by a factor which is an inner automorphism in K. Therefore, for each $a\in G$ there exists a $\phi(a)\in K$ so that (11a) with $\phi(e)=\epsilon$ is valid. Using these $\phi(a)$, we define $\omega'(a,b)$ by means of (11b). The group E' of the form (7) with k' and $\omega'(a,b)$ in its composition law is equivalent to E via (9) due to Lemmas 2 and 1. QED

3. HOMOMORPHISM OF G INTO AUT(K)

From now on we confine ourselves to those $\mathrm{ext}_g(G,K)$ which contain at least one semidirect product $K\circledS G$. For such an extension, $k\circ g$ is a homomorphism which we denote by σ, and $\mathrm{ext}_g(G,K)$ we write as $\mathrm{ext}_\sigma(G,K)$.

It is a consequence of Corollary 2 that in every class of equivalent extensions there is one with σ in its composition law.

Lemma 3: In each extension with $k\circ g=\sigma$ the factor system satisfies $\omega(a,b)\in C$, $\forall a,b\in G$.

Proof: The factor system is usually defined by

$$h(a)h(b) = \omega(a,b)h(ab) \quad (15)$$

(see Diag. 1). We now apply the homomorphism f to (15): $f\circ h[a]f\circ h[b] = f(\omega(a,b))f\circ h[ab]$. Since $f\circ h = k\circ g=\sigma$, and σ is a homomorphism, one has $f(\omega(a,b))=I_K$, i.e., $\omega(a,b)\in C_E(K)$. As by definition $\omega(a,b)\in K$, $\forall a,b\in G$, one has $\omega(a,b)\in C$. QED

Lemma 4: In every extension $E\in\mathrm{ext}_\sigma(G,K)$ there is a subgroup H containing $i(C)$, which is an extension of G by C belonging to $\mathrm{ext}_{j\circ\sigma}(G,C)$.

Proof: For each E there is an equivalent $E'=\mu(E)$ with $k\circ g=\sigma$. We define the following subset in E':

$$H' = \{(\gamma,a)|\gamma\in C, a\in G\}. \quad (16a)$$

The composition law (7b) in it becomes

$$(\gamma,a)(\delta,b) = (\gamma j\circ\sigma[a](\delta)\omega(a,b),ab), \quad (16b)$$

because $\forall\delta\in C$: $\sigma[a](\delta)=j\circ\sigma[a](\delta)$ (see Diag. 1). It is shown in Lemma 3 that, in E', $\omega(a,b)\in C$, $\forall a,b\in G$, so that $\gamma j\circ\sigma[a](\delta)\omega(a,b)\in C$, which makes H' closed to multiplication. Furthermore, $(\gamma,a)^{-1}=(\omega(a^{-1},a)^{-1}$ $j\circ\sigma[a^{-1}](\gamma^{-1}),a^{-1})\in H'$, which means that H' is a subgroup. Equations (16a), (16b) entail $H'\in\mathrm{ext}_{j\circ\sigma}(G,C)$. Also the subgroup $H=\mu^{-1}(H')\subset E$ belongs to $\mathrm{ext}_{j\circ\sigma}(G,C)$. QED

Lemma 5: Whenever a homomorphism $\sigma: G\to \mathrm{Aut}(K)$ and H defined by (16) are given, there exists E, an extension of G by K:

$$E = \{(\alpha,a)|\alpha\in K, a\in G\}, \quad (17a)$$

$$(\alpha,a)(\beta,b) = (\alpha\sigma[a](\beta)\omega(a,b),ab), \quad (17b)$$

where $\omega(a,b)$ is the factor system of H.

Proof: It is easy to verify that Eqs. (8a), (8b) are satisfied for E, i.e., that E is an extension. QED

Proposition 1: Equations (16) and (17) establish a one-to-one correspondence between the E's given with

σ and the H's having $j\circ\sigma$; the corresponding E and H have the same $\omega(a,b)$. This correspondence preserves equivalence in both directions.

Proof: Since equivalent E and E' have in our case the same σ, Corollary 1 implies that the set $\{\phi(a)|\ a\in G\}$, through which this equivalence is realized [cf. (9)], is a subset of C. This property of the $\phi(a)$'s makes, through Lemmas 1 and 2, the corresponding H and H' equivalent. It is straightforward to invert this argument, because $\{\phi(a)|\ a\in G\}$ are common to E and its corresponding H [see the definition of $\phi(a)$, Eq. (10)]. QED

4. GENERALIZED SEMIDIRECT PRODUCT

We assume that a homomorphism $\sigma: G\to \mathrm{Aut}(K)$ is given, such that $\sigma = k\circ g$. Let (l,H,n) be an arbitrary element from $\mathrm{ext}_{j\circ\sigma}(G,C)$ (see Diag. 2). Obviously, $\tau=\sigma\circ n$ is a homomorphism $H\to\mathrm{Aut}(K)$, and it can be used to define the semidirect product $K\textcircled{T}H$.

Lemma 6: The antidiagonal of $C\otimes l(C)$

$$C'=\{(\gamma,l(\gamma^{-1}))|\gamma\in C\} \tag{18}$$

is an invariant subgroup of $K\textcircled{T}H$.

Proof: The group $K\textcircled{T}H$ can be written as

$$K\textcircled{T}H\cong\{(\alpha,(\gamma,a))|\alpha\in K,\gamma\in C,a\in G\}, \tag{19a}$$

with the composition law

$$(\alpha,(\gamma,a))(\beta,(\delta,b))=(\alpha\sigma[a](\beta),(\gamma\sigma[a](\delta)\omega(a,b),ab)), \tag{19b}$$

because of $\tau[(\gamma,a)]=\sigma[a]$ and $j\circ\sigma[a](\delta)=\sigma[a](\delta)$, $\forall\delta\in C$.

The subset C' now reads

$$C'=\{(\gamma,(\gamma^{-1},e))|\gamma\in C\}. \tag{20}$$

It follows from (19b) that C' is a subgroup of $K\textcircled{T}H$ since $\sigma[e]=I_K$ and $\omega(e,e)=\epsilon$. It is also invariant if, $\forall\delta\in C$, there exists a $\gamma\in C$ such that $(\alpha,(\eta,a))(\delta,(\delta^{-1},e))=(\gamma,(\gamma^{-1},e))(\alpha,(\eta,a))$, for an arbitrary $(\alpha,(\eta,a))\in K\textcircled{T}H$. It is easy to see that such a γ is $\sigma[a](\delta)$. QED

Theorem 1: The factor group

$$E=(K\textcircled{T}H)/C', \tag{21}$$

i.e., the GSP, is an extension (i,E,s) of G by K and belongs to $\mathrm{ext}_\sigma(G,K)$. In (21) the extension $(l,H,n)\in\mathrm{ext}_{j\circ\sigma}(G,C)$, $\tau=\sigma\circ n$, and C' is defined by (18). The isomorphism i and the homomorphism s are given by $i(\alpha)=(\alpha,1)C'$, $s((\alpha,x)C')=n(x)$ (see Diag. 2), where $\alpha\in K$, and $1,x\in H$.

Proof: We are going to demonstrate the existence of an isomorphism between E given by (21) and that defined by (17). This isomorphism immediately implies the statement of the Theorem.

The cosets of C' in $K\textcircled{T}H$ can be conveniently written so that E of (21) takes the form

$$E\cong\{\{(\alpha\gamma,(\gamma^{-1},a))|\gamma\in C\}|\alpha\in K,a\in G\}. \tag{22}$$

To show this, let us take an arbitrary element of E

$$\{(\delta,(\delta^{-1},e))(\beta,(\eta,a))|\delta\in C\}=\{(\delta\beta,(\delta^{-1}\eta,a))|\delta\in C\}$$

[cf. (20) and (21)]. Replacing $\delta^{-1}\eta=\gamma^{-1}$, $\eta\beta=\alpha$, we obtain (22). Clearly, indexing of the classes by (α,a) is unique. Thus a one-to-one correspondence between (21) and (17) is achieved. It is an isomorphism because

$$(\alpha\gamma,(\gamma^{-1},a))(\beta\delta,(\delta^{-1},b))=(\alpha\sigma[a](\beta)\omega(a,b)\eta,(\eta^{-1},ab)), \tag{23}$$

where $\eta=\omega(a,b)^{-1}\sigma[a](\delta)\gamma$. The last claim of the Theorem can be seen immediately either from (22) or from Diagram 2. QED

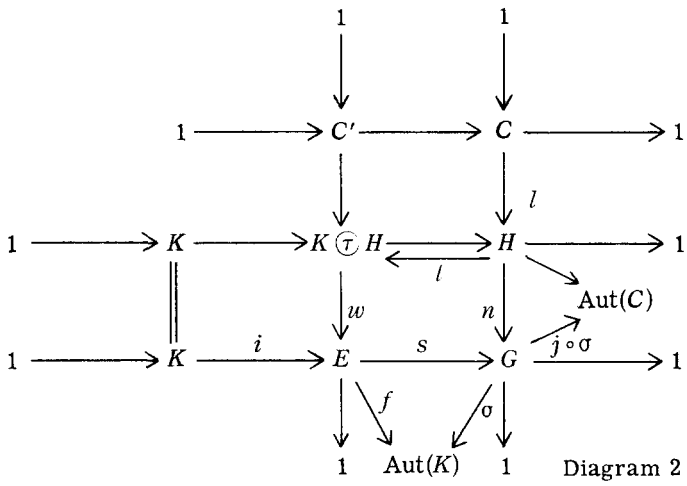


Diagram 2

The considered extensions are related to each other as exhibited on the diagram.

Remark 1: The extensions E and H in (21), when written in the respective isomorphic forms (17) and (16), have a common factor system. This is an immediate consequence of (23).

Corollary 3: If, in the GSP (21), H is itself a semidirect product of C with G (via the homomorphism $j\circ\sigma$), then E is isomorphic to $K\textcircled{@}G$.

Proof: Selecting from each coset in (22) a representative of the form $(\alpha,(\epsilon,a))$, the resulting set is a group isomorphic to $K\textcircled{@}G$. QED

Corollary 4: When $K=C\otimes D$, and D is a subgroup invariant under $\sigma[a]$, $\forall a\in G$, then for every extension H expression (21) takes on the simplified form

$$E\cong D\textcircled{T}H. \tag{24}$$

Proof: In this case the elements of K are uniquely written as $\alpha=\gamma x$, $\gamma\in C, x\in D$. Hence the coset representatives can be chosen in the form $(x,(\gamma,a))$. The set of representatives is a group isomorphic to $D\textcircled{T}H$. QED

Corollary 5: In case when there exists a subgroup C_0 of C which is invariant under $\sigma[a]$, $\forall a\in G$, and when for some extension H one has $\omega(a,b)\in C_0$, $\forall a,b\in G$, then the corresponding E becomes

$$E\cong(K\textcircled{T}H_0)/C_0', \tag{25}$$

where H_0 is a subgroup of H which is, in its turn, the extension of G by C_0 with $j\circ\sigma$ and the same $\omega(a,b)$ and where $C_0' = \{(\lambda, l(\lambda^{-1}))\,|\,\lambda \in C_0\}$.

Proof: Analogous to that of Theorem 1 restricting the elements of C to C_0. QED

It should be noticed that expression (25) may simplify (21) considerably (see Example 1 below).

Theorem 2: Let us single out from each class of equivalent extensions in $\mathrm{ext}_{j\circ\sigma}(G,C)$ one representative H_i in an arbitrary fashion. For each of these we construct the corresponding E_i according to (21). These E_i turn out to be representatives one from every class of equivalent extensions in $\mathrm{ext}_{\sigma}(G,K)$.

Proof: The groups H and E in (21) have their equivalents H' and E' of the form (16) and (17) respectively. The latter have a common factor system (cf. Remark 1), so that Proposition 1 is valid for them. Hence, nonequivalence of $H_i, H_j \in \mathrm{ext}_{j\circ\sigma}(G,C)$ implies nonequivalence of the corresponding $E_i, E_j \in \mathrm{ext}_{\sigma}(G,K)$. The fact that no class of $\mathrm{ext}_{\sigma}(G,K)$ is left out in this way can be established by an ab contrario argument from Proposition 1. QED

5. DISCUSSION

(A) Our extension procedure (GSP) is applicable to most of the groups used in theoretical physics. For instance, here belong all compact Lie groups, all semisimple Lie groups, the Poincaré group, etc. Namely, a sufficient condition for the existence of a homomorphism σ, which is $k\circ g$, is the splitting of $\mathrm{Aut}(K)$:

$$1 \to I(K) \to \mathrm{Aut}(K) \rightleftarrows A(K) \to 1,$$

and G may be arbitrary. All of the mentioned groups satisfy this condition.[3]

(B) The problem of finding all irreducible representations (IR's) of the extension E obtained as a GSP [cf. (21)] can be solved in three steps: (1) One finds all IR's of H. This task is facilitated by the fact that H is an extension of G by an Abelian group, and all the IR's of the latter are one-dimensional. (2) One determines all IR's of the semidirect product $K \textcircled{+} H$.[5,6] (3) One finally selects out the subset of all those IR's of $K \textcircled{+} H$ whose kernels contain C'. The whole procedure is illustrated in Example 2.

(C) The group E given as a generalized direct product (4) (a central extension) can be viewed as the product of two subgroups both of which are invariant, the first being isomorphic to K and the second the centralizer of the first in E and isomorphic to H. Their intersection is the center of the first subgroup.

If E is a GSP, it can be regarded as the product of its invariant subgroup $i(K)$, isomorphic to K, and a second subgroup $w\circ t(H)$, which is isomorphic to H (see Diag. 2). The two subgroups intersect in the center of the first $i(C)$. In general, $w\circ t(H)$ need be neither the centralizer nor invariant in E.

(D) Since in the case of a central extension the elements of $i(K)$ and those of $w\circ t(H)$ commute, the quantum numbers of the IR's of K and H are compatible. In the more general case of a GSP not all elements of $w\circ t(H)$ commute with all elements of $i(K)$, which means that the corresponding quantum numbers are incompatible. For instance, in example (2) below, where the group $\{1, \hat{C}\}$ is extended by $SU(6)/Z_3$, the quantum number of charge conjugation $\hat{C}$ is incompatible with those of this group of internal symmetries. In this particular case this fact has been known for some time,[7] viz., for the groups $SU(n)$, $n > 2$, G-parity cannot be defined.

(E) If the reader is familiar with the application of cohomology theory in group extensions,[2] he may wonder about the relation between the Eilenberg–MacLane theory and the GSP. If in $\mathrm{ext}_g(G,K)$ there exists a semidirect product with $\sigma = k\circ g$, and it is multiplied with all extensions from $\mathrm{ext}_{j\circ\sigma}(G,C)$ in the fashion given in Ref. 2, then an equivalence can be established between the extensions obtained in this way and those given as GSP's.

6. EXAMPLES

(1) As an illustration of Corollary 5, we may take $G = Z_2(\hat{C})$, $\hat{C}$ being the charge conjugation operator, and $K = U(1) \otimes SU(3)/Z_3$ (the eightfold way model of Gell-Mann and Ne'eman). The center of K is isomorphic to $U(1)$, and the automorphism $\sigma[\hat{C}]$ is the complex conjugation of the elements of K. Since $H^2(Z_2(\hat{C}), U(1)) \cong Z_2$, there are only two inequivalent extensions E_1 and E_2. The factor systems of these two extensions have to satisfy (8b), which in this case reduces to $\omega(\hat{C},\hat{C}) = \omega^*(\hat{C},\hat{C})$, implying $\omega(\hat{C},\hat{C}) = \pm 1$ (three-dimensional unity). Evidently in the nontrivial case of E_2 the factor system belongs to a subgroup of the center, $C_0 = \{1, -1\}$, which is invariant under $\sigma[\hat{C}]$. The extension H_0 in (25) is isomorphic to Z_4. Therefore the simplest form of E_2 is

$$E_2 = \{[U(1) \otimes SU(3)/Z_3] \textcircled{+} Z_4\}/C_0', \tag{26}$$

where $\tau = \sigma\circ n$, n being the homomorphism $Z_4 \to Z_2(\hat{C})$.

(2) To illustrate the method of constructing IR's of a GSP [see Discussion (B)], we consider $G = Z_2(\hat{C}) = \{e, \hat{C}\}$ and $K = SU(6)/Z_3$, $\hat{C}$ and $\sigma[\hat{C}]$ as in Example 1. The center of K is now isomorphic to Z_2, so that there are no more than two inequivalent extensions of the center: $H_1 = Z_2 \otimes Z_2(\hat{C})$ and $H_2 = Z_4$. According to (21), the GSP's are: $E_1 = [SU(6)/Z_3] \textcircled{@} Z_2(\hat{C})$ (cf. Corollary 3), and

$$E_2 = B/Z_2', \quad \text{where} \quad B = [SU(6)/Z_3] \textcircled{+} Z_4. \tag{27}$$

As a simple example of an IR for E_2, let us choose the IR $D^{(35)}$ of $SU(6)/Z_3$, which is used to classify mesons. First we find the IR's of B. Since the little group of $D^{(35)}$ is the whole B, and Z_4 is cyclic, the IR's of B are[5,6]

$$\Delta^{(i)}(y) \otimes U(y) D^{(35)}(\alpha), \tag{28}$$

where y runs through Z_4, $\Delta^{(i)}$ are the IR's of $Z_4 = \{1, x, x^2, x^3\}$, and U is a 35-dimensional representation of Z_4, so that

$$D^{(35)}(\alpha^*) = U(x) D^{(35)}(\alpha) U^{-1}(x). \tag{29}$$

Since the kernel of $D^{(35)}$ is just Z_2, Sec. 5(B) implies that those and only those of the IR's given by (28) are

also IR's of E_2 for which Z_2 is contained in the kernels of both $\Delta^{(i)}$ and U. In this way from the four IR's of $\Delta^{(i)}$ only two remain: the identity representation $\Delta^{(1)}$ and the nontrivial real one $\Delta^{(2)}$. Due to $\sigma \circ n[x^2](\alpha) = \sigma[e](\alpha) = \alpha$ and $U(x^2) = e^{i\varphi}U^2(x)$ implied by (29), it is possible to select $U = U_0$ so that $U_0(x^2) = 1$, i.e., the kernel of U_0 is Z_2. Thus, all the IR's of E_2 associated with the IR $D^{(35)}$ are

$$\Delta^{(k)}(z) \otimes U_0(z) D^{(35)}(\alpha), \quad k = 1, 2, \tag{30}$$

where $z = n(y)$, i.e., $z = e, \hat{C}$ [n being the homomorphism: $Z_4 \to Z_2(\hat{C})$].

[1] L. Michel, "Invariance in Quantum Mechanics and Group Extensions," in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).
[2] S. Eilenberg and S. MacLane, Ann. Math. **48**, 326 (1947).
[3] L. Michel, "Relativistic Invariance," in *Particle Symmetries and Axiomatic Field Theory*, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).
[4] A. G. Kurosh, *The Theory of Groups* (Chelsea, New York, 1955), Chap. XII.
[5] A. H. Clifford, Ann. Math. **38**, 533 (1937).
[6] L. Jansen and M. Boon, *Theory of Finite Groups, Applications in Physics* (North-Holland, Amsterdam, 1967).
[7] K. Tanabe and K. Shima, J. Math. Phys. **8**, 657 (1967).

Relation between the Statistical Representations of Real and Associated Complex Fields in Optical Coherence Theory*

G. S. Agarwal† and E. Wolf
Department of Physics and Astronomy, University of Rochester, Rochester, New York 11627

(Received 22 December 1971; Revised Manuscript Received 30 May 1972)

In the general theory of optical coherence, the following problem discussed in the present paper, arises: to determine the statistical properties of a field represented by an analytic signal from the knowledge of the statistical properties of the corresponding real field. It is shown by the use of the characteristic functionals that in order to determine the joint probability distributions of the complex field at N space–time points, the knowledge of the complete statistical description of the real field is required; on the other hand, the moments of the complex field up to that order can be determined from the knowledge of the moments of the real field up to the same order. The results are illustrated by explicit calculations relating to the Gaussian random process, which, as is well known, characterizes the fluctuations of thermal light. A converse of a well-known theorem of Kac and Siegert relating to a Gaussian random process is derived as an immediate consequence of our analysis.

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

The classical theory of optical coherence of statistical fields, whether the field is stationary or nonstationary, whether it is generated by a thermal source or some other source is today well formulated[1,2] and is being applied to a wider and wider class of optical problems. There is, however, one problem in the foundation of the theory, which has so far not been treated. In this paper we will formulate this problem and present a solution of it for a wide class of nonstationary processes.

Let us denote by $X(\mathbf{r}, t)$ a real field variable, representing the optical field at a point $\mathbf{r}$, at time t. For the sake of simplicity we consider X to be a scalar, e.g., a Cartesian component of the electric field. (Generalization to a vector field is straightforward.) For any realistic field, $X(\mathbf{r}, t)$ will fluctuate in the course of time in a manner that is not strictly predictable. It is, therefore, appropriate to regard X as a member of an ensemble of different realizations of the field. The statistical properties of the field may then be specified by a sequence of probability densities

$$p_1^{(X)}(X_1; R_1),\ p_2^{(X)}(X_1, X_2; R_1, R_2),\ p_3^{(X)}(X_1, X_2, X_3; R_1, R_2, R_3), \ldots, \tag{1.1}$$

where $R_j \equiv \mathbf{r}_j, t_j$ denotes a typical space–time point. To illustrate the meaning of these probability densities let us consider $p_2^{(X)}$: The quantity $p_2^{(X)}(X_1, X_2; R_1, R_2)dX_1dX_2$ denotes the probability that at the space–time points R_1 and R_2, X will take on values that are in the intervals $X_1, X_1 + dX_1$ and X_2 and $X_2 + dX_2$, respectively.

It is customary and useful to introduce a complex representation of the field.[3] This representation, originally due to Gabor,[4] arises naturally in the theory of photo-electric detection of light fluctuations. The associated complex field also corresponds to the eigenvalues of an operator used in the theory of quantized fields, to represent the annihilation of a photon[2] at a space–time point R. Some minimal properties of this complex representation have been established by Mandel.[5]

To introduce this complex representation, we assume, that for each $\mathbf{r}$, the typical realization $X(\mathbf{r}, t)$ is square integrable, with respect to time[6] and, hence, may be represented as a Fourier integral

$$X(\mathbf{r}, t) = \int_{-\infty}^{+\infty} \xi(\mathbf{r}, \nu) \exp(-2\pi i\nu t)\, d\nu. \tag{1.2}$$

Since $X(\mathbf{r}, t)$ is real, $\xi(r, -\nu) = \xi^*(r, \nu)$, where the asterisk denotes the complex conjugate. Because of this relation it is clear that the negative frequency components of each realization of the field do not contain any information that is not contained in the positive ones; hence, in place of the real function $X(\mathbf{r}, t)$ we may employ the complex function

$$Z(\mathbf{r}, t) = 2\int_0^{\infty} \xi(\mathbf{r}, \nu) \exp(-2\pi i\nu t)\, d\nu, \tag{1.3}$$

known as the *complex analytic signal*, associated with $X(\mathbf{r}, t)$. This terminology arises from the fact that by a well-known theorem[7] $Z(\mathbf{r}, t)$ considered as a function of a complex t, is regular and analytic in the lower half of the complex t plane. It is trivial to show that the real part of $Z(\mathbf{r}, t)$ is precisely the function $X(\mathbf{r}, t)$ and, on using the analytic property of Z that we just mentioned, one may readily show that

also IR's of E_2 for which Z_2 is contained in the kernels of both $\Delta^{(i)}$ and U. In this way from the four IR's of $\Delta^{(i)}$ only two remain: the identity representation $\Delta^{(1)}$ and the nontrivial real one $\Delta^{(2)}$. Due to $\sigma \circ n[x^2](\alpha) = \sigma[e](\alpha) = \alpha$ and $U(x^2) = e^{i\varphi}U^2(x)$ implied by (29), it is possible to select $U = U_0$ so that $U_0(x^2) = 1$, i.e., the kernel of U_0 is Z_2. Thus, all the IR's of E_2 associated with the IR $D^{(35)}$ are

$$\Delta^{(k)}(z) \otimes U_0(z) D^{(35)}(\alpha), \quad k = 1, 2, \tag{30}$$

where $z = n(y)$, i.e., $z = e, \hat{C}$ [n being the homomorphism: $Z_4 \to Z_2(\hat{C})$].

[1] L. Michel, "Invariance in Quantum Mechanics and Group Extensions," in *Group Theoretical Concepts and Methods in Elementary Particle Physics*, edited by F. Gürsey (Gordon and Breach, New York, 1964).
[2] S. Eilenberg and S. MacLane, Ann. Math. **48**, 326 (1947).
[3] L. Michel, "Relativistic Invariance," in *Particle Symmetries and Axiomatic Field Theory*, edited by M. Chretien and S. Deser (Gordon and Breach, New York, 1966).
[4] A. G. Kurosh, *The Theory of Groups* (Chelsea, New York, 1955), Chap. XII.
[5] A. H. Clifford, Ann. Math. **38**, 533 (1937).
[6] L. Jansen and M. Boon, *Theory of Finite Groups, Applications in Physics* (North-Holland, Amsterdam, 1967).
[7] K. Tanabe and K. Shima, J. Math. Phys. **8**, 657 (1967).

Relation between the Statistical Representations of Real and Associated Complex Fields in Optical Coherence Theory*

G. S. Agarwal† and E. Wolf
Department of Physics and Astronomy, University of Rochester, Rochester, New York 11627

(Received 22 December 1971; Revised Manuscript Received 30 May 1972)

In the general theory of optical coherence, the following problem discussed in the present paper, arises: to determine the statistical properties of a field represented by an analytic signal from the knowledge of the statistical properties of the corresponding real field. It is shown by the use of the characteristic functionals that in order to determine the joint probability distributions of the complex field at N space–time points, the knowledge of the complete statistical description of the real field is required; on the other hand, the moments of the complex field up to that order can be determined from the knowledge of the moments of the real field up to the same order. The results are illustrated by explicit calculations relating to the Gaussian random process, which, as is well known, characterizes the fluctuations of thermal light. A converse of a well-known theorem of Kac and Siegert relating to a Gaussian random process is derived as an immediate consequence of our analysis.

1. INTRODUCTION AND FORMULATION OF THE PROBLEM

The classical theory of optical coherence of statistical fields, whether the field is stationary or nonstationary, whether it is generated by a thermal source or some other source is today well formulated[1,2] and is being applied to a wider and wider class of optical problems. There is, however, one problem in the foundation of the theory, which has so far not been treated. In this paper we will formulate this problem and present a solution of it for a wide class of nonstationary processes.

Let us denote by $X(\mathbf{r}, t)$ a real field variable, representing the optical field at a point $\mathbf{r}$, at time t. For the sake of simplicity we consider X to be a scalar, e.g., a Cartesian component of the electric field. (Generalization to a vector field is straightforward.) For any realistic field, $X(\mathbf{r}, t)$ will fluctuate in the course of time in a manner that is not strictly predictable. It is, therefore, appropriate to regard X as a member of an ensemble of different realizations of the field. The statistical properties of the field may then be specified by a sequence of probability densities

$$p_1^{(X)}(X_1; R_1),\ p_2^{(X)}(X_1, X_2; R_1, R_2),\ p_3^{(X)}(X_1, X_2, X_3; R_1, R_2, R_3), \ldots, \tag{1.1}$$

where $R_j \equiv \mathbf{r}_j, t_j$ denotes a typical space–time point. To illustrate the meaning of these probability densities let us consider $p_2^{(X)}$: The quantity $p_2^{(X)}(X_1, X_2; R_1, R_2)dX_1dX_2$ denotes the probability that at the space–time points R_1 and R_2, X will take on values that are in the intervals $X_1, X_1 + dX_1$ and X_2 and $X_2 + dX_2$, respectively.

It is customary and useful to introduce a complex representation of the field.[3] This representation, originally due to Gabor,[4] arises naturally in the theory of photo-electric detection of light fluctuations. The associated complex field also corresponds to the eigenvalues of an operator used in the theory of quantized fields, to represent the annihilation of a photon[2] at a space–time point R. Some minimal properties of this complex representation have been established by Mandel.[5]

To introduce this complex representation, we assume, that for each $\mathbf{r}$, the typical realization $X(\mathbf{r}, t)$ is square integrable, with respect to time[6] and, hence, may be represented as a Fourier integral

$$X(\mathbf{r}, t) = \int_{-\infty}^{+\infty} \xi(\mathbf{r}, \nu) \exp(-2\pi i\nu t)\, d\nu. \tag{1.2}$$

Since $X(\mathbf{r}, t)$ is real, $\xi(r, -\nu) = \xi^*(r, \nu)$, where the asterisk denotes the complex conjugate. Because of this relation it is clear that the negative frequency components of each realization of the field do not contain any information that is not contained in the positive ones; hence, in place of the real function $X(\mathbf{r}, t)$ we may employ the complex function

$$Z(\mathbf{r}, t) = 2\int_0^{\infty} \xi(\mathbf{r}, \nu) \exp(-2\pi i\nu t)\, d\nu, \tag{1.3}$$

known as the *complex analytic signal*, associated with $X(\mathbf{r}, t)$. This terminology arises from the fact that by a well-known theorem[7] $Z(\mathbf{r}, t)$ considered as a function of a complex t, is regular and analytic in the lower half of the complex t plane. It is trivial to show that the real part of $Z(\mathbf{r}, t)$ is precisely the function $X(\mathbf{r}, t)$ and, on using the analytic property of Z that we just mentioned, one may readily show that

the real and imaginary parts of Z form a Hilbert transform pair, i.e.,

$$Z(\mathbf{r},t) = X(\mathbf{r},t) + iY(\mathbf{r},t), \tag{1.4}$$

with

$$Y(\mathbf{r},t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{X(\mathbf{r},t')\,dt'}{(t'-t)}, \tag{1.5a}$$

$$X(\mathbf{r},t) = -\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{Y(\mathbf{r},t')\,dt'}{(t'-t)}, \tag{1.5b}$$

where the strokes on the integral signs indicate that one takes the Cauchy principal value of the integrals at $t' = t$.

The statistical properties of the complex field $Z(\mathbf{r},t)$ may also be specified by an infinite sequence of probability densities

$$p_1^{(Z)}(Z_1;R_1), p_2^{(Z)}(Z_1,Z_2;R_1,R_2), \\ p_3^{(Z)}(Z_1,Z_2,Z_3;R_1,R_2,R_3), \cdots. \tag{1.6}$$

The quantity $p_2^{(Z)}(Z_1,Z_2;R_1,R_2)$, for example, has the following meaning: If d^2Z_k denotes the product $dX_k dY_k$, then $p_2^{(Z)}(Z_1,Z_2;R_1,R_2)\,d^2Z_1 d^2Z_2$ is the joint probability that at the space–time point R_1, the real and imaginary parts of Z will lie in the intervals $X_1, X_1 + dX_1$ and $Y_1, Y_1 + dY_1$, and at the space–time point R_2, they will lie in the intervals $X_2, X_2 + dX_2$ and $Y_2, Y_2 + dY_2$, respectively.

Although Eq. (1.4), together with Eq. (1.5a), shows how to calculate the complex analytic signal $Z(\mathbf{r},t)$ that is associated with any particular realization $X(\mathbf{r},t)$ of the real field, it is not clear how one determines the statistical properties of the ensemble of the Z's from the knowledge of the statistical properties of the ensemble of the X's. The present paper is concerned with this question. More precisely we will show how one may determine the sequence (1.6) of the probability densities $\{p_n^{(Z)}\}$ from the knowledge of the sequence (1.1) of the probability densities $\{p_n^{(X)}\}$. The problem is not a trivial one, since according to Eqs. (1.4) and (1.5) the relation between the real and the complex fields is nonlocal in time. In solving this problem, it will be useful to work explicitly with the characteristic functionals[8] rather than with the infinite sequence of probability densities. We will, therefore, reformulate the problem in terms of such functionals.

2. REFORMULATION OF THE PROBLEM IN TERMS OF CHARACTERISTIC FUNCTIONALS

Let $g = g(\mathbf{r},t)$ be an arbitrary real function. The characteristic functional of the real field $X(R) \equiv X(\mathbf{r},t)$ is then defined by the formulas

$$C^{(X)}[g(\cdot)] = \left\langle \exp\left\{ i \int g(R)X(R)\,d^4R \right\} \right\rangle \tag{2.1a}$$

$$= \sum_{n=0}^{\infty} \frac{(i)^n}{n!} \left\langle \left\{ \int g(R)X(R)\,d^4R \right\}^n \right\rangle. \tag{2.1b}$$

In Eqs. (2.1a) and (2.1b) the sharp brackets denote statistical average. From this functional one obtains the Nth order characteristic function $C_N^{(X)}$ and the Nth order probability density of the real field in the usual way. One chooses

$$g(R) = \sum_{n=1}^{N} \alpha_n \delta^{(4)}(R - R_n), \tag{2.2}$$

where the α_n's are real parameters, the R_n's are arbitrary space–time points, and $\delta^{(4)}$ is the four-dimensional Dirac delta function. Then

$$C^{(X)}[g(\cdot)] \to C_N^{(X)}(\alpha_1,\alpha_2,\ldots,\alpha_n;R_1,R_2,\ldots,R_n) \\ = \left\langle \exp\left(i \sum_{n=0}^{N} \alpha_n X(R_n) \right) \right\rangle \\ = \int \cdots \int \exp\left(i \sum_{n=1}^{N} \alpha_n X_n \right) \\ \times p_n^{(X)}(X_1,X_2;\ldots,X_N;R_1,R_2,\ldots,R_N) \\ \times dX_1 dX_2 \cdots dX_N, \tag{2.3}$$

and on Fourier inversion,

$$p^{(X)}(X_1,X_2,\ldots,X_N;R_1,R_2,\ldots,R_N) \\ = \frac{1}{(2\pi)^N} \int \cdots \int C_N^{(X)}(\alpha_1,\alpha_2,\ldots,\alpha_N; \\ R_1,R_2,\ldots,R_N) \exp\left(-i \sum_{n=1}^{N} \alpha_n X_n \right) \\ \times d\alpha_1 d\alpha_2 \cdots d\alpha_N. \tag{2.4}$$

Each of the integrals on the rhs of Eqs. (2.3) and (2.4) extends from $-\infty$ to $+\infty$.

In a similar way we may define the characteristic functional of the complex field $Z(R) \equiv Z(\mathbf{r},t)$ as

$$C^{(Z)}[h(\cdot)] = \left\langle \exp\left\{ i \int [h^*(R)Z(R) + h(R)Z^*(R)]\,d^4R \right\} \right\rangle, \tag{2.5}$$

where $h(R) \equiv h(\mathbf{r},t)$ is an arbitrary, generally complex function of $\mathbf{r}$ and t. The Nth order characteristic function and the Nth order probability density of the complex field Z is then formally obtained from (2.5) by setting

$$h(R) = \sum_{n=1}^{N} \beta_n \delta^{(4)}(R - R_n), \tag{2.6}$$

where the β_n's are complex parameters. Then

$$C^{(Z)}[h(\cdot)] \to C_N^{(Z)}(\beta_1,\beta_2,\ldots,\beta_N;R_1,R_2,\ldots,R_N) \\ = \left\langle \exp\left(i \sum_{n=1}^{N} [\beta_n^* Z(R_n) + \beta_n Z^*(R_n)] \right) \right\rangle \\ = \int \cdots \int \exp\left(i \sum_{n=1}^{N} [\beta_n^* Z_n + \beta_n Z_n^*] \right) \\ \times p_N^{(Z)}(Z_1,Z_2,\ldots,Z_N;R_1,R_2,\ldots,R_N) \\ \times d^2Z_1 \cdots d^2Z_n, \tag{2.7}$$

and, on Fourier inversion

$$p_N^{(Z)}(Z_1,Z_2,\ldots,Z_N;R_1,R_2,\ldots,R_N) \\ = \frac{1}{\pi^{2N}} \int \cdots \int C_N^{(Z)}(\beta_1,\beta_2,\ldots,\beta_n;R_1,R_2,\ldots,R_n) \\ \times \exp\left(-i \sum_{n=1}^{N} [\beta_n^* Z_n + \beta_n Z_n^*] \right) d^2\beta_1 \cdots d^2\beta_N. \tag{2.8}$$

Each of the integrals on the rhs of Eqs. (2.7) and (2.8) extends over a complete complex β_n plane.

The correlation functions (moments) of the fields may be obtained from the characteristic functionals in the usual way by functional differentiation. Thus

$$\langle X(R_1) \cdots X(R_n) \rangle = \left. \frac{(-i)^n \delta^n C^{(X)}[g(\cdot)]}{\delta g(R_1) \cdots \delta g(R_n)} \right|_{g=0}, \tag{2.9}$$

$$\langle Z^*(R_1)\cdots Z^*(R_n)Z(R_{n+1})\cdots Z(R_{n+m})\rangle = \frac{(-i)^{n+m}\,\delta^{n+m}C^{(Z)}[h(\cdot)]}{\delta h(R_1)\cdots\delta h(R_n)\,\delta h^*(R_{n+1})\cdots\delta h^*(R_{n+m})}\Bigg|_{h=0}. \tag{2.10}$$

It is clear now that the problem posed in Sec. 1 is equivalent to the following one: To determine the characteristic functional $C^{(Z)}[h(\cdot)]$ of the complex field $Z(R)$ from the knowledge of the characteristic functional $C^{(X)}[g(\cdot)]$ of the real field $X(R)$.

3. FORMAL SOLUTION OF THE PROBLEM

As preliminary to solving the general problem that we just posed, let us consider first the more restricted one, of determining the statistical behavior of the complex field variable $Z(\mathbf{r}, t)$ at some fixed point $\mathbf{r}$ in space from the knowledge of the statistical properties of the real field variable $X(\mathbf{r}, t)$, at the same point. Since the fixed point $\mathbf{r}$ plays no essential role in our analysis, we will suppress the explicit dependence on $\mathbf{r}$, i.e., we will write $X(t)$ in place of $X(\mathbf{r}, t)$, etc.

We recall that we assumed that each realization of the real field is square integrable with respect to time. Under these circumstances the random variable $X(t)$ (which, for the sake of simplicity we assume to have zero mean value), may be expressed in the form of the generalized Karhunen–Loeve expansion,[9] valid for all $t(-\infty < t < \infty)$,

$$X(t) = \sum_{n=1}^{\infty} c_n \phi_n(t), \tag{3.1}$$

where the $\phi_n(t)$ form an orthonormal set

$$\int_{-\infty}^{\infty} \phi_n(t)\,\phi_m(t)\,dt = \delta_{n,m}, \tag{3.2}$$

$\delta_{n,m}$ being the Kronecker symbol, and the coefficients c_n are uncorrelated real variables

$$\langle c_n c_m\rangle = \lambda_n \delta_{n,m}, \tag{3.3}$$

with the λ_n being real and nonnegative. As is well known[9] the ϕ_n and the λ_n are the eigenvalues and the eigenfunctions of the integral equation

$$\int_{-\infty}^{\infty} R(t_1, t_2)\,\phi_n(t_2)\,dt_2 = \lambda_n \phi_n(t_1), \tag{3.4}$$

with the kernel

$$R(t_1, t_2) = \langle X(t_1)X(t_2)\rangle. \tag{3.5}$$

According to (3.1) and (2.1a), the characteristic functional of the real random field variable $X(t)$, which we will denote by $C_1^{(X)}[g(\cdot)]$ (subscript 1 indicating that we now deal with behavior at *one* fixed point in space) may be expressed in the form

$$C_1^{(X)}[g(\cdot)] = \left\langle \exp\left(i \sum_{n=1}^{\infty} c_n u_n\right)\right\rangle, \tag{3.6}$$

where

$$u_n = \int_{-\infty}^{+\infty} g(t)\,\phi_n(t)\,dt. \tag{3.7}$$

Let us now turn to the "conjugate" field variable $Y(t)$ [Eq. (1.5a)]. If we substitute in Eq. (1.5a) for $X(t)$ the orthogonal expansion (3.1), we may express $Y(t)$ in the form

$$Y(t) = \sum_{n=1}^{\infty} c_n \psi_n(t), \tag{3.8}$$

where

$$\psi_n(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\phi_n(t')\,dt'}{(t' - t)}. \tag{3.9}$$

One can readily show that the ψ_n, just like the ϕ_n, form an orthonormal set.[10] In fact, Eq. (3.8) is nothing but the generalized Karhunen–Loeve orthogonal expansion of the random field variable $Y(t)$. It follows on using (3.8), that the characteristic functional $Y(t)$ may be expressed in a form strictly analogous to (3.6):

$$C_1^{(Y)}[g(\cdot)] = \left\langle \exp\left[i \int g(t)Y(t)\,dt\right]\right\rangle \tag{3.10}$$

$$= \left\langle \exp\left(i \sum_{n=1}^{\infty} c_n v_n\right)\right\rangle, \tag{3.11}$$

where

$$v_n = \int_{-\infty}^{+\infty} g(t)\psi_n(t)\,dt. \tag{3.12}$$

The average in (3.10) is defined in a manner similar to that employed in connection with Eq. (2.1a).

If we compare (3.11) with (3.6) we see that the right-hand sides are of the same functional form, except that the u_n have been replaced by the v_n's. Thus if we express $C_1^{(X)}[g(\cdot)]$ in the form

$$C_1^{(X)}[g(\cdot)] = f(u_1, u_2, \ldots, u_n, \cdots), \tag{3.13a}$$

then

$$C_1^{(Y)}[g(\cdot)] = f(v_1, v_2, \ldots, v_n, \cdots), \tag{3.13b}$$

i.e., *the characteristic functional of the conjugate random variable $Y(t)$ may be obtained from the characteristic functional of the original random variable $X(t)$ by simply replacing all the u_n's by the v_n's*, where the u_n's are the projections of $g(t)$ onto the set $\{\phi_n\}$ and the v_n's are the projections of $g(t)$ onto the set $\{\psi_n\}$ [cf. Eqs. (3.7) and (3.12)].

Next let us consider the complex random field variable $Z(t) = X(t) + iY(t)$. Using the expansions (3.1) and (3.8) it follows that $Z(t)$ may be expressed in the form

$$Z(t) = \sum_{n=1}^{\infty} c_n \chi_n(t), \tag{3.14}$$

where

$$\chi_n(t) = \phi_n(t) + i\psi_n(t). \tag{3.15}$$

From (3.15) and (2.5) we then obtain the following expression for the characteristic functional of Z:

$$C_1^{(Z)}[h(\cdot)] = \left\langle \exp\left(i \sum_{n=1}^{\infty} c_n w_n\right)\right\rangle, \tag{3.16}$$

where

$$w_n = \int_{-\infty}^{+\infty} [h^*(t)\chi_n(t) + h(t)\chi_n^*(t)]\,dt. \tag{3.17}$$

If again we express $C_1^{(X)}[g(\cdot)]$ in the form (3.13a) we see, on comparing (3.16) with (3.6), that $C_1^{(Z)}[h(\cdot)]$ may be expressed in the form

$$C_1^{(Z)}[h(\cdot)] = f(w_1, w_2, \ldots, w_n, \cdots), \tag{3.18}$$

i.e., the *characteristic functional of the complex random field variable $Z(t)$ may be obtained from the characteristic functional of the real random field*

variable X(t) by replacing the u_n by the w_n. In view of the remarks follows Eqs. (3.10) this result provides a formal solution to our problem, for the special case when only one spatial point is considered.

For the purpose of later discussion, we will also write down an expression for the joint probability $p(\{c_n\}) \equiv p(c_1, c_2, \ldots, c_n \cdots)$ of all the coefficients c_n occurring in the generalized Karhunen–Loeve expansion (3.1). It is clear that the characteristic functional $C_1^{(X)}[g(\cdot)]$, given by (3.6), may also be expressed in the form

$$C_1^{(X)}[g(\cdot)] = \int \cdots \int \exp\left(i \sum_{n=1}^{\infty} c_n u_n\right) p(\{c_n\})\, d(\{c_n\}). \tag{3.19}$$

The integral on the right-hand side is taken over all the c_n ($-\infty \leq c_n \leq \infty$) and is, therefore, infinite dimensional. Formal Fourier inversion of (3.19) gives the required joint probability $p(\{c_n\})$ in terms of the characteristic functional $C_1^{(X)}[g(\cdot)]$:

$$p(\{c_n\}) = \int \cdots \int \prod_{n=1}^{\infty} [(1/2\pi) e^{-ic_n u_n}] C_1^{(X)}[g(\cdot)]\, d(\{u_n\}). \tag{3.20}$$

Returning to expressions (3.13a) and (3.18) we see that the function $f(w_1, w_2, \ldots, w_n, \cdots)$ is specified by the complete statistical behavior of the real field variable $X(t)$. *Hence in order to determine from our formulas the joint probability density $p_N^{(Z)}$ of the complex variable Z(t) at N instants of time, one must know the statistical behavior of the real field variable X(t) for all times.* On the other hand, on using the Hilbert transform relations between the real and imaginary parts of the complex field variable [Eqs. (1.5)], one can readily see that in order to determine an Nth order correlation function (which may be an equal-time correlation function) of the complex field variable, one only needs to know the Nth order correlation function, (which, in general, is a multitime correlation function), of the real field variable. To illustrate this, consider the second-order correlation function $\langle Z(t_1) Z^*(t_2)\rangle$. We have

$$\begin{aligned}\langle Z(t_1)Z^*(t_2)\rangle &= \langle [X(t_1) + iY(t_1)][X(t_2) - iY(t_2)]\rangle \\ &= \langle X(t_1)X(t_2)\rangle + \langle Y(t_1)Y(t_2)\rangle - i\langle X(t_1)Y(t_2)\rangle \\ &\quad + i\langle Y(t_1)X(t_2)\rangle . \end{aligned} \tag{3.21}$$

In view of (1.5) one evidently has

$$\begin{aligned}\langle Y(t_1)Y(t_2)\rangle &= \frac{1}{\pi^2} \fint_{-\infty}^{+\infty} dt_1' \fint_{-\infty}^{+\infty} dt_2' \frac{\langle X(t_1')X(t_2')\rangle}{(t_1' - t_1)(t_2' - t_2)}, \\ \langle X(t_1)Y(t_2)\rangle &= \frac{1}{\pi} \fint_{-\infty}^{+\infty} dt_2' \frac{\langle X(t_1)X(t_2')\rangle}{(t_2' - t_2)}, \\ \langle Y(t_1)X(t_2)\rangle &= \frac{1}{\pi} \fint_{-\infty}^{+\infty} dt_1' \frac{\langle X(t_1')X(t_2)\rangle}{(t_1' - t_1)}. \end{aligned} \tag{3.22}$$

Equation (3.21), together with the relations (3.22), expresses the second-order correlation function $\langle Z(t_1)Z^*(t_2)\rangle$ in terms of the second-order correlation function $\langle X(t_1)X(t_2)\rangle$. Of course to determine $\langle Z(t_1)Z^*(t_2)\rangle$ for a fixed pair of values t_1 and t_2, one must know $\langle X(t_1')X(t_2')\rangle$ for all values of its arguments t_1' and t_2'.

Up to now we have confined our attention to the behavior of the field at one space point only. Generalization to the complete field is straightforward. In place of Eq. (3.1) we now have the expansion

$$X(\mathbf{r}, t) = \sum_{n=1}^{\infty} c_n(\mathbf{r})\, \phi_n(\mathbf{r}, t), \tag{3.23}$$

where the $\phi_n(\mathbf{r}, t)$ are solutions of the integral equation

$$\int_{-\infty}^{+\infty} R(\mathbf{r}, t_1; \mathbf{r}, t_2)\, \phi_n(\mathbf{r}, t_2)\, dt_2 = \lambda_n(\mathbf{r})\, \phi_n(\mathbf{r}, t_1), \tag{3.24}$$

with the kernel

$$R(\mathbf{r}, t_1; \mathbf{r}, t_2) = \langle X(\mathbf{r}, t_1) X(\mathbf{r}, t_2)\rangle . \tag{3.25}$$

The functions $\phi_n(\mathbf{r}, t)$ and $c_n(\mathbf{r}, t)$ satisfy, for each $\mathbf{r}$, orthogonality relations analogous to (3.2) and (3.3). The characteristic functional $C^{(X)}[g(\cdot)]$ of the real field $X(\mathbf{r}, t)$ defined by Eq. (2.1) may, by analogy with Eq. (3.13a), be expressed in the form

$$C^{(X)}[g(\cdot)] = f[u_1(\cdot), \ldots, u_n(\cdot), \cdots], \tag{3.26}$$

where

$$u_n(\cdot) = \int \cdots \int g(\mathbf{r}, t)\, \phi_n(\mathbf{r}, t)\, d^3r dt. \tag{3.27}$$

The characteristic functional for the associated complex field $Z(\mathbf{r}, t)$ may then be expressed in the form

$$C^{(Z)}[h(\cdot)] = f[w_1(\cdot), \ldots, w_n(\cdot), \cdots], \tag{3.28}$$

where

$$w_n(\cdot) = \int \cdots \int [h^*(\mathbf{r}, t)\, \chi_n(\mathbf{r}, t) + h(\mathbf{r}, t)\, \chi_n^*(\mathbf{r}, t)]\, d^3r dt, \tag{3.29}$$

and the χ_n are defined, for each $\mathbf{r}$, in a similar way as before.

4. EXAMPLE: GAUSSIAN RANDOM PROCESS

To illustrate our results we will now determine the characteristic functional $C_1^{(Z)}[h(\cdot)]$ and the probability density $p_1^{(Z)}[\mathbf{r}, t]$ for the case when the real field variable $X(t)$ represents a Gaussian random process with zero mean.

The characteristic functional for a real Gaussian random process with zero mean may be expressed in the form (see Appendix)

$$C_1^{(X)}[g(\cdot)] = \exp\left(-\tfrac{1}{2} \iint_{-\infty}^{+\infty} g(t_1) g(t_2) \times \langle X(t_1)X(t_2)\rangle\, dt_1 dt_2\right). \tag{4.1}$$

Now according to (3.1) and (3.5), the correlation function $\langle X(t_1)X(t_2)\rangle$ may be expressed in the form

$$\begin{aligned}\langle X(t_1)X(t_2)\rangle &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \langle c_n c_m\rangle\, \phi_n(t_1)\, \phi_m(t_2) \\ &= \sum_{n=1}^{\infty} \lambda_n \phi_n(t_1)\, \phi_n(t_2). \end{aligned} \tag{4.2}$$

If we substitute from Eq. (4.2) into Eq. (4.1) and recall the definitions of u_n, given by (3.7), the characteristic functional $C_1^{(X)}$ may be expressed in the simple form

$$C_1^{(X)}[g(\cdot)] = \exp\left(-\tfrac{1}{2} \sum_{n=1}^{\infty} \lambda_n u_n^2\right). \tag{4.3}$$

According to the theorem following Eqs. (3.13) and (3.18), the characteristic functionals $C_1^{(Y)}[g(\cdot)]$ and

$C_1^{(Z)}[h(\cdot)]$ are obtained at once from (4.3) on replacing the u_n by the v_n and w_n [defined by Eqs. (3.12) and (3.17)], respectively,

$$C_1^{(Y)}[g(\cdot)] = \exp\left(-\tfrac{1}{2}\sum_{n=1}^{\infty}\lambda_n v_n^2\right), \tag{4.4}$$

$$C_1^{(Z)}[g(\cdot)] = \exp\left(-\tfrac{1}{2}\sum_{n=1}^{\infty}\lambda_n w_n^2\right). \tag{4.5}$$

We may readily express (4.4) and (4.5) in forms analogous to (4.1). We have from (3.12), (3.8), and (3.3)

$$\begin{aligned}\sum_{n=1}^{\infty}\lambda_n v_n^2 &= \sum_{n=1}^{\infty}\lambda_n \iint_{-\infty}^{+\infty} g(t_1)g(t_2)\psi_n(t_1)\psi_n(t_2)\,dt_1dt_2\\ &= \iint_{-\infty}^{+\infty} g(t_1)g(t_2)\sum_{n=1}^{\infty}\lambda_n\psi_n(t_1)\psi_n(t_2)\,dt_1dt_2\\ &= \iint_{-\infty}^{+\infty} g(t_1)g(t_2)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\langle c_n\psi_n(t_1)c_m\psi_m(t_2)\rangle\\ &\quad\times dt_1dt_2\\ &= \iint_{-\infty}^{+\infty} g(t_1)g(t_2)\langle Y(t_1)Y(t_2)\rangle\,dt_1dt_2.\end{aligned} \tag{4.6}$$

Similarly, we have from (3.17), (3.14), and (3.3)

$$\begin{aligned}\sum_{n=1}^{\infty}\lambda_n w_n^2 &= \sum_{n=1}^{\infty}\lambda_n\iint_{-\infty}^{+\infty}[h(t_1)h(t_2)\chi_n^*(t_1)\chi_n^*(t_2)\\ &\quad + h(t_1)h^*(t_2)\chi_n^*(t_1)\chi_n(t_2) + \text{c.c.}]\,dt_1dt_2\\ &= \iint_{-\infty}^{+\infty} h(t_1)h(t_2)\langle Z^*(t_1)Z^*(t_2)\rangle\,dt_1dt_2\\ &\quad + \iint_{-\infty}^{+\infty} h(t_1)h^*(t_2)\langle Z^*(t_1)Z(t_2)\rangle\\ &\quad\times dt_1dt_2 + \text{c.c.},\end{aligned} \tag{4.7}$$

where c.c. denotes the complex conjugate. On substituting from Eqs. (4.6) and (4.7) into Eqs. (4.4) and (4.5), respectively, we obtain the following expressions for the characteristic functionals $C_1^{(Y)}$ and $C_1^{(Z)}$:

$$C_1^{(Y)}[g(\cdot)] = \exp\left(-\tfrac{1}{2}\iint_{-\infty}^{+\infty} g(t_1)g(t_2)\langle Y(t_1)Y(t_2)\rangle \times dt_1dt_2\right), \tag{4.8}$$

$$\begin{aligned}C_1^{(Z)}[h(\cdot)] = \exp\Big(-\tfrac{1}{2}\iint_{-\infty}^{+\infty}&[h(t_1)h(t_2)\langle Z^*(t_1)Z^*(t_2)\rangle\\ &+ h(t_1)h^*(t_2)\langle Z^*(t_1)Z(t_2)\rangle + \text{c.c.}]\,dt_1dt_2\Big).\end{aligned} \tag{4.9}$$

Equation (4.8) implies that the conjugate process $Y(t)$ is also a real Gaussian process with zero mean; and by a similar argument as given in the Appendix in connection with the real Gaussian process, one can show that (4.9) implies that the process $Z(t)$ is a complex Gaussian random process with zero mean. The correlation functions occurring in (4.8) and (4.9) may, of course, be expressed in terms of the correlation functions of the real process $X(t)$, with the help of relations of the form (3.21) and (3.22).

Next let us determine the first probability density $p_1^{(Z)}(Z,t)$ of the complex process, by following the procedure outlined in Sec. 2. We choose

$$h(t) = \beta\delta(t - t'), \tag{4.10}$$

where t' is a real and β is a complex parameter. Then from Eqs. (4.9) and (4.10) we obtain the formula

$$\begin{aligned}C_1^{(Z)}(\beta;t') = \exp\{&-\tfrac{1}{2}[\beta^2\langle Z^*(t')Z^*(t')\rangle\\ &+ \beta\beta^*\langle Z(t')Z^*(t')\rangle + \text{c.c.}]\},\end{aligned} \tag{4.11}$$

and from Eq. (2.8), we find with $N = 1$ and with a trivial change in notation

$$\begin{aligned}p_1^{(Z)}(Z;t) &= (1/\pi^2)\\ &\times\int\exp\{-\tfrac{1}{2}[\beta^2\langle Z^*(t)Z^*(t)\rangle + \beta\beta^*\langle Z(t)Z^*(t)\rangle\\ &+ \text{c.c.}]\}\exp\{-i(\beta^*Z + \beta Z^*)\}\,d^2\beta.\end{aligned} \tag{4.12}$$

The integral occurring on the rhs of Eqs. (4.1) may be evaluated with the help of a formula derived by Bargmann (Ref. 11, Sec. 1h). The result is

$$p_1^{(Z)}(Z,t) = \frac{1}{\pi(b^2 - 4a^*a)^{1/2}} \times\exp\left(-\frac{(-a^*Z^2 - aZ^{*2} + bZ^*Z)}{(b^2 - 4a^*a)}\right), \tag{4.13}$$

where

$$a = \langle Z(t)Z(t)\rangle, \quad b = \langle Z^*(t)Z(t)\rangle. \tag{4.14}$$

In general, both a and b are, of course, functions of t and are derived from the real correlation function $\langle X(t_1)X(t_2)\rangle$ via formulas of the type (3.21) and (3.22).

Finally we note that from Eqs. (4.3) and (3.20) we can readily obtain the following expression for the joint probability distribution for the coefficients $c_1, c_2, \cdots$ in the generalized Karhunen–Loeve expansion of a real Gaussian random process

$$\begin{aligned}p(\{c_n\}) = \int\cdots\int\prod_n[&(1/2\pi)\\ &\times\exp(-\tfrac{1}{2}\lambda_n u_n^2 - ic_nu_n)]\,d(\{u_n\}).\end{aligned} \tag{4.15}$$

Each of the integrals on the right may readily be evaluated, and one obtains the result

$$p(\{c_n\}) = \prod_{n=1}^{\infty}\left[\frac{1}{\sqrt{2\pi\lambda_n}}\exp\left(\frac{-c_n^2}{2\lambda_n}\right)\right]. \tag{4.16}$$

The joint probability of the first N coefficients c_n may then be obtained by integrating (4.16) over all possible values of each of the coefficients c_j ($-\infty \leq c_j \leq \infty$) for $j = N+1, N+2, \cdots$. One then obtains the formula

$$p(c_1, c_2, \ldots, c_N) = \prod_{n=1}^{N}\left[\frac{1}{\sqrt{2\pi\lambda_n}}\exp\left(\frac{-c_n^2}{2\lambda_n}\right)\right]. \tag{4.17}$$

The result expressed by Eq. (4.17) is essentially the converse of a well-known theorem of Kac and Siegert.[12]

APPENDIX: THE CHARACTERISTIC FUNCTIONAL OF A REAL GAUSSIAN RANDOM PROCESS

By definition the characteristic functional of a real random process $X(t)$ is

$$C_1^{(X)}[g(\cdot)] = \left\langle\exp\left[i\int X(t)g(t)\,dt\right]\right\rangle. \tag{A1}$$

The average on the right-hand side of (A1) is to be interpreted with the help of the Taylor expansion, i.e.,

$$C_1^{(x)}[g(\cdot)] = \left\langle \sum_{n=0}^{\infty} \frac{i^n}{n!} \left[\int X(t)g(t)\,dt\right]^n \right\rangle$$
$$= \sum_{n=0}^{\infty} \frac{i^n}{n!} \int \cdots \int \langle X(t_1) \cdots X(t_n)\rangle\, g(t_1) \cdots g(t_n) \times dt_1 \cdots dt_n, \quad \text{(A2)}$$

where all the integrations are taken from $-\infty$ to $+\infty$.

Suppose now that $X(t)$ is a Gaussian random process. Then according to the moment theorem for such a process[13] we have, for every nonnegative integer K,

$$\langle X(t_1)X(t_2) \cdots X(t_{2K+1})\rangle = 0, \quad \text{(A3)}$$

$$\langle X(t_1)X(t_2) \cdots X(t_{2K})\rangle = \sum_{\text{II}} \langle X(t_{i_1})X(t_{i_2})\rangle \cdots \times \langle X(t_{i_{2K-1}})X(t_{i_{2K}})\rangle, \quad \text{(A4)}$$

where the symbol $\sum_{\text{II}}$ denotes summation over all possible permutation of the indices $1, 2, \ldots, 2K$ labeling the time arguments. There are $(2K)!/(2^K K!)$ terms in this summation.

From (A2)–(A4) it follows that for a Gaussian random process $X(t)$ of zero mean

$$C_1^{(X)}[g(\cdot)] = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} \int \cdots \int \sum_{\text{II}} \langle X(t_{i_1})X(t_{i_2})\rangle \cdots \times \langle X(t_{i_{2n-1}})\, X(t_{i_{2n}})\rangle\, g(t_1)g(t_2) \cdots g(t_{2n}) \times dt_1 dt_2 \cdots dt_{2n}. \quad \text{(A5)}$$

We interchange the multiple integration and the summation over all the permutations. Since all the t_i integrations extend over the range from $-\infty$ to $+\infty$, each of the $2n$-folded integrals will give the same contribution and, since there are $(2n)!/2^n n!$ terms in the summation $\sum_{\text{II}}$, (A5) reduces to

$$C_1^{(X)}[g(\cdot)] = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} \frac{(2n)!}{2^n n!} \left(\iint_{-\infty}^{+\infty} \langle X(t_1)X(t_2)\rangle \times g(t_1)g(t_2)\,dt_1 dt_2 \right)^n$$
$$= \exp\left(-\tfrac{1}{2} \iint_{-\infty}^{+\infty} \langle X(t_1)X(t_2)\rangle \times g(t_1)g(t_2)\,dt_1 dt_2 \right). \quad \text{(A6)}$$

* Research supported by the U.S. Air Force Office at Scientific Research and the U.S. Army Research Office (Durham). A preliminary account of this work was presented at the OSA meeting, Washington, D.C., March 1968 [Abstract WB 13, J. Opt. Soc. Amer. **58**, 714 (1968)].

† Presently on leave of absence at Institut für Theoretische Physik, Universität Stuttgart, 7 Stuttgart 1, Germany.

1 E. Wolf, *Proceedings of the Symposium on Optical Masers* (Wiley, New York, 1963) p. 29.

2 L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

3 M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford and New York, 1970), 4th ed., Chap. X.

4 D. Gabor, J. Inst. Elec. Engrs. (London) **93**, 429 (1946).

5 L. Mandel, J. Opt. Soc. Amer. **57**, 613 (1967).

6 This assumption excludes the important case of stationary fields. Presumably this case could be treated in a similar manner within the framework of the theory of generalized functions. In this connection see also Ref. 9 below.

7 See, for example, E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon, Oxford, 1948), 2nd ed., Chap. 5.

8 For discussion of functionals see, for example, the classic work of V. Voltera, *Theory of Functions and Integrals and Integro-Differential Equations* (Dover, New York, 1959). An introduction to the theory of functionals may be found in many books, e.g. M. J. Beran, *Statistical Continuum Theory* (Interscience, New York, 1968).

9 Cf. J. L. Lumley, *Stochastic Tools in Turbulence* (Academic, New York, 1970), pp. 54–59. It is also shown in Sec. 3.12 of this reference that in the case of stationary fields, the Karhunen–Loeve expansion becomes the Fourier integral representation, provided it is interpreted in terms of generalized function theory. Thus our results can be expected to have a direct analog for stationary fields. In particular, the analogs of Eqs. (3.13a) and (3.18) are $C_1^{(X)}[g(\cdot)] = \mathcal{F}[\{\hat{g}(\nu)\}, \{\hat{g}(-\nu)\}]$, $C_1^{(Z)}[h(\cdot)] = \mathcal{F}[\{2\hat{h}^*(\nu)\}, \{2h(\nu)\}]$, where $\hat{g}(\nu)$ and $\hat{h}(\nu)$ are the Fourier transforms of the functions $g(t)$ and $h(t)$, respectively, i.e., $\hat{g}(\nu) = \int_{-\infty}^{\infty} g(t) \exp(-2\pi i \nu t)\,dt$, etc., interpreted as generalized functions.

10 This fact follows at once from the theorem that the convolution of any two real functions is equal to the convolution of their Hilbert transforms. For a proof of this theorem see, for example, M. J. Beran and G. B. Parrent, *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N.J., 1964), p. 19.

11 V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

12 M. Kac and A. J. F. Siegert, Ann. Math. Stat. **18**, 438 (1947). It should be noted that Kac and Siegert derived the theorem for stationary Gaussian random processes whereas, strictly speaking, our method of derivation applies to nonstationary Gaussian random processes whose sample functions are square integrable.

13 See, e.g., C. L. Mehta, in *Lectures in Theoretical Physics*, edited by W. E. Brittin (U. of Colorado Press, Boulder, Colorado, 1965), Vol. VII C, p. 398.

Existence and Uniqueness Questions for the Unitarity Equation

Michael Tortorella

Division of Mathematical Sciences, Purdue University, Lafayette, Indiana 47907

(Received 30 May 1972)

In early articles on this subject, it was asserted that the transformation which arises from the generalized optical theorem for inverse scattering of scalar waves is completely continuous in certain spaces. In this article we show this is not correct: This transformation is not compact in these spaces. We also obtain an improved uniqueness result for the case where there is no spherical symmetry. The article concludes with a discussion of the proper setting of the unitarity equation (generalized optical theorem) in the larger context of the inverse scattering problem.

1. INTRODUCTION

In previous papers,[1,2] hereinafter referred to as I, II, Newton and Martin have asserted that the transformation (called $\mathfrak{M}$ in I) which arises from the generalized optical theorem for inverse scattering of scalar waves is completely continuous in certain spaces. In this article it is shown that this is not correct: This transformation is not compact in these spaces, although it is continuous. In Sec. 3, an improved uniqueness result is obtained for the case where there is no spherical symmetry. The article concludes with a discussion of the proper setting of the generalized optical theorem in the larger context of the inverse scattering problem.

$$C_1^{(x)}[g(\cdot)] = \left\langle \sum_{n=0}^{\infty} \frac{i^n}{n!} \left[\int X(t)g(t)\,dt\right]^n \right\rangle$$
$$= \sum_{n=0}^{\infty} \frac{i^n}{n!} \int \cdots \int \langle X(t_1) \cdots X(t_n)\rangle\, g(t_1) \cdots g(t_n) \times dt_1 \cdots dt_n, \quad \text{(A2)}$$

where all the integrations are taken from $-\infty$ to $+\infty$.

Suppose now that $X(t)$ is a Gaussian random process. Then according to the moment theorem for such a process[13] we have, for every nonnegative integer K,

$$\langle X(t_1)X(t_2) \cdots X(t_{2K+1})\rangle = 0, \quad \text{(A3)}$$

$$\langle X(t_1)X(t_2) \cdots X(t_{2K})\rangle = \sum_{\text{II}} \langle X(t_{i_1})X(t_{i_2})\rangle \cdots \times \langle X(t_{i_{2K-1}})X(t_{i_{2K}})\rangle, \quad \text{(A4)}$$

where the symbol $\sum_{\text{II}}$ denotes summation over all possible permutation of the indices $1, 2, \ldots, 2K$ labeling the time arguments. There are $(2K)!/(2^K K!)$ terms in this summation.

From (A2)–(A4) it follows that for a Gaussian random process $X(t)$ of zero mean

$$C_1^{(X)}[g(\cdot)] = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} \int \cdots \int \sum_{\text{II}} \langle X(t_{i_1})X(t_{i_2})\rangle \cdots \times \langle X(t_{i_{2n-1}})\, X(t_{i_{2n}})\rangle\, g(t_1)g(t_2) \cdots g(t_{2n}) \times dt_1 dt_2 \cdots dt_{2n}. \quad \text{(A5)}$$

We interchange the multiple integration and the summation over all the permutations. Since all the t_i integrations extend over the range from $-\infty$ to $+\infty$, each of the $2n$-folded integrals will give the same contribution and, since there are $(2n)!/2^n n!$ terms in the summation $\sum_{\text{II}}$, (A5) reduces to

$$C_1^{(X)}[g(\cdot)] = \sum_{n=0}^{\infty} \frac{i^{2n}}{(2n)!} \frac{(2n)!}{2^n n!} \left(\iint_{-\infty}^{+\infty} \langle X(t_1)X(t_2)\rangle \times g(t_1)g(t_2)\,dt_1 dt_2 \right)^n$$
$$= \exp\left(-\tfrac{1}{2} \iint_{-\infty}^{+\infty} \langle X(t_1)X(t_2)\rangle \times g(t_1)g(t_2)\,dt_1 dt_2 \right). \quad \text{(A6)}$$

* Research supported by the U.S. Air Force Office at Scientific Research and the U.S. Army Research Office (Durham). A preliminary account of this work was presented at the OSA meeting, Washington, D.C., March 1968 [Abstract WB 13, J. Opt. Soc. Amer. **58**, 714 (1968)].

† Presently on leave of absence at Institut für Theoretische Physik, Universität Stuttgart, 7 Stuttgart 1, Germany.

1 E. Wolf, *Proceedings of the Symposium on Optical Masers* (Wiley, New York, 1963) p. 29.

2 L. Mandel and E. Wolf, Rev. Mod. Phys. **37**, 231 (1965).

3 M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford and New York, 1970), 4th ed., Chap. X.

4 D. Gabor, J. Inst. Elec. Engrs. (London) **93**, 429 (1946).

5 L. Mandel, J. Opt. Soc. Amer. **57**, 613 (1967).

6 This assumption excludes the important case of stationary fields. Presumably this case could be treated in a similar manner within the framework of the theory of generalized functions. In this connection see also Ref. 9 below.

7 See, for example, E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Clarendon, Oxford, 1948), 2nd ed., Chap. 5.

8 For discussion of functionals see, for example, the classic work of V. Voltera, *Theory of Functions and Integrals and Integro-Differential Equations* (Dover, New York, 1959). An introduction to the theory of functionals may be found in many books, e.g. M. J. Beran, *Statistical Continuum Theory* (Interscience, New York, 1968).

9 Cf. J. L. Lumley, *Stochastic Tools in Turbulence* (Academic, New York, 1970), pp. 54–59. It is also shown in Sec. 3.12 of this reference that in the case of stationary fields, the Karhunen–Loeve expansion becomes the Fourier integral representation, provided it is interpreted in terms of generalized function theory. Thus our results can be expected to have a direct analog for stationary fields. In particular, the analogs of Eqs. (3.13a) and (3.18) are $C_1^{(X)}[g(\cdot)] = \mathcal{F}[\{\hat{g}(\nu)\}, \{\hat{g}(-\nu)\}]$, $C_1^{(Z)}[h(\cdot)] = \mathcal{F}[\{2\hat{h}^*(\nu)\}, \{2h(\nu)\}]$, where $\hat{g}(\nu)$ and $\hat{h}(\nu)$ are the Fourier transforms of the functions $g(t)$ and $h(t)$, respectively, i.e., $\hat{g}(\nu) = \int_{-\infty}^{\infty} g(t) \exp(-2\pi i \nu t)\,dt$, etc., interpreted as generalized functions.

10 This fact follows at once from the theorem that the convolution of any two real functions is equal to the convolution of their Hilbert transforms. For a proof of this theorem see, for example, M. J. Beran and G. B. Parrent, *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, N.J., 1964), p. 19.

11 V. Bargmann, Commun. Pure Appl. Math. **14**, 187 (1961).

12 M. Kac and A. J. F. Siegert, Ann. Math. Stat. **18**, 438 (1947). It should be noted that Kac and Siegert derived the theorem for stationary Gaussian random processes whereas, strictly speaking, our method of derivation applies to nonstationary Gaussian random processes whose sample functions are square integrable.

13 See, e.g., C. L. Mehta, in *Lectures in Theoretical Physics*, edited by W. E. Brittin (U. of Colorado Press, Boulder, Colorado, 1965), Vol. VII C, p. 398.

Existence and Uniqueness Questions for the Unitarity Equation

Michael Tortorella

Division of Mathematical Sciences, Purdue University, Lafayette, Indiana 47907

(Received 30 May 1972)

In early articles on this subject, it was asserted that the transformation which arises from the generalized optical theorem for inverse scattering of scalar waves is completely continuous in certain spaces. In this article we show this is not correct: This transformation is not compact in these spaces. We also obtain an improved uniqueness result for the case where there is no spherical symmetry. The article concludes with a discussion of the proper setting of the unitarity equation (generalized optical theorem) in the larger context of the inverse scattering problem.

1. INTRODUCTION

In previous papers,[1,2] hereinafter referred to as I, II, Newton and Martin have asserted that the transformation (called $\mathfrak{M}$ in I) which arises from the generalized optical theorem for inverse scattering of scalar waves is completely continuous in certain spaces. In this article it is shown that this is not correct: This transformation is not compact in these spaces, although it is continuous. In Sec. 3, an improved uniqueness result is obtained for the case where there is no spherical symmetry. The article concludes with a discussion of the proper setting of the generalized optical theorem in the larger context of the inverse scattering problem.

2. NONCOMPACTNESS

In this section an example is given which shows that the transformation which arises from the generalized optical theorem is not compact in a space of continuous functions with the uniform topology, contrary to Theorem 1 of I and Sec. 3 of II. The example is given in a canonical form, namely we do not assume spherical symmetry, and no change of variable is used in the integral. Furthermore, the same example can be made to work in the special case where spherical symmetry is assumed.

We shall generally follow the notation of I. The equation under consideration is

$$\mathrm{Im}A(\mathbf{n}',\mathbf{n}) = (k/4\pi)\int_S A(\mathbf{n}'',\mathbf{n})\overline{A}(\mathbf{n}'',\mathbf{n}')d\Omega(\mathbf{n}''), \quad \mathbf{n}',\mathbf{n}\in S, \tag{1}$$

where $A(\mathbf{n}',\mathbf{n})$ [which is assumed to equal $A(\mathbf{n},\mathbf{n}')$] is the scattering amplitude for incident direction $\mathbf{n}$ and scattered direction $\mathbf{n}'$; k is the wavenumber; and $d\Omega$ is the solid angle measure on the unit sphere S in $\mathbb{R}^3$. (1) is sometimes known as the generalized optical theorem and leads immediately to two equations:

$$\mathrm{Im}A(\mathbf{n}',\mathbf{n}) = (k/4\pi)\int_S [\mathrm{Re}A(\mathbf{n}'',\mathbf{n})\,\mathrm{Re}A(\mathbf{n}'',\mathbf{n}') + \mathrm{Im}A(\mathbf{n}'',\mathbf{n})\,\mathrm{Im}A(\mathbf{n}'',\mathbf{n}')]d\Omega, \tag{2}$$

$$0 = (k/4\pi)\int_S \mathrm{Im}A(\mathbf{n}'',\mathbf{n}')\,\mathrm{Re}A(\mathbf{n}'',\mathbf{n}) \times [-\mathrm{Re}A(\mathbf{n}'',\mathbf{n}')\,\mathrm{Im}A(\mathbf{n}'',\mathbf{n})]d\Omega. \tag{2'}$$

That (2′) is identically satisfied for any solution A of (1) is a consequence of the fact that $A(\mathbf{n}',\mathbf{n}) = A(\mathbf{n},\mathbf{n}')$. Thus our attention focuses on (2). Write $A(\mathbf{n}',\mathbf{n}) = k^{-1}G(\mathbf{n}',\mathbf{n})e^{i\phi(\mathbf{n}',\mathbf{n})}$ and put

$$H(\mathbf{n}'',\mathbf{n}',\mathbf{n}) = G(\mathbf{n}'',\mathbf{n}')G(\mathbf{n}'',\mathbf{n})/4\pi G(\mathbf{n}',\mathbf{n})$$

and

$$Q(\mathbf{n}',\mathbf{n}) = \int_S H(\mathbf{n}'',\mathbf{n}',\mathbf{n})d\Omega. \tag{3}$$

Then (2) becomes

$$\sin\phi(\mathbf{n}',\mathbf{n}) = \int_S H(\mathbf{n}'',\mathbf{n}',\mathbf{n})\cos[\phi(\mathbf{n}'',\mathbf{n}') - \phi(\mathbf{n}'',\mathbf{n})]d\Omega. \tag{4}$$

Let Y_0 be the Banach space of continuous functions on $S\times S$ into $\mathbb{R}$ with $\|f\| = \sup\{|f(\mathbf{n}_1,\mathbf{n}_2)|: \mathbf{n}_1,\mathbf{n}_2\in S\}$. Put $Y = \{f\in Y_0 : f(\mathbf{n}_1,\mathbf{n}_2) = f(\mathbf{n}_2,\mathbf{n}_1), \forall \mathbf{n}_1,\mathbf{n}_2\in S\}$. Y is a closed subspace of Y_0 and so is itself a Banach space. We now define a transformation $\mathfrak{M}$ on Y: for $\phi\in Y, \mathbf{n}_1,\mathbf{n}_2\in S$, we define

$$\mathfrak{M}\phi(\mathbf{n}',\mathbf{n}) = \mathrm{Arcsin}\int_S H(\mathbf{n}'',\mathbf{n}',\mathbf{n}) \times \cos[\phi(\mathbf{n}'',\mathbf{n}') - \phi(\mathbf{n}'',\mathbf{n})]d\Omega. \tag{5}$$

Then to solve (1), we seek a fixed point of $\mathfrak{M}$. Let us assume the analogs of the hypotheses of Theorem 1 of I, namely

$$\begin{aligned} &\text{H1.}\quad G: S\times S\to[0,\infty) \text{ is continuous;}\\ &\text{H2.}\quad \sup\{Q(\mathbf{n}_1,\mathbf{n}_2): \mathbf{n}_1,\mathbf{n}_2\in S\} = M < 1. \end{aligned} \tag{6}$$

We note that H2 forces $\min\{G(\mathbf{n}_1,\mathbf{n}_2): \mathbf{n}_1,\mathbf{n}_2\in S\} > 0$, so H is continuous (hence bounded), and the same is true of Q, so $Q\in Y$ and H2 may be replaced by the statement $\|Q\| = M < 1$.

Convergence in $S\times S$ is defined in terms of the Euclidean norm $|\cdot|_E$ in $\mathbb{R}^6$; an easy application of the Lebesgue bounded convergence theorem shows $\phi\in Y \Rightarrow \mathfrak{M}\phi$ is continuous. Noting that $\mathfrak{M}\phi(\mathbf{n}',\mathbf{n}) = \mathfrak{M}\phi(\mathbf{n},\mathbf{n}')$, we conclude $\mathfrak{M}\phi\in Y$. In particular, because of H2, range $(\mathfrak{M}) \subset B'_Y(0, \mathrm{Arcsin}M)$, the closed ball in Y about 0 with radius $\mathrm{Arcsin}M$.

We get the same results in the case of spherical symmetry when the appropriate modifications of the equation and the underlying function space are made [for example, one might use (6) of I, being careful to note that (6) only holds in the open interval $-1 < x < 1$; for $x = \pm 1$ the correct expression is (6′)]. We only remark that in this case, convergence in S is convergence in the metric $d(\mathbf{n}_1,\mathbf{n}_2) = |\angle(\mathbf{n}_1,\mathbf{n}_2)|$ where $\angle(\mathbf{n}_1,\mathbf{n}_2)$ is the angle between $\mathbf{n}_1$ and $\mathbf{n}_2$ in radians, and convergence in $S\times S$ is convergence in any product metric.

We are now ready to state the main result of this section.

Theorem 2.1: Range $(\mathfrak{M}) \subset Y$ is not compact.

Proof: Suppose the contrary. Then by Ascoli's theorem, Range $(\mathfrak{M})$ is an equicontinuous uniformly bounded set (this is Newton's claim); in particular, Range $(\mathfrak{M})$ is equicontinuous at $(\mathbf{n}_0,\mathbf{n}_0)\in S\times S$. Then given $\epsilon > 0$, $\exists\delta' = \delta'(\epsilon,\mathbf{n}_0) > 0$ such that for all $\mathbf{n}$ with $|\mathbf{n}-\mathbf{n}_0|_E < \delta'$ and for all $\phi\in Y = \mathrm{Domain}(\mathfrak{M})$, we have $|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0) - \mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n})| < \epsilon$. Let us choose $\epsilon < \frac{5}{16}Q(\mathbf{n}_0,\mathbf{n}_0)$, noting that $Q(\mathbf{n}_0,\mathbf{n}_0) = 0$ if and only if $G\equiv 0$, a possibility which we shall rule out.

Let $\phi\in Y$; we wish to bound $|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0) - \mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n})|$ below:

$$\begin{aligned}
&|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0) - \mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n})|\\
&\quad= |\mathrm{Arcsin}\int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)d\Omega(\mathbf{n}'')\\
&\qquad - \mathrm{Arcsin}\int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n})\cos[\phi(\mathbf{n}'',\mathbf{n}_0)\\
&\qquad - \phi(\mathbf{n}'',\mathbf{n})]d\Omega(\mathbf{n}'')|\\
&\quad= (1-\mu^2)^{-1/2}|\int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)d\Omega\\
&\qquad - \int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n})\cos[\phi(\mathbf{n}'',\mathbf{n}_0)\\
&\qquad - \phi(\mathbf{n}'',\mathbf{n})]d\Omega| \quad (0\le\mu\le M)\\
&\quad\ge |\int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)\{1-\cos[\phi(\mathbf{n}'',\mathbf{n}_0)\\
&\qquad - \phi(\mathbf{n}'',\mathbf{n})]\}d\Omega - \int_S [H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n})\\
&\qquad - H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)]\cos[\phi(\mathbf{n}'',\mathbf{n}_0) - \phi(\mathbf{n}'',\mathbf{n})]d\Omega|\\
&\quad\ge \int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)\{1-\cos[\phi(\mathbf{n}'',\mathbf{n}_0)\\
&\qquad - \phi(\mathbf{n}'',\mathbf{n})]\}d\Omega - |\int_S [H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n})\\
&\qquad - H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)]\cos[\phi(\mathbf{n}'',\mathbf{n}_0) - \phi(\mathbf{n}'',\mathbf{n})]d\Omega|.
\end{aligned}$$

Considering the rhs term for a moment, we see that $H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n})$ is a continuous function of $\mathbf{n}$ and is bounded, and so we may apply to

$$\begin{aligned}
&|\int_S [H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}) - H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)]\\
&\qquad\times\cos[\phi(\mathbf{n}'',\mathbf{n}_0) - \phi(\mathbf{n}'',\mathbf{n})]d\Omega|\\
&\quad\le \int_S |H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}) - H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)|d\Omega,
\end{aligned}$$

the Lebesgue bounded convergence theorem. This gives $\exists\delta'' > 0$ such that $|\mathbf{n}-\mathbf{n}_0|_E < \delta'' \Longrightarrow$ this quantity is $< \frac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0)$.

Thus for $|\mathbf{n}-\mathbf{n}_0|_E < \delta''$,

$$|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0)-\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n})| \geq \int_S H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0) \times\{1-\cos[\phi(\mathbf{n}'',\mathbf{n}_0)-\phi(\mathbf{n}'',\mathbf{n})]\}d\Omega - \tfrac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0).$$

Let $\delta = \min\{\delta',\delta''\}$ and choose $\mathbf{n}_1 \in S$ with $|\mathbf{n}_1-\mathbf{n}_0|_E < \delta$. We now show the existence of a function in $B'_Y(0,\frac{1}{2}\pi)$ for which $|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0)-\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_1)|$ is $>\epsilon$. Let B_i be an open ball on S centered at $\mathbf{n}_i$ for which

$$\int_{B_i} d\Omega < Q(\mathbf{n}_0,\mathbf{n}_0)/16P, \qquad i=0,1,$$

where

$$P = \max_{\mathbf{n}''\in S} H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0).$$

Put $T = S\setminus(B_0 \cup B_1)$. Then clearly

$$|\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n}_0)-\mathfrak{M}\phi(\mathbf{n}_0,\mathbf{n})| \geq \int_T H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0) \times\{1-\cos[\phi(\mathbf{n}'',\mathbf{n}_0)-\phi(\mathbf{n}'',\mathbf{n})]\}d\Omega - \tfrac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0) \tag{7}$$

and

$$\int_T H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)d\Omega = Q(\mathbf{n}_0,\mathbf{n}_0) - \int_{B_0\cup B_1} H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)d\Omega \geq Q(\mathbf{n}_0,\mathbf{n}_0) - P\left(\int_{B_0} d\Omega + \int_{B_1} d\Omega\right) \geq \tfrac{7}{8}Q(\mathbf{n}_0,\mathbf{n}_0). \tag{8}$$

Let ψ be a Urysohn function for the disjoint closed subsets $T_0 = T\times\{\mathbf{n}_0\}\cup\{\mathbf{n}_0\}\times T$ and $T_1 = T\times\{\mathbf{n}_1\}\cup\{\mathbf{n}_1\}\times T$ of the (normal) space $S\times S$ which sends T_0 onto $\frac{1}{6}\pi$ and T_1 onto 0. Put $\phi_0(\mathbf{n},\mathbf{m}) = \psi(\mathbf{m},\mathbf{n}) + \psi(\mathbf{m},\mathbf{n})$. Then clearly ϕ_0 is continuous, $0 \leq \phi_0 \leq \frac{1}{3}\pi$, and $\phi_0(\mathbf{m},\mathbf{n}) = \phi_0(\mathbf{n},\mathbf{m})$, so $\phi_0 \in B'_Y(0,\frac{1}{2}\pi)$. Furthermore, for all $\mathbf{n}'' \in T$, we have $|\phi_0(\mathbf{n}'',\mathbf{n}_0)-\phi_0(\mathbf{n}'',\mathbf{n}_1)| = \frac{1}{3}\pi$.

Then finally,

$$\begin{aligned}|\mathfrak{M}\phi_0(\mathbf{n}_0,\mathbf{n}_0)-\mathfrak{M}\phi_0(\mathbf{n}_0,\mathbf{n}_1)| &\geq \int_T H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)\{1-\cos[\phi_0(\mathbf{n}'',\mathbf{n}_0) - \phi_0(\mathbf{n}'',\mathbf{n}_1)]\}d\Omega - \tfrac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0)\\ &= \tfrac{1}{2}\int_T H(\mathbf{n}'',\mathbf{n}_0,\mathbf{n}_0)d\Omega - \tfrac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0)\\ &\geq \tfrac{7}{16}Q(\mathbf{n}_0,\mathbf{n}_0) - \tfrac{1}{8}Q(\mathbf{n}_0,\mathbf{n}_0) = \tfrac{5}{16}Q(\mathbf{n}_0,\mathbf{n}_0) > \epsilon\end{aligned}$$

which is the desired contradiction.

Remarks: (1) A very similar technique yields the same results for the spherically symmetric case. The main difference is that instead of cutting out small balls about $\mathbf{n}_0$ and $\mathbf{n}_1$, one has to cut out an Ω-small open strip Σ surrounding the equator of the sphere on which $\angle(\mathbf{n}_0,\mathbf{n}'') = \angle(\mathbf{n}_1,\mathbf{n}'')$ for every $\mathbf{n}''$, and of course ϕ_0 now needs to be constructed in accordance with this change: Here T is $S\setminus\Sigma$, T_0, T_1, and ψ are as before, with the new T.

(2) This theorem shows that we cannot use Schauder's theorem to conclude the existence of a fixed point of $\mathfrak{M}$ in Y without some additional conditions. For example, one could restrict the domain of $\mathfrak{M}$ to some compact convex subset of Y and add some conditions on G so that $\mathfrak{M}$ leaves this set invariant; however, it is felt that these conditions become much too severe.

(3) The theorem also shows we cannot get $\mathfrak{M}$ to be compact by restricting the domain of $\mathfrak{M}$ to the positive cone in Y, i.e., $\{\phi \in Y : \phi(\mathbf{n}_1,\mathbf{n}_2) \geq 0, \mathbf{n}_1,\mathbf{n}_2 \in S\}$.

(4) In II, the attempt is made to show $\mathfrak{M}$ is compact by showing $\mathfrak{M}^2$ is compact. Without additional conditions, one may not conclude this, however: For example, even if T is a bounded linear operator in a Hilbert space, to assert T^2 compact $\Rightarrow$ T compact one needs at least the additional condition that T be normal.

At any rate, even if the demonstration that $\mathfrak{M}^2$ is compact were successful (which it is not, as we shall see below), we could only conclude the existence of a fixed point $\phi = \mathfrak{M}^2\phi$, and only if it were unique (which cannot be concluded *a priori*) could we get $\phi = \mathfrak{M}\phi$.

To see that $\mathfrak{M}^2$ is not compact either, we note first of all that one cannot expect to control $|P_l(\cos\theta_1) - P_l(\cos\theta_2)|$ for all values of θ_1,θ_2 with a bound that does not grow with l. However, and more convincingly, an explicit counterexample can be constructed along the lines of the theorem. $\mathfrak{M}\phi \in Y$ so we have (7) and (8) are true for functions in Range $(\mathfrak{M})$ as well. So if we can show there is a $\psi_0 \in Y$ such that $\phi_0 = \mathfrak{M}\psi_0$, where ϕ_0 is the function of Theorem 2.1, we are done. This can be done easily if we are willing to sacrifice the equality sign and replace it by the condition that $\|\phi_0 - \mathfrak{M}\psi_0\|$ be suitably small. The technique used to obtain ψ_0 and the remainder of the argument is quite similar to that used in Theorem 2.1, and since there are no new ideas here, the detailed computation will not be produced.

3. UNIQUENESS IN THE CASE OF NO SPHERICAL SYMMETRY

The motivation for this section is to try to extend the results of the first part of Sec. 4 of II to the case where there is no spherical symmetry. Throughout we will assume (6) and denote by X_0 the Hilbert space $L^2_G(S\times S, d\Omega) = L^2(S\times S, Gd\Omega)$, that is, the space of all real-valued, square-integrable $[d\Omega]$ functions on $S\times S$ with weight function G, and $\|\phi\|_G = \|G\phi\|_2$. We put $X = \{\phi \in X_0 : \phi(\mathbf{n}_1,\mathbf{n}_2) = \phi(\mathbf{n}_2,\mathbf{n}_1), \forall \mathbf{n}_1,\mathbf{n}_2 \in S\}$. X is a closed subspace of X_0 and so is itself a Hilbert space. As an aid to establishing the principal result of this section, we prove the following simple lemma:

Lemma 3.1: Let (E,d) be a complete metric space, $A: E\to E$, and let $h: E\to E$ be a homeomorphism such that $d(h\circ Ax_1, h\circ Ax_2) \leq \gamma d(hx_1,hx_2)$ for $x_1,x_2 \in E$, with $\gamma < 1$. Then A has a unique fixed point which can be constructed by successive approximation.

Proof: Apply the Banach contraction mapping theorem to the map $T: h(E)\to h(E)$ given by $T(hx) = h\circ Ax$. This gives a unique $x_0 \in E$ such that $h\circ Ax_0 = hx_0$. Apply h^{-1} to get $Ax_0 = x_0$.

Theorem 3.1: Under hypotheses (6) and if in addition $\sup\{Q(\mathbf{n}_1,\mathbf{n}_2): \mathbf{n}_2 \in S\} = M < (1 + (1/4\pi^2)\|G\|_2)^{-1/2}$, then (1) has a unique solution in X, and the solution can be found by successive approximation.

Proof: For $\phi \in X$, $\mathbf{n}', \mathbf{n} \in S$, define a transformation $\mathfrak{M}: X\to X$ by expression (5). We prove an esti-

mate which allows us to use Lemma 3.1 with $A = \mathfrak{M}$, $E = \{\phi \in X: 0 \le \phi \le \arcsin M < \frac{1}{2}\pi \text{ a.e. } [d\Omega]\}$, and h the mapping $t \to \sin t$, which is a homeomorphism of the interval $[0, \arcsin M]$ onto the interval $[0, M]$. H2 guarantees that $\mathfrak{M}: E \to E$. Let $\phi_1, \phi_2 \in E$; we begin with (19) of II, which in our context may be written

$$
\begin{aligned}
&|G(\mathbf{n}', \mathbf{n})[\sin\mathfrak{M}\phi_1(\mathbf{n}', \mathbf{n}) - \sin\mathfrak{M}\phi_2(\mathbf{n}', \mathbf{n})]| \\
&\quad \le (1/4\pi)M(1 - M^2)^{-1/2} \int_S G(\mathbf{n}'', \mathbf{n}') \\
&\quad \times [\sin\mathfrak{M}\phi_1(\mathbf{n}'', \mathbf{n}') - \sin\mathfrak{M}\phi_2(\mathbf{n}'', \mathbf{n}')]G(\mathbf{n}'', \mathbf{n}) \\
&\quad \times [\cos\mathfrak{M}\phi_1(\mathbf{n}'', \mathbf{n}) + \cos\mathfrak{M}\phi_2(\mathbf{n}'', \mathbf{n})]d\Omega.
\end{aligned}
$$

Use the Schwarz inequality, square both sides, and integrate over $d\Omega(\mathbf{n}')d\Omega(\mathbf{n})$. Noting that $\|\cos\phi_1 + \cos\phi_2\|_G \le 2\|G\|_2$, we get $\|\sin\mathfrak{M}\phi_1 - \sin\mathfrak{M}\phi_2\|_G^2 \le (1/4\pi^2)M^2(1 - M^2)^{-1}\|G\|_2^2\|\sin\phi_1 - \sin\phi_2\|_G^2$, and now the result follows from Lemma 3.1.

Remarks: (1) The theorem still holds if H1 of (6) is weakened to $G \in L^\infty(S \times S)$.

(2) The result is obviously true in the spherically symmetric case as well; in particular, if $\|G\|_2 < 2\pi$ this theorem is an improvement on the first part of Sec. 4 of II. Of course, since we are in L^2, the fixed point need no longer be a continuous function.

(3) It is also possible to get a unique fixed point for $\mathfrak{M}$ in L_G^2 under the condition $M < (1 + \frac{1}{2\pi}\|G\|_\infty)^{-\frac{1}{2}}$; however, in this case, as in Sec. 4 of II, we do not get the construction of the fixed point. This assertion follows from the following lemma, which, together with some elementary estimates, shows that if ϕ is a fixed point of $\mathfrak{M}$, $\|\cos\phi\|_G^2 \le 2\pi\|G\|_\infty$ (use this estimate in place of $\|\cos\phi_1 + \cos\phi_2\| \le 2\|G\|_2$ in the estimate in Theorem 3.1, and proceed as in II).

Lemma 3.2: If ϕ is a fixed point of $\mathfrak{M}$, then $0 \le \phi \le \text{Arcsin}M$,

$$0 \le \int_S \int_S G(\mathbf{n}', \mathbf{n}) \cos\phi(\mathbf{n}', \mathbf{n})d\Omega d\Omega \le 2\pi,$$

and

$$0 \le \int_S \int_S G(\mathbf{n}', \mathbf{n}) \sin\phi(\mathbf{n}', \mathbf{n})d\Omega d\Omega \le 4\pi.$$

Proof: $0 \le \phi \le \text{Arcsin}M$ is proved in Sec. 2 of II. Define $A(\mathbf{n}) = \int_S G(\mathbf{n}', \mathbf{n}) \cos\phi(\mathbf{n}', \mathbf{n})d\Omega$ and $B(\mathbf{n}) = \int_S G(\mathbf{n}', \mathbf{n}) \sin\phi(\mathbf{n}', \mathbf{n})d\Omega$. Integrate (4) $d\Omega(\mathbf{n})$ to get

$$
\begin{aligned}
B(\mathbf{n}') &= (1/4\pi) \int_S A(\mathbf{n}'')G(\mathbf{n}', \mathbf{n}'') \cos\phi(\mathbf{n}', \mathbf{n}'')d\Omega \\
&\quad + (1/4\pi) \int_S B(\mathbf{n}'')G(\mathbf{n}', \mathbf{n}'') \sin\phi(\mathbf{n}', \mathbf{n}'')d\Omega. \quad (9)
\end{aligned}
$$

Now integrate (9) $d\Omega(n')$ to get

$$
\begin{aligned}
\int_S B(\mathbf{n})d\Omega &= (1/4\pi) \int_S [A(\mathbf{n})^2 + B(\mathbf{n})^2]d\Omega \\
&\ge (1/4\pi)[\int_S A(\mathbf{n})d\Omega]^2 + (1/4\pi)[\int_S B(\mathbf{n})d\Omega]^2.
\end{aligned}
$$

Putting $a = \int_S A(\mathbf{n})d\Omega$, $b = \int_S B(\mathbf{n})d\Omega$, we have $b^2 - 4\pi b + a^2 \le 0$, which is satisfied for $2\pi - (4\pi^2 - a^2)^{1/2} \le b \le 2\pi + (4\pi^2 - a^2)^{1/2}$, which forces $|a| \le 2\pi$ and $0 \le b \le 4\pi$.

(4) The estimates obtained here are somehow satisfying since as $\|G\|$ gets small, M increases to 1, and so as $G \to 0$ (and we certainly know A for $G \equiv 0$!) we get closer to the condition $M < 1$ for a unique solution.

4. CONCLUSION

The condition H2 introduced by Newton is felt to be too severe for the mere existence of a solution. Of course, $\|Q\| \le 1$ is an inescapable *a priori* estimate if one wishes to apply any of the classical fixed point theorems (e.g., Schauder, Banach contraction), for it is this condition that guarantees that range $(\mathfrak{M})$ stays in the right place. However, we now have to consider what we really mean when we ask existence and uniqueness questions about (1). Let us recall that for any given (sufficiently regular) bounded obstacle and boundary conditions, the function $A(\mathbf{n}', \mathbf{n})$ exists and is unique (use, for example, the uniqueness theorem of Wilcox[3]). The generalized optical theorem arises only as a necessary condition on this function A; in other words, (1) is always satisfied by the function A which is the scattering amplitude for the particular problem under consideration. So the existence of a solution to (1) is really not in question (Schauder's theorem is not much help after all!). We view the generalized optical theorem, then, as a computational device for recovering the function A from the experimentally determined $|A|^2$, and as such, the interesting question is that of uniqueness of solution, for it is this question which will settle the usefulness of (1) as a computational tool. We need to know under what conditions on $|A|$ are we guaranteed that (1) has only one solution.

Note added in manuscript: The form of the generalized optical theorem used here, and the property $A(\mathbf{n}, \mathbf{n}') = A(\mathbf{n}'\mathbf{n})$, depend on the hypothesis $A(\mathbf{n}, \mathbf{n}') = A(-\mathbf{n}, -\mathbf{n}')$, which is used in the original derivation of the equation. This hypothesis alone does not force spherical symmetry; however, Professor V. Weston has suggested that perhaps $A(\mathbf{n}, \mathbf{n}') = A(-\mathbf{n}, -\mathbf{n}')$ together with some other conditions on A (e.g., A analytic) might imply spherical symmetry. Also, there are perhaps cases with no spherical symmetry where we do not even have $A(\mathbf{n}, \mathbf{n}') = A(-\mathbf{n}, -\mathbf{n}')$; and, in that case, (1) is no longer valid, and the considerations for (2′) do not apply.

1 R. G. Newton, J. Math. Phys. 9, 2050 (1968).
2 A. Martin, Nuovo Cimento 59A, 131 (1969).
3 C. H. Wilcox, Proc. Amer. Math. Soc. 7, 271 (1965).

Note on Nonlinear Representations

I. V. V. Raghavacharyulu*
Department of Physics, Indian Institute of Technology, Madras-36, India
(Received 3 March 1971)

In the theory of non-linear representations of a continuous group G with respect to a closed subgroup H, the peculiar transformation behavior of the reduction matrix L_ϕ is found to be identical with the transformation behavior of a set of coset representative elements of G with respect to H. The limitations of the extended definition of the boost by Salam and Strathdee are discussed.

The most significant recent achievement[1] in the study of renormalizable field theories lies in broadening the concept of boost. This leads immediately to the setting up of nonlinear representations of a continuous group G through the introduction of the *reducing matrix* L_ϕ. However, this matrix is found to have a "peculiar transformation behavior under the operations of G".[2] In view of many possible applications of the nonlinear group representation technique, a critical study of the transformation of L_ϕ is worth while. Let H be some specified closed subgroup of a continuous group G for which nonlinear representations are to be set up which become both linear and possibly reducible when restricted to H.

Let 1 be the unit element of G. Following the prescription of Salam and Strathdee, let L_ϕ be the reducing matrix. From the prescribed transformation properties[2] of L_ϕ, we have

$$L_{g\phi} = gL_\phi h^{-1}(\phi,g). \tag{1}$$

Hence $L_{g\phi} \in G$ for all $g \in G$ and $h(\phi,g) \in H$.

For a fixed ϕ let

$$L_{g\phi} = K(g) \tag{2}$$

in the self-representation of G. From (2) we have

$$L_\phi = K(1) = K_1 \quad \text{(say)}. \tag{3}$$

Hence, given L_ϕ, the element $K(1)$ is uniquely fixed. By writing $h_{K_1}(g)$ for $h(\phi,g)$, the transformation law (1) becomes

$$gK(1) = K(g)h_{K_1}(g). \tag{4}$$

Obviously, it follows from (4) that

$$g = K(gK_1^{-1})\,h_{K_1}(gK_1^{-1}). \tag{5}$$

Hence every element $g \in G$ can be represented as $K(gK_1^{-1})h$ which is nothing but an element in the right coset KH of G with respect to H. Now let $p \in H$; then we have from (5)

$$gp = K(gK_1^{-1})h_{K_1}(gK_1^{-1})p = K(gpK_1^{-1})h_{K_1}(gpK_1^{-1}). \tag{6}$$

From the uniqueness of the right coset decomposition it immediately follows from (6) that

$$\begin{aligned} &K(gK_1^{-1}) = K(gpK_1^{-1}) \\ &\text{and} \\ &h_{K_1}(gpK_1^{-1}) = h_{K_1}(gK_1^{-1})p. \end{aligned} \tag{7}$$

Using $1 = K(K_1^{-1})h_{K_1}(K_1^{-1})$, we also have $K(K_1^{-1}) = h_{K_1}(K_1^{-1}) = 1$.

Further, from (7),

$$\begin{aligned} &K(K_1^{-1}) = K(pK_1^{-1}) = 1 \\ &\text{and} \\ &h_{K_1}(pK_1^{-1}) = h_{K_1}(K_1^{-1})p = p \end{aligned} \tag{8}$$

for all $p \in H$.

Now consider $pK_1 = K(p)h_{K_1}(p)$. Then

$$K(p) = pK_1h_{K_1}^{-1}(p) \tag{9}$$

is the most general way in which K_p transforms for all $p \in H$. If in particular

$$h_{K_1}(p) = p, \tag{10}$$

then the conditions imposed on L_ϕ by Salam and Strathdee are satisfied. However, obviously it is not necessarily satisfied for a general reduction matrix L_ϕ. Now let us consider the transformation properties of $K(g)$. From (4) and (7) it follows that

(i) $K(g)$ transforms like a set of coset representative elements of the group G with respect to the closed subgroup H.

(ii) Given $h_K(g)$, which immediately defines the transformation properties of L_ϕ over G, the L_ϕ is unique to within multiplication by a scalar and vice versa.

From the above analysis it is obvious that given a set of algebraic relations satisfied by L_ϕ, it may not be possible to set up nonlinear representations of G over H in which L_ϕ satisfies prescribed analytic conditions. For example, in the group-theoretical approach to the study of kinematical details of the multi-Regge model, it is found necessary to construct[3] boosts in the complex Lorentz group such that the amplitudes introduced by Bali, Chew, and Pignotti[4] are free of kinematic singularities and constraints.

ACKNOWLEDGMENT

The author expresses his gratitude to Dr. C. Ramasastry, Professor and Head of the Department of Physics, for providing facilities to be associated with the Physics Department, Indian Institute of Technology, Madras.

---

* Present address: Nuclear Physics Division, BARC, Trombay, Bombay-85, India.

1 Abdus Salam and J. Strathdee, Phys. Rev. **184**, 1750 (1969).
2 Abdus Salam and J. Strathdee, Phys. Rev. **184**, 1760 (1969).
3 N. Toller, CERN Preprint Th. 975 (1968).
4 N. F. Bali, G. F. Chew, and A. Pignotti, Phys. Rev. Letters **19**, 614, (1967); Phys. Rev. **163**, 1572 (1967)

A Study of Relaxing Waves

D. Y. Hsieh
Division of Applied Mathematics, Brown University, Rhode Island
(Received 25 February 1971)

The equation $u_t + (1+u)u_x - \int_0^t (1+u)u_x f(t-\tau)\,d\tau = 0$ is studied when an initial finite pulse $u(x,0)$ is given. For the linearized equation, general solutions in terms of Laplace transforms are obtained and more explicit expression for exponential kernels is given. An iteration expansion scheme is established for general kernels. For positive kernels, it is found that the stability condition for the solution is $\int_0^\infty f(t)dt < 1$. Then the large time solution as well as the solution representing the main disturbance is obtained. For the nonlinear equation, the condition for the shock formation is obtained for the special case $f(t) = \mu e^{-\mu t}$, or when the nonlinearity is weak.

1. INTRODUCTION

In this paper we like to study the property of the following integro-differential equation

$$u_t + (1+u)u_x - \int_0^t (1+u)u_x f(t-\tau)d\tau = 0. \quad (1.1)$$

In many physical problems we are interested in the study of wave propagation in a medium which exhibits relaxation behavior. A notable example is the wave propagation in a viscoelastic material, whose stress–strain relation may be represented by the following relation:

$$\sigma = F(t) + \int_0^\infty G[\epsilon(t-\tau), \tau]d\tau. \quad (1.2)$$

When the strain is small, i.e., when the linear viscoelastic theory can apply, the wave equation can be put in the form

$$u_{tt} - u_{xx} - \int_0^t u_{xx}(x,\tau)f(t-\tau)d\tau = 0. \quad (1.3)$$

This equation has been briefly discussed by Volterra.[1] A simpler version derivable from (1.3) will then be the linearized equation of (1.1),

$$u_t + u_x - \int_0^t u_x(x,\tau)f(t-\tau)d\tau = 0, \quad (1.4)$$

which will represent waves propagating along one characteristic direction. Equations slightly different from (1.1) have also been mentioned by Whitham in connection to the study of water waves.[2]

Equation (1.1) represents almost the simplest type of wave equations that incorporate the effect of both nonlinearity and relaxation or heredity. For most physical problems, there are usually also present the effect of dissipation or diffusion. Often the effect of dissipation will overshadow the effect of relaxation. Here we purposely neglect any explicit dissipative effects in order to see more clearly the role played by the relaxation effect and how the mechanism of relaxation interacts with the nonlinearity.

In the following, after a very brief general discussion, we shall first study the linearized equation (1.4). Although the problem can in principle be solved by the method of Laplace transformation, not much can be said about the formal solution. The emphases then are concentrated on the development of a convergent iteration expansion and the study of the asymptotic expansions for large t. The latter study also leads to the establishment of the condition for the stability of the solution. Both as probe and illustration, particular cases with exponential kernels are studied in some detail.

For the nonlinear equation, solution may not exist for all the time, since among other things shock will form. Our emphasis then is placed on the establishment of the condition of shock formation. Although we have definite results only for some particular exponential kernel, and when the nonlinearity is weak, it is clearly demonstrated that both the length of the pulse and the nature of the kernel as well as the maximum slope of the pulse will play an important role in the formation of shock.

We may also mention that a technical report[3] bearing the same title has also been prepared, where more detailed analyses can be found.

2. GENERAL PROPERTY OF THE EQUATION

We like to investigate the equation

$$u_t + (1+u)u_x - \int_0^t (1+u)u_x f(t-\tau)d\tau = 0, \quad (2.1)$$

with the initial condition

$$u(x,0) = u_0(x). \quad (2.2)$$

$u_0(x)$ is supposed to be nonvanishing only in a finite interval of x; i.e., we are interested in the propagation of a pulse.

Since $u = 0$ as $x \to \pm\infty$ for any t, we obtain by integrating (2.1)

$$\frac{d}{dt}\int_{-\infty}^{+\infty} u(x,t)dx = 0, \quad (2.3)$$

or

$$\int_{-\infty}^{+\infty} u(x,t)dx = \text{const} = \int_{-\infty}^{+\infty} u_0(x)dx, \quad (2.4)$$

so long as u remains integrable, in particular, single valued in x.

If we treat x in $u(x,t)$ as a parameter, then (2.1) represents a Volterra integral equation of the second kind for the unknown $(1+u)u_x$. Let $H(t-\tau)$ be the resolvent of $f(t-\tau)$; then we obtain[4]

$$u_t + (1+u)u_x + \int_0^t \frac{\partial u}{\partial \tau}(x,\tau)H(t-\tau)d\tau = 0. \quad (2.5)$$

3. LINEAR EQUATION

When u is small compared with 1, we may consider the linearized equation

$$u_t + u_x - \int_0^t u_x(x,\tau)d\tau = 0. \quad (3.1)$$

Define the Laplace transform of $u(x,t)$ and $f(t)$ by

$$\hat{u}(x,s) = \int_0^\infty e^{-st}u(x,t)dt$$

and

$$\hat{f}(s) = \int_0^\infty e^{-st}f(t)dt.$$

We obtain

$$u(x,t) = \int_{-\infty}^{x} dy\, u_0(y)F(x-y,t), \quad (3.2)$$

where

$$F(x-y,t)=\frac{1}{2\pi i}\int_c ds\, e^{st}F(x-y,s), \tag{3.3}$$

and

$$\hat{F}(z,s)=e^{-zs/1-\hat{f}(s)}/1-\hat{f}(s), \tag{3.4}$$

with c, as usual, the contour to the right of all singularities.

Now $\hat{f}(s)\to 0$ as $s\to\infty$ to the right of the contour c. Therefore,

$$F(x-y,t)=0 \quad \text{for } x-y-t>0.$$

Hence

$$u(x,t)=\int_{x-t}^{x} F(x-y,t)u_0(y)dy. \tag{3.5}$$

If $u_0(x)=0$ for $x<x_1$ and $x>x_2$, then it is clear that $u(x,t)=0$ for $x>x_2+t$ and $x<x_1$. This is indeed what we should expect, since the wavefront is traveling in x direction with speed 1. Nothing happens ahead of the wavefront and behind the initial disturbance.

Let $z=x-y$; we can also rewrite (5) as

$$u(x,t)=\int_0^t F(z,t)u_0(x-z)dz. \tag{3.6}$$

4. EXPONENTIAL KERNELS

The result in Sec. 3 is not very useful for practical purposes, since the Laplace transform of a general $f(t)$ is not easy to obtain and to invert the transform in (3.3) is even more difficult. However, when $f(t)$ is an exponential function, or even a sum of exponential functions, an explicit Laplace transform and its inverse can be obtained.

Take

$$f(t)=\sum_{i=1}^{n}\nu_i e^{-\mu_i t}, \quad \nu_i,\mu_i \text{ real and } \mu_i>0. \tag{4.1}$$

Then

$$\hat{f}(s)=\sum_{i=1}^{n}\frac{\nu_i}{s+\mu_i}.$$

If we now substitute in (3.3), an explicit inverse transform can indeed be obtained, although in terms of complicated multiple integrals, which again is not very useful.

Also we may remark that with the exponential kernels, the integro-differential equation can be transformed into a partial differential equation by successive differentiation with respect to t. For the kernel given in (4.1), the differential equation is of the form

$$\begin{vmatrix} 1 & 1 & \cdots & 1 & U_0 \\ \mu_1 & \mu_2 & \cdots & \mu_n & U_1 \\ \mu_1^2 & \mu_2 & \cdots & \mu_n^2 & \vdots \\ \vdots & & & & \vdots \\ \mu_1^n & \mu_2^n & \cdots & \mu_n^n & U_n \end{vmatrix} = 0,$$

where

$$U_l=(-1)^l\left(\frac{\partial^l}{\partial t^l}(u_t+u_x)+\sum_{m=0}^{l-1}\sum_{i=1}^{n}(-1)^{l-m}\nu_i\mu_i^{l-m-1}\frac{\partial^m}{\partial t^m}u_x\right). \tag{4.2}$$

The simplest exponential kernel is the case of $n=1$. This case, though simple, illustrates many points which may be of more general validity. For this case, the original linear equation becomes

$$[u_t+u_x]_t+\mu u_t+(\mu-\nu)u_x=0. \tag{4.3}$$

Thus the original differentio-integral equation is equivalent to the differential equation (4.2). In particular, if $\mu=\nu$, then (4.2) can be integrated immediately once to yield

$$u_t+u_x+\mu[u-u_0(x)]=0. \tag{4.4}$$

Equation (4.3) may be solved by the method of Laplace transformation to give

$$\begin{aligned}F(z,t)=e^{-\nu z}\{&H(t-z)H(z)e^{-(\mu-\nu)(t-z)}\\&\times[\nu I_0(2\sqrt{\nu(\mu-\nu)z(t-z)})+\sqrt{\nu(\mu-\nu)z/(t-z)}\\&\times I_1(2\sqrt{\nu(\mu-\nu)z(t-z)})]+\delta(t-z)\}.\end{aligned} \tag{4.5}$$

Hence

$$\begin{aligned}u(x,t)=&u_0(x-t)e^{-\nu t}\\&+\int_0^t dz u_0(x-z)e^{-(\mu-\nu)t+(\mu+2\nu)z}\\&\times[\nu I_0(2\sqrt{\nu(\mu-\nu)z(t-z)})+\sqrt{\nu(\mu-\nu)z/(t-z)}\\&\times I_1(2\sqrt{\nu(\mu-\nu)z(t-z)})].\end{aligned} \tag{4.6}$$

For the particular case $\nu=\mu$, the solution reduces to

$$u(x,t)=u_0(x-t)e^{-\mu t}+\mu\int_0^t u_0(x-z)e^{-\mu z}.$$

Thus the first term represents the decay of the wave-packet, while the second term represents the smearing-out effect of the relaxation.

Now, let us study specifically the small time and large time behavior of the solution (4.6).

(i) Small time behavior: We obtain from (4.6) by straightforward Taylor expansion

$$u(x,t)=u_0(x)-tu_0'(x)+O(t^2).$$

Thus, the behavior is essentially governed by the wave operator, since there is not sufficient time for the relaxation mechanism to take effect.

(ii) Large time behavior: Making use of the asymptotic expansion of Bessel function, we may obtain from integration by parts that, for $\mu>\nu>0$,

$$\begin{aligned}u(x,t)\sim\ &u_0(x-t)e^{-\nu t}\\&+\frac{\nu}{\sqrt{\pi}}\frac{\exp\{-(\mu-\nu)t+2[\nu(\mu-\nu)tx]^{1/2}\}}{[\nu(\mu-\nu)t]^{1/4}}\\&\times\left\{\sum_{n=0}\frac{\phi^{(n)}(\sqrt{x})}{[4\nu(\mu-\nu)t]^{n+1/2}}\left[1+O\left(\frac{1}{t^{1/2}}\right)\right]\right\},\end{aligned}$$

where $\phi(\eta)=\eta^{1/2}u_0(x-\eta^2)e^{(\mu-2\nu)\eta^2}$.

If $u_0^{(n)}(0)\neq 0$, but $u_0^{(m)}(0)=0$ for $m<n$, then

$$u(x,t) \sim u_0(x-t)e^{-\nu t} + \frac{\nu}{\sqrt{\pi}} \frac{(-1)^n}{[\nu(\mu-\nu)t]^{1/4}} \frac{2^n x^{\frac{1}{2}n+\frac{1}{4}} u_0^{(n)}(0) \exp\{-(\mu-\nu)t + (\mu-2\nu)x + 2[\nu(\mu-\nu)tx]^{1/2}\}}{[4\nu(\mu-\nu)t]^{(n+1)/2}}. \quad (4.7)$$

On the other hand, if $\mu > 0$, $\nu < 0$, or $\nu > \mu > 0$, then we obtain

$$u(x,t) \sim u_0(x-t)e^{-\nu t} + \frac{2\nu}{\pi^{1/2}} \frac{e^{-(\mu-\nu)t}}{[\nu(\mu-\nu)t]^{1/4}} \times \left(\sum_{n=0} \frac{\phi^n(\sqrt{x}) \cos\{2[\nu(\mu-\nu)tx]^{1/2} + (\frac{1}{2}n - \frac{3}{4})\pi\}}{[4\nu(\mu-\nu)t]^{n+1/2}} + \sum_{n=0} \frac{\Gamma(n+\frac{1}{2})}{n!} \frac{\psi^n(0)\cos\frac{1}{2}n\pi}{[4\nu(\mu-\nu)t]^{n+1/2}} \right)[1 + O(t^{-1/2})],$$

where $\psi(\eta) = \eta^{1/2}\phi(\eta)$.

In particular, if $u_0(0) \neq 0$, then

$$u(x,t) \sim u_0(x-t)e^{-\nu t} + \frac{2\nu}{\pi^{1/2}} \frac{1}{[\nu(\mu-\nu)t]^{1/4}} \times \frac{x^{1/4} u_0(0) e^{-(\mu-\nu)t + (\mu-2\nu)x} \cos\{2[\nu(\mu-\nu)tx]^{1/2} - \frac{3}{4}\pi\}}{[4\nu(\mu-\nu)t]^{1/2}} \quad (4.8)$$

For finite x, the first term retained in (4.7) and (4.8) is in fact zero for t large enough, since we have assumed at the outset that $u_0(x)$ is nonvanishing only in a finite interval. However, this term clearly shows that $u(x,t)$ will be exponentially large at $x \sim t$ for large t, if $\nu > 0$. The amplitude of $u(x,t)$ also becomes exponentially large for the case $\nu > \mu$, as may be seen from (4.8). These two cases are unstable. Indeed, they violate the stability condition established by Whitham.[5] Therefore, the only physically meaningful case is the case $\mu > \nu > 0$, which appropriately represents a mechanism of relaxation.

For the particular case of $\mu = \nu$, we obtain

$$u(x,t) \sim \int_0^x u_0(x-z)e^{-\mu z} dz \qquad \text{as } t \to \infty, \quad x \text{ finite};$$

hence u is independent of t.

5. SOLUTION BY ITERATION EXPANSION

For the general linear equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = \int_0^t u_x(x, t-\tau) f(\tau) d\tau \quad (5.1)$$

a convergent iteration expansion can be established. To do this, let us introduce new variables (ξ, η) such that

$$\xi = x - t, \qquad \eta = t, \quad (5.2)$$

also let us define $v(\xi,\eta) = u(\xi+\eta, \eta)$, then Eq. (5.1) becomes

$$\frac{\partial v}{\partial \eta} = \frac{\partial}{\partial \xi} \int_\xi^{\xi+\eta} v(s, \xi+\eta-s) f(s-\xi) ds, \quad (5.3)$$

and the initial condition becomes

$$v(\xi, 0) = u_0(\xi). \quad (5.4)$$

Let us now formally write

$$v(\xi,\eta) = \sum_{n=0}^{\infty} v_n(\xi,\eta). \quad (5.5)$$

Then Eq. (5.3) will be satisfied if

$$\frac{\partial v_0}{\partial \eta} = 0, \quad (5.6)$$

$$\frac{\partial v_n}{\partial \eta} = \frac{\partial}{\partial \xi} \int_\xi^{\xi+\eta} v_{n-1}(s, \xi+\eta-s) f(s-\xi) ds = F_n(\xi,\eta), \qquad n \geq 1, \quad (5.7)$$

and if the expansion (5.5) is uniformly convergent.

Now it can be shown that if $|f^{(n)}(t)| \leq M$ for all t and n and $|u_0(x)| \leq U$, where M and U are two constants, then the series (5.5) is uniformly convergent for any finite η.

To prove this, we observe from (5.7) that $F_n(\xi,\eta)$ is majorized by the following expression:

$$|F_n(\xi,\eta)| \leq \sum_{m=0}^{n} a_{nm} \frac{\eta^{n+m-1}}{(n+m-1)!} \qquad \text{for all } \xi,$$

where

$$a_{10} \leq 2UM, \qquad a_{nm} \leq \tfrac{4}{3} UM(3M)^{n-1}, \qquad \text{other } n, m \geq 0.$$

Now in the series on the right-hand side of the following inequality:

$$\sum_{n=0}^{\infty} |F_n(\xi,\eta| \leq \sum_{n=0}^{\infty} \sum_{m=0}^{n} a_{nm} \frac{\eta^{n+m-1}}{(n+m-1)!},$$

there are $(n+2)/2$ or $(n+3)/2$ terms of $\eta^n/n!$. The coefficient of each term is smaller than $MU(3M)^n$, if $3M$ is chosen to be larger than 1. Therefore, the coefficient of η^n will be less than $MU(3M)^n/(n-1)!$ for $n > 1$. Hence the series $\sum_{n=0}^{\infty} |F_n(\xi,\eta)|$ is convergent. A direct integration of Eq. (5.7) leads to the convergence of $\sum_{n=0}^{\infty} v_n(\xi,\eta)$ also. This expansion converges rapidly if the relaxation kernel f is small.

6. LARGE TIME SOLUTION

The behavior of the solution (3.6) for t large can be viewed from two directions. One is the solution for $t \to \infty$, keeping x finite. The other is the behavior of the main disturbance as $t \to \infty$. The latter is often physically more significant.

The behavior of the solution for $t \to \infty$, while keeping x finite, may be studied in a similar approach as we did in Sec. 4. Without loss of generality, we can take $u_0(x)$ to be nonvanishing only in the interval $(0, x_0)$. Then the solution (3.6) can be written as

$$u(x,t) = \int_0^x u_0(x-z)F(z,t)dz, \qquad \text{for } t \geq x. \quad (6.1)$$

Therefore, for any finite x, z is always small in comparison with t, when t is large enough. Hence in (6.1), we can evaluate the asymptotic expression of $F(z,t)$ for t large by treating z as a finite parameter. To do this, let us write $h(s) = 1 - \hat{f}(s)$, and let the λ_j's be the zeros of $h(s)$, i.e., $h(\lambda_j) = 0$.

Let us order $\lambda_j = a_j + ib_j$ by $a_0 \geq a_1 \geq a_2 \geq \cdots$. Assume $1/h(s)$ can be expanded in the neighborhood of $s = \lambda_0$, with

$$[h(s)]^{-1} = \sum_{n=0}^{\infty} \alpha_n (s-\lambda_0)^{n-1}.$$

Then the asymptotic expression of $F(z,t)$ may be obtained by a similar approach as given in Carslaw and Jaeger.[6] It is found that when $\alpha_0\lambda_0$ is not real and positive,

$$F(z,t)\sim\alpha_0\exp[\lambda_0 t-(\alpha_0+\lambda_0\alpha_1)z]\frac{e^{i(2a\sqrt{zt}-\pi/4)+2b\sqrt{zt}}}{2\sqrt{\pi}(\alpha_0\lambda_0 zt)^{1/4}}, \tag{6.2}$$

where the branch of $\sqrt{\alpha_0\lambda_0}=a-ib$, $b>0$ is chosen. When $\alpha_0\lambda_0$ is real and positive, take $a=|\sqrt{\alpha_0\lambda_0}|$; it is found that

$$F(z,t)\sim\alpha_0\exp[\lambda_0 t-(\alpha_0+\lambda_0\alpha_1)z]\times\frac{\cos(2\sqrt{\alpha_0\lambda_0 zt}-\pi/4)}{\sqrt{\pi}(\alpha_0\lambda_0 zt)^{1/4}}. \tag{6.3}$$

The asymptotic expression of $u(x,t)$ for large t may be obtained in a manner similar to that of Sec. 4. Let

$$\phi(\eta)=\eta^{1/2}u_0(x-\eta^2)\exp[-(\alpha_0+\lambda_0\alpha_1)\eta^2],\qquad \psi(\eta)=\eta^{1/2}\phi(\eta);$$

then, the asymptotic expression of $u(x,t)$ is given by the leading term of the following expansion:

$$\begin{aligned}u(x,t)&\sim\frac{\alpha_0}{\sqrt{\pi}}\frac{e^{\lambda_0 t}e^{2b\sqrt{xt}}e^{i(2a\sqrt{xt}-\pi/4)}}{(\alpha_0\lambda_0 t)^{1/4}}\\&\quad\times\left(\sum_{n=0}\phi^{(n)}(\sqrt{x})(4\alpha_0\lambda_0 t)^{-(n+1)/2}\right)[1+O(t^{-1/2})]\\&=\frac{\alpha_0}{\sqrt{\pi}}\frac{\exp(\lambda_0 t)\exp[i(2\sqrt{\alpha_0\lambda_0 xt}-\pi/4)]}{(\alpha_0\lambda_0 t)^{1/4}}\\&\quad\times\sum_{n=0}\phi^n(\sqrt{x})[4\alpha_0\lambda_0 t]^{-(n+1)/2}\}[1+O(t^{-1/2})]\end{aligned}$$

$$\text{for }\sqrt{\alpha_0\lambda_0}=a-ib,\quad b>0, \tag{6.4}$$

and

$$\begin{aligned}u(x,t)&\sim\frac{2\alpha_0}{\sqrt{\pi}}e^{\lambda_0 t}\left(\sum_{n=0}\frac{\dot\phi^{(n)}\sqrt{x}\cos[2\sqrt{\alpha_0\lambda_0 xt}+(\frac{1}{2}n-\frac{3}{4})\pi]}{(4\alpha_0\lambda_0 t)^{(n+1)/2}}\right.\\&\quad\left.+\sum_{n=0}\frac{\Gamma(n+\frac{1}{2})\psi^{(n)}(0)\cos\frac{1}{2}n\pi}{n!(4\alpha_0\lambda_0 t)^{(n+1)/2}}\right)[1+O(t^{-1/2})]\end{aligned}$$

$$\text{for }\sqrt{\alpha\lambda_0}>0\text{ real.} \tag{6.5}$$

From these asymptotic expansions, we can conclude that the stability of the system is governed by the sign of the real part of λ_0. From the definition of λ_0, we can then establish the following stability criterion for positive kernels: The solution is stable if

$$\int_0^\infty f(t)dt<1. \tag{6.6}$$

To prove the last statement, we note that for positive kernels i.e., $f(t)\geq 0$ for all $t\geq 0$, the function $\hat f(s)=\int_0^\infty e^{-st}f(t)dt$ is a monotonously decreasing function of s for real s. Hence, there always exists a real s_0 such that $\hat f(s_0)=1$. Furthermore, $s_0<0$ if and only if $\int_0^\infty f(t)dt<1$. If we can further identify s_0 to be λ_0, then our proof is complete.

Let us take any $\lambda_j=a_j+ib_j$, such that $a_j>s_0$; then

$$\hat f(\lambda_j)=\int_0^\infty e^{-a_j t+ib_j t}f(t)dt$$

and

$$|\operatorname{Re}\hat f(\lambda_j)|=\left|\int_0^\infty e^{-a_j t}\cos b_j t\,f(t)dt\right|\leq\int_0^\infty e^{-a_j t}f(t)dt<1.$$

Therefore, $\hat f(\lambda_j)=0$ is possible only if $a_j\leq s_0$.

We may also notice that $\hat f'(\lambda_0)$ is real and negative; hence α_0 is real and positive. Thus the stable solution is represented by Eq. (6.4) with $a=0$.

Very often we are more interested in the lasting main disturbance after the passing of some initial time interval. The main disturbance will then propagate not with the speed 1 as determined by the real characteristics of the problem, but with some other speed less than 1. This is equivalent to finding the asymptotic expression of (3.6) for large t and x, with the parameter x/t kept fixed.

Let us rewrite (3.6) as

$$u(x,t)=\int_0^1 F(ty,t)u_0(x-ty)t\,dy \tag{6.7}$$

where

$$F(ty,t)=(1/2\pi i)\int_c\{e^{tw(y,s)}/[1-\hat f(s)]\}ds,$$

and

$$w(y,s)=s-y\,s/[1-\hat f(s)]\equiv s-y\,g(s). \tag{6.8}$$

The asymptotic expression for large t then can be obtained by the method of steepest descent. For the asymptotic expression of $F(ty,t)$, let a saddle point be

$$s=k_1(y),$$

which is obtained from

$$\frac{\partial w}{\partial s}=1-y\,g'(s)=0. \tag{6.9}$$

Then, this contribution to $F(ty,t)$ is

$$F(ty,t)\sim\frac{e^{i\gamma_1}}{2\pi i}\frac{1}{1-\hat f(k_1(y))}\left[\frac{2\pi}{t\,|\,yg''(s)|_{s=k_1(y)}}\right]^{1/2}e^{tJ(y)}, \tag{6.10}$$

where $J(y)=w(y,k_1(y))$, and γ_1 is some real constant arising from the possible deformation of the contour.

Substituting (6.10) into (6.7), we can once more perform the integration by the method of the steepest descent. Via (6.8) and (6.9), it may be shown[3] that, from $dJ/dy=0$, the saddle point is given by $k_1(y)=0$, or

$$y_0=1/g'(0)=1/[1-\hat f(0)]. \tag{6.11}$$

Moreover, at this saddle point,

$$\frac{d^2J}{dy^2}=\frac{g'(0)^3}{g''(0)} \tag{6.12}$$

Let us confine our problem to stable positive kernels so that

$$\hat f(0)=\int_0^\infty f(t)dt<1;$$

then both $g'(0)$ and $g''(0)$ are real, and we have $0<y_0<1$. With these observations, it is then found[3] that

$$u(x,t)\sim u_0[x-(1-c)t]\quad\text{as } t\to\infty, \tag{6.13}$$

where

$$c=\int_0^\infty f(t)dt. \tag{6.14}$$

Thus we may conclude that the lasting main disturbance will propagate with a speed reduced from the real characteristic speed by an amount c, and retaining essentially the initial shape, if $f(t)\geq 0$ and $\int_0^\infty f(t)dt=c<1$.

As an example, let us consider the case

$$f(t) = \nu_1 e^{-\mu_1 t} + \nu_2 e^{-\mu_2 t}. \tag{6.15}$$

Then

$$c = \nu_1/\mu_1 + \nu_2/\mu_2,$$
$$1 - c = (1/\mu_1\mu_2)(\mu_1\mu_2 - \nu_1\mu_2 - \nu_2\mu_1);$$

therefore,

$$u(x,t) \sim u_0[x - (1/\mu_1\mu_2)(\mu_1\mu_2 - \nu_1\mu_2 - \nu_2\mu_1)t], \quad \text{as } t \to \infty, \tag{6.16}$$

if

$$\nu_1/\mu_1 + \nu_2/\mu_2 < 1 \quad \text{or}$$
$$(1/\mu_1\mu_2)(\mu_1\mu_2 - \nu_1\mu_2 - \nu_2\mu_1) < 1, \tag{6.17}$$

$\nu_1, \mu_1, \nu_2, \mu_2$ all positive.

Now for $f(t)$ given by (6.15), the equation

$$u_t + u_x - \int_0^t u_x(t,\tau) f(t-\tau)d\tau = 0.$$

can be transformed, by successive differentiation, into the following differential equation:

$$[u_t + u_x]_{tt} + (\mu_1 + \mu_2) \times [u_t + [(\mu_1 + \mu_2 - \nu_1 - \nu_2)/(\mu_1 + \mu_2)]u_x]_t + \mu_1\mu_2[u_t + [(\mu_1\mu_2 - \mu_1\nu_2 - \nu_2\mu_1)/\mu_1\mu_2]u_x] = 0. \tag{6.18}$$

An intuitive approach will lead us to conclude that, for large t, the main disturbance will be governed by the lowest order terms in Eq. (6.18):

$$u_t + [(\mu_1\mu_2 - \mu_1\nu_2 - \nu_2\mu_1)/\mu_1\mu_2]u_x = 0, \tag{6.19}$$

which will yield the solution (6.16). The condition (6.17) is essentially that the characteristic direction of Eq. (6.19) should lie between the characteristic directions of the higher order terms, an extension of Whitham's result.[5] We may note that the asymptotic solution (6.13) can also be applied to rapidly decaying kernels. For instance, take

$$f(t) = f(0)e^{-\mu h(\tau)},$$

where $h(\tau)$ is a monotonously increasing function of τ, $h(0) = 0$, $h'(0) = 1$, and μ is a large parameter. In this case time is scaled by $1/\mu$. Therefore, when μ is large, any finite t will correspond to asymptotically large time on this relevant scale.

7. THE NONLINEAR EQUATION

Let us consider now the nonlinear equation (2.1),

$$u_t + (1+u)u_x = \int_0^t (1+u)u_x f(t-\tau)d\tau = \frac{\partial}{\partial x}\int_0^t uf(t-\tau)d\tau + \frac{1}{2}\frac{\partial}{\partial x}\int_0^t u^2 f(t-\tau)d\tau, \tag{7.1}$$

with the initial condition

$$u(x,0) = u_0(x). \tag{7.2}$$

It is clear, from our knowledge about the nonlinear partial differential equation without the integral term, that a solution may not exist for all $t > 0$, even with $u_0(x)$, which is continuously differentiable to any desired order, because shocks may appear.

The natural way to deal with (7.1) is to make use of the characteristic coordinate. Let us introduce the set of new variables (ξ, η),

$$\xi = \xi(x,t), \tag{7.3}$$

$$\eta = t, \tag{7.4}$$

such that

$$\frac{dx}{dt} = (1+u), \quad \text{on } \xi(x,t) = \text{const}, \tag{7.5}$$

or

$$\xi_t + (1+u)\xi_x = 0. \tag{7.6}$$

Thus, Eq. (7.1) can be rewritten as

$$\frac{\partial u}{\partial \eta} = \xi_x \frac{\partial}{\partial \xi}\left(\int_0^\eta \{u[x(\xi,\eta),\tau] + \tfrac{1}{2}u^2[x(\xi,\eta),\tau]\}f(\eta-\tau)d\tau\right) = \xi_x \frac{\partial}{\partial \xi}\int_0^\eta \{u[x(\xi,\eta),\eta-\tau] + \tfrac{1}{2}u^2[x(\xi,\eta),\eta-\tau]\}f(\tau)d\tau. \tag{7.7}$$

For the linear case, we have $\xi = x - t$; then (7.7) becomes identical with (5.3). With initial data prescribed, we can attempt to integrate step by step Eq. (7.7). The integration process, however, may not in general be carried on indefinitely. On the one hand, the nonlinear term in the integral may cause divergence in the expansion for large enough η. On the other hand, the characteristics given by (7.5) may intersect and multiple-valued solution, or shock, will appear. It is, however, also possible that the relaxation terms in the integral may help preventing the characteristics from intersecting with each other, which would otherwise intersect if there is no relaxation.

The nonlinearity enters both in the characteristic equation and the integral. Even when the nonlinearity is present only in the integral, whether an iteration procedure similar to that presented in Sec. 5 can be carried out for any finite t is not certain. To explore along this direction, let us consider the following equation:

$$u_t + u_x = \int_0^t (1+u)u_x f(t-\tau)d\tau. \tag{7.8}$$

This implies

$$\xi(x,t) = x - t.$$

Now, as in Sec. 5, denote

$$v(\xi,\eta) = u(\xi+\eta,\eta);$$

then Eq. (7.7) becomes

$$\frac{\partial v}{\partial \eta} = \frac{\partial}{\partial \xi}\int_\xi^{\xi+\eta}[v(s,\xi+\eta-s) + \tfrac{1}{2}v^2(s,\xi+\eta-s)]f(s-\xi)ds. \tag{7.9}$$

Let us now try to solve (7.9) by an iteration procedure, i.e.,

$$v = \lim_{n\to\infty} v_n = v_0 + \sum_{n=1}^{\infty}(v_n - v_{n-1}).$$

Hence, we obtain from (7.9)

$$\frac{\partial v_0}{\partial \eta} = 0, \tag{7.10}$$

$$\frac{\partial v_n}{\partial \eta} = \frac{\partial}{\partial \xi}\int_{\xi}^{\xi+\eta}[v_{n-1}(s, \xi + \eta - s) + \tfrac{1}{2} v_{n-1}^2(s, \xi + \eta - s)]f(s - \xi)ds, \tag{7.11}$$

or, with the definition $v_{-1} = 0$,

$$\frac{\partial (v_n - v_{n-1})}{\partial \eta} = \frac{\partial}{\partial \xi}\int_0^{\xi+\eta}\{v_{n-1} - v_{n-2} + \tfrac{1}{2}v_{n-1}^2 - \tfrac{1}{2}v_{n-2}^2\}_{(s,\xi+\eta-s)}f(s - \xi)ds, \quad n \geq 1. \tag{7.12}$$

The initial conditions are

$$v_n(\xi, 0) - v_{n-1}(\xi, 0) = 0, \text{ all } \xi. \tag{7.13}$$

Now let us consider only the domain such that $|v_n| \leq N$ for all n, i.e., consider only for those $\eta < \eta_0$ such that $|v_n(\xi, \eta)| \leq N$. Let $\max |u_0(x)| \leq U_0$ and $|f^{(n)}(t)| \leq M$ for all t and n. It is possible [3] to obtain a rough estimate from (12) that this iteration procedure can be carried out for all η such that

$$\tfrac{4}{3}(1 + N)M\eta\, e^{3(1+N)M\eta} \leq N/U_0. \tag{7.14}$$

Although we have only established the sufficient condition for the existence of the solution of Eq. (7.8), these results indicate that the iteration procedure may not in general be carried out indefinitely. At certain $\eta = \eta_0$, this procedure may break down.

8. A PARTICULAR EXPONENTIAL KERNEL

Let us consider the case that $f(t) = e^{-\mu t}$. Then Eq. (7.1) becomes

$$u_t + (1 + u)u_x = \mu \int_0^t (1 + u)u_x\, e^{-\mu(t-\tau)}d\tau. \tag{8.1}$$

Differentiate (8.1) with respect to t, and, using the initial condition, we obtain

$$u_t + (1 + u)u_x + \mu u = \mu u_0(x). \tag{8.2}$$

Let us introduce the variable (ξ, η) as in Sec. 7; then Eq. (8.2) becomes

$$\frac{\partial u}{\partial \eta} + \mu u = \mu u_0[x(\xi, \eta)], \tag{8.3}$$

where

$$\frac{\partial x}{\partial \eta} = (1 + u), \quad \xi = \xi(x, t), \quad \eta = t, \tag{8.4}$$

and we can assign the initial condition as

$$x(\xi, 0) = \xi. \tag{8.5}$$

Equation (8.4) can be integrated, and we obtain

$$u(\xi, \eta) = u_0(\xi)e^{-\mu\eta} + \mu e^{-\mu\eta}\int_0^{\eta} e^{\mu\tau}u_0[x(\xi, \tau)]d\tau. \tag{8.6}$$

Thus

$$\frac{\partial x}{\partial \eta} = 1 + u_0(\xi)e^{-\mu\eta} + \mu e^{-\mu\eta}\int_0^{\eta} e^{\mu\tau}u_0[x(\xi, \tau)]d\tau. \tag{8.7}$$

Hence

$$x(\xi, \eta) = \xi + \eta + \frac{u_0(\xi)}{\mu}(1 - e^{-\mu\eta}) + \mu\int_0^{\eta} ds\, e^{-\mu s}\int_0^{s} e^{\mu\tau}u_0[x(\xi, \tau)]d\tau. \tag{8.8}$$

Shock will appear as soon as the characteristics will intersect, i.e., when

$$x(\xi_i, \eta) = x(\xi_j, \eta), \quad \text{for any pair of } \xi_i \text{ and } \xi_j,$$

or when

$$\frac{\partial x}{\partial \xi} = 0. \tag{8.9}$$

From (8.8), we obtain

$$\frac{\partial x}{\partial \xi} = 1 + \frac{1}{\mu}(1 - e^{-\mu\eta})u_0'(\xi) + \mu\int_0^{\eta} ds e^{-\mu s}\int_0^{s} e^{\mu\tau}u_0'[x(\xi, \tau)]\frac{\partial x}{\partial \xi}(\xi, \tau)d\tau. \tag{8.10}$$

When there is no relaxation effect, i.e., when $\mu \to 0$, we have

$$\frac{\partial x}{\partial \xi} = 1 + \eta u_0'(\xi). \tag{8.11}$$

Thus, the criterion (8.9) means

$$\eta = -1/u_0'(\xi). \tag{8.12}$$

Hence shock will appear near the characteristic $\xi = \xi_m$, where $u_0'(\xi)$ is minimum, and it will occur at

$$t = \eta = 1/-u_0'(\xi_m). \tag{8.13}$$

For this case, we see that so long as there exists some ξ_m where $u_0'(\xi_m) < 0$, then shock will appear sooner or later. Since for any pulse of finite duration $u_0'(\xi)$ cannot be always positive, we expect that there will eventually be a shock.

When the relaxation effect is present, whether there will be shock or not at all depends very much on the magnitude of $\mu, u_0(\xi_m)$ and the length of the pulse.

Again consider the case that the length of the pulse is finite. Assume

$$u_0(x) = 0, \quad \text{for } x < 0 \text{ and } x > l,$$
$$u_0(x) \geq 0, \quad \text{for } 0 < x < l.$$

Also let $\max u_0'(x) = m_1 > 0$, and $\min u_0'(x) = m_2 < 0$.

Let us take M to be the least upper bound of $|\partial x/\partial \xi|$. The existence of an upper bound of $|\partial x/\partial \xi|$ can be established if m_1 is small enough. From (8.10), a sufficient condition[3] can be found to be

$$(m_1/\mu)(e^{\mu l} - 1) < 1.$$

Then, it can be derived from Eq. (8.10) that no shock will appear if

$$(1 - e^{-\mu\eta})m_2/\mu < 1/[1 + M(e^{\mu l} - 1)] \tag{8.14}$$

The condition (8.14) is only a sufficient condition; better conditions can, of course, be obtained if we know more about the shape of the initial pulse. However, it clearly demonstrates that the formation of shock will depend not only on the maximum slope of the pulse but also on the length of the pulse, and, when

$$m_2/\mu < 1/[1 + M(e^{\mu l} - 1)], \tag{8.15}$$

no shock will appear at all.

9. THE EXPONENTIAL KERNEL $f(t) = \nu e^{-\mu t}$; WEAK NONLINEARITY

For the case $f(t) = \nu e^{-\mu t}$, Eq. (7.1) becomes

$$u_t + (1+u)u_x = \nu \int_0^t (1+u)u_x\, e^{-\mu(t-\tau)}\, d\tau. \quad (9.1)$$

Differentiate (9.1) with respect to t; we obtain

$$[u_t + (1+u)u_x]_t = -\mu\{u_t + [(\mu-\nu)/\mu](1+u)u_x\}. \quad (9.2)$$

Equation (9.2) also belongs to the category that has been discussed by Whitham.[5] Thus, when $\mu > \nu > 0$, the main disturbance will be associated with the lower-order terms; i.e., as far as the main disturbance is concerned, we have approximately

$$u_t + [(\mu-\nu)/\mu](1+u)u_x = 0. \quad (9.3)$$

The shock associated with (9.3) will be smoothed out by the higher-order terms. The real shock of the problem, if they exist, of course, is still associated with the higher-order terms, and they will in general be damped by the lower-order terms. The lower-order terms may also modify the higher-order terms to such an extent that the real shocks will not form at all.

If we integrate (9.2) with respect to t, we obtain

$$u_t + (1+u)u_x = -\mu[u - u_0(x)] - (\mu-\nu)\int_0^t (1+u)u_x d\tau. \quad (9.4)$$

From (9.1) and (9.4), we obtain

$$\mu[u - u_0(x)] = \int_0^t (1+u)u_x[\nu e^{-\mu(t-\nu)} - (\mu-\nu)]d\tau. \quad (9.5)$$

Thus, for $\mu \gg \nu$, we obtain

$$u - u_0(x) \approx -\int_0^t (1+u)u_x\, d\tau. \quad (9.6)$$

Then

$$u_t \approx -(1+u)u_x. \quad (9.7)$$

Hence the relaxation effect is minimal. Also the lower-order equation (9.3) is essentially the same as (9.7).

When $\mu - \nu \ll \mu$, as far as the higher-order terms and the real shocks are concerned, the results from Sec. 8 may apply essentially. On the other hand, the main disturbance will be propagated with the slow speed $[(\mu-\nu)/\mu](1+u)$ as given by (9.3).

For weakly nonlinear cases, a perturbation iteration scheme can be devised. Take (7.5) and (7.7):

$$\frac{dx}{dt} = (1+u), \quad \text{on } \xi(x,t) = \text{const}, \quad (9.8)$$

and

$$\frac{\partial}{\partial\eta} u[x(\xi,\eta),\eta] = \xi_x \frac{\partial}{\partial\xi}\int_0^\eta \{u[x(\xi,\eta),\eta-\tau] + \tfrac{1}{2}u^2[x(\xi,\eta),\eta-\tau]\}f(\tau)\,d\tau. \quad (9.9)$$

In (9.8), we shall first take $u = 0$ and obtain

$$\xi_1 = x - t, \quad x = \xi_1 + \eta. \quad (9.10)$$

Then in (9.9) we shall try to solve the linear equation

$$\frac{\partial}{\partial\eta} u_1(\xi+\eta,\eta) = \frac{\partial}{\partial\xi_1}\int_0^\eta \{u_1(\xi_1+\eta,\eta-\tau)\}f(\tau)\,d\tau. \quad (9.11)$$

From (9.11) we obtain

$$u_1 = u_1(\xi_1+\eta,\eta) \equiv v_1(\xi_1,\eta). \quad (9.12)$$

Then substitute u_1 for u in (9.8); i.e., we shall solve

$$\frac{\partial x}{\partial\eta} = 1 + v_1(\xi,\eta), \quad (9.13)$$

and obtain ξ as function of (x,t), which we call $\xi_2(x,t)$, i.e.,

$$\xi_2 = \xi_2(x,t) \quad \text{or} \quad x = x(\xi_2,\eta). \quad (9.14)$$

Then we write (9.7) as

$$\frac{\partial v_2(\xi_2,\eta)}{\partial\eta} = \frac{\partial u_2[x(\xi_2,\eta),\eta]}{\partial\eta} = \frac{\partial\xi_2}{\partial x}\frac{\partial}{\partial\xi_2}\int_0^\eta \{u_1[x(\xi_2,\eta),\eta-\tau] + \tfrac{1}{2}u_1^2[x(\xi_2,\eta),\eta-\tau]\}f(\tau)\,d\tau, \quad (9.15)$$

which we can integrate right away to obtain $v_2(\xi_2,\eta)$. So the process continues. Namely, we have

$$\frac{\partial v_n}{\partial\eta} = \frac{\partial\xi_n}{\partial x}\frac{\partial}{\partial\xi_n}\int_0^\eta \{u_{n-1}[x(\xi_n,\eta),\eta-\tau] + \tfrac{1}{2}u_{n-1}^2[x(\xi_n,\eta),\eta-\tau]\}f(\tau)\,d\tau \quad (9.16)$$

to obtain $v_n(\xi_n,\eta)$, which is $u_n[x(\xi_n,\eta),\eta]$. Then solve

$$\frac{\partial x}{\partial\eta} = 1 + v_n(\xi_{n+1},\eta) \quad (9.17)$$

to obtain

$$x = x(\xi_{n+1},\eta) \quad \text{or} \quad \xi_{n+1} = \xi_{n+1}(x,t). \quad (9.18)$$

Then we have the equation for u_{n+1} like (9.16) with n replaced by $n+1$.

For weakly nonlinear case, we hope to obtain good approximation with only one or two steps. In particular, $x = x(\xi_2,\eta)$ should give us much information about whether and when the shock will develop. As an example let us consider the case of $f(t) = \nu e^{-\mu t}$.

For this case, we have from (4.6)

$$u_1(x,t) = u_0(x-t)e^{-\nu t} + \int_0^t dz\, u_0(x-z)\, e^{-(\mu-\nu)t+(\mu-2\nu)z} \times [\nu I_0(2\sqrt{\nu(\mu-\nu)z(t-z)}) + \sqrt{\nu(\mu-\nu)z/(t-z)} \times I_1(2\sqrt{\nu(\mu-\nu)z(t-z)})].$$

Or

$$u_1(\xi_1,\eta) = u_0(\xi_1)e^{-\nu\eta} + \int_0^\eta dz\, u_0(\xi_1+\eta-z)\, e^{-(\mu-\nu)\eta+(\mu-2\nu)z} \times [\nu I_0(2\sqrt{\nu(\mu-\nu)z(\eta-z)}) + \sqrt{\nu(\mu-\nu)z/(\eta-z)} \times I_1(2\sqrt{\nu(\mu-\nu)z(\eta-z)})] \equiv u_0(\xi_1)e^{-\nu\eta} + \int_0^\eta dz\, F(\xi_1,\eta,z). \quad (9.19)$$

Thus from

$$\frac{\partial x}{\partial\eta} = 1 + u_1(\xi,\eta),$$

we obtain to the second approximation that, since the initial condition is chosen such that $x(\xi, 0) = \xi$,

$$x(\xi,\eta) = \eta + \xi + \frac{u_0(\xi)}{\nu}[1 - e^{-\nu\eta}] + \int_0^\eta d\tau \int_0^\tau F(\xi,\tau,z)\,dz.$$

Notice that

$$\int_0^\eta d\tau \int_0^\tau dz\ F(\xi,\tau,z) = \int_0^\eta dz \int_z^\eta d\tau\ F(\xi,\tau,z) = \int_0^\eta dz \int_0^{\eta-z} F(\xi,y+z,z)\,dy.$$

Hence we obtain

$$\begin{aligned} x(\xi,\eta) = \eta + \xi + \frac{u_0(\xi)}{\nu}(1 - e^{-\nu\eta}) &+ \int_0^\eta dz \int_0^{\eta-z} dy u_0(\xi+y)\, e^{-\nu z-(\mu-\nu)y} \\ &\times [\nu I_0(2\sqrt{\nu(\mu-\nu)yz}) + \sqrt{\nu(\mu-\nu)z/y} \\ &\times I_1(2\sqrt{\nu(\mu-\nu)yz})]. \end{aligned} \tag{9.20}$$

Thus

$$\begin{aligned} \frac{\partial x}{\partial \xi} = \frac{u_0'(\xi)}{\nu}(1 - e^{-\nu\eta}) &+ \int_0^\eta dz\, e^{-\nu z} \int_0^{\eta-z} dy\ u_0'(\xi+y)\, e^{-(\mu-\nu)y} \\ &\times [\nu I_0(2\sqrt{\nu(\mu-\nu)yz}) + \sqrt{\nu(\mu-\nu)z/y} \\ &\times I_1(2\sqrt{\nu(\mu-\nu)yz})] \end{aligned} \tag{9.21}$$

Again, when $\partial x/\partial\xi = 0$, the characteristics will intersect and shock will form. We may note that in (8.10), if we also make the approximation that

$$x(\xi,\eta) \approx \xi + \eta,$$

then it can also be written as

$$\frac{\partial x}{\partial \xi} = 1 + \frac{1}{\mu}(1 - e^{-\mu\eta})u_0'(\xi) + \mu \int_0^\eta dz\, e^{-\mu z} \int_0^{\eta-z} u_0'(\xi+y)\,dy, \tag{9.22}$$

which is identical to what we would obtain from (9.21) if we set $\mu = \nu$.

From (22), we have

$$\frac{\partial x}{\partial \xi} = 1 + \frac{1}{\mu}(1 - e^{-\mu\eta})u_0'(\xi) + \mu \int_0^\eta dz\, e^{-\mu z}[u_0(\xi+\eta-z) - u_0(\xi)]. \tag{9.23}$$

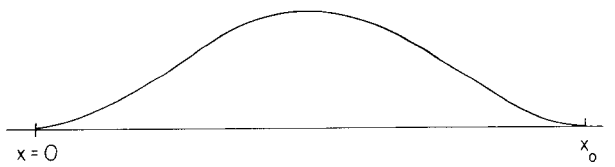


FIG. 1

For an initial pulse of finite duration like that shown in the Fig. 1. The greatest negative contribution from the second and third term in (9.23) occurs when $\eta \to \infty$, and since $u_0(x)$ is zero for $x > x_0$. We obtain, for the most critical case,

$$\frac{\partial x}{\partial \xi} = 1 + \frac{1}{\mu}u_0'(\xi) - u_0(\xi). \tag{9.24}$$

Thus, so long as

$$1 > u_0(\xi) - (1/\mu)u_0'(\xi),$$

for any ξ, the shock will not form. The most critical ξ is given by

$$u_0'(\xi) - (1/\mu)u_0''(\xi) = 0, \tag{9.25}$$

which lies between the maximum and the point of inflection of $u_0(\xi)$.

If μ is a large parameter, more can be said about (9.21). Rewrite (9.21) in the form

$$\begin{aligned} \frac{\partial x}{\partial \xi} = 1 + \frac{u_0'(\xi)}{\nu}(1 - e^{-\nu\eta}) &+ \int_0^\eta d\tau \int_0^\tau dz\ u_0'(\xi+\tau-z)\, e^{-(\xi-\nu)\tau+(\mu-2\nu)z} \\ &\times [\nu I_0(2\sqrt{\nu(\mu-\nu)z(\tau-z)}) + \sqrt{\nu(\mu-\nu)z/(\tau-z)} \\ &\times I_1(2\sqrt{\nu(\mu-\nu)z(\tau-z)})]. \end{aligned} \tag{9.26}$$

Using the asymptotic expansion of $I_n(x)$, we obtain for large μ

$$\begin{aligned} \frac{\partial x}{\partial \xi} &\sim 1 + \frac{u_0'(\xi)}{\nu}(1 - e^{-\nu\eta}) + \int_0^\eta d\tau\ u_0'\left(\xi + \tau - \frac{\mu-\nu}{\mu}\tau\right) \\ &= 1 + \frac{u_0'(\xi)}{\nu}(1 - e^{-\nu\eta}) + \frac{\mu}{\nu}\left[u_0\left(\xi + \frac{\nu}{\mu}\eta\right) - u_0(\xi)\right]. \end{aligned} \tag{9.27}$$

Hence the most critical case will be when

$$\frac{\partial x}{\partial \xi} \sim 1 + \frac{1}{\nu}u_0'(\xi) - \frac{\mu}{\nu}u_0(\xi). \tag{9.28}$$

The similarity between Eqs. (9.28) and (9.24) is worth noting.

Our study of the interaction between the nonlinearity and relaxation on the propagation of a simple wave is far from complete. However, it is clear that the effect of relaxation is distinct from that of ordinary dissipation. When the dissipation mechanism is also explicitly taken into account, whether the relaxation effect would then be completely overshadowed by the dissipation is the next question worth studying.

ACKNOWLEDGMENT

The results present here were obtained in the course of research sponsored in part by the Advanced Research Projects Agency under Contract No. DAHC-15-67-C-0217.

1 V. Volterra, *Theory of Functionals* (Dover, New York, 1959), p. 154.
2 G. B. Whitham, Proc. Roy. Soc. (London) **A299**, 6 (1967).
3 D. Y. Hsieh, "A Study of Relaxing Waves", Technical Report No. ARPA/AM-42 (Div. of Appl. Math., Brown University, 1970).
4 F. G. Tricomi, *Integral Equation* (Interscience, New York, 1957), p. 10.
5 G. B. Whitham, Commun. Pure Appl. Math. **12**, 113 (1959).
6 H. S. Carslaw and J. C. Jaeger, *Operational Methods in Applied Mathematics* (Oxford U. P., Oxford, 1953), p. 280.

Polarized Elastic Materials with Electronic Spin—A Relativistic Approach*

Gérard A. Maugin† and A. Cemal Eringen
Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, New Jersey 08540
(Received 27 May 1971; Revised Manuscript Received 12 August 1971)

In this article, we derive the balance laws and constitutive equations of polarized elastic solids with electronic spin by use of a relativistic (special) variational principle. The theory is fully dynamical and nonlinear. It is shown that this approach encompasses several previous works in micromagnetism (magnetoelastic interactions) and elastic dielectric theory.

1. INTRODUCTION

The peculiar form of the electromagnetic equations is definitively linked to the property of invariance under the Lorentz group. The special relativistic treatment assures of the good transformation properties of electromagnetic quantities. In Ref. 1 we have shown that the correct definition of the ponderomotive force acting upon a magnetized medium follows from the relativistic covariant formulation. For instance, if we examine the theories of deformable dielectrics developed during the last decade (cf. Toupin,[2] Eringen,[3] Toupin,[4] Dixon and Eringen,[5] and Grot and Eringen[6]), the fully satisfying one is that of Ref. 6, which employs the smallest number of hypotheses (practically only the Maxwell's equations written in four-dimensional formalism).

In the preceding article we tried to develop a theory of deformable magnetized materials in which account was taken of the electronic spin and the associated effects. However, to avoid cumbersome algebra, the theory developed was restricted to the case of quasi-magnetostatics, which forbids the large material velocities and fast propagations of discontinuities. The next step is to develop a fully dynamical theory in which both the magnetization and the polarization are taken into account. Thus there is need for a synthesized theory of both fields (deformable dielectrics and deformable magnetized bodies with electronic spin).

It is well known that a rigorous theory of magnetized and polarized media can only be achieved in the realm of special relativity theory. Indeed it is necessary to consider both magnetization and polarization. For instance, consider the transformation formulas for the magnetization $\mathbf{M}$ and the polarization $\mathbf{P}$ for a Lorentz mapping $\Lambda(\mathbf{v})$ (cf. Anderson[7]):

$$\mathbf{P}' = \gamma[\mathbf{P} + (1/c)\mathbf{v}\times\mathbf{M} + (\mathbf{v}/v^2)\mathbf{v}\cdot\mathbf{P}(\gamma^{-1} - 1)],$$

$$\mathbf{M}' = \gamma[\mathbf{M} - (1/c)\mathbf{v}\times\mathbf{P} + (\mathbf{v}/v^2)\mathbf{v}\cdot\mathbf{M}(\gamma^{-1} - 1)],$$

where

$$\gamma \equiv (1 - v^2/c^2)^{-1/2}.$$

The second equation tells us that a polarized moving body will appear to be magnetized. This is not surprising since moving charge distributions produce currents. What is more surprising is that a magnetized moving body will appear to be electrically polarized. Unfortunately few observable conclusions can be drawn due to the difficulty of obtaining sufficiently high velocities for material media. Of course, for practical calculations, one only needs the equations which may be deduced from the relativistic ones in the rest frame to within terms of the magnitude of $1/c^2$.

In this article we give a variational principle for nondissipative polarized and magnetized materials whose material points are equipped with electronic spins.

The present work generalizes works of Frenkel,[8] Taub,[9] Halbwachs,[10] Grot and Eringen,[6] Grot,[11] and Sedov[12] to include the electronic spin and a properly invariant theory of nonlinear solids. It encompasses the nonrelativistic works of Toupin,[13] Tiersten,[14] Brown,[15] and Maugin and Eringen.[1] The special relativistic treatment assures the correct transformation properties essential to electromagnetic fields while giving the nonrelativistic theory in the limit of small velocities. The complete set of field equations and related jump conditions are obtained and thermodynamics is given. By use of Lagrange's multipliers, certain constraints are duly included. The constitutive equations for the nondissipative electro-elastic solids are obtained. A reduced form following the application of the objectivity requirement is given and nonrelativistic limits are deduced.

The bases of our attempt are:

(i) a four-dimensional (Minkowskian) treatment,

(ii) a variational principle (Lagrangian) as a starting point,

(iii) an invariance (Lorentzian) requirement.

Our final goal is a dynamical theory of deformable polarized and magnetized media. The body is not necessarily saturated (as in Refs. 1 and 16), but we may have a variable magnetization amplitude in space and time.

2. KINEMATICS[17] IN V^4

Consider the four-dimensional manifold V^4 of Minkowski equipped with a hyperbolic normal metric of signature $(+,+,+,-)$. In an inertial frame, the square of an arc length is given by:

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2 - c^2(dt)^2 = \delta_{\alpha\beta}dz^\alpha dz^\beta, \tag{2.1}$$

where

$$(z^1, z^2, z^3, z^4) = (x, y, z, ict), \qquad i \equiv \sqrt{-1}$$

Here, (x, y, z) are rectangular coordinates, t is the time, c is the velocity of light in vacuum, and $\delta_{\alpha\beta}$ is the Kronecker symbol. Greek indices are assumed to take the values, 1, 2, 3, 4 and Latin indices (small or capital) the values 1, 2, 3. The proper time τ of an infinitesimal element of continuum is defined by

$$(d\tau)^2 = -(ds)^2/c^2. \tag{2.2}$$

In curvilinear coordinates, (2.1) reads

$$ds^2 = g_{\alpha\beta}dx^\alpha dx^\beta, \tag{2.3}$$

where $g_{\alpha\beta}$ is the metric tensor with signature $(+,+,+,-)$ and $g^{\alpha\beta}$ is its reciprocal given by

$$g_{\alpha\beta}g^{\beta\gamma} = \delta^\gamma_\alpha. \tag{2.4}$$

The motion of a continuous medium in V^4 is described by the set of relations

$$x^\alpha = x^\alpha(X^\Delta), \quad \alpha = 1,2,3,4, \qquad X^\Delta \equiv (X^K, \tau), \quad \Delta = 1,2,3,4, \quad K = 1,2,3, \tag{2.5}$$

where X^K are the Lagrangian coordinates of a material particle in E^3 and τ is a monotonically increasing timelike parameter along the world line ($\mathcal{C}$) of a particle in V^4, defined by (2.2).

The 4-velocity and 4-acceleration are respectively defined by:

$$u^\alpha = \frac{\partial x^\alpha}{\partial \tau} \quad \text{and} \quad \dot{u}^\alpha = \frac{\partial^2 x^\alpha}{\partial \tau^2} \tag{2.6}$$

It is not difficult to show that the operator $\partial/\partial\tau$ generalizes the notion of material derivative so that we can write

$$\frac{\partial \mathbf{A}}{\partial \tau} = \dot{\mathbf{A}} = u^\alpha \nabla_\alpha \mathbf{A}, \tag{2.7}$$

where ∇_α (or sometimes an index following a semicolon, e.g., $A_{\beta;\alpha}$) denotes the covariant partial derivative with respect to x^α. The modulus of the 4-velocity is constant, i.e.,

$$g_{\alpha\beta} u^\alpha u^\beta = -c^2. \tag{2.8}$$

In an inertial frame, we write

$$u^\alpha = (v^k/(1-\beta^2)^{1/2}, ic/(1-\beta^2)^{1/2}), \quad \beta = |\mathbf{v}|/c. \tag{2.9}$$

In a rest frame (inertial frame in which $v^k = 0$ at z^α), we have

$$u^\alpha = (0, ic), \quad \dot{u}^\alpha = (\dot{v}^k, 0), \quad u_{\alpha;\beta} = \begin{pmatrix} v_{k;l} & -i\dot{v}_k/c \\ 0 & 0 \end{pmatrix} \tag{2.10}$$

The motion (5) is postulated to be invertible so that

$$X^K = X^K(x^\alpha), \quad \tau = \tau(x^\alpha); \tag{2.11}$$

thus the following quantities are well-defined:

$$x^\alpha{}_{,K} = \frac{\partial x^\alpha}{\partial X^K}, \quad X^K{}_{,\alpha} = \frac{\partial X^K}{\partial x^\alpha}, \quad \tau_{,\alpha} = \frac{\partial \tau}{\partial x^\alpha}. \tag{2.12}$$

It is easily shown (Kafadar and Eringen[17]) that

$$X^K{}_{,\mu} x^\mu{}_{,L} = \delta^K{}_L, \quad \frac{\partial \tau}{\partial X^K} = 0, \quad \frac{\partial X^K}{\partial \tau} = X^K{}_{,\mu} u^\mu = 0, \qquad x^\mu{}_{,K} X^K{}_{,\lambda} = \delta^\mu_\lambda - u^\mu \tau_{,\lambda}. \tag{2.13}$$

In the sequel we need the *projection operator* used extensively in the literature. Let $V_\perp^3$ be the hypersurface orthogonal to the world line ($\mathcal{C}$) of a particle at the point **M** of the particle history in V^4. Clearly, every vector A^α such that $A^\alpha \subset V_\perp^3$ is spacelike. Any tensorial quantity associated with a point **M** in the Minkowski manifold may be projected onto $V_\perp^3$ by use of the *projector* **P** defined by [18-20]

$$P^\alpha{}_\beta \equiv \delta^\alpha_\beta + c^{-2} u^\alpha u_\beta, \quad P^\alpha{}_\alpha = 3. \tag{2.14}$$

For instance, any 4-vector F^α may be decomposed in a unique way into a 4-vector $f^\alpha \subset V_\perp^3$ and a component parallel to the 4-velocity[21]

$$F^\alpha = f^\alpha + u^\alpha f, \tag{2.15}$$

where

$$f^\alpha = P^\alpha{}_\beta F^\beta, \quad f = -c^{-2} F^\alpha u_\alpha.$$

It may be verified that

$$f^\alpha u_\alpha = 0, \tag{2.16a}$$

$$P^\alpha{}_\beta (u^\beta f) = 0. \tag{2.16b}$$

A 4-vector satisfying (2.16a) is said to be in $V_\perp^3$ and a 4-vector satisfying (2.16b) is parallel to the 4-velocity. A 4-vector $B^\alpha \subset V_\perp^3$ satisfies the identity

$$P^\alpha{}_\beta B^\beta \equiv B^\alpha. \tag{2.17}$$

The projector defined by (2.14) has two important properties:

$$P^\alpha{}_\beta P^\beta{}_\gamma = P^\alpha{}_\gamma \quad \text{(idempotence)}, \tag{2.18}$$

$$P^\alpha{}_\beta u^\beta = 0 \quad (\mathbf{P}^\alpha \subset V_\perp^3). \tag{2.19}$$

Given a 4-vector A^α, the relation $A^\alpha u_\alpha = 0$ assures that, in a rest frame, A^α reduces to $A^\alpha = (A^k, 0)$.

The deformation field: Using the projector, we can form the so-called direct gradient of the deformation field,

$$x^\alpha{}_K \equiv P^\alpha{}_\beta x^\beta{}_{,K}, \quad u_\alpha x^\alpha{}_K = 0, \tag{2.20}$$

that reduces to the classical gradient $x^k{}_{,K}$ in a rest frame. Using (2.20), we see that (2.13d) reads

$$x^\alpha{}_K X^K{}_{,\lambda} = P^\alpha{}_\lambda. \tag{2.21}$$

Thus $X^K{}_{,\alpha}$ will be called the inverse deformation gradient. The relativistic Green and Cauchy deformation tensors and their inverses are defined according to the relations:

$$C_{KL} = g_{\alpha\beta} x^\alpha{}_K x^\beta{}_L, \quad c_{\mu\lambda} = G_{KL} X^K{}_{,\mu} X^L{}_{,\lambda}, \qquad \overset{-1}{C}{}^{MN} = g^{\alpha\beta} X^M{}_{,\alpha} X^N{}_{,\beta}, \quad \overset{-1}{c}{}^{\mu\lambda} = G^{KL} x^\mu{}_K x^\lambda{}_L. \tag{2.22}$$

Finally we recall some useful expressions (cf. Grot and Eringen[6] and Kafadar and Eringen[17]:

$$\frac{\partial}{\partial \tau} X^K{}_{,\alpha} = -X^K{}_{,\beta} u^\beta{}_{;\alpha}, \quad \frac{\partial}{\partial \tau} x^\mu{}_{,K} = u^\mu{}_{;\lambda} x^\lambda{}_{,K}, \tag{2.23}$$

$$\frac{\partial}{\partial \tau} x^\mu{}_K = (u^\mu{}_{;\lambda} + c^{-2} u^\mu \dot{u}_\lambda) x^\lambda{}_K, \tag{2.24}$$

$$\tau_{,\mu} = (6ic)^{-1} \epsilon_{\alpha\beta\gamma\mu} x^\alpha{}_{,K} x^\beta{}_{,L} x^\gamma{}_{,M} \epsilon^{KLM}, \tag{2.25}$$

and

$$J \equiv (6ic)^{-1} \epsilon_{\alpha\beta\gamma\mu} x^\alpha{}_{,K} x^\beta{}_{,L} x^\gamma{}_{,M} u^\mu \epsilon^{KLM} \tag{2.26}$$

is the Jacobian which satisfies the relations

$$\frac{\partial J}{\partial \tau} = J u^\alpha{}_{;\alpha}, \quad J = [\det(C^K{}_L)]^{1/2}. \tag{2.27}$$

If we set

$$\overset{*}{u}{}^\alpha{}_\beta = P^\sigma{}_\beta u^\alpha{}_{;\sigma}, \quad \overset{*}{d}_{\alpha\beta} = \overset{*}{u}_{(\alpha\beta)}, \tag{2.28}$$

then $\overset{*}{d}_{\alpha\beta}$ represents the relativistic generalization of the deformation rate tensor as can be verified by

$$\frac{\partial}{\partial \tau} C_{KL} = 2 \overset{*}{d}_{\alpha\beta} x^\alpha{}_K x^\beta{}_L. \tag{2.29}$$

Given a material body (B) of boundary (∂B) in E^3, we shall call $(\mathcal{B})$ the tube swept out by (B) in V^4 as the proper time increases and $(\partial\mathcal{B})$ its boundary.

3. THE VARIATIONAL PRINCIPLE

The action: With any tube $(\mathcal{B}) \subset \mathbb{V}^4$, of boundary $(\partial\mathcal{B})$ swept out by a material body (B) in the 4-dimensional manifold of Minkowski V^4 and with V^4 itself, we associate the following action:

$$A = \int_{(\mathcal{B})} [\tfrac{1}{2}\rho\mathfrak{M}(u^\alpha u_\alpha + c^2) + \rho a_\alpha S^{\alpha\beta} u_\beta - \rho\psi(X^K{}_{,\alpha}, \tilde{\pi}^{\alpha\beta}, \Pi^{\alpha\beta}{}_K, \theta) - \tfrac{1}{2}\pi_{\alpha\beta}F^{\beta\alpha}]d^4v + \int_{V^4} \tfrac{1}{4}F_{\alpha\beta}F^{\beta\alpha}d^4v, \tag{3.1}$$

in which $\mathfrak{M}$ is a Lagrange multiplier introduced to take account of the constraint (2.8) and a_α are four Lagrange multipliers introduced to take account of the constraint

$$S^{\alpha\beta}u_\beta = 0, \tag{3.2}$$

which is the so-called Frenkel condition. The constraint (3.2) must hold for the following reason.

Given $\pi^{\alpha\beta}$ the magnetization tensor (or polarization tensor, or magneto-electric moment), it is supposed that the internal angular momentum (or spin) of the "particle" is linked to $\pi^{\alpha\beta}$ by the classical formula

$$\pi^{\alpha\beta} = \Gamma S^{\alpha\beta}, \qquad \Gamma \equiv -ge/2m_0c, \tag{3.3}$$

where $S^{\alpha\beta}$ is the spin per unit volume, e is the electric charge, m_0 is the rest mass of the electron, and g is a coupling constant (equal to 2 for electrons). The quantity Γ is called the gyromagnetic ratio. In a rest frame, the magnetization 3-vector **M** (an axial vector) and the polarization (or electric moment) 3-vector **P** (a polar vector) can be expressed as

$$M_k \overset{*}{=} \tfrac{1}{2}\epsilon_{ijk}\pi^{ij}, \qquad iP_k \overset{*}{=} \pi_{k4}. \tag{3.4}$$

Following the hypothesis of Uhlenbeck and Goudsmit, the moment is purely magnetic (i.e., $\mathbf{P} = \mathbf{0}$) in the rest frame of the electron. In covariant form this assertion reads [cf. remark following Eq. (2.19)]

$$\pi^{\alpha\beta}u_\beta = 0; \tag{3.5}$$

hence (3.2). For a continuous medium built up of electrons, in agreement with the definitions of Weyssenhoff,[22] we write

$$\pi^{\alpha\beta} = \rho\tilde{\pi}^{\alpha\beta} = \rho\Gamma S^{\alpha\beta}, \tag{3.6}$$

where, from here on, $S^{\alpha\beta}$ is considered to be the spin per unit of proper mass and ρ is the relativistic mass density equal to the classical mass density of continuum mechanics in the rest frame.

In (3.1), $X^K{}_{,\alpha}$ is the inverse deformation gradient and $F_{\alpha\beta}$ is the magnetic flux tensor which, in an Euclidean frame of reference, reads

$$F_{\alpha\beta} \overset{*}{=} [\text{dual}\mathbf{B}, -i\mathbf{E}]. \tag{3.7}$$

The quantity $\frac{1}{2}\pi_{\alpha\beta}F^{\beta\alpha}$ is the energy of a doublet in a magnetic field and $\frac{1}{4}F_{\alpha\beta}F^{\beta\alpha}$ is the classical self-energy of the magnetic field. The quantity

$$\Pi^{\alpha\beta}{}_K \overset{\text{DEF}}{=} P^\alpha{}_{[\gamma}P^\beta{}_{\delta]}\tilde{\pi}^{\gamma\delta}{}_{;\rho}\chi^\rho{}_K \tag{3.8}$$

appearing in the argument of the expression of the action (3.1) represents the spin interaction of neighboring points. This form assures the spacelike character and in a rest frame, it reduces to $\mu^k{}_{,K}$ (where μ^k is the magnetization 3-vector per unit mass). Hence $\Pi^{\alpha\beta}{}_K$ is truly a quantity which generalizes the gradient appearing in earlier works (Tiersten,[14] Brown,[15] and Ref. 1). In ψ, the argument θ is the proper temperature measured in a rest frame.

Kinetic energy of spin: The most obvious four-dimensional generalization of the quantities $\boldsymbol{\mu}$ and $\delta\boldsymbol{\theta}$ of Ref. 1 are $\tilde{\pi}^{\alpha\beta}$, a second-order skew-symmetric tensor, and $\delta\Omega^{\alpha\beta}$, a second-order skew-symmetric tensor representing an infinitesimal four-dimensional rotation in V^4. Therefore, following Frenkel[8] in the generalization of the classical mechanics counterpart [formula (2.19) of Ref. 1], the term to be included in the variational principle is of the form

$$\delta W = -\int_{(\mathcal{B})} \tfrac{1}{2}\rho\dot{S}^{\alpha\beta}\delta\Omega_{\beta\alpha}d^4v = -\int_{(\mathcal{B})} (1/2\Gamma)\rho\dot{\tilde{\pi}}^{\alpha\beta}\delta\Omega_{\beta\alpha}d^4v. \tag{3.9}$$

The relations between $\delta\tilde{\pi}^{\alpha\beta}$ and $\delta\Omega^{\alpha\beta}$, generalizing (2.16) of Ref. 1, are given by

$$\delta\tilde{\pi}_{\alpha\beta} = 2\delta\Omega_{[\alpha}{}^\gamma\tilde{\pi}_{\beta]\gamma}, \qquad \delta\Omega_{\alpha\beta} = -\delta\Omega_{\beta\alpha}, \qquad \delta\Omega^{\alpha\beta}u_\beta = 0. \tag{3.10}$$

We remark that $\delta\Omega^{\alpha\beta}$ and $\delta\tilde{\pi}^{\alpha\beta}$ are anholonomic variations. Therefore the integrand of (3.9) does not represent an exact differential.

Variational principle: Following the tradition of Lagrange and Piola, we introduce indeterminate multipliers for the basic arguments varied in the Lagrangian in $(\mathcal{B})$ and on $(\partial\mathcal{C})$. Hence the proposed variational principle is

$$\delta A + \delta W + \delta W^* = 0, \tag{3.11}$$

where δW is given by (3.9) and

$$\delta W^* = \int_{(\mathcal{B}-\Gamma)} \rho f^\alpha\delta x_\alpha d^4v - \int_{(\partial\mathcal{B}-\Gamma)} T^\alpha\delta x_\alpha d^3s - \int_{(\mathcal{B}-\Gamma)} (1/c)J^\alpha\hat{\delta}A_\alpha d^4v + \int_{(\Gamma)} (1/c)K^{\alpha\beta}\hat{\delta}A_\beta n_\alpha d^3s_\Gamma - \int_{(\mathcal{B}-\Gamma)} \rho\eta\hat{\delta}\theta d^4v. \tag{3.12}$$

In (3.12), f^α is the 4-force due to nonelectromagnetic causes, per unit mass, J^α is the 4-volume current, $K^{\alpha\beta}$ is the 4-surface current density prescribed on the discontinuity hypersurface (Γ) which splits $(\mathcal{B})$ into two parts, T^α is the stress 4-vector acting upon $(\partial\mathcal{B})$, η is the proper density of entropy per unit mass, and n_α is the positive unit normal to (Γ). A dot superposed on letters indicates partial differentiation with respect to the proper time τ.

The variation is effected with respect to a parameter λ, i.e., we write

$$x^\alpha = x^\alpha(X^K, \tau, \lambda), \tag{3.13}$$

with $\lambda = \lambda_0$ a fixed value on the nonvaried particle trajectory and $\lambda = \lambda_0 + \delta\lambda$ on the slightly changed trajectory. Thus, for example,

$$\delta x^\alpha = \frac{\partial x^\alpha}{\partial \lambda}\Big|_{X^K,\tau,\lambda=\lambda_0}\delta\lambda.$$

Henceforward the variations are carried out without explicitly indicating the foregoing definition. We note that the operators δ and $\partial/\partial X^K$ and $\partial/\partial\tau$ commute, i.e.,

$$\delta(\phi_{,K}) = (\delta\phi)_{,K}, \qquad \delta(\dot{\phi}) = \dot{\overline{\delta\phi}}$$

Using this property, we find the following useful expressions:

$$\delta X^K{}_{,\alpha} = -X^K{}_{,\rho}(\delta x^\rho)_{;\alpha}, \tag{3.14a}$$

$$\delta u^\alpha = (\delta x^\alpha)_{;\beta}u^\beta, \tag{3.14b}$$

$$\delta\rho = -\rho P^\alpha{}_\beta(\delta x^\beta)_{;\alpha}, \tag{3.14c}$$

$$\delta J = J\delta^\alpha_\beta(\delta x^\beta)_{;\alpha}, \tag{3.14d}$$

$$\delta(\boldsymbol{\phi}_{;\alpha}) = (\delta\boldsymbol{\phi})_{;\alpha} - \boldsymbol{\phi}_{;\beta}(\delta x^\beta)_{;\alpha}, \tag{3.14e}$$

the last expression being valid for any arbitrary tensor $\boldsymbol{\phi}$.

We recall that $F_{\alpha\beta}$ is derivable from a 4-potential, i.e.,

$$F_{\alpha\beta} = 2\nabla_{[\alpha}A_{\beta]}, \tag{3.15}$$

which satisfies the Maxwell's equations

$$\epsilon^{\alpha\beta\gamma\delta}F_{\gamma\delta;\beta} = 0 \quad \text{in } (\mathfrak{B}-\Gamma), \tag{3.16a}$$

$$\epsilon^{\alpha\beta\gamma\delta}[F_{\gamma\delta}]n_\beta = 0 \quad \text{on } (\Gamma). \tag{3.16b}$$

Equation (3.15) is a constraint in the variational process; therefore, we introduce the Weiss-gauge invariant variation (cf. Grot[11])

$$\hat{\delta}A_\alpha = \delta A_\alpha - A_{\gamma;\alpha}\delta x^\gamma. \tag{3.17}$$

The variation deduced from (3.16) and (3.15) reads

$$\delta F_{\alpha\beta} = -2(\hat{\delta}A_{[\alpha})_{;\beta]} + 2F_{\gamma[\alpha}(\delta x^\gamma)_{;\beta]}. \tag{3.18}$$

Similarly the total variation of the temperature is the sum of a proper variation $\hat{\delta}\theta$ and of another one due to the spatial dependence, i.e.,

$$\delta\theta = \hat{\delta}\theta + \tilde{\delta}\theta. \tag{3.19}$$

The last term, following Taub,[9] may be expressed in terms of a new variable Θ through the definition

$$\tilde{\delta}\theta \overset{\text{DEF}}{=} \frac{\partial}{\partial\tau}(\delta\Theta) = (\delta\Theta)_{;\beta}u^\beta; \tag{3.20}$$

hence

$$\delta\theta = \hat{\delta}\theta + (\delta\Theta)_{;\beta}u^\beta. \tag{3.21}$$

The variations: In the sequel we need the following expressions of variations:

a. Mass:

$$\delta(\rho\, d^4v) = -\rho(u^\alpha u_\beta/c^2)(\delta x^\beta)_{;\alpha}d^4v. \tag{3.22}$$

The proof of this follows by passing to the reference frame (X^K, τ), i.e.,

$$\delta(\rho\, d^4v) = \delta(\rho J d^4V) = (J\delta\rho + \rho\delta J)d^4V.$$

Now use (3.14c) and (3.14d); this gives (3.22).

b. Magnetization: We recall the variation

$$\delta\tilde{\pi}_{\alpha\beta} = 2\delta\Omega_{[\alpha}{}^\gamma\tilde{\pi}_{\beta]\gamma}. \tag{3.10}$$

Taking account of (3.18) and (3.10), we have

$$\begin{aligned}\delta(\tfrac{1}{2}\pi_{\alpha\beta}F^{\beta\alpha}) = &-\pi^{\gamma\beta}F^\alpha{}_\gamma(\delta x_\alpha)_{;\beta} + \pi^{\alpha\beta}(\hat{\delta}A_\alpha)_{;\beta}\\ &-\rho\Gamma F^{\alpha\gamma}S_\gamma{}^\beta\delta\Omega_{\alpha\beta}\\ &-\tfrac{1}{2}\pi_{\alpha\beta}F^{\beta\alpha}P^\rho{}_\sigma(\delta x^\sigma)_{;\rho}.\end{aligned} \tag{3.23}$$

c. Polarization gradient: The potential ψ depends on $\tilde{\pi}^{\gamma\delta}{}_{;\rho}$ only through $\Pi^{\alpha\beta}{}_K$. Therefore, we set

$$\psi = \tilde{\psi}(\Pi^{\alpha\beta}{}_K) = \hat{\psi}(\tilde{\pi}^{\gamma\delta}{}_{;\rho}), \tag{3.24}$$

subject to the restrictions

$$\frac{\partial\hat{\psi}}{\partial\tilde{\pi}^{\gamma\delta}{}_{;\rho}}u_\rho = 0, \qquad \frac{\partial\hat{\psi}}{\partial\tilde{\pi}^{\gamma\delta}{}_{;\rho}}u^\gamma = 0. \tag{3.25}$$

Thus,

$$\rho\frac{\partial\hat{\psi}}{\partial\tilde{\pi}^{\alpha\beta}{}_{;\gamma}}\delta(\tilde{\pi}^{\alpha\beta}{}_{;\gamma}) = \mathfrak{M}_{\alpha\beta}{}^\gamma\,\delta(\tilde{\pi}^{\alpha\beta}{}_{;\gamma}), \tag{3.26}$$

where the tensor defined by

$$\mathfrak{M}_{\alpha\beta}{}^\gamma \overset{\text{DEF}}{=} \rho\frac{\partial\tilde{\psi}}{\partial\pi^{\rho\sigma}{}_K}P^\rho{}_{[\alpha}P^\sigma{}_{\beta]}x^\gamma{}_K \tag{3.27}$$

will be called the "*electromagnetic hyperstress tensor.*" It is subject to the restrictions

$$\mathfrak{M}_{\alpha\beta}{}^\gamma u^\alpha = 0, \qquad \mathfrak{M}_{\alpha\beta}{}^\gamma u^\beta = 0, \qquad \mathfrak{M}_{\alpha\beta}{}^\gamma u_\gamma = 0. \tag{3.28}$$

Now we carry out the variation of the action:

$$\begin{aligned}\delta A = \int_{(\mathscr{B}-\Gamma)}\Big(&\rho\mathfrak{M}u^\alpha u^\gamma(\delta x_\alpha)_{;\gamma} + \rho\psi u^\alpha u^\beta c^{-2}(\delta x_\alpha)_{;\beta}\\ &+ (2c)^{-2}\pi_{\rho\sigma}F^{\sigma\rho}u^\alpha u^\beta(\delta x_\alpha)_{;\beta}\\ &+ \rho a^\alpha u^\beta(\delta\Omega_{\alpha\gamma}S_\beta{}^\gamma - \delta\Omega_{\beta\gamma}S_\alpha{}^\gamma)\\ &+ \rho a_\alpha S^{\alpha\beta}u^\gamma(\delta x_\beta)_{;\gamma} + \rho\frac{\partial\psi}{\partial X^K{}_{,\alpha}}X^K{}_{,\beta}(\delta x^\beta)_{;\alpha}\\ &+ \pi^{\gamma\beta}F^\alpha{}_\gamma(\delta x_\alpha)_{;\beta} - \pi^{\alpha\beta}(\hat{\delta}A_\alpha)_{;\beta} + \rho\Gamma F^{\alpha\gamma}S_\gamma{}^\beta\delta\Omega_{\alpha\beta}\\ &- 2\rho\Gamma\frac{\partial\psi}{\partial\tilde{\pi}_{\alpha\beta}}\delta\Omega_{[\alpha}{}^\gamma S_{\beta]\gamma} + \tfrac{1}{2}F^{\beta\alpha}[(\hat{\delta}A_\beta)_{;\alpha} - (\hat{\delta}A_\alpha)_{;\beta}\\ &- F_{\gamma\beta}(\delta x^\gamma)_{;\alpha} + F_{\gamma\alpha}(\delta x^\gamma)_{;\beta}] - \rho\frac{\partial\psi}{\partial\theta}\hat{\delta}\theta\\ &- \rho\frac{\partial\psi}{\partial\theta}u^\beta(\delta\Theta)_{;\beta}\Big)d^4v + \int_{(V^4-\mathscr{B})}\tfrac{1}{2}F^{\alpha\beta}\delta F_{\beta\alpha}d^4v\\ &- \int_{(\mathscr{B}-\Gamma)}\mathfrak{M}_{\alpha\beta}{}^\gamma\delta(\tilde{\pi}^{\alpha\beta}{}_{;\gamma})d^4v\\ &+ \int_{V^4}\tfrac{1}{4}F_{\alpha\beta}F^{\beta\alpha}\delta^\sigma_\rho(\delta x^\rho)_{;\sigma}d^4v.\end{aligned} \tag{3.29}$$

We carry (3.29) into (3.11) and write for various product terms of the form

$$\mathbf{F}(\delta\mathbf{G})_{;\alpha} = (\mathbf{F}\delta\mathbf{G})_{;\alpha} - \mathbf{F}_{;\alpha}\delta\mathbf{G}.$$

Upon using (3.10) and the generalized Green–Gauss theorem (see Eringen[23]), we obtain after some lengthy manipulations

$$
\begin{aligned}
&-\int_{(\mathcal{B}-\Gamma)}(T^{\alpha\beta}{}_{;\beta}-\rho f^{\alpha})\delta x_{\alpha}\,d^4v\\
&+\int_{(\partial\mathcal{B}-\Gamma)}([T^{\alpha\beta}]n_{\beta}-T^{\alpha})\delta x_{\alpha}\,d^3s\\
&-\int_{(\Gamma)}[T^{\alpha\beta}]n_{\beta}\delta x_{\alpha}\,d^3s_{\Gamma}-\int_{(V^4-\mathcal{B})}T^{\alpha\beta}{}_{;\beta}\delta x_{\alpha}\,d^4v\\
&+\int_{(\mathcal{B}-\Gamma)}(G^{\alpha\beta}{}_{;\beta}-c^{-1}J^{\alpha})\hat{\delta}A_{\alpha}\,d^4v\\
&+\int_{(\Gamma)}([G^{\alpha\beta}]n_{\beta}-c^{-1}K^{\alpha\beta}n_{\beta})\hat{\delta}A_{\alpha}\,d^3s_{\Gamma}\\
&+\int_{(V^4-\mathcal{B})}F^{\alpha\beta}{}_{;\beta}\hat{\delta}A_{\alpha}\,d^4v-\int_{(\mathcal{B}-\Gamma)}\left(\rho\frac{\partial\psi}{\partial\theta}+\eta\right)\hat{\delta}\theta\,d^4v\\
&+\int_{(\mathcal{B}-\Gamma)}\left(\rho\frac{\partial\psi}{\partial\theta}u^{\beta}\right)_{;\beta}\delta\Theta\,d^4v-\int_{(\partial\mathcal{B}-\Gamma)}\rho\frac{\partial\psi}{\partial\theta}u^{\beta}n_{\beta}\delta\Theta d^3s\\
&+\int_{(\Gamma)}\left[\rho\frac{\partial\psi}{\partial\theta}u^{\beta}\right]n_{\beta}\delta\Theta\,d^3s_{\Gamma}+\int_{(\mathcal{B}-\Gamma)}\Big\{\tfrac{1}{2}\rho\dot{S}^{\alpha\beta}\\
&+\rho(a^{\alpha}u^{\gamma}-a^{\gamma}u^{\alpha})S_{\gamma}{}^{\beta}+\tfrac{1}{2}\Gamma(F^{\alpha\gamma}S_{\gamma}{}^{\beta}-F^{\beta\gamma}S_{\gamma}{}^{\alpha})\\
&-2\rho\Gamma\frac{\partial\psi}{\partial\tilde{\pi}_{\alpha\gamma}}S_{\gamma}{}^{\beta}+2\mathfrak{M}^{\alpha\mu\gamma}{}_{;\gamma}\tilde{\pi}_{\mu}{}^{\beta}\Big\}\delta\Omega_{\alpha\beta}d^4v\\
&-\int_{(\partial\mathcal{B}-\Gamma)}2\mathfrak{M}^{\alpha\mu\gamma}\tilde{\pi}_{\mu}{}^{\beta}n_{\gamma}\delta\Omega_{\alpha\beta}\,d^3s\\
&+\int_{(\Gamma)}[2\mathfrak{M}^{\alpha\mu\gamma}\tilde{\pi}_{\mu}{}^{\beta}]n_{\gamma}\delta\Omega_{\alpha\beta}d^3s_{\Gamma}=0, \qquad (3.30)
\end{aligned}
$$

where we have defined the following quantities

$$T^{\alpha\beta}=\rho\omega u^{\alpha}u^{\beta}+p^{\alpha}u^{\beta}-t^{\beta\alpha}+\mathfrak{M}_{\mu\nu}{}^{\beta}\tilde{\pi}^{\mu\nu;\alpha}+T^{\alpha\beta}_{(\mathrm{em})}\quad\text{in }(\mathcal{B}), \qquad (3.31)$$

$$T^{\alpha\beta}\equiv T^{\alpha\beta}_{(\mathrm{em})}\quad\text{in }(V^4-\mathcal{B}), \qquad (3.32)$$

$$\omega\equiv\mathfrak{M}+\psi c^{-2}+(2c)^{-2}\tilde{\pi}_{\rho\sigma}F^{\sigma\rho}, \qquad (3.33)$$

$$p^{\alpha}\equiv\rho a_{\gamma}S^{\gamma\alpha}, \qquad (3.34)$$

$$t^{\beta\alpha}\equiv-\rho\frac{\partial\psi}{\partial X^{K}{}_{,\beta}}X^{K,\alpha}, \qquad (3.35)$$

$$T^{\alpha\beta}_{(\mathrm{em})}\equiv-F^{\alpha}{}_{\gamma}G^{\gamma\beta}+\tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta}\quad\text{in }(\mathcal{B}), \qquad (3.36)$$

$$T^{\alpha\beta}_{(\mathrm{em})}\equiv-F^{\alpha}{}_{\gamma}F^{\gamma\beta}+\tfrac{1}{4}F_{\mu\nu}F^{\nu\mu}g^{\alpha\beta}\quad\text{in }(V^4-\mathcal{B}), \qquad (3.37)$$

and used the definition of the *electric displacement–magnetic field intensity* $G^{\alpha\beta}$, given by

$$G^{\alpha\beta}\equiv F^{\alpha\beta}-\pi^{\alpha\beta}\quad\text{in }(\mathcal{B}). \qquad (3.38)$$

Here $T^{\alpha\beta}$ is the total *stress–energy–momentum tensor*, ω is the density of energy per unit mass, p^{α} is called the *nonmechanical momentum 4-vector* and $T^{\alpha\beta}_{(\mathrm{em})}$ is the electromagnetic stress–energy–momentum tensor introduced by Grot and Eringen.[6] (See also the Appendix of Ref. 1). They have shown that this corresponds to an electromagnetic force $f^{\alpha}_{(\mathrm{em})}$ through the relation:

$$T^{\alpha\beta}_{(\mathrm{em})};_{\beta}=-f^{\alpha}_{(\mathrm{em})}=-(\pi^{\beta\gamma}F^{\alpha}{}_{\gamma;\beta}+c^{-1}J^{\gamma}F^{\alpha}{}_{\gamma}), \qquad (3.39)$$

where the last term

$$f^{\alpha}_{L}\equiv c^{-1}J^{\gamma}F^{\alpha}{}_{\gamma} \qquad (3.40)$$

is the Lorentz force and the first one is the force arising from the presence of magnetization in matter. One can show that this term is none other than the *Stern–Gerlach force* used by Halbwachs.[10] Indeed

$$\pi^{\beta\gamma}F_{\alpha\gamma;\beta}=\tfrac{1}{2}\pi^{\beta\gamma}(F_{\alpha\gamma;\beta}-F_{\alpha\beta;\gamma})$$

from the skew-symmetry of $\pi^{\beta\gamma}$. The set of Maxwell's equations (3.16) may be written as

$$F_{\alpha\gamma;\beta}=F_{\beta\gamma;\alpha}+F_{\alpha\beta;\gamma}\quad\text{in }(\mathcal{B}-\Gamma);$$

hence

$$\pi^{\beta\gamma}F_{\alpha\gamma;\beta}=\tfrac{1}{2}(F_{\beta\gamma;\alpha}+F_{\alpha\beta;\gamma}-F_{\alpha\beta;\gamma})\pi^{\beta\gamma}=\tfrac{1}{2}\rho\Gamma S^{\beta\gamma}F_{\beta\gamma;\alpha}. \qquad (3.41)$$

The definition (3.35) corresponds to the expression of the stress tensor due to the deformation field according to Grot and Eringen.[6]

In writing (3.30) we assumed that, across the discontinuity hypersurface (Γ)

$$[\delta x^{\alpha}]=[\hat{\delta}A^{\alpha}]=[\delta\Theta]=[\delta\Omega_{\alpha\beta}]=0. \qquad (3.42)$$

The expression (3.30) is posited to be valid for any volume and any surface and for any variations δx^{α}, $\delta\theta$, $\delta\Theta$, $\hat{\delta}A_{\alpha}$, $\delta\Omega_{\alpha\beta}$ in $(\mathcal{B}-\Gamma)$, on $(\partial\mathcal{B}-\Gamma)$ and on (Γ). Hence we obtain the local field equations

$$T^{\alpha\beta}{}_{;\beta}=\rho f^{\alpha}\quad\text{in }(\mathcal{B}-\Gamma),\qquad T^{\alpha\beta}{}_{;\beta}=0\quad\text{in }(V^4-\mathcal{B}), \qquad (3.43a)$$

$$[T^{\alpha\beta}]n_{\beta}=T^{\alpha}\quad\text{on }(\partial\mathcal{B}-\Gamma),\qquad [T^{\alpha\beta}]n_{\beta}=0\quad\text{on }(\Gamma), \qquad (3.43b)$$

$$G^{\alpha\beta}{}_{;\beta}=c^{-1}J^{\alpha}\text{ in }(\mathcal{B}-\Gamma),\qquad F^{\alpha\beta}{}_{;\beta}=0\quad\text{in }(V^4-\mathcal{B}), \qquad (3.44a)$$

$$[G^{\alpha\beta}]n_{\beta}=c^{-1}K^{\alpha}\equiv c^{-1}K^{\alpha\beta}n_{\beta}\quad\text{on }(\Gamma), \qquad (3.44b)$$

$$\eta=-\frac{\partial\psi}{\partial\theta}, \qquad (3.45a)$$

$$(\rho\eta u^{\beta})_{;\beta}=\rho\dot{\eta}=0\quad\text{in }(\mathcal{B}-\Gamma), \qquad (3.45b)$$

$$\rho\eta u^{\beta}n_{\beta}=0\quad\text{on }(\partial\mathcal{B}-\Gamma), \qquad (3.45c)$$

$$[\rho\eta u^{\beta}]n_{\beta}=0\quad\text{on }(\Gamma), \qquad (3.45d)$$

$$\rho\dot{S}^{\alpha\beta}=2\rho\Gamma\overset{*}{F}{}^{[\alpha}{}_{\gamma}S^{\beta]\gamma}\quad\text{in }(\mathcal{B}-\Gamma), \qquad (3.46)$$

$$\begin{aligned}&\mathfrak{M}^{[\alpha|\gamma\lambda|}\tilde{\pi}_{\gamma}{}^{\beta]}n_{\lambda}=0\quad\text{on }(\partial\mathcal{B}-\Gamma),\\&[\mathfrak{M}^{[\alpha|\gamma\lambda|}\tilde{\pi}_{\gamma}{}^{\beta]}]n_{\lambda}=0\quad\text{on }(\Gamma).\end{aligned} \qquad (3.47)$$

In (3.46), we have defined the *effective electromagnetic field* by

$$\overset{*}{F}{}^{\alpha\gamma}=F^{\alpha\gamma}+\frac{1}{\Gamma}(a^{\alpha}u^{\gamma}-a^{\gamma}u^{\alpha})-2\frac{\partial\psi}{\partial\tilde{\pi}_{\alpha\gamma}}+\frac{2}{\rho}\mathfrak{M}^{\alpha\gamma\rho}{}_{;\rho}, \qquad (3.48)$$

which is a skew-symmetric second-order tensor.

Equations (3.43) are Cauchy's laws of motion in $(\mathcal{B}-\Gamma)$, on $(\partial\mathcal{B}-\Gamma)$ and on (Γ). Equations (3.44) are the Maxwell's equations in matter and vacuum which must be supplemented with (3.16), its jump across (Γ), and the conservation of charge equations:

$$J^{\alpha}{}_{;\alpha}=0\quad\text{in }(\mathcal{B}-\Gamma), \qquad (3.49a)$$

$$[J^{\alpha}]n_{\alpha}=0\quad\text{on }(\Gamma). \qquad (3.49b)$$

Equations (3.45) are the definition of entropy and the entropy conservation law for a nondissipative process. Equation (3.46) is the electronic spin equation that generalizes previous works by Frenkel[8] and Halbwachs[10] in taking account of the presence of the deformation gradients, the magnetization and its grad-

ients in ψ. It is also a generalization of the classical known equation (Brown,[15] Tiersten,[14] and Maugin and Eringen[1]).

Multiplying both sides of (3.46) by $S_{\alpha\beta}$ and summing over all indices, we get

$$\rho S_{\alpha\beta}\dot{S}^{\alpha\beta} = 2\rho\Gamma S_{\alpha\beta}\overset{*}{F}{}^{[\alpha}{}_{\gamma}S^{\beta]\gamma} = 0. \tag{3.50}$$

from the skew-symmetry of $\overset{*}{F}{}^{\alpha\gamma}$. Hence integrating (3.50) over proper time, we obtain

$$\tfrac{1}{2}S_{\alpha\beta}S^{\beta\alpha} = S_0^2 = \text{const}, \tag{3.51}$$

i.e., the modulus of the spin tensor is constant. If we introduce in a unique way the 4-vector spin s_α by the relation

$$s_\alpha \equiv (2ic)^{-1}\epsilon_{\alpha\beta\gamma\delta}S^{\beta\gamma}u^\delta, \qquad S^{\beta\gamma} = (ic)^{-1}\epsilon^{\beta\gamma\alpha\rho}s_\alpha u_\rho, \tag{3.52}$$

we have

$$s^\alpha u_\alpha = 0. \tag{3.53}$$

Therefore, s_α is in $V_1{}^3$ and reduces to its classical analog in the rest frame. Thus Eq. (3.51) may be written as

$$S_0^2 = s_0^2 = s_\alpha s^\alpha = \text{const} \tag{3.54}$$

and is none other than the generalization of the classical constraint imposed on the magnetization vector in a saturated medium (cf. Ref. 1).

It remains to find the values of the Lagrange multipliers $\mathfrak{M}$ and a_α.

4. DETERMINATION OF THE LAGRANGE MULTIPLIERS

In order to determine a_α, we multiply (3.46) with u_β and use the property

$$\rho\dot{S}^{\alpha\beta}u_\beta = -\rho S^{\alpha\beta}\dot{u}_\beta, \tag{4.1}$$

which follows from (3.2). Hence,

$$0 = \rho S^{\alpha\beta}\Big(\dot{u}_\beta + \Gamma F_\beta{}^\gamma u_\gamma + 2\Gamma\frac{\partial\psi}{\partial\tilde{\pi}_\gamma{}^\beta}u_\gamma - \frac{2\Gamma}{\rho}\mathfrak{M}^\gamma{}_\beta{}^\rho{}_{;\rho}u_\gamma - c^2 a_\beta\Big), \tag{4.2}$$

which is posited to be valid for any spin $S^{\alpha\beta}$. Thus it follows that

$$a_\beta = \frac{\dot{u}_\beta}{c^2} + \frac{\Gamma}{c^2}F_\beta{}^\gamma u_\gamma + \frac{2\Gamma}{c^2}\frac{\partial\psi}{\partial\tilde{\pi}_\gamma{}^\beta}u_\gamma - \frac{2\Gamma}{\rho c^2}\mathfrak{M}^\gamma{}_\beta{}^\rho{}_{;\rho}u_\gamma. \tag{4.3}$$

From this expression we see that a_β is in $V_1{}^3$, i.e.,

$$a_\beta u^\beta \equiv 0. \tag{4.4}$$

If we neglect the presence of the magnetization and its gradients in ψ, we obtain

$$a_\beta = \frac{\dot{u}_\beta}{c^2} + \frac{\Gamma}{c^2}F_\beta{}^\gamma u_\gamma = 0. \tag{4.5}$$

This has the same form as the equation of motion for one electron in a magnetic field $F^{\alpha\beta}$, i.e., (cf. Anderson[7])

$$m_0\dot{u}_\beta = -(e/c)F_\beta{}^\gamma u_\gamma. \tag{4.6}$$

Hence as postulated in (3.1), the introduction of a_α reflects the presence of the magnetization. From (4.3), assuming that *the current is only due to moving particles*,[24] we get the Lorentz force

$$f_{L\beta} = -(e/c)F_\beta{}^\gamma u_\gamma = (1/c)F_\beta{}^\gamma J_\gamma, \qquad J_\gamma = -eu_\gamma, \tag{4.7}$$

$$f_{L\beta} = -m_0c^2a_\beta + m_0\dot{u}_\beta + \frac{2e}{c}\frac{\partial\psi}{\partial\tilde{\pi}_\gamma{}^\beta}u_\gamma - \frac{2e}{\rho c}\mathfrak{M}^\gamma{}_\beta{}^\rho{}_{;\rho}u_\gamma. \tag{4.8}$$

Value of $\mathfrak{M}$: We perform the differentiation in the lhs of (3.43a) and use (3.39), (3.41), (3.31)–(3.35), and the continuity equation

$$(\rho u^\alpha)_{;\alpha} = 0 \quad \text{in } (\mathcal{B}-\Gamma), [\rho u^\alpha]n_\alpha = 0 \quad \text{on } (\Gamma). \tag{4.9}$$

Upon contraction of the resulting equation with u_α, we obtain

$$\begin{aligned}-\rho c^2\frac{\partial}{\partial\tau}\Big(\mathfrak{M} + \frac{\psi}{c^2} + \frac{1}{2c^2}\tilde{\pi}^{\rho\sigma}F_{\sigma\rho}\Big) &+ \mathfrak{M}_{\mu\nu}{}^\beta{}_{;\beta}\dot{\tilde{\pi}}^{\mu\nu} + \mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta} \\ &= -\rho\dot{S}^{\gamma\alpha}a_\gamma u_\alpha + t^{\beta\alpha}{}_{;\beta}u_\alpha + \rho f^\alpha u_\alpha + \tfrac{1}{2}\rho\Gamma S^{\beta\gamma}\dot{F}_{\beta\gamma}.\end{aligned} \tag{4.10}$$

Now consider the particular case for which the 4-force f^α is derivable from a potential Φ. Hence we have

$$f^\alpha u_\alpha = -\Phi^{,\alpha}u_\alpha = -\dot{\Phi}. \tag{4.11}$$

Thus (4.10) yields

$$\begin{aligned}-\rho c^2\dot{\mathfrak{M}} - \rho\dot{\psi} + \mathfrak{M}_{\mu\nu}{}^\beta{}_{;\beta}\dot{\tilde{\pi}}^{\mu\nu} + \mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta} \\ = t^{\beta\alpha}{}_{;\beta}u_\alpha - \rho\dot{\Phi} - \tfrac{1}{2}\rho\Gamma\dot{S}^{\beta\gamma}[F_{\beta\gamma} + (1/\Gamma)(a_\beta u_\gamma - a_\gamma u_\beta)].\end{aligned} \tag{4.12}$$

Upon using the definition (3.48), the last term in the rhs of (4.12) may be written as

$$\tfrac{1}{2}\rho\dot{S}^{\beta\gamma}\Big(F_{\beta\gamma} + \frac{1}{\Gamma}(a_\beta u_\gamma - a_\gamma u_\beta)\Big) \tag{4.13}$$

$$= \rho\Gamma\dot{S}^{\beta\gamma}\Big(\frac{\partial\psi}{\partial\tilde{\pi}^{\beta\gamma}} - \frac{1}{\rho}\mathfrak{M}_{\beta\gamma}{}^\rho{}_{;\rho}\Big),$$

where we took account of the following result due to the skew-symmetry of $S^{\gamma\mu}$:

$$\dot{S}^{\beta\gamma}\overset{*}{F}_{\beta\gamma} = 2\Gamma\overset{*}{F}{}^{[\beta}{}_{\mu}S^{\gamma]\mu}\overset{*}{F}_{\beta\mu} = 0. \tag{4.14}$$

Upon use of (3.35) it is easily verified that

$$t^{\beta\alpha}{}_{;\beta}u_\alpha = -t^{\beta\alpha}u_{\alpha;\beta} = -\rho\frac{\partial\psi}{\partial X^K{}_{,\beta}}\dot{\overline{X^K{}_{,\beta}}} \tag{4.15}$$

Note also that

$$\mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta} = \rho\frac{\partial\psi}{\partial\tilde{\pi}^{\mu\nu;\beta}}\dot{\tilde{\pi}}^{\mu\nu;\beta} \tag{4.16}$$

and that

$$\dot{\psi} = \frac{\partial\psi}{\partial X^K{}_{,\beta}}\dot{\overline{X^K{}_{,\beta}}} + \frac{\partial\psi}{\partial\tilde{\pi}_{\beta\gamma}}\dot{\tilde{\pi}}_{\beta\gamma} + \frac{\partial\psi}{\partial\tilde{\pi}^{\mu\nu;\beta}}\dot{\overline{\tilde{\pi}^{\mu\nu;\beta}}} + \frac{\partial\psi}{\partial\theta}\dot{\theta}. \tag{4.17}$$

We introduce the internal energy ϵ per unit mass by the definition

$$\epsilon = \psi + \eta\theta. \tag{4.18}$$

By using the conservation of entropy flux (3.45b) and the definition (3.45a), (4.18) gives, by differentiation

$$\dot{\psi} = \dot{\epsilon} - \eta\dot{\theta}. \tag{4.19}$$

Now carrying the expressions (4.15)–(4.17) into (4.12) and using (4.13) and (4.19), we get

$$c^2\dot{\mathfrak{M}} = \dot{\Phi} + \dot{\epsilon} - \dot{\psi}. \tag{4.20}$$

Upon integrating over proper time, we finally obtain

$$c^2\mathfrak{M} = \Phi + \epsilon - \psi + c^2, \tag{4.21}$$

where c^2 is a constant of integration which represents the rest energy. Now we can write (3.33) and (3.34) in their definitive forms

$$\omega = 1 + \frac{\epsilon + \Phi}{c^2} + \frac{1}{2c^2}\tilde{\pi}_{\rho\sigma}F^{\sigma\rho}, \tag{4.22}$$

$$p^\alpha = \rho\frac{S^{\gamma\alpha}}{c^2}\left(\dot{u}_\gamma - \frac{f_{L\gamma}}{m_0}\right) + \frac{2\Gamma\rho}{c^2}\left(\frac{\partial\psi}{\partial\tilde{\pi}_\mu{}^\gamma} - \frac{1}{\rho}\mathfrak{M}\mu_\gamma{}^\rho{}_{;\rho}\right)S^{\gamma\alpha}u_\mu. \tag{4.23}$$

Here ω is the total kinetic energy per unit mass. In the nonrelativistic limit, it is the sum of the rest energy, the classical kinetic energy of translation, the kinetic energy of rotation of the spin (or energy of a magnetic doublet), the internal energy, and the potential of the force **f**. In the expression of the nonmechanical momentum p^α, the first term was already found by Halbwachs[10] and the second term which involves the electromagnetic anisotropy effects and the neighboring-spin interaction phenomena is believed to be new.

5. INCOMPRESSIBILITY

It would not be difficult to consider an incompressible material and draw the consequences as to the form of the stress-energy-momentum tensor $T^{\alpha\beta}$. In that case, when we follow the motion, the following constraint is imposed:

$$\delta\rho = 0. \tag{5.1}$$

From (3.14c), this is equivalent to:

$$P^\alpha{}_\beta(\delta x^\beta)_{;\alpha} = 0. \tag{5.2}$$

Therefore we introduce a Lagrange multiplier p referred to as the *mechanical pressure* (note that it is not introduced through the potential ψ and therefore, it is not to be confused with the thermodynamical pressure π appearing in the treatment of fluids, which is defined as $\pi = -\partial\psi/\partial_\rho^{-1}$). We must add to δW a term

$$-\int_{(\mathfrak{B}\cdot\Gamma)} pP^\alpha{}_\beta(\delta x^\beta)_{;\alpha}d^4v. \tag{5.3}$$

It is then easily verified that this results in adding a term $-pP^{\alpha\beta}$ to the expression (3.31).

Equations (3.43)–(3.49), (3.16), and (4.9) constitute the complete set of field equations for *nondissipative polarized elastic solids with electronic spin* in special relativity theory. In the next section, we give a slightly different derivation of these field equations. For this, we need the following expression:

$$T^{[\alpha\beta]} = \rho a_\gamma S^{\gamma[\alpha}u^{\beta]} - t^{[\beta\alpha]} + \rho\Gamma F[^\alpha{}_\gamma S^{|\gamma|\beta]} + \mathfrak{M}^{\mu\mu[\beta}\tilde{\pi}_{\mu\nu}{}^{i\alpha]}. \tag{5.4}$$

6. THE LORENTZ INVARIANCE REQUIREMENT

Now we propose to recover the basic laws governing the behavior of the material, i.e., the balance of energy–momentum and the balance of moment of energy–momentum, by applying the Lorentz invariance to the variational principle (3.11). We postulate the following,

> The balance laws follow from the invariance of the variational principle (3.11) under the inhomogeneous proper group of Lorentz or Poincaré group (Λ_p).

A. The Group of Lorentz

A Lorentz referential change is a real linear transformation of coordinates in V^4 which conserves the norm of intervals in this spacetime manifold of normal hyperbolic metric. New coordinates $x^{*\alpha}$ are deduced from the old ones x^α according to the set of relations

$$x^{*\mu} = \Lambda^\mu{}_\nu x^\nu + b^\mu, \tag{6.1}$$

where b^μ is a constant 4-vector. The conditions of reality and invariance of the norm are expressed by

$$\overline{\Lambda}_{\mu\nu} = \Lambda_{\mu\nu}, \qquad \Lambda_{\mu\nu}\Lambda^{\nu\lambda} = \Lambda_{\nu\mu}\Lambda^{\lambda\mu} = \delta^\lambda_\nu, \tag{6.2}$$

where a superposed bar indicates the complex conjugate. From (6.2), it follows that

$$\det(\Lambda^\mu{}_\nu) = \pm 1. \tag{6.3}$$

The inverse of (6.1) reads:

$$x^\mu = x^{*\nu}\overset{-1}{\Lambda}{}_\nu^\mu + c^\mu. \tag{6.4}$$

The changes of referential (6.1) subject to (6.2) form the Poincaré group (Λ) or inhomogeneous group of Lorentz.

If $\det(\Lambda^\mu{}_\nu) = +1$, the sense of the triad formed by the three spatial axes is conserved under the transformation (6.1), and we have the *proper* Lorentz group (Λ_p). If $b^\mu = 0$, we obtain the homogeneous Lorentz group (Λ_0). The proper group is a connected Lie group, i.e., all the transformations that belong to (Λ_p) can be considered as resulting from successions of infinitesimal transformations. Thus one can confine the study of invariance to the effects of infinitesimal transformations.

An infinitesimal mapping generated by (Λ_p) can be written

$$x^{*\mu} = R^\mu{}_\nu x^\nu + d^\mu, \qquad R^\mu{}_\nu = \delta^\mu{}_\nu + \epsilon\mathfrak{Q}^\mu{}_\nu, \tag{6.5}$$

where ϵ is an infinitesimally small constant, $\mathfrak{Q}_{\mu\nu}$ is a skew-symmetric constant 4×4 matrix and d^μ is an infinitesimal 4-vector. Therefore there exist six independent $\mathfrak{Q}_{\mu\nu}$ and four d^μ. The group (Λ_p) so induced is said to be a ten-parameter Lie group. According to Noether's theory of invariants,[25] there corresponds a conservation law to each symmetry of the system. For (Λ_p), we obtain four equations giving the conservation of energy–momentum and six equations

giving the conservation of moment of energy–momentum.

B. Balance of Energy–Momentum

Making $\epsilon = 0$ in (6.5), we get an infinitesimal shift of coordinates:

$$x^{*\mu} = x^\mu + d^\mu, \tag{6.6a}$$

$$\delta x^\mu = d^\mu. \tag{6.6b}$$

Upon carrying (6.6b) into (3.30) and postulating the expressions to be valid for any d^μ and any volume and hypersurface, we obtain the Cauchy's laws of motion (3.43) in V^4 if and only if the Lagrangian density or equivalently ψ does note depend explicitly on the coordinates, i.e.,

$$\frac{\partial \psi}{\partial x^\alpha} = 0. \tag{6.7}$$

C. Balance of Moment of Energy–Momentum

Making $d^\mu = 0$ in (6.5) we get an infinitesimal rotation of the coordinates

$$x^{*\alpha} = (\delta^\alpha_\beta + \epsilon \mathcal{Q}^\alpha{}_\beta)x^\beta, \tag{6.8a}$$

$$\delta x^\alpha = \epsilon \mathcal{Q}^\alpha{}_\beta x^\beta. \tag{6.8b}$$

For a skew-symmetric tensor $M^{\alpha\beta}$, neglecting the term in ϵ^2, we obtain

$$\delta M^{\alpha\beta} = 2\epsilon \mathcal{Q}^{[\alpha}{}_\gamma M^{|\gamma|\beta]}. \tag{6.9}$$

Note the similarity of (6.9) with (3.10). Thus, since $\delta\Omega_{\alpha\beta}$ is arbitrary, we can select

$$\delta\Omega_{\alpha\beta} = \epsilon \mathcal{Q}_{\alpha\beta}. \tag{6.10}$$

We carry the variations (6.8b) and (6.10) into (3.30), add the null quantity $\epsilon(T^{[\alpha\beta]} - T^{[\alpha\beta]})\mathcal{Q}_{\alpha\beta}$ to the integrand and use (5.4):

$$\int_{(\mathcal{B}-\Gamma)}(-T^{[\alpha|\gamma|}{}_{;\gamma} + \rho f^{[\alpha}x^{\beta]})\mathcal{Q}_{\alpha\beta}d^4v - \int_{(\partial\mathcal{B}-\Gamma)}M^{\alpha\beta\lambda}n_\lambda \times \mathcal{Q}_{\alpha\beta}d^3s + \int_{(\mathcal{B}-\Gamma)}(\tfrac{1}{2}\rho\dot{S}^{\alpha\beta} - T^{[\alpha\beta]} - L^{\alpha\beta} - M^{\alpha\beta\gamma}{}_{,\gamma} - Z^{[\alpha\beta]})\mathcal{Q}_{\alpha\beta}d^4v - \int_{(\Gamma)}[T^{[\alpha|\lambda|}x^{\beta]} + M^{\alpha\beta\lambda}]n_\lambda \mathcal{Q}_{\alpha\beta}d^3s_\Gamma = 0, \tag{6.11}$$

in which we have used the definitions

$$L^{\alpha\beta} \equiv 2F^{[\alpha}{}_\gamma \pi^{\beta]\gamma}, \tag{6.12}$$

$$M^{\alpha\beta\gamma} \equiv 2\mathfrak{M}^{[\alpha|\mu\gamma|}\tilde{\pi}^{\beta]}{}_\mu, \tag{6.13}$$

$$Z^{\alpha\beta} \equiv t^{\alpha\beta} + 2\rho\frac{\partial\hat{\psi}}{\partial\tilde{\pi}_{\alpha\gamma}}S_\gamma{}^\beta + 2\mathfrak{M}^{\alpha\mu\gamma}\tilde{\pi}_\mu{}^\beta{}_{;\gamma} - \mathfrak{M}^{\mu\nu\beta}\tilde{\pi}_{\mu\nu}{}^{;\alpha} \tag{6.14}$$

$L^{\alpha\beta}$ is the electromagnetic body couple per unit volume. We call $M^{\alpha\beta\lambda}$ the electromagnetic couple stress tensor.

The first term of (6.11) vanishes because of the balance of energy–momentum. If the remainder is posited to be valid for any volume $(\mathcal{B} - \Gamma)$ and any hypersurface (Γ) and for any arbitrary constant $\mathcal{Q}_{\alpha\beta}$, then we obtain the local field equations:

$$\tfrac{1}{2}\rho\dot{S}^{\alpha\beta} - M^{\alpha\beta\gamma}{}_{;\gamma} - T^{[\alpha\beta]} = L^{\alpha\beta} + Z^{[\alpha\beta]} \quad \text{in } (\mathcal{B} - \Gamma), \tag{6.15}$$

$$[T^{[\alpha|\lambda|}x^{\beta]} + M^{\alpha\beta\lambda}]n_\lambda = 0 \quad \text{on } (\Gamma), \tag{6.16}$$

$$M^{\alpha\beta\lambda}n_\lambda = 0 \quad \text{on } (\partial\mathcal{B} - \Gamma), \tag{6.17}$$

Upon use of formulas (3.27) and (3.35), $Z^{\alpha\beta}$ can be written in the form

$$Z^{\alpha\beta} = -\rho\left(\frac{\partial\hat{\psi}}{\partial X^K{}_{,\alpha}}X^{K,\beta} + 2\frac{\partial\hat{\psi}}{\partial\tilde{\pi}_{\alpha\gamma}}\tilde{\pi}^\beta{}_\gamma + 2\frac{\partial\hat{\psi}}{\partial\tilde{\pi}_{\alpha\mu;\gamma}}\tilde{\pi}^\beta{}_{\mu;\gamma} + \frac{\partial\hat{\psi}}{\partial\tilde{\pi}_{\mu\nu;\alpha}}\tilde{\pi}_{\nu\mu}{}^{;\beta}\right) \tag{6.18}$$

with

$$\psi = \hat{\psi}(X^K{}_{,\alpha}, \tilde{\pi}_{\alpha\gamma}, \tilde{\pi}_{\alpha\mu;\gamma}, \theta). \tag{6.19}$$

We say that the moment of energy–momentum is balanced if and only if $Z^{\alpha\beta}$ is a symmetric tensor, i.e.,

$$Z^{[\alpha\beta]} = 0. \tag{6.20}$$

This equation is referred to as the *Lorentz invariance requirement*. It results from the requirement of form invariance of ψ under Lorentz mappings (6.1). Alternatively Eq. (6.20) may be considered as a constitutive equation for the antisymmetric part of $t^{\alpha\beta}$. We shall give a solution of (6.20) while studying the relativistic objectivity in Sec. 10. We remark that (6.7) and (6.20) appear to be the relativistic counterparts of the relations derived by Toupin[26] and Maugin[27] (see also Ref. 1) as a consequence of Euclidean invariance.

Finally we give another form for Eq. (6.15). We introduce the *total spin* third-order tensor $\mathcal{S}^{\alpha\beta\gamma}$ through the definition:

$$\mathcal{S}^{\alpha\beta\gamma} \equiv \tfrac{1}{2}\rho S^{\alpha\beta}u^\gamma - M^{\alpha\beta\gamma}. \tag{6.21}$$

It satisfies the following restrictions

$$\mathcal{S}^{\alpha\beta\gamma}u_\beta = 0, \quad \mathcal{S}^{\alpha\beta\gamma}u_\alpha = 0, \quad \mathcal{S}^{\alpha\beta\gamma}u_\gamma/c^2 = \tfrac{1}{2}\rho S^{\alpha\beta}. \tag{6.22}$$

Taking account of (6.20) and using the continuity equation (4.9) and the definition (6.21), we write (6.15) in the form

$$\mathcal{S}^{\alpha\beta\gamma}{}_{;\gamma} - T^{[\alpha\beta]} = L^{\alpha\beta}. \tag{6.23}$$

This is the canonical form given by Grot and Eringen[6] for the equation of balance of moment of energy–momentum. Upon use of (6.20), it is trivial to show that (6.23) and (3.46) are two equivalent forms of this conservation law.

7. ENERGY EQUATION

This equation has been arrived at during the evaluation of the Lagrange multiplier $\mathfrak{M}$. Indeed, taking account of the result (4.21), we can write (4.12) in the form

$$\rho\dot{\epsilon} + t^{\beta\alpha}{}_{;\beta}u_\alpha - \tfrac{1}{2}\rho\Gamma\dot{S}^{\beta\gamma}[F_{\beta\gamma} + (1/\Gamma)(a_\beta u_\gamma - a_\gamma u_\beta)] - \mathfrak{M}_{\mu\nu}{}^\beta{}_{;\beta}\dot{\tilde{\pi}}^{\mu\nu} - \mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta} = 0. \tag{7.1}$$

Upon using (4.13) and (4.15), then we obtain the final form of the equation of balance of energy,

$$\rho\dot{\epsilon} = t^{\beta\alpha}u_{\alpha;\beta} - \tfrac{1}{2}\rho\dot{\tilde{\pi}}^{\beta\gamma}\mathcal{F}_{\beta\gamma} + \mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta}. \tag{7.2}$$

Here we introduced the local magnetic flux tensor $\mathcal{F}_{\beta\gamma}$ by the definition

$$\mathcal{F}_{\beta\gamma} \equiv -2\frac{\partial\Psi}{\partial\tilde{\pi}^{\beta\gamma}}. \tag{7.3}$$

Equation (7.2) is the relativistic counterpart of Eq. (8.9) of Ref. 1.

8. CAUCHY'S EQUATIONS

A. Balance of Momentum

We perform the differentiation in (3.43a) and, taking account of the continuity equation, we get

$$\rho\dot{\omega}u^{\alpha} + \rho\omega\dot{u}^{\alpha} + \rho\,\dot{\overline{(p^{\alpha}/\rho)}} - t^{\beta\alpha}{}_{;\beta} + \mathfrak{M}_{\mu\nu}{}^{\beta}{}_{;\beta}\tilde{\pi}^{\mu\nu;\alpha} + \mathfrak{M}_{\mu\nu}{}^{\beta}\dot{\tilde{\pi}}^{\mu\nu;\alpha}{}_{;\beta} = f_{(\mathrm{em})}^{\alpha} + \rho f^{\alpha}. \tag{8.1}$$

Note that

$$\begin{aligned} &\dot{u}_{\alpha} \subset V_{\perp}^{3}, \quad u^{\alpha}\perp V_{\perp}^{3}, \\ &t^{\beta\alpha}{}_{;\beta}P^{\gamma}{}_{\alpha} \equiv t^{\beta\gamma}{}_{;\beta} - c^{-2}t^{\beta\alpha}u_{\alpha;\beta}u^{\gamma}, \\ &f^{\alpha}P^{\gamma}{}_{\alpha} = -\Phi^{,\gamma} - c^{-2}\dot{\Phi}u^{\gamma}. \end{aligned} \tag{8.2}$$

Applying the projector $P^{\gamma}{}_{\alpha}$ to (8.1), we find

$$\begin{aligned} &\rho\omega\dot{u}^{\gamma} + \rho\,\dot{\overline{(p^{\alpha}/\rho)}}P^{\gamma}{}_{\alpha} - t^{\beta\gamma}{}_{;\beta} + c^{-2}t^{\beta\alpha}u_{\alpha;\beta}u^{\gamma} \\ &+ c^{-2}\mathfrak{M}_{\mu\nu}{}^{\beta}{}_{;\beta}\dot{\tilde{\pi}}^{\mu\nu}u^{\alpha} + c^{-2}\mathfrak{M}_{\mu\nu\beta}\dot{\tilde{\pi}}^{\mu\nu;\beta}u^{\gamma} \\ &+ (\mathfrak{M}_{\mu\nu}{}^{\beta}\tilde{\pi}^{\mu\nu;\gamma})_{;\beta} = -\rho\Phi^{,\gamma} - \rho c^{-2}\dot{\Phi}u^{\gamma} + f_{(\mathrm{em})}^{\alpha}P^{\gamma}{}_{\alpha}, \end{aligned} \tag{8.3}$$

which constitutes a set of three independent equations. This is the relativistic counterpart of Cauchy's first law of motion.

B. Balance of Moment of Momentum

The analog of Cauchy's second law is found by multiplying (6.23) by $P^{\mu}{}_{\alpha}P^{\nu}{}_{\beta}$ and using (5.4) and (6.21):

$$\begin{aligned} &\tfrac{1}{2}\rho\dot{S}^{\mu\nu} - c^{-2}\rho S^{[\mu}{}_{\beta}u^{\nu]}\dot{u}^{\beta} - M^{\mu\nu\gamma}{}_{;\gamma} + 2c^{-2}M^{[\mu|\beta\gamma|}u^{\nu]}u_{\beta;\gamma} \\ &- \mathfrak{M}^{\zeta\sigma[\nu}\tilde{\pi}_{\zeta\sigma}{}^{;\mu]} - c^{-2}\mathfrak{M}^{\zeta\sigma[\nu}u^{\mu]}\dot{\tilde{\pi}}_{\zeta\sigma} \\ &+ t^{[\nu\mu]} = \mathcal{P}^{[\nu}\mathcal{E}^{\mu]} + \mathfrak{M}^{[\nu}\mathcal{B}^{\mu]}, \end{aligned} \tag{8.4}$$

which constitutes a set of three independent equations. In (8.4) we have made use of the definitions of the following 4-vectors:

$$\mathcal{P}^{\alpha} \equiv c^{-1}\pi^{\beta\alpha}u_{\beta}, \quad \mathfrak{M}_{\alpha} \equiv (2c)^{-1}\epsilon_{\alpha\beta\gamma\delta}\pi^{\beta\gamma}u^{\delta}, \tag{8.5}$$

$$\mathcal{E}^{\alpha} \equiv c^{-1}F^{\alpha\beta}u_{\beta}, \quad \mathcal{B}^{\alpha} \equiv (2c)^{-1}\epsilon^{\alpha\beta\gamma\delta}F_{\beta\gamma}u_{\delta}. \tag{8.6}$$

In absence of magnetic spin and polarization gradients Eq. (8.4) reduces to Eq. (5.19) of Grot and Eringen.[6]

It is also of interest to project (6.23) along u_{β}. Upon use of (6.21) and (5.4), we obtain

$$-\tfrac{1}{2}\rho S^{\alpha\beta}\dot{u}_{\beta} + M^{\alpha\beta\gamma}u_{\beta;\gamma} + c^{2}\rho a_{\gamma}S^{\gamma\alpha} = 2F^{[\alpha}{}_{\gamma}\pi^{\beta]\gamma}u_{\beta} \tag{8.7}$$

or, with the expression (4.3) for a_{γ},

$$\tfrac{1}{2}\rho S^{\alpha\beta}\dot{u}_{\beta} + M^{\alpha\beta\gamma}u_{\beta;\gamma} + \left(F_{\mu\gamma} - 2\frac{\partial\Psi}{\partial\tilde{\pi}^{\mu\gamma}} + \frac{2}{\rho}\mathfrak{M}_{\mu\gamma}{}^{\zeta}{}_{;\zeta}\right)\pi^{\gamma\alpha}u^{\mu} = 0. \tag{8.8}$$

In the absence of magnetic spin and polarization gradients, Eq. (8.8) reduces to

$$\left(F_{\mu\gamma} - 2\frac{\partial\Psi}{\partial\tilde{\pi}^{\mu\gamma}}\right)\pi^{\gamma\alpha}u^{\mu} = 0.$$

The significance of the latter result will appear when we go to the limit of small velocities. Because of the skew-symmetries, Eq. (8.9) represents only three independent equations if it is to be valid for any velocity field and any polarization tensor.

9. SOLVABILITY OF A PROBLEM

The number of scalar unknowns represented by the set of variables $[\rho, u^{\alpha}$(3 independent components), $t^{\beta\alpha}$, $\eta, J^{\alpha}, F^{\alpha\beta}, \tilde{\pi}^{\alpha\beta}, \mathcal{F}^{\alpha\beta}, M^{\alpha\beta\gamma}]$ amounts to $1+3+16+1+4+6+6+6+24 = 67$ and the number of components of Eqs. (4.9), (3.43a), (3.46), (3.49a), (3.44a), (3.16a), (3.45b), (3.35), (7.3), and (6.13) is precisely $1+4+6+1+4+4+1+16+6+24 = 67$. Thus, if we assume that the functional form of the free energy ψ and the nonelectromagnetic force f^{α} are given and that what happens in the remainder of the universe [outside $(\mathcal{B})$] is known, then any problem for the *nonlinear relativistic theory of polarized elastic solids with electronic spin* can be solved. Ultimately, theorems of existence and uniqueness for the system of partial differential equations given above have to be proved.

We must however remark that a solution to such a problem, even well posed with "ad hoc" boundary conditions on $(\partial\mathcal{B})$ and initial conditions on a spacelike hypersurface [e.g., the initial configuration of the material body (B)] is clearly unmanageable.

Remark: According to (3.45b), η = const along world lines. Thus (3.45c) and (3.45d) are satisfied identically since mass is conserved, i.e., $\rho\eta u^{\beta}n_{\beta} = 0$ on $(\partial\mathcal{B} - \Gamma)$ and, on (Γ), $[\rho\eta u^{\beta}]n_{\beta} = [\eta]\rho u^{\beta}n_{\beta} = 0$, i.e., $[\eta] = 0$ on (Γ). Therefore, no reference need be made to entropy change in problem solving.

10. RELATIVISTIC OBJECTIVITY

Herein we follow the line of Bressan,[28,29] Kafadar and Eringen[17], and use the results due to Söderholm.[17]

A. Polar Decomposition

Let $\mathbf{F}$ and $\overset{-1}{\mathbf{F}}$ respectively denote the direct and inverse deformation gradients of the motion introduced in Sec. 2:

$$(\mathbf{F})^{\mu}{}_{K} \equiv P^{\mu}{}_{\lambda}x^{\lambda}{}_{,K}, \quad (\overset{-1}{\mathbf{F}})^{K}{}_{\alpha} \equiv X^{K}{}_{,\alpha}. \tag{10.1}$$

The Green and Cauchy strain tensors are then defined by

$$\mathbf{C} = \mathbf{F}^{T}\cdot\mathbf{F}, \quad \mathbf{c} = (\overset{-1}{\mathbf{F}})^{T}\overset{-1}{\mathbf{F}}, \tag{10.2}$$

where the superscript T denotes transposition.

We seek $\mathbf{R}$ isometric and $\mathbf{U}$ symmetric definite positive such that

$$\mathbf{F} = \mathbf{R}\cdot\mathbf{U} \tag{10.3}$$

with

$$\mathbf{R}: V^{4} \to V_{K_R}, \mathbf{U}: V_{K_R} \to V_{K_R},$$

where V_{K_R} is the three-dimensional affine space of the reference configuration ($V_{K_R} \equiv E^{3}$). The isometry of $\mathbf{R}$ means

$$\mathbf{R}^{T}\cdot\mathbf{R} = \mathbf{I}, \quad \text{i.e.,} \quad R^{\mu}{}_{K}R_{\mu}{}^{L} = \delta_{K}^{L}. \tag{10.4}$$

Moreover, $\mathbf{R}$ satisfies the following properties

$$\mathbf{R}\cdot\mathbf{R}^T = \mathbf{P}, \quad \text{i.e.,} \quad R^{\mu}{}_{K}R_{\lambda}{}^{K} = P^{\mu}{}_{\lambda}, \tag{10.5}$$

$$\mathbf{R}\cdot\mathbf{u} = 0, \quad \text{i.e.,} \quad R^{\mu}{}_{K}u_{\mu} = 0. \tag{10.6}$$

From the latter property, we see that $\mathbf{R}$ reduces to its classical analog (rotation tensor) in an instantaneous rest frame, i.e.,

$$R^{\mu}{}_{K} \overset{*}{=} (R^{k}{}_{K}, 0). \tag{10.7}$$

$\mathbf{R}$ represents the application

$$\mathbf{R}: \mathrm{Ei}(\mathbf{C}) \to \mathrm{Ei}(\mathbf{c}), \tag{10.8}$$

where $\mathrm{Ei}(\mathbf{C})$ is the set of orthogonal eigenvectors $\mathbf{N}$ of $\mathbf{C}$ and $\mathrm{Ei}(\mathbf{c})$ is the set of orthogonal eigenvectors $\mathbf{n}$ of $\mathbf{c}$ (see Bressan[28] and Kafadar and Eringen[17]), i.e.,

$$n^{\mu}{}_{(\alpha)} = R^{\mu}{}_{K}N^{K}{}_{(\alpha)}, \quad \alpha = 1,2,3, \quad n^{\mu}{}_{(\alpha)}u_{\mu} = 0. \tag{10.9}$$

$\mathbf{R}$ is therefore the *conventional rotation tensor*.

The only solution to (10.3) is

$$\mathbf{U} = (\mathbf{F}^T\cdot\mathbf{F})^{1/2} \equiv \mathbf{C}^{1/2}, \tag{10.10}$$

$$\mathbf{U} = \mathbf{R}^T\cdot\mathbf{F}. \tag{10.11}$$

B. Equivalent Motions (Söderholm[17])

Given $X^K = \mathfrak{X}^K(x^\alpha)$ an inverse motion of domain of definition $\mathbf{M} \in V^4$ and $(\mathcal{C})$ the particle trajectory through $\mathbf{M}$[$(\mathcal{C})$ being defined on $\mathbb{R}$ by the timelike parameter τ according to the equation $\mathcal{C}(\tau = 0) = \mathbf{M}$; $\mathfrak{X}$ is of class C^{n+2} on $(\mathcal{C})$], let the same apply to $\overset{*}{\mathbf{X}}, \overset{*}{\mathbf{M}}$, and $(\overset{*}{\mathcal{C}})$. The pairs $(\mathfrak{X}, \mathbf{M})$ and $(\overset{*}{\mathfrak{X}}, \overset{*}{\mathbf{M}})$ are said to be *equivalent* if there exists a mapping φ such that the following *equivalence relation* holds true:

a. φ is defined in a neighborhood of $(\mathcal{C})$, its range being a neighborhood of $(\overset{*}{\mathcal{C}})$. φ is of class C^{n+1} and C^{n+2} on $(\mathcal{C})$. It has an inverse $\bar{\varphi}^1$ which is of class C^{n+1} and $\mathcal{C}^{n+2}$ on $(\overset{*}{\mathcal{C}})$.

b. For any τ, $\varphi[\mathcal{C}(\tau)] = \overset{*}{\mathcal{C}}(\tau)$.

c. φ maps $V^3_\perp$ at $\mathbf{M}$ into $\overset{*}{V}{}^3_\perp$ at $\overset{*}{\mathbf{M}}$.

d. The reduction of φ to $V^3_\perp$ is isometric for any τ.

e. In a neighborhood of $(\overset{*}{\mathcal{C}})$

$$\overset{*}{\mathfrak{X}} = \mathfrak{X}\cdot\bar{\varphi}^1.$$

It is shown by Söderholm that such a φ is given by

$$\varphi: \mathcal{C}(\tau) + \mathbf{x} \to \overset{*}{\mathcal{C}}(\tau) + \mathcal{L}(\tau)\cdot\mathbf{x}, \tag{10.12}$$

where $\mathcal{L}$ is an automorphism of class C^{n+1} defined on the interval $(-\infty, 0]$. We call $\mathbf{Q}$ the restriction of $\mathcal{L}(\tau = 0)$ to $\dot{V}^4$ (the domain of V^4 where $\partial\mathcal{C}/\partial\tau|_{\tau=0}$ is defined, i.e., a neighborhood of $\mathbf{M}$. $\mathbf{Q}$ is thus time dependent and isometric.

C. Objectivity (Söderholm[17])

Let $\mathcal{F}(\mathfrak{X}, \mathbf{M})$ be a functional with values in $\mathbb{R}$. Such a functional is said to be *objective* if

$$\mathcal{F}(\mathfrak{X}, \mathbf{M}) = \mathcal{F}(\overset{*}{\mathfrak{X}}, \overset{*}{\mathbf{M}}), \tag{10.13}$$

when $(\mathfrak{X}, \mathbf{M})$ and $(\overset{*}{\mathfrak{X}}, \overset{*}{\mathbf{M}})$ are *equivalent*.

Thus according to (10.12) the function ψ introduced in (3.1) is objective if and only if

$$\psi(\mathbf{F}, \tilde{\boldsymbol{\pi}}, \boldsymbol{\Pi}) = \psi(\mathcal{Q}\mathbf{F}, \mathcal{Q}\tilde{\boldsymbol{\pi}}\mathcal{Q}^T, \mathcal{Q}\boldsymbol{\Pi}\mathcal{Q}^T) \tag{10.14}$$

If in particular we select

$$\mathcal{Q}^{\alpha}{}_{K} \equiv R^{\alpha}{}_{K}, \tag{10.15}$$

then we can write

$$\psi = \psi(\overset{-1}{C}{}^{KL}, \Pi^{KL}, \Pi^{KL}{}_{M}), \tag{10.16}$$

where

$$\begin{aligned} &\overset{-1}{C}{}^{KL} \equiv g^{\alpha\beta}X^{K}{}_{,\alpha}X^{L}{}_{,\beta}, \\ &\Pi^{KL} \equiv X^{K}{}_{,\alpha}\tilde{\pi}^{\alpha\beta}X^{L}{}_{,\beta}, \\ &\Pi^{KL} \equiv X^{K}{}_{,\alpha}\Pi^{\alpha\beta}{}_{M}X^{L}{}_{,\beta} \end{aligned} \tag{10.17}$$

These three material tensors provide a set of 18 independent quantities which form a minimal function basis for ψ. Upon use of (10.16), the constitutive equations (3.35), (7.3), and (6.13) read

$$t^{\beta\alpha} = -2\rho\left(\frac{\partial\psi}{\partial\overset{-1}{C}{}^{KL}}X^{L,\beta} + \frac{\partial\psi}{\partial\Pi^{KL}}\tilde{\pi}^{\beta\gamma}X^{L}{}_{,\gamma} + \frac{\partial\psi}{\partial\Pi^{KL}{}_{M}}\Pi^{\beta\gamma}{}_{M}X^{L}{}_{,\gamma}\right) \times X^{K,\alpha}, \tag{10.18}$$

$$\mathcal{F}^{\alpha\beta} = -2\frac{\partial\psi}{\partial\Pi^{KL}}X^{K,[\alpha}X^{L,\beta]}, \tag{10.19}$$

$$M^{\alpha\beta\gamma} = 2\rho\frac{\partial\psi}{\partial\Pi^{KL}{}_{M}}X^{K,[\alpha}\tilde{\pi}^{\beta]}{}_{\mu}X^{L,\mu}X^{\gamma}{}_{M}. \tag{10.20}$$

Remarks: (a) We can check that (10.16) constitutes one solution for the system of partial differential equations (6.20). Therefore, in the present case, the *Lorentz invariance requirement* and the *relativistic objectivity* lead to the same functional form for ψ.

(b) We could have started with a function ψ in the form:

$$\psi = \psi(X^{K}{}_{,\alpha}, \tilde{\mathcal{P}}^{\alpha}, \tilde{\mathfrak{M}}^{\alpha}, P^{\alpha}{}_{K}, M^{\alpha}{}_{K}) \tag{10.21}$$

with

$$\begin{aligned} &P^{\alpha}{}_{K} \equiv P^{\alpha}{}_{\gamma}\tilde{\mathcal{P}}^{\gamma}{}_{;\lambda}x^{\lambda}{}_{K}, \quad M^{\alpha}{}_{K} \equiv P^{\alpha}{}_{\gamma}\tilde{\mathfrak{M}}^{\gamma}{}_{;\lambda}x^{\lambda}{}_{K}, \\ &\tilde{\mathcal{P}}^{\gamma} = \mathcal{P}^{\gamma}/\rho, \quad \tilde{\mathfrak{M}}^{\gamma} = \mathfrak{M}^{\gamma}/\rho, \end{aligned} \tag{10.22}$$

$\mathcal{P}$ and $\mathfrak{M}$ being defined by (8.5). Instead of the constraint (3.2), we should have imposed

$$\tilde{\mathfrak{M}}^{\alpha}u_{\alpha} = 0. \tag{10.23}$$

Then, the objectivity requirement would lead to the reduced form

$$\psi = \psi(\overset{-1}{C}{}^{KL}, \tilde{\mathcal{P}}^{K}, \tilde{\mathfrak{M}}^{K}, P^{L}{}_{K}, M^{L}{}_{K}) \tag{10.24}$$

with

$$\tilde{\mathcal{P}}^{K} \equiv X^{K}{}_{,\alpha}\tilde{\mathcal{P}}^{\alpha}, \quad \tilde{\mathfrak{M}}^{K} \equiv X^{K}{}_{,\alpha}\tilde{\mathfrak{M}}^{\alpha}\,\mathrm{sgn}(x^{i}/X^{K}), \tag{10.25}$$

$$P^{L}{}_{K} \equiv X^{L}{}_{,\alpha}P^{\alpha}{}_{K}, \quad M^{L}{}_{K} \equiv X^{L}{}_{,\alpha}M^{\alpha}{}_{K}. \tag{10.26}$$

11. NONRELATIVISTIC LIMIT

Relativistic theories are self-consistent and do not need formally to be written in three-dimensional formalism. In the present case the full formulas written in such a way would be somewhat cumbersome. Nevertheless, one would like to compare the results

obtained above with those of previous works. In the frame of special relativity theory, the present theory is in agreement with the general presentation of relativistic continuum mechanics given earlier by Grot and Eringen[6]. It generalizes the works of Frenkel[8], Weyssenhoff and Raabe, [22] and Halbwachs[10] in introducing the tensorial fields $t^{\beta\alpha}$, $\mathfrak{F}^{\alpha\beta}$, and $M^{\alpha\beta\gamma}$. However, one has to go to three-dimensional formalism and to the slow motion limit to be able to compare with classical theories such as those of Brown[15], Tiersten[14], Toupin[13], Eringen[3], Suhubi[30], Mindlin[31] and the theory developed in Refs. 1 and 16. Cases for which quasimagnetostatics or quasielectrostatics alone are considered, are of particular interest.

In order to obtain these three-dimensional expressions, the limiting process is carried out in two steps.

(1) Every tensorial equation in four-dimensional formalism is projected (a) onto the hyperplane $V_\perp^3$, (b) along the 4-velocity u^α, and written uniquely in terms of 4-vectorial quantities. The splitting of space and time that were synthesized in Minkovskian tensorial formalism is then accomplished.

(2) In an inertial frame, the 4-vectors are expressed in terms of their spacelike and timelike components. Finally terms of order of magnitude smaller than or equal to $1/c^2$ are neglected and electromagnetic quantities are expressed in the rest frame in order to arrive at a quasistatic theory.

Calculations are lengthy and only results are given in the sequel. The first step has already been carried out as far as the mechanical equations are concerned. It resulted in Eqs. (8. 3), (7. 2), (8. 4), and (8. 8). The projection of Maxwell's equations has been given in Grot and Eringen.[6] In absence of current and for quasimagnetostatics, the only equations left are

$$\nabla \times \mathbf{H} = 0, \quad \nabla\cdot\mathbf{B} = 0 \quad \text{in } (B - \Gamma), \tag{11.1}$$

$$\mathbf{n} \times [\mathbf{H}] = 0, \quad [\mathbf{B}]\cdot\mathbf{n} = 0 \quad \text{on } (\Gamma), \tag{11.2}$$

and expressions similar to (11. 2) on $(\partial B - \Gamma)$. In absence of charge and for quasielectrostatics, we get

$$\nabla \times \mathbf{E} = 0, \quad \nabla\cdot\mathbf{D} = 0 \quad \text{in } (B - \Gamma) \tag{11.3}$$

$$\mathbf{n} \times [\mathbf{E}] = 0, \quad [\mathbf{D}]\cdot\mathbf{n} = 0 \quad \text{on } (\Gamma) \text{ and } (\partial B - \Gamma) \tag{11.4}$$

Note the expression of the derivative with respect to the proper time τ:

$$\frac{\partial}{\partial\tau} \equiv \frac{1}{\gamma}\frac{d}{dt}, \quad \gamma \equiv \sqrt{1-\beta^2}, \quad \beta = |\mathbf{v}|/c, \tag{11.5}$$

where $\mathbf{v}$ is the three-dimensional velocity of the material and d/dt is the material derivative.

It is easily shown (cf. Appendix of Ref. 1) that, in a *rest frame*, the electromagnetic force (3. 39) reads

$$f^k_{(\mathrm{em})} = qE^k + (1/c)(\mathbf{J} \times \mathbf{B})^k + (\nabla^k\mathbf{E})\cdot\mathbf{P} + (\nabla^k\mathbf{B})\cdot\mathbf{M}, \tag{11.6}$$

where q is the charge density, $\mathbf{J}$ is the 3-vector current, $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic intensity, and $\mathbf{P}$ and $\mathbf{M}$ are, respectively, the 3-vectors polarization and magnetization per unit volume.

For the case of quasimagnetostatics, in absence of currents, Eqs. (8. 3), (7. 2), (8. 4), and (6. 20) reduce to

$$\rho\frac{dv^k}{dt} = t^{ik}{}_{;i} + \rho f^k + (\nabla^k\mathbf{B})\cdot\mathbf{M}, \tag{11.7}$$

$$\rho\frac{d\psi}{dt} = t^{kl}v_{l;k} - \rho\,{}_LB^k\frac{d}{dt}\mu_k + t^{(\mu)km}\frac{d}{dt}(\mu_{k;m}) \times \text{in } (B - \Gamma) \tag{11.8}$$

$$\frac{\rho}{2\Gamma}\frac{d}{dt}S^{kl} = \left(B^{[k} + {}_LB^{[k} + \frac{1}{\rho}t^{(\mu)m[k}{}_{;m}\right)M^{l]}, \tag{11.9}$$

$$\frac{\partial\psi}{\partial X^K{}_{,[k}}X^{K,l]} + \frac{\partial\psi}{\partial\mu_{[k}}\mu^{l]} + \frac{\partial\psi}{\partial\mu_{[k,K}}\mu^{l]}{}_{,K} = 0, \tag{11.10}$$

with

$$\psi = \psi(X^K{}_{,k}, \mu^k, \mu^k{}_{,K}), \tag{11.11}$$

$$t^{ik} = -\rho\frac{\partial\psi}{\partial X^K{}_{,i}}X^{K,k}, \quad {}_LB^k = -\frac{\partial\psi}{\partial\mu_k},$$

$$t^{(\mu)ki} = \rho\frac{\partial\psi}{\partial\mu_{i,K}}x^k{}_{,K}. \tag{11.12}$$

Here μ^k is the magnetization per unit mass, S^{kl} is its dual, and $\mu^k{}_{,K}$ is its material gradient. Equations (11. 7)–(11. 12) are similar to equations found in Ref. 1.[32] The continuity equation, boundary conditions, and jump relations reduced accordingly.

We now examine the case of quasi-electrostatics. The condition (3. 5) eliminates the occurrence of polarization in the rest frame of a particle. Thus, in order to recover the theory of elastic dielectrics, we must neglect all gyromagnetic phenomena and consequently micromagnetic phenomena. With this assumption, for quasi-electrostatics, in the absence of charge, Eqs. (8. 3), (7. 2), (8. 8), and (6. 20) reduce to

$$\rho\frac{dv^k}{dt} = t^{ik}{}_{;i} + \rho f^k + (\nabla^k\mathbf{E})\cdot\mathbf{P}, \tag{11.13}$$

$$\rho\frac{d\psi}{dt} = t^{kl}v_{l;k} + \rho E_k\frac{d}{dt}\left(\frac{P^k}{\rho}\right) \quad \text{in } (B - \Gamma), \tag{11.14}$$

$$E^k + {}_LE^k + \frac{1}{\rho}\,{}_LE^{lk}{}_{;l} = 0, \tag{11.15}$$

$$\frac{\partial\psi}{\partial X^K{}_{,[k}}X^{K,l]} + \frac{\partial\psi}{\partial(P_{[k}/\rho)}\frac{P^{l]}}{\rho} + \frac{\partial\psi}{\partial(P_{[k}/\rho)_{;K}}\left(\frac{P^{l]}}{\rho}\right)_{;K} = 0, \tag{11.16}$$

with

$$\psi = \psi(X^K{}_{,k}, P^k/\rho, (P^k/\rho)_{,K}), \tag{11.17}$$

$$t^{ik} = -\rho\frac{\partial\psi}{\partial X^K{}_{,i}}X^{K,k}, \quad {}_LE^k = \rho\frac{\partial\psi}{\partial(P_k/\rho)},$$

$${}_LE^{lk} = \rho\frac{\partial\psi}{\partial(P_k/\rho)_{,K}}x^l{}_{,K}. \tag{11.18}$$

In order to obtain (11. 15) we have posited (8. 8) to be valid for any velocity field and any polarization field. Equation (11. 15) is referred to as the equation of *molecular equilibrium* (Eringen[3]). Equations (11. 13) to (11. 18) are similar (cf. Footnote 32) to those of Suhubi[2] and Mindlin[5]. In absence of polarization gradients, and for *statics*, they reduce to the equations of electroelasticity given by Eringen.[3] Boundary conditions and jump relations reduce accordingly. Equation (11. 16) is none other than the constitutive equation for the antisymmetric part of the stress tensor. Thus we can state that the theory developed in this article reduces satisfactorily to the classical approach to micromagnetism as enunciated in Ref. 1 for the magnetostatic part and to the classical treatment of elastic dielectrics with polarization gradients for the electrostatic part.

12. CONCLUSION

A variational principle has been constructed that led to the complete set of differential equations, boundary conditions, jump relations, and constitutive equations for the *dynamical theory of nondissipative polarized and magnetized solids which exhibit an internal spin of electronic (i.e., magnetic) origin*. The present work is clearly a generalization and unification (in four-dimensional formalism) of previous works on the theories of *magnetoelastic interactions* (in the sense of Brown[15]) and of *elastic dielectrics*. We may thus consider that it gives a sound treatment of these different phenomena while, from the pragmatic point of view, it allows us to write equations and more particularly electromagnetic entities in their best frame (the special theory of relativity).

The materials studied here exhibit extra degrees of freedom of quantum-electromagnetic origin. Further, there exist materials for which extra degrees of freedom of purely mechanical origin can intervene (e.g., micromorphic materials and polar materials) and where the magnetic effects studied in this article are of importance. Therefore, it is the purpose of a subsequent article to provide a variational treatment for the mechanical effects in the frame of four-dimensional formalism. A unification of both aspects is expected and left for further researches.

* This work was supported by the National Science Foundation.
† Now at Université de Paris, Faculté des Sciences, Mécanique théorique, Tour 66-9, quai Saint-Bernard, 75005 Paris, France.

1 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 143 (1972).
2 R. A. Toupin, J. Ratl. Mech. Anal. **5**, 849 (1956).
3 A. C. Eringen, *Non-Linear Theory of Continuous Media* (McGraw-Hill, New York, 1962).
4 R. A. Toupin, Intern. J. Eng. Sci. **1**, 101 (1963).
5 R. G. Dixon and A. C. Eringen, Intern. J. Eng. Sci. **3**, 359 (1965).
6 R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 611 (1966).
7 J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).
8 Frenkel, Z. Physik **37**, 243 (1926).
9 A. H. Taub, Phys. Rev. **94**, 1468 (1954).
10 F. Halbwachs, *Théorie relativiste des fluides à spin* (Gauthier-Villars, Paris, 1960).
11 R. A. Grot, J. Math. Phys. **11**, 109 (1970).
12 L. I. Sedov, Prikl. Mat. Mekh. **29**, 4 (1965).
13 R. A. Toupin, Arch. Ratl. Mech. Anal. **5**, 440 (1960).
14 H. Tiersten, J. Math. Phys. **6**, 779 (1965).
15 W. F. Brown, *Magnetoelastic Interactions* (Springer, New York, 1966).
16 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 1334 (1972).
17 See L. Söderholm, Arch. Ratl. Mech. Anal. **39**, 89 (1970), for an axiomatic approach including topological properties. Geometry and kinematics of polar media in V^4 were given by C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 271 (1971).
18 The projector **P**, as defined by (2. 14) and considered as a transformation, belongs to the class of orthochronous improper Lorentz transformations. It is a particular case of *plane projection or symmetry* $\sum_{(\mathbf{n})}$, where $\sum_{(\mathbf{n})}\mathbf{x}$ (**n** and **x** being 4-vectors at the point **M** in V^4) is the reflection of **x** through a hyperplane orthogonal to **n** at **M**. If **n** is along the time axis, we are dealing with a *time reflection* transformation. Generally **n** is only timelike as in the case $\mathbf{n} = \mathbf{u}$. For more details, see Refs. 19 and 20.
19 L. Michel, Nuovo Cimento **S14**, 95 (1959).
20 A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).
21 For the decomposition of any tensorial quantity, see T. Takabayasi, Progr. Theoret. Phys. (Kyoto) Suppl. **4**, (1957).
22 J. Y. Weyssenhoff and Raabe, Acta Phys. Polon. **9**, 8 (1947).
23 A. C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967).
24 I.e., in a rest frame, the Joule term vanishes: $\mathbf{J}\cdot\mathbf{E} = 0$; there is no electric dissipation.
25 E. Noether, Nachr. Ges. Wiss. Göttingen, 211 (1918).
26 R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).
27 G. A. Maugin, Compt. Rend. **271A**, 807 (1970).
28 A. Bressan, Ann. Mat. Pura Appl. **62**, 99 (1963).
29 A. Bressan, Rend. Sem. Mat. Univ. Padova **34**, 74 (1964).
30 E. Suhubi, Intern. J. Eng. Sci. **7**, 993 (1969).
31 R. D. Mindlin, Intern. J. Solids Structures **4**, 637 (1968).
32 Note that in Ref. 1, Brown[15] and Tiersten,[14] the divergence of stress tensors is taken with respect to the last index, hence a discrepancy of sign in the expressions involving $t^{[kl]}$.

Relativistic Continua with Directors*

Gérard A. Maugin† and A. Cemal Eringen
Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, New Jersey 08540
(Received 27 May 1971; Revised Manuscript Received 19 January 1972)

In this article, we give an axiomatic introduction of directors in the space-time continuum of Minkowski. A variational principle is presented that leads to the complete set of field equations, boundary conditions, and jump relations of the Kafadar-Eringen theory of relativistic polar media. The constitutive equations follows from the variational formulation.

1. INTRODUCTION

Generalized continuum mechanics, a part of mechanical science dealing with oriented media is now fairly well accepted. The idea suggested by Duhem[1] was to attach to each point of a three-dimensional continuum a number of directions, later called *directors*, which can rotate independently of the displacements of material points. The same idea was also conceived by Voigt[2] in his study of crystal elasticity. In a remarkable monograph, the Cosserat brothers[3] laid down the mathematical foundations of polar media with rigid directors. They introduced the notion of "trièdre caché", a 3-tuple of unit rigid directors, and constructed the theory now known as the *Cosserat continuum*.

The notion of *generalized Cosserat continuum* was made clear by Eriksen and Truesdell.[4] They considered deformable directors of which the number can exceed three (see also Toupin[5]). On physical grounds using the concept of *microstructure*, Eringen and Suhubi[6] and Eringen[7] constructed the theory of *micromorphic media*. The connection between micromorphic mechanics and the theory of three deformable directors was established by Maugin.[8] Eringen[9,10] also gave an axiomatic and unified theory of micromorphic media of grade greater than 1. As a particular case of micromorphic media of grade 1, Eringen[11] deduced the theory of *micropolar media* and gave various theories on directed fluids, solids and memory dependent materials (cf. Refs. 12–15).

12. CONCLUSION

A variational principle has been constructed that led to the complete set of differential equations, boundary conditions, jump relations, and constitutive equations for the *dynamical theory of nondissipative polarized and magnetized solids which exhibit an internal spin of electronic (i.e., magnetic) origin*. The present work is clearly a generalization and unification (in four-dimensional formalism) of previous works on the theories of *magnetoelastic interactions* (in the sense of Brown[15]) and of *elastic dielectrics*. We may thus consider that it gives a sound treatment of these different phenomena while, from the pragmatic point of view, it allows us to write equations and more particularly electromagnetic entities in their best frame (the special theory of relativity).

The materials studied here exhibit extra degrees of freedom of quantum-electromagnetic origin. Further, there exist materials for which extra degrees of freedom of purely mechanical origin can intervene (e.g., micromorphic materials and polar materials) and where the magnetic effects studied in this article are of importance. Therefore, it is the purpose of a subsequent article to provide a variational treatment for the mechanical effects in the frame of four-dimensional formalism. A unification of both aspects is expected and left for further researches.

* This work was supported by the National Science Foundation.
† Now at Université de Paris, Faculté des Sciences, Mécanique théorique, Tour 66-9, quai Saint-Bernard, 75005 Paris, France.

1 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 143 (1972).
2 R. A. Toupin, J. Ratl. Mech. Anal. **5**, 849 (1956).
3 A. C. Eringen, *Non-Linear Theory of Continuous Media* (McGraw-Hill, New York, 1962).
4 R. A. Toupin, Intern. J. Eng. Sci. **1**, 101 (1963).
5 R. G. Dixon and A. C. Eringen, Intern. J. Eng. Sci. **3**, 359 (1965).
6 R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 611 (1966).
7 J. L. Anderson, *Principles of Relativity Physics* (Academic, New York, 1967).
8 Frenkel, Z. Physik **37**, 243 (1926).
9 A. H. Taub, Phys. Rev. **94**, 1468 (1954).
10 F. Halbwachs, *Théorie relativiste des fluides à spin* (Gauthier-Villars, Paris, 1960).
11 R. A. Grot, J. Math. Phys. **11**, 109 (1970).
12 L. I. Sedov, Prikl. Mat. Mekh. **29**, 4 (1965).
13 R. A. Toupin, Arch. Ratl. Mech. Anal. **5**, 440 (1960).
14 H. Tiersten, J. Math. Phys. **6**, 779 (1965).
15 W. F. Brown, *Magnetoelastic Interactions* (Springer, New York, 1966).
16 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 1334 (1972).
17 See L. Söderholm, Arch. Ratl. Mech. Anal. **39**, 89 (1970), for an axiomatic approach including topological properties. Geometry and kinematics of polar media in V^4 were given by C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 271 (1971).
18 The projector **P**, as defined by (2. 14) and considered as a transformation, belongs to the class of orthochronous improper Lorentz transformations. It is a particular case of *plane projection or symmetry* $\sum_{(\mathbf{n})}$, where $\sum_{(\mathbf{n})}\mathbf{x}$ (**n** and **x** being 4-vectors at the point **M** in V^4) is the reflection of **x** through a hyperplane orthogonal to **n** at **M**. If **n** is along the time axis, we are dealing with a *time reflection* transformation. Generally **n** is only timelike as in the case $\mathbf{n} = \mathbf{u}$. For more details, see Refs. 19 and 20.
19 L. Michel, Nuovo Cimento **S14**, 95 (1959).
20 A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (Macmillan, New York, 1964).
21 For the decomposition of any tensorial quantity, see T. Takabayasi, Progr. Theoret. Phys. (Kyoto) Suppl. **4**, (1957).
22 J. Y. Weyssenhoff and Raabe, Acta Phys. Polon. **9**, 8 (1947).
23 A. C. Eringen, *Mechanics of Continua* (Wiley, New York, 1967).
24 I.e., in a rest frame, the Joule term vanishes: $\mathbf{J}\cdot\mathbf{E} = 0$; there is no electric dissipation.
25 E. Noether, Nachr. Ges. Wiss. Göttingen, 211 (1918).
26 R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).
27 G. A. Maugin, Compt. Rend. **271A**, 807 (1970).
28 A. Bressan, Ann. Mat. Pura Appl. **62**, 99 (1963).
29 A. Bressan, Rend. Sem. Mat. Univ. Padova **34**, 74 (1964).
30 E. Suhubi, Intern. J. Eng. Sci. **7**, 993 (1969).
31 R. D. Mindlin, Intern. J. Solids Structures **4**, 637 (1968).
32 Note that in Ref. 1, Brown[15] and Tiersten,[14] the divergence of stress tensors is taken with respect to the last index, hence a discrepancy of sign in the expressions involving $t^{[kl]}$.

Relativistic Continua with Directors*

Gérard A. Maugin† and A. Cemal Eringen
Department of Aerospace and Mechanical Sciences, Princeton University, Princeton, New Jersey 08540
(Received 27 May 1971; Revised Manuscript Received 19 January 1972)

In this article, we give an axiomatic introduction of directors in the space-time continuum of Minkowski. A variational principle is presented that leads to the complete set of field equations, boundary conditions, and jump relations of the Kafadar-Eringen theory of relativistic polar media. The constitutive equations follows from the variational formulation.

1. INTRODUCTION

Generalized continuum mechanics, a part of mechanical science dealing with oriented media is now fairly well accepted. The idea suggested by Duhem[1] was to attach to each point of a three-dimensional continuum a number of directions, later called *directors*, which can rotate independently of the displacements of material points. The same idea was also conceived by Voigt[2] in his study of crystal elasticity. In a remarkable monograph, the Cosserat brothers[3] laid down the mathematical foundations of polar media with rigid directors. They introduced the notion of "trièdre caché", a 3-tuple of unit rigid directors, and constructed the theory now known as the *Cosserat continuum*.

The notion of *generalized Cosserat continuum* was made clear by Eriksen and Truesdell.[4] They considered deformable directors of which the number can exceed three (see also Toupin[5]). On physical grounds using the concept of *microstructure*, Eringen and Suhubi[6] and Eringen[7] constructed the theory of *micromorphic media*. The connection between micromorphic mechanics and the theory of three deformable directors was established by Maugin.[8] Eringen[9,10] also gave an axiomatic and unified theory of micromorphic media of grade greater than 1. As a particular case of micromorphic media of grade 1, Eringen[11] deduced the theory of *micropolar media* and gave various theories on directed fluids, solids and memory dependent materials (cf. Refs. 12–15).

Kafadar and Eringen[16] gave later on an axiomatic approach to the nonlinear theory of micropolar media with its relativistic extensions.[17]

In his theory of liquid crystals, Eriksen[18] used a three-dimensional continuum with one director which represents the orientation of cigarlike molecules. Other applications were made by a number of writers, e.g., Green and Laws,[19] Green and Naghdi,[20] and Alblas.[21] Maugin and Eringen[22] considered the rotating magnetization vector at a point of a deformable magnetically saturated medium as being a director.[23]

Thus we see that, except for the Cosserats' work, the main developments in the three-dimensional theory took place during the last decade. Surprisingly enough, a concept almost identical to that of *directors*[24] has been developed in the frame of special and general relativity theories. In 1928, Einstein,[24] in one of his attempts to construct a unified theory (for which he considered a spacetime continuum of null curvature and nonvanishing torsion), used fields of parallel *tetrapods*[30] since then called Einstein–Kramers variables (see Kramers[31]). The use of these variables was taken over by a group of physicists working on the hydrodynamical interpretation of the wavefunction of quantum theory (the so-called "causal re-interpretation of quantum mechanics" (deBroglie); see particularly Aymart[32], Unal and Vigier,[33] Takabayasi,[34] and Halbwachs[35]. They consider the quantum fluid to be a field of microscopic spinning tops viewed, of course, in a continuous way at our observation scale. The kinematical description of the motion requires essentially the knowledge of the velocities and proper rotations. In particular, Gursey[36] used the Einstein–Kramers parameters for such a description.

Our goal is to look at tetrapods (for convenience, we call them *directors*) of varying orientation in the Minkowskian space–time of special relativity. Through the vehicle of a variational principle, we obtain the field equations, constitutive equations, and jump conditions for a polar elastic solid. One expects that the field equations based on this approach will be somewhat similar to those of Kafadar and Eringen.[17]

A remark is in order on the evolution in the axiomatization of relativistic continuum mechanics: Most of the classical works in this field start from a nonrelativistic object defined in a rest frame and, then, by guess make passage to the proper covariant four-dimensional relativistic analog (cf. the construction of the energy–momentum tensor for a perfect fluid in Landau and Lifshitz[37] or in Adler *et al.*[38]). We prefer to start with four-dimensional axiomatically set objects and verify in the limit of small velocities that we have not created "monsters" which have no equivalent in classical continuum physics.

2. KINEMATICS OF ORIENTED MEDIA IN V^4

A. Classical Motion in V^4

For a complete description of the classical motion of continuous media in V^4 we refer the reader to Grot and Eringen,[39] Kafadar and Eringen,[17] and Maugin and Eringen.[40] The following brief account is needed in the sequel.

We consider the Minkowski four-dimensional manifold V^4. In rectangular coordinates, the square of an arc length is written

$$(ds)^2 = (dx)^2 + (dy)^2 + (dz)^2 - c^2dt^2 = dz_\mu dz^\mu, \qquad (z^1, z^2, z^3, z^4) = (x, y, z, ict), \tag{2.1}$$

where (x, y, z) are three rectangular coordinates, t is the time, and c is the velocity of light. Henceforward, Greek indices (small or capital) are assumed to take the values 1, 2, 3, and 4, and all Latin indices the values 1, 2, 3. The Einstein summation convention is used throughout the article.

In curvilinear coordinates (2.1) is written

$$ds^2 = g_{\alpha\beta}\, dx^\alpha dx^\beta \tag{2.2}$$

The reciprocal $g^{\beta\gamma}$ of the metric tensor is defined according to:

$$g_{\alpha\beta}g^{\beta\gamma} = \delta^\gamma_\alpha,$$

where δ^γ_α is the Kronecker symbol. The proper time τ, a timelike parameter monotonically increasing along the world line $(\mathcal{C}_{XK})$ of a material particle originally situated at the Lagrangian coordinates X^K, $K = 1, 2, 3$, in E^3, is defined by

$$(d\tau)^2 = -(ds)^2/c^2. \tag{2.3}$$

The classical motion of a continuous medium in V^4 is described entirely by the set of relations[41]

$$x^\alpha = x^\alpha(X^\Delta) \quad \text{with} \quad X^\Delta = (X^K, ic\tau), \quad \Delta = 1, 2, 3, 4. \tag{2.4}$$

The operator $\partial/\partial\tau$ generalizes the notion of material derivative. Given a tensorial object $\mathbf{A}$, we note

$$\frac{\partial \mathbf{A}}{\partial \tau} = \dot{\mathbf{A}} = \mathbf{A}_{;\alpha}u^\alpha \quad \text{with} \quad u^\alpha = \frac{\partial x^\alpha}{\partial \tau}. \tag{2.5}$$

u^α is the 4-velocity the modulus of which is constant, i.e.,

$$g_{\alpha\beta}u^\alpha u^\beta = -c^2. \tag{2.6}$$

The projection operator or *projector* P^α_β is defined by

$$P^\alpha{}_\beta = \delta^\alpha_\beta + c^{-2}u^\alpha u_\beta \tag{2.7}$$

and satisfies the properties

$$P^\alpha{}_\beta P^\beta{}_\gamma = P^\alpha{}_\gamma, \quad P^\alpha{}_\beta u_\alpha = 0. \tag{2.8}$$

A 4-vector A^α verifying the relation

$$A^\alpha u_\alpha \equiv 0 \quad \text{or} \quad P^\alpha{}_\beta A^\beta \equiv A^\alpha \tag{2.9}$$

reduces to

$$A^\alpha \triangleq (A^k, 0), \quad k = 1, 2, 3, \tag{2.10}$$

in a rest frame.

We assume that (2.4) possesses the unique inverse

$$X^K = X^K(x^\alpha), \quad \tau = \tau(x^\alpha). \tag{2.11}$$

The following quantities are thus well defined:

$$\left(\overset{-1}{\mathbf{F}}\right)_{K}{}^{\alpha} \equiv X^{K}{}_{,\alpha} = \frac{\partial X^{K}}{\partial x^{\alpha}}, \quad x^{\alpha}{}_{,K} = \frac{\partial x^{\alpha}}{\partial X^{K}}, \quad \tau_{,\alpha} = \frac{\partial \tau}{\partial x^{\alpha}}. \tag{2.12}$$

$\overset{-1}{\mathbf{F}}$ is the inverse gradient of the motion while the direct gradient $\mathbf{F}$ is defined by

$$(\mathbf{F})^{\alpha}{}_{K} \equiv x^{\alpha}{}_{K} = P^{\alpha}{}_{\beta} x^{\beta}{}_{,K}. \tag{2.13}$$

One can verify that

$$X^{K}{}_{,\mu} x^{\mu}{}_{,L} = \delta^{K}_{L}, \quad x^{\mu}{}_{K} X^{K}{}_{,\lambda} = P^{\mu}{}_{\lambda}. \tag{2.14}$$

Since $\partial X^{K}/\partial \tau = 0$ and from the definition of $\mathbf{F}$, we have

$$X^{K}{}_{,\mu}\, u^{\mu} = 0, \quad x^{\alpha}{}_{K}\, u_{\alpha} = 0. \tag{2.15}$$

Thus in a rest frame, according to (2.9) and (2.10), the gradients reduce to their classical analogs

$$X^{K}{}_{,\mu} \overset{*}{=} (X^{K}{}_{,k}, 0), \quad x^{\alpha}{}_{K} \overset{*}{=} (x^{k}{}_{,K}, 0), \quad K \text{ fixed}. \tag{2.16}$$

Finally we note that the Green strain tensor C_{KL} and the Jacobian J of the motion are defined as

$$C_{KL} = g_{\alpha\beta} x^{\alpha}{}_{K} x^{\beta}{}_{L}, \tag{2.17a}$$

$$J = (\det C^{K}{}_{L})^{1/2}. \tag{2.17b}$$

In the sequel, commas, semicolons, and colons are used to denote partial, covariant partial, and covariant total differentiations, respectively.

B. Directors in V^4

To each point M of coordinates x^{α} in the 4-dimensional continuum, we attach four non-coplanar 4-vector fields $\mathbf{d}^{(\xi)}$, $(\xi) = 1, 2, 3, 4$. The index (ξ) is a number identifying the director and has no tensorial character. We have

$$\mathbf{d}^{(\xi)} = \mathbf{d}^{(\xi)}(x^{\alpha}) = \mathbf{d}^{(\xi)}(X^{K}, \tau). \tag{2.18}$$

The symmetric, nonsingular metric $g^{(\xi)(\zeta)}$ and its reciprocal $g_{(\zeta)(\gamma)}$ are defined by:

$$g^{(\xi)(\zeta)} \equiv g_{\alpha\beta}\, d^{(\xi)\alpha} d^{(\zeta)\beta}, \quad g^{(\xi)(\zeta)} g_{(\zeta)(\gamma)} = \delta^{(\xi)}_{(\gamma)}, \tag{2.19}$$

where $\delta^{(\xi)}_{(\gamma)}$ is the unit matrix in the *Euclidean* space E^4.

A system of four reciprocal directors $\mathbf{e}_{(\xi)}$ exists such that

$$\mathbf{d}^{(\xi)} \cdot \mathbf{e}_{(\zeta)} = \delta^{(\xi)}_{(\zeta)}. \tag{2.20}$$

In fact, the unique solution to the sixteen linear equations (2.20) is given by

$$e_{(\xi)}{}^{\mu} = \frac{1}{3!\, \det|g^{(\alpha)(\beta)}|}\, \epsilon_{(\xi)(\sigma)(\rho)(\zeta)} \epsilon^{\mu\nu\alpha\beta} d^{(\sigma)}{}_{\nu}\, d^{(\sigma)}{}_{\alpha}\, d^{(\zeta)}{}_{\beta}, \tag{2.21}$$

where $\epsilon_{(\xi)(\omega)(\rho)(\zeta)}$ and $\epsilon^{\mu\nu\alpha\beta}$ are permutation symbols. If we require the 4-vectors $\mathbf{d}^{(\xi)}$ to have a *unit length* and to form a tetrad of *orthogonal* 4-vectors at a point M of V^4, then the following constraint holds:

$$\mathbf{d}^{(\xi)} \cdot \mathbf{d}^{(\zeta)} = \delta^{(\xi)(\zeta)}. \tag{2.22}$$

The 4-vectors $\mathbf{d}^{(\xi)}$ and $g^{(\xi)(\mu)}\mathbf{e}_{(\mu)}$ are no longer distinguishable and the matrix $g^{(\xi)(\mu)}$ becomes a unit matrix. Therefore, we may write $\mathbf{d}_{(\xi)}$ so that

$$d_{(\omega)}{}^{\mu} d^{(\xi)}{}_{\mu} = \delta^{(\xi)}_{(\omega)}, \tag{2.23}$$

$$d_{(\xi)}{}^{\gamma} d^{(\xi)}{}_{\mu} = \delta^{\gamma}_{\mu}. \tag{2.24}$$

With the foregoing assumptions, we have

$$\det(d^{(\xi)}{}_{\lambda}) = 1, \quad \det(d_{(\omega)}{}^{\lambda}) = 1. \tag{2.25}$$

Thus, $d^{(\xi)}{}_{\lambda}$ *behaves like an orthochronous Lorentz* transformation, though, in general, it is not.

Now each point M of V^4 is equipped with a tetrad of *unit rigid vectors* or *directors*. We are therefore dealing with a *4-dimensional Cosserat continuum*.

The 4-vectors $\mathbf{d}^{(\xi)}$ (ξ *fixed*) may be decomposed into space- and timelike components by use of the projector $P^{\alpha\beta}$, i.e.,

$$d^{(\xi)}{}_{\alpha} = d^{(\xi)} u_{\alpha} + \bar{d}^{(\xi)}{}_{\alpha}, \tag{2.26}$$

where

$$d^{(\xi)} = -\, d^{(\xi)}{}_{\alpha} u^{\alpha}/c^2, \bar{d}^{(\xi)}{}_{\alpha} = P_{\alpha}{}^{\beta} d^{(\xi)}{}_{\beta}. \tag{2.27}$$

The following identities are satisfied:

$$\bar{d}^{(\xi)}{}_{\alpha} u^{\alpha} = 0, \bar{d}^{(\xi)}{}_{\alpha} \equiv P_{\alpha}{}^{\beta} \bar{d}^{(\xi)}{}_{\beta}. \tag{2.28}$$

We now propose to select

$$d^{(4)}{}_{\alpha} = u_{\alpha}/ic, \tag{2.29}$$

i.e., $\mathbf{d}^{(4)}$ is timelike and the remaining three directors are contained in the hypersurface $V^3_{\perp}$ orthogonal to the worldline $(\mathcal{C}_{X^K})$ at a point M of V^4 where $\mathbf{d}^{(\alpha)}$ are defined. We have

$$d_{(\xi)}{}^{\alpha} d^{(4)}{}_{\alpha} = (ic)^{-1} d_{(\xi)}{}^{\alpha} u_{\alpha} = 0, \quad \xi \neq 4. \tag{2.30}$$

Thus we set

$$d^{(\xi)}{}_{\alpha} = (d^{(K)}{}_{\alpha}, d^{(4)}{}_{\alpha}), \quad K = 1, 2, 3, \tag{2.31}$$

which is a special case of the decomposition (2.26). Introducing the new symbols $\chi_{\alpha}{}^{K}$, we have:

$$\begin{aligned} &\bar{d}^{(K)}{}_{\alpha} \equiv \chi_{\alpha}{}^{K} \neq 0, \quad d^{(K)} = 0, \\ &\bar{d}^{(4)}{}_{\alpha} = 0, \quad d^{(4)} = i^{-1}, \end{aligned} \tag{2.32}$$

and

$$\begin{aligned} &d^{(K)\alpha} = P^{\alpha}{}_{\beta} d^{(K)\beta} = \bar{d}^{(K)\alpha} = \chi^{\alpha K}, \quad u^{\alpha} d^{(K)}{}_{\alpha} = 0, \\ &P^{\alpha}{}_{\beta} d^{(4)\beta} = (ic)^{-1} P^{\alpha}{}_{\beta} u^{\beta} = 0, \quad d^{(4)\beta} u_{\beta} = ic. \end{aligned} \tag{2.33}$$

According to (2.9) and (2.10), in a rest frame, we obtain the following reduction:

$$d^{(K)}{}_{\alpha} \overset{*}{=} (\chi_{k}{}^{K}, 0), \quad K = 1, 2, 3. \tag{2.34}$$

We note that (2.23) can be written

$$d_{(K)}{}^{\gamma} d^{(K)}{}_{\mu} = \delta^{\gamma}_{\mu} + c^{-2} u^{\gamma} u_{\mu},$$

i.e.,

$$\chi^{\gamma}{}_{K} \chi_{\mu}{}^{K} = P^{\gamma}{}_{\mu} \tag{2.35}$$

with the definition

$$\bar{d}_{(K)}{}^{\gamma} \equiv \chi^{\gamma}{}_{K}, \quad d_{(K)} \equiv 0.$$

The motion of the directors: We consider tetrapods of unit directors of varying orientation from point to point in V^4. It must be noted that the instantaneous motion of a "*structured particle*" i.e., a material point M in V^4 equipped with a set of four directors) is characterized by the *infinitesimal Lorentz transformation* which defines the evolution of the system as the time goes on. Hence the proper time rate of the directors can be written

$$\frac{\partial}{\partial\tau} d^{(\xi)}{}_{\alpha} = \Omega_{\alpha\beta} d^{(\xi)\beta}, \tag{2.36}$$

where $\Omega_{\alpha\beta}$ is a skew-symmetric 4×4 matrix which has the general expression

$$\Omega^{\alpha\beta} = c^{-1}(A^{\alpha}u^{\beta} - A^{\beta}u^{\alpha}) + \overset{*}{\Omega}{}^{\alpha\beta} \tag{2.37}$$

with

$$\overset{*}{\Omega}{}^{\alpha\beta} \equiv P^{\alpha}{}_{\gamma}P^{\beta}{}_{\delta}\Omega^{\gamma\delta} = c^{-1}\epsilon^{\alpha\beta\gamma\delta}\omega_{\gamma}u_{\delta}, \quad \overset{*}{\Omega}{}^{\alpha\beta}u_{\beta} = 0, \tag{2.38}$$

$$A^{\alpha} = -c^{-1}\Omega^{\alpha\beta}u_{\beta}, \quad A^{\alpha}u_{\alpha} = 0. \tag{2.39}$$

In (2.38), we have defined the 4-vector ω^{α} by[43]

$$\omega^{\alpha} \equiv (2ic)^{-1}\epsilon^{\alpha\beta\gamma\delta}\Omega_{\beta\gamma}u_{\delta}, \quad \omega^{\alpha}u_{\alpha} = 0. \tag{2.40}$$

Dotting (2.36) with $d_{(\zeta)\gamma}$, we get

$$\Omega_{\alpha\gamma} = \frac{\partial}{\partial\tau} d^{(\xi)}{}_{[\alpha} \cdot d_{(\xi)\gamma]} = -\Omega_{\gamma\alpha}. \tag{2.41}$$

Note that this definition is slightly different from that of Kafadar and Eringen.[17] For the special choice (2.29), we see that

$$\Omega_{\alpha\beta} = \nu_{\alpha\beta} - c^{-2}\dot{u}_{[\alpha}u_{\beta]}, \tag{2.42}$$

where we have set

$$\nu_{\alpha\beta} = \frac{\partial}{\partial\tau} d^{(K)}{}_{[\alpha} \cdot d_{(K)\beta]} \equiv \frac{\partial}{\partial\tau} \chi_{[\alpha}{}^{K} \cdot \chi_{\beta]K} = -\nu_{\beta\alpha}, \tag{2.43}$$

while the following identities are satisfied:

$$\nu_{\alpha\beta}u^{\beta} = 0, \quad \Omega_{\alpha\beta}u^{\beta} = \dot{u}_{\alpha}, \quad u^{\alpha}\Omega_{\alpha\beta} = \dot{u}_{\beta}. \tag{2.44}$$

It follows that the elements of the decomposition (2.37) read

$$A^{\alpha} = -\dot{u}^{\alpha}/c, \quad \omega_{\gamma} = \nu_{\gamma} \equiv (2ic)^{-1}\epsilon_{\gamma\alpha\beta\lambda}\nu^{\alpha\beta}u^{\lambda}. \tag{2.45}$$

Reciprocally,

$$\nu_{\alpha\beta} = (ic)^{-1}\epsilon_{\alpha\beta\mu\lambda}\nu^{\mu}u^{\lambda}. \tag{2.46}$$

According to (2.9) and (2.10), in a rest frame, we obtain the reduction

$$\nu_{\mu} \overset{*}{=} (\nu_{k}, 0), \quad \nu_{\alpha\beta} \overset{*}{=} \begin{pmatrix} \nu_{kl} & 0 \\ 0 & 0 \end{pmatrix}. \tag{2.47}$$

We call $\Omega_{\alpha\beta}$ the *angular velocity* of the tetrapod and $\nu_{\alpha\beta}$ the *relativistic gyration tensor*. Note that, in a rest frame, $\Omega_{\alpha\beta}$ reduces to the classical gyration tensor (See Ref. 16) for the space–space components and to the linear acceleration for the space–time components, i.e.,

$$\Omega_{\alpha\beta} \overset{*}{=} \begin{pmatrix} \nu_{kl} & \dot{v}_{k}/ic \\ -\dot{v}_{k}/ic & 0 \end{pmatrix}. \tag{2.48}$$

An elementary derivation of (2.41) may be found in Halbwachs[35].

C. Introduction of the Spin

In order to measure the rotation of the tetrad $\mathbf{d}_{(\xi)}$, we need a reference.[44] We consider a 4-vector σ_{α} that will be called *spin*. In conformity with our ideas on quantum mechanical spin, we assume that σ_{α} is an *axial* vector which is *spacelike in nature* (cf. Uhlenbeck and Goudsmit). Hence, we take

$$\sigma_{\alpha}u^{\alpha} = 0. \tag{2.49}$$

With every axial vector σ_{α}, we may associate, in a *unique* way, a skew-symmetric second-order tensor, the *spin tensor* $\sigma_{\alpha\beta}$ defined by

$$\sigma_{\alpha\beta} \equiv (ic)^{-1}\epsilon_{\alpha\beta\lambda\mu}\sigma^{\lambda}u^{\mu}, \quad \sigma_{\alpha\beta}u^{\beta} = u^{\gamma}\sigma_{\gamma\alpha} = 0. \tag{2.50}$$

From (2.50) we solve for

$$\sigma_{\alpha} = (2ic)^{-1}\epsilon_{\alpha\beta\gamma\delta}\sigma^{\beta\gamma}u^{\delta}. \tag{2.51}$$

No restrictions being imposed on the directors, we can postulate that there exists an operator $\Sigma_{(\xi)\gamma\alpha}$ such that σ_{α} is expressed linearly as a function of the components of the directors, i.e.,

$$\sigma_{\alpha} = d^{(\xi)\gamma}\Sigma_{(\xi)\gamma\alpha}. \tag{2.52}$$

The explicit form of this operator is given in Sec. 3.

If the 4-vectors $\mathbf{d}^{(\xi)}$ form a *tetrad of unit rigid* directors and if the assumption (2.29) is used, then σ_{α}, which is rigidly attached to the tetrapod, can be expressed in the simplest linear combination of the $\mathbf{d}^{(\xi)}$. For instance, with σ colinear to $\mathbf{d}^{(3)}$, we could take

$$\sigma_{\alpha} = \rho\Sigma_{0}d^{(3)}{}_{\alpha}, \tag{2.53}$$

where ρ is the so-called invariant relativistic density defined later on and Σ_0 is the modulus of a standard particle spin per unit of proper mass. In this case, the meaning of the two remaining directors $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$ is left free, and any couple of unit vectors linked to the particle in its proper frame and orthogonal to both spin and 4-velocity is acceptable. Yet we shall not use a relation as particular as (2.53), and we shall consider the general relation (2.52) even if the assumption (2.29) holds.

3. MASS, INERTIA

A. Density

Let ρ_R be the material density of mass in the reference configuration of a material body (B) in E^3. Then, the so-called *invariant relativistic mass density* ρ is defined by

$$\rho = \rho_{R}J^{-1}, \tag{3.1}$$

where J is given by (2.17b) or, alternatively, by either one of the following formulas:

$$J \equiv (6ic)^{-1}\epsilon_{\alpha\beta\gamma\mu}x^{\alpha}{}_{,K}x^{\beta}{}_{,L}x^{\gamma}{}_{,M}u^{\mu}\epsilon^{KLM},$$
$$J = [\det(X^{\Delta}{}_{,\alpha})]^{-1}. \tag{3.2}$$

As time goes on, the material body (B) enclosed within a surface (∂B), sweeps out the 4-dimensional region

($\mathcal{B}$) of V^4. In ($\mathcal{B}$), ρ verifies the known *continuity equation*

$$(\rho u^\alpha)_{;\alpha} = 0, \quad \text{or} \quad \frac{\partial \rho}{\partial \tau} + \rho u^\alpha{}_{;\alpha} = 0, \tag{3.3}$$

while it satisfies the jump relation

$$[\rho u^\alpha]\Gamma_{,\alpha} = 0 \tag{3.4}$$

across a discontinuity hypersurface (Γ) of given equation $\Gamma(x^\alpha) = 0$. Here, the symbolism $[\cdots]$ denotes the jump.

B. Generalized Inertia

Generalizing the formula given by Eringen and Suhubi[6] and Eringen,[7] we posit that the *generalized kinetic energy of rotation in* V^4 assumes the form[4,5]

$$K = \tfrac{1}{2} I^{(\xi)(\zeta)} \dot{\mathbf{d}}_{(\xi)} \cdot \dot{\mathbf{d}}_{(\zeta)} \tag{3.5}$$

when no hypotheses are made concerning the $\mathbf{d}_{(\xi)}$. The object $I^{(\xi)(\zeta)}$ generalizes the notion of inertia; it is symmetric and, therefore, represents a set of ten independent quantities.

Upon use of (2.36), Eq. (3.5) yields

$$K = \tfrac{1}{2} \mathcal{I}_{\beta\gamma} \Omega^{\alpha\beta} \Omega_\alpha{}^\alpha, \tag{3.6}$$

where we have set

$$\mathcal{I}_{\beta\gamma} \equiv I^{(\xi)(\zeta)} d_{(\xi)\beta} d_{(\zeta)\gamma} \tag{3.7}$$

The symmetric second-order tensor $\mathcal{I}_{\beta\gamma}$ is called the *generalized inertia*. It admits a decomposition of the form

$$\mathcal{I}_{\alpha\delta} = j_{\alpha\delta} + u_\alpha b_\delta + c_\alpha u_\delta - e u_\alpha u_\delta \tag{3.8}$$

with

$$j_{\alpha\delta} \equiv \mathcal{I}_{\beta\gamma} P^\beta{}_\alpha P^\gamma{}_\delta, \tag{3.9a}$$

$$j_{\alpha\delta} u^\alpha = j_{\delta\alpha} u^\alpha = 0, \tag{3.9b}$$

$$b_\delta \equiv - P_\delta{}^\gamma \mathcal{I}_{\beta\gamma} u^\beta = c_\delta, \tag{3.9c}$$

$$b_\alpha u^\alpha = c_\alpha u^\alpha = 0, \tag{3.9d}$$

$$e \equiv - \mathcal{I}_{\beta\gamma} u^\beta u^\gamma. \tag{3.9e}$$

Now, by analogy with rational mechanics, we postulate the relation between the spin tensor and the angular velocity:

$$\sigma_{\gamma\alpha} = \mathcal{I}_{\beta[\gamma} \Omega_{\alpha]}{}^\beta. \tag{3.10}$$

Equivalently, with (2.51) and (2.40), we have, in 4-vector form,

$$\sigma_\gamma = \mathcal{I}_{\beta\gamma} \omega^\beta. \tag{3.11}$$

Then, Eq. (3.6) takes the form

$$K = \tfrac{1}{2} \sigma_{\gamma\alpha} \Omega^{\alpha\gamma} = \tfrac{1}{2} \sigma_\gamma \omega^\gamma \tag{3.12}$$

It is easily shown that

$$\sigma_{\gamma\alpha} = I_{(\xi)(\zeta)} d^{(\zeta)}{}_{[\gamma} \dot{d}^{(\xi)}{}_{\alpha]} \tag{3.13}$$

or, via (2.51),

$$\sigma_\mu = (2ic)^{-1} \epsilon_{\mu\gamma\alpha\beta} I_{(\xi)(\zeta)} d^{(\zeta)[\gamma} \dot{d}^{(\xi)\alpha]} u^\beta. \tag{3.14}$$

Thus, the operator $\Sigma_{(\zeta)\gamma\mu}$ defined in (2.52), is

$$\Sigma_{(\zeta)\gamma\mu} \equiv (2ic)^{-1} \epsilon_{\mu\gamma\alpha\beta} I_{(\xi)(\zeta)} \dot{d}^{(\zeta)\alpha} u^\beta. \tag{3.15}$$

Hence, Eqs. (2.52) and (3.11) are consistent.

Henceforth, we consider the case for which the directors $\mathbf{d}^{(\xi)}$ form a tetrad of unit rigid orthogonal vectors, $\mathbf{d}^{(4)}$ being defined according to (2.29). Then, upon use of (2.23), the relation (3.7) is invertible. Thus,

$$I^{(\xi)(\zeta)} = \mathcal{I}_{\beta\gamma} d^{(\xi)\beta} d^{(\zeta)\gamma}. \tag{3.16}$$

With the symbolism (2.32), we can write (3.7) as

$$\mathcal{I}_{\alpha\gamma} = I^{KL} \chi_{\alpha K} \chi_{\gamma L} + (2/ic) I^{K(4)} \chi_{(\alpha K} u_{\gamma)} - c^{-2} I^{(4)(4)} u_\alpha u_\gamma. \tag{3.17}$$

An identification term by term, with the general decomposition (3.8), yields

$$j_{\alpha\gamma} = I^{KL} \chi_{\alpha K} \chi_{\gamma L}, \tag{3.18a}$$

$$b_\gamma = c_\gamma = (ic)^{-1} I^{K(4)} \chi_{\gamma K}, \tag{3.18b}$$

$$e = c^{-2} I^{(4)(4)}. \tag{3.18c}$$

Note that I^{KL} is a quantity fixed for a given "structured" particle; it is referred to as the *material inertia density* tensor by Kafadar and Eringen.[17] It follows that $\partial I^{KL}/\partial \tau = 0$. Thus, by differentiating (3.18a) with respect to τ and using (2.43), we obtain

$$\frac{\partial}{\partial \tau} j_{\beta\gamma} = \dot{\overline{j_{\beta\gamma}}} = 2 j_{\sigma(\gamma} \nu_{\beta)}{}^\sigma. \tag{3.19}$$

This is the relativistic expression of the equations of conservation of inertia generalizing those first given by Eringen.[7]

From (3.9b) and (3.19), we get the useful result

$$j_{\alpha\beta} \dot{u}^\alpha = - \dot{\overline{j_{\alpha\beta}}} u^\alpha = 0. \tag{3.20}$$

A straightforward calculation, using (2.42) and (3.8), leads to an equivalent form for (3.6):

$$K = \tfrac{1}{2} j_{\beta\gamma} \nu^{\alpha\beta} \nu_\alpha{}^\gamma + \tfrac{1}{2} b_\gamma \nu^{\alpha\gamma} \dot{u}_\alpha - \tfrac{1}{8} e g_{\alpha\beta} \dot{u}^\alpha \dot{u}^\beta, \tag{3.21}$$

where we used (3.20).

Since, in the nonrelativistic limit, we cannot find any classical equivalents to the quantities b_γ and e, we shall set

$$b_\gamma \equiv 0, \quad e \equiv 0, \tag{3.22}$$

or, equivalently,

$$I^{K(4)} = 0, \quad I^{(4)(4)} = 0, \tag{3.23}$$

which assures the spacelike character of the inertia.

Thus we can write (3.6) and (3.12) in the forms

$$K = \tfrac{1}{2} j_{\beta\gamma} \nu^{\alpha\beta} \nu_\alpha{}^\gamma, \tag{3.24}$$

$$K = \tfrac{1}{2} \sigma_{\alpha\beta} \nu^{\beta\alpha} = \tfrac{1}{2} \sigma_\alpha \nu^\alpha = \tfrac{1}{2} j_{\alpha\beta} \nu^\alpha \nu^\beta. \tag{3.25}$$

Finally, we note that the derivative of K with respect to the proper time τ is (cf. Kafadar and Eringen[17])

$$\frac{\partial K}{\partial \tau} = \frac{\partial}{\partial \tau}(\sigma_{\mu\lambda})\nu^{\lambda\mu}. \tag{3.26}$$

4. VARIATIONAL PRINCIPLE FOR RELATIVISTIC POLAR MEDIA

A. The Action

With any tube $(\mathcal{B}) \subset V^4$, swept out by the material body (B) as time goes on, we associate the following action:

$$\mathcal{A} = \int_{(\mathcal{B})} \Lambda \, d^4v, \qquad \Lambda = -\rho\psi, \tag{4.1}$$

in which Λ is the Lagrangian density and ψ is the relativistic strain energy function. For a nonlinear polar elastic medium, a natural choice for the set of arguments of ψ is

$$X^{\Delta}{}_{,\beta}, \qquad d^{(\xi)}{}_{\beta}, \qquad d^{(\xi)}{}_{\beta,\Delta}. \tag{4.2}$$

The arguments of ψ must, of course, reduce to their classical analogs in an instantaneous rest frame. Thus, on account of (2.9) and (2.10), a reasonable set of arguments seems to be

$$X^{K}{}_{,\alpha}, \qquad \mathcal{K}^{\mu}{}_{K}, \qquad \mathfrak{X}^{\mu}{}_{KL} \tag{4.3}$$

with

$$X^{K}{}_{,\alpha}u^{\alpha} = 0, \qquad P_{\alpha}{}^{\beta}X^{K}{}_{,\beta} \equiv X^{K}{}_{,\alpha},$$

$$\mathcal{K}^{\mu}{}_{K} \overset{\text{DEF}}{=} P^{\mu}{}_{\lambda}\chi^{\lambda}{}_{K}, \qquad \mathcal{K}^{\mu}{}_{K}u_{\mu} = 0, \qquad P^{\sigma}{}_{\mu}\mathcal{K}^{\mu}{}_{K} \equiv \mathcal{K}^{\sigma}{}_{K} \tag{4.4}$$

$$\mathfrak{X}^{\mu}{}_{KL} \overset{\text{DEF}}{=} P^{\mu}{}_{\alpha}\chi^{\alpha}{}_{K:\lambda}\chi^{\lambda}{}_{L}, \quad \mathfrak{X}^{\mu}{}_{KL}u_{\mu} = 0, P^{\sigma}{}_{\mu}\mathfrak{X}^{\mu}{}_{KL} \equiv \mathfrak{X}^{\sigma}{}_{KL}.$$

In a rest frame, $\mathcal{K}^{\mu}{}_{K}$ and $\mathfrak{X}^{\mu}{}_{KL}$ reduce to $\chi^{k}{}_{K}$ and $\chi^{k}{}_{K:L}$ of Ref. 16.

With the choice (4.3) of constitutive arguments we see that ψ depends on $\chi^{\mu}{}_{K}$ and $\chi^{\mu}{}_{K:\alpha}$ only through $\mathcal{K}^{\mu}{}_{K}$ and $\mathfrak{X}^{\mu}{}_{KL}$, i.e., we have the

$$\hat{\psi}(X^{K}{}_{,\alpha}, \chi^{\mu}{}_{K}, \chi^{\mu}{}_{K:\alpha}) \equiv \psi(X^{K}{}_{,\alpha}, \mathcal{K}^{\mu}{}_{K}, \mathfrak{X}^{\mu}{}_{KL}) \tag{4.5}$$

if the following identities are satisfied:

$$\begin{aligned} u^{\mu}\frac{\partial\hat{\psi}}{\partial\chi^{\mu}{}_{K:\alpha}} &= 0, \qquad u_{\alpha}\frac{\partial\hat{\psi}}{\partial\chi^{\mu}{}_{K:\alpha}} = 0, \\ u^{\alpha}\frac{\partial\hat{\psi}}{\partial\chi^{\alpha}{}_{K}} &= 0, \qquad u_{\alpha}\frac{\partial\hat{\psi}}{\partial X^{K}{}_{,\alpha}} = 0. \end{aligned} \tag{4.6}$$

It follows that

$$\frac{\partial\hat{\psi}}{\partial\chi^{\mu}{}_{K}} = \frac{\partial\psi}{\partial\mathcal{K}^{\alpha}{}_{K}}P^{\alpha}{}_{\mu}, \qquad \frac{\partial\hat{\psi}}{\partial\chi^{\mu}{}_{K:\alpha}} = \frac{\partial\psi}{\partial\mathfrak{X}^{\nu}{}_{KL}}P^{\nu}{}_{\mu}\chi^{\alpha}{}_{L}. \tag{4.7}$$

B. Constraints

The directors $\mathbf{d}_{(\omega)}$ are mutually orthogonal and have unit lengths. These constraints are taken into account by introducing the ten Lagrange multipliers $\mathfrak{M}^{(\omega)}{}_{(\xi)}$ in the supplementary term to be included in the Lagrangian:

$$A = \tfrac{1}{2}\rho\mathfrak{M}^{(\omega)}{}_{(\xi)}(d_{(\omega)}{}^{\mu}d^{(\xi)}{}_{\mu} - \delta^{(\xi)}_{(\omega)}).$$

Note that the matrix $\mathfrak{M}_{(\omega)(\xi)}$ is symmetric.

With the choice (2.29) and the notations (2.32) a more explicit form of A is

$$A = \tfrac{1}{2}\rho\mathfrak{M}^{(K)}{}_{(L)}(\chi^{\mu}{}_{K}\chi_{\mu}{}^{L} - \delta^{L}_{K}) + \tfrac{1}{2}\rho\mathfrak{M}(u^{\alpha}u_{\alpha} + c^2), \tag{4.8}$$

in which we have set

$$\mathfrak{M}^{(4)}{}_{(4)} \equiv -c^2\mathfrak{M}. \tag{4.9}$$

The expression (4.8) generalizes the term $\frac{1}{2}\rho\mathfrak{M}(u^{\alpha}u_{\alpha} + c^2)$ introduced by Maugin and Eringen.[40]

We introduce the spin through an already-varied term.[46] The form of (2.43) suggests the introduction of an anholonomic, i.e., nonintegrable, four-angle variation $\delta\omega^{\alpha\beta}$ by

$$\begin{aligned} &\delta\omega^{\alpha\beta} \overset{\text{DEF}}{=} P^{[\alpha}{}_{\gamma}\delta\chi^{\gamma}{}_{K}\chi^{\beta]K}, \\ &\delta\omega^{\alpha\beta}u_{\beta} = 0, \qquad u_{\alpha}\delta\omega^{\alpha\beta} = 0, \qquad \delta\omega^{\alpha\beta} = -\delta\omega^{\beta\alpha}. \end{aligned} \tag{4.10}$$

To take account of the spin, we then insert the following integral:

$$\delta W = \int_{(\mathcal{B}-\Gamma)} \rho\dot{\sigma}_{\alpha\beta}\delta\omega^{\alpha\beta}d^4v \tag{4.11}$$

into the variational principle.

For each basic argument varied in the Lagrangian density, i.e., the classical motion and the χ's (or equivalently $\omega^{\alpha\beta}$), we introduce indeterminate multipliers f^{α}, T^{α}, $L_{\mu\gamma}$ and $m_{\mu\gamma}$ in $(\mathcal{B})$ and on $(\partial\mathcal{B})$ by

$$\begin{aligned} \delta W^* = &\int_{(\mathcal{B}-\Gamma)} \rho f^{\alpha}\delta x_{\alpha}\, d^4v - \int_{(\partial\mathcal{B}-\Gamma)} T^{\alpha}\delta x_{\alpha}\, d^3s \\ &- \int_{(\mathcal{B}-\Gamma)} \rho L_{\mu\gamma}\delta w^{\gamma\mu}d^4v - \int_{(\partial\mathcal{B}-\Gamma)} m_{\mu\gamma}\delta w^{\gamma\mu}d^3s. \end{aligned} \tag{4.12}$$

Here (Γ) is a discontinuity surface in $(\mathcal{B})$ whose unit positive normal is denoted by N_{α}. The terminology associates a physical significance to each of these multipliers namely: f^{α} is the four-body force, T^{α} is the stress 4-vector, $L_{\alpha\beta}$ is the body couple, and $m_{\alpha\beta}$ is the microstress tensor (or bivector). We have

$$\begin{aligned} &f^{\alpha} = fu^{\alpha} + \bar{f}^{\alpha}, \qquad f = -c^{-2}f^{\alpha}u_{\alpha}, \\ &\bar{f}^{\alpha} = P^{\alpha}{}_{\gamma}f^{\gamma}, \qquad \bar{f}^{\alpha}u_{\alpha} = 0, \end{aligned} \tag{4.13}$$

$$L_{\alpha\beta} = -L_{\beta\alpha}, \qquad L^{\alpha\beta}u_{\beta} = 0, \tag{4.14}$$

$$m_{\alpha\beta} = -m_{\beta\alpha}. \tag{4.15}$$

Finally the variational principle reads

$$\delta\mathcal{A} + \delta W + \delta W^* = 0, \tag{4.16}$$

where $\mathcal{A}$ is given by

$$\begin{aligned} &\mathcal{A} = \int_{(\mathcal{B}-\Gamma)} \Lambda^* d^4v, \\ &\Lambda^* = -\rho\psi + \tfrac{1}{2}\rho\mathfrak{M}^{(K)}{}_{(L)}(\chi^{\mu}{}_{K}\chi_{\mu}{}^{L} - \delta^{L}_{K}) \\ &\qquad\qquad + \tfrac{1}{2}\rho\mathfrak{M}(u^{\alpha}u_{\alpha} + c^2), \end{aligned} \tag{4.17}$$

and δW and δW^* are respectively given by (4.11) and (4.12).

5. THE VARIATION

The δ-*variation* is defined as follows: The fields associated with material points in V^4 are considered to depend on a parameter λ, e.g.,

$$x^{\alpha} = x^{\alpha}(X^K, \tau, \lambda), \qquad \text{etc.} \tag{5.1}$$

It is assumed that for $\lambda = \lambda_0$ these quantities describe a given state of the body. The variation is then defined as:

$$\delta x^\alpha = \frac{\partial x^\alpha}{\partial \lambda}\Big|_{X^K,\tau}\delta\lambda, \quad \text{etc.} \tag{5.2}$$

Of particular interest are the following results:

$$\delta X^K{}_{,\alpha} = -X^K{}_{,\beta}(\delta x^\beta)_{;\alpha}, \tag{5.3a}$$

$$\delta\rho = -\rho P^\alpha{}_\beta(\delta x^\beta)_{;\alpha}, \tag{5.3b}$$

$$\delta(d^4v) = \delta^\alpha_\beta(\delta x^\beta)_{;\alpha}d^4v, \tag{5.3c}$$

$$\delta u^\alpha = (\delta x^\alpha)_{;\lambda}u^\lambda, \tag{5.3d}$$

$$\delta P_{\alpha\mu} = c^{-2}(\delta x^\gamma)_{;\lambda}u^\lambda(P_{\gamma\mu}u_\alpha + P_{\alpha\gamma}u_\mu), \tag{5.3e}$$

$$\delta(\phi_{;\alpha}) = (\delta\phi)_{;\alpha} - \phi_{;\alpha}(\delta x^\gamma)_{;\alpha}. \tag{5.3f}$$

The latter relation valid for any tensor ϕ is derived as follows: Consider $\delta(\phi_{;\alpha})$ where ϕ is any tensorial object. We have

$$\delta(\phi_{;\alpha}) = \delta(\phi_{;\Delta}X^\Delta{}_{,\alpha}),$$

where

$$X^\Delta{}_{,\alpha} = (X^K{}_{,\alpha}, ic\tau_{,\alpha}), \qquad \Delta = 1,2,3,4, \quad K = 1,2,3.$$

Carrying out the variation, we obtain

$$\delta(\phi_{;\alpha}) = \delta(\phi_{;\Delta})X^\Delta{}_{,\alpha} + \phi_{;\Delta}\delta(X^\Delta{}_{,\alpha})$$

upon interchange of the δ variation and the derivative with respect to the generalized material coordinates $X^\Delta = (X^K, ic\tau)$, and, using (5.3a), we obtain (5.3f).

We also note that

$$\begin{aligned}
&\delta\chi^\alpha{}_L = \chi_{\beta L}\delta\omega^{\alpha\beta},\\
&\delta(\chi^\mu{}_{K:\beta}) = (\chi_{\gamma K}\delta\omega^{\mu\gamma})_{;\beta} - \chi^\mu{}_{K:\lambda}(\delta x^\lambda)_{;\beta},\\
&\delta(\chi^\mu{}_K\chi_\mu{}^L - \delta^L_K) = 0.
\end{aligned} \tag{5.4}$$

In performing the variation of various integrals we also need the generalized Green–Gauss theorem:

$$\int_{(\mathfrak{B}-\Gamma)}\mathbf{A}^\alpha{}_{;\alpha}d^4v = \int_{(\partial\mathfrak{B}-\Gamma)}\mathbf{A}^\alpha n_\alpha d^3s - \int_{(\Gamma)}[\mathbf{A}^\alpha]N_\alpha d^3s_\Gamma. \tag{5.5}$$

We now carry out the variations indicated by (4.16). Upon using (5.3), (5.4), and integrations by parts, with the help of (5.5), we obtain

$$\begin{aligned}
&\int_{(\mathfrak{B}-\Gamma)}(T^{\alpha\beta}{}_{;\beta} - \rho f^\alpha)\delta x_\alpha d^4v + \int_{(\Gamma)}[T^{\alpha\beta}\delta x_\alpha]N_\beta d^3s_\Gamma\\
&- \int_{(\partial\mathfrak{B}-\Gamma)}(T^{\alpha\beta}n_\beta - T^\alpha)\delta x_\alpha d^3s - \int_{(\partial\mathfrak{B}-\Gamma)}(m^\alpha{}_{\mu\beta}n_\alpha - m_{\mu\beta})\\
&\times\delta\omega^{\mu\beta}d^3s + \int_{(\mathfrak{B}-\Gamma)}(m^\alpha{}_{\mu\beta;\alpha} - \rho\dot\sigma_{\mu\beta} + \rho L_{\mu\beta} - \rho s^*_{\mu\beta})\delta\omega^{\mu\beta}d^4v\\
&+ \int_{(\Gamma)}[m^\alpha{}_{\mu\beta}\delta\omega^{\mu\beta}]N_\alpha d^3s_\Gamma = 0,
\end{aligned} \tag{5.6}$$

where we have defined

$$T^{\alpha\beta} = \omega u^\alpha u^\beta - \bar t^{\beta\alpha}, \qquad \bar t^{\beta\alpha} = t^{\beta\alpha} + \tilde t^{\beta\alpha}, \tag{5.7}$$

$$\omega \equiv \rho(\mathfrak{M} + \frac{\psi}{c^2}), \tag{5.8a}$$

$$t^{\beta\alpha} \equiv -\rho\frac{\partial\hat\psi}{\partial X^K{}_{,\beta}}X^{K,\alpha} \tag{5.8b}$$

$$\tilde t^{\beta\alpha} \equiv -\rho\frac{\partial\hat\psi}{\partial\chi^\mu{}_{K:\beta}}\chi^\mu{}_K{}^{:\alpha} \equiv -m^\beta{}_{\mu\lambda}\chi^{\lambda L}\chi^\mu{}_L{}^{:\alpha}, \tag{5.9}$$

$$m^\alpha{}_{\mu\beta} \equiv \rho\frac{\partial\hat\psi}{\partial\chi^\mu{}_{K:\alpha}}\chi_{\beta K} = \rho\frac{\partial\psi}{\partial\mathfrak{X}^\nu{}_{KL}}P^\nu{}_\mu\chi_{\beta K}x^\alpha{}_L, \tag{5.10}$$

$$\overset{*}{s}{}_{\mu\beta} \equiv \frac{\partial\hat\psi}{\partial\chi^\mu{}_K}\chi_{\beta K} + \frac{\partial\hat\psi}{\partial\chi^\mu{}_{K:\alpha}}\chi_{\beta K:\alpha}. \tag{5.11}$$

Some of these quantities are granted a physical significance: $T^{\alpha\beta}$ is the *total stress–energy–momentum tensor*, $t^{\beta\alpha}$ is the *stress–energy–momentum tensor due to the deformation field*, $m^\alpha{}_{\mu\beta}$ is the *relativistic stress moment tensor*, and ω is the *density of energy* (of any kind) per unit of proper volume.

Note that

$$\bar t^{\beta\alpha}u_\alpha = t^{\beta\alpha}u_\alpha - m^\beta{}_{\mu\lambda}\nu^{\mu\lambda}, \tag{5.12}$$

$$\bar t^{\beta\alpha}u_\beta = 0. \tag{5.13}$$

Equation (5.12) follows from (5.8b), (5.9), and (2.43). Equation (5.13) is a consequence of (2.15) and the fact that

$$m^\alpha{}_{\mu\beta}u_\alpha = 0, \qquad m^\alpha{}_{\mu\beta}u^\mu = 0, \qquad m^\alpha{}_{\mu\beta}u^\beta = 0, \tag{5.14}$$

which follows from (2.15), (4.6), and (2.33).

Thus $t^{\beta\alpha}$ can be written in the equivalent form

$$\begin{aligned}
&\bar t^{\beta\alpha} = t^{\beta\alpha} = t^{\beta\alpha} - u^\alpha\mathcal{Q}^\beta,\\
&\mathcal{Q}^\beta = -c^{-2}m^\beta{}_{\mu\lambda}\nu^{\lambda\mu},
\end{aligned} \tag{5.15}$$

and with (5.13) we see that u_β is a right eigenvector for $T^{\alpha\beta}$ with the corresponding eigenvalue $-\omega c^2$ since

$$T^{\alpha\beta}u_\beta = -\omega c^2u^\alpha. \tag{5.16}$$

If the expression (5.6) is posited to be valid for any volume in $(\mathfrak{B})$ and any hypersurface and for any variations δx_α and $\delta\omega^{\mu\beta}$ such that

$$[\delta x^\alpha] = [\delta\omega^{\mu\beta}] = 0 \quad \text{across } (\Gamma),$$

then we obtain the local field equations

$$T^{\alpha\beta}{}_{;\beta} = \rho f^\alpha \quad \text{in } (\mathfrak{B}-\Gamma), \tag{5.17a}$$

$$T^{\alpha\beta}n_\beta = T^\alpha \quad \text{on } (\partial\mathfrak{B}-\Gamma), \tag{5.17b}$$

$$\rho\dot\sigma_{\mu\beta} - m^\alpha{}_{[\mu\beta];\alpha} = \rho L_{\mu\beta} + \rho s^*_{[\mu\beta]} \quad \text{in } (\mathfrak{B}-\Gamma), \tag{5.18}$$

$$m^\alpha{}_{[\mu\beta]}n_\alpha = m_{\mu\beta} \quad \text{on } (\partial\mathfrak{B}-\Gamma), \tag{5.19}$$

$$[T^{\alpha\beta}]N_\beta = 0, \tag{5.20a}$$

$$[m^\alpha{}_{[\mu\beta]}]N_\alpha = 0 \quad \text{on } (\Gamma). \tag{5.20b}$$

Equations (5.17) and (5.20a) represent the balance of energy–momentum while (5.18), (5.19), and (5.20b) represent the balance of moment of energy–momentum. An alternative elegant form for the latter may be obtained by introducing the *total spin–energy–momentum tensor* $\mathcal{S}$ by:

$$\mathcal{S}^{\mu\beta\alpha} = \rho\sigma^{\mu\beta}u^\alpha - m^{\alpha[\mu\beta]} \tag{5.21}$$

such that

$$\mathcal{S}^{\mu\beta\alpha}u_\mu = 0, \qquad \mathcal{S}^{\mu\beta\alpha}u_\beta = 0, \qquad \mathcal{S}^{\mu\beta\alpha}u_\alpha = -\rho c^2\sigma^{\mu\beta}. \tag{5.22}$$

Thus

$$S^{\mu\beta\alpha}{}_{;\alpha} \equiv \rho\dot{\sigma}^{\mu\beta} - m^{\alpha[\mu\beta]}{}_{;\alpha} \tag{5.23}$$

from the equation of continuity (3.3) and the definition (2.5). If we now add and subtract $T_{[\mu\beta]}$ to the left-hand side of (5.18), taking account of (5.7), (5.8), and (5.21) and raising the indices, we obtain

$$S^{\mu\beta\alpha}{}_{;\alpha} - T^{[\mu\beta]} = \rho L^{\mu\beta} + \rho s^{[\mu\beta]}, \tag{5.24}$$

where we have defined

$$s^{\mu\beta} = \overset{*}{s}{}^{\mu\beta} - \frac{\partial\hat{\psi}}{\partial X^K{}_{,\beta}} X^{K,\mu} - \frac{\partial\hat{\psi}}{\partial\chi_{\alpha K:\mu}} \chi_{\alpha K}{}^{:\beta}, \tag{5.25}$$

but we must have

$$s^{[\mu\beta]} = 0 \quad \text{and} \quad \frac{\partial\hat{\psi}}{\partial x^\alpha} = 0. \tag{5.26}$$

These two equations are consequences of the *Lorentz invariance* requirement (equivalent to the *Euclidean invariance* requirement used by Toupin,[5] Maugin,[8] Maugin and Eringen[22] in classical continuum mechanics) postulated as follows:

The balance laws follow from the invariance of the variational principle under the inhomogeneous proper group of Lorentz (or Poincaré group) Λ_P. (For more details see Ref. 40)

Indeed, an infinitesimal mapping generated by the Lie group Λ_P is written

$$x^{*\mu} = (\delta^\mu_\nu + \epsilon\mathcal{Q}^\mu{}_\nu)x^\nu + d^\mu, \tag{5.27}$$

where ϵ is an infinitesimally small, $\mathcal{Q}$ is a constant skew-symmetric 4×4 matrix, and d is an infinitesimally small constant 4-vector. If one requires the relativistic strain energy function $\hat{\psi}$ to be invariant under the mapping (5.27), then Eqs. (5.26) follow. The first-order partial differential equations (5.26) can be integrated. A solution is given in Sec. 8 on relativistic objectivity. With (5.26) satisfied, equation (5.24) takes the canonical form of a balance law of moment of energy–momentum (see Grot and Eringen[39]):

$$S^{\mu\beta\alpha}{}_{;\alpha} - T^{[\mu\beta]} = \rho L^{\mu\beta}.$$

6. DETERMINATION OF THE LAGRANGE MULTIPLIER $\mathfrak{M}$

We perform the differentiation in (5.17a) and multiply the result with u_α. Upon using (3.3) and (4.13), after some manipulations, we obtain

$$-\rho c^2\dot{\mathfrak{M}} - \rho\dot{\psi} - t^{\beta\alpha}{}_{;\beta}u_\alpha - m^{\beta\mu\lambda}{}_{;\beta}\nu_{\mu\lambda} + \frac{\partial\hat{\psi}}{\partial\chi^\mu{}_{K:\beta}}\dot{\overline{\chi^\mu{}_{K:\beta}}} + \rho\frac{\partial\hat{\psi}}{\partial\chi^\mu{}_{L:\beta}}\chi^\lambda{}_L\chi_\lambda{}^K{}_{:\beta}\dot{\overline{\chi^\mu{}_K}} = -\rho c^2 f. \tag{6.1}$$

Contraction of (5.18) with $\nu^{\beta\mu}$ yields

$$\rho\dot{\sigma}_{\mu\beta}\nu^{\beta\mu} - m^\alpha{}_{\mu\beta;\alpha}\nu^{\beta\mu} = \rho L_{\mu\beta}\nu^{\beta\mu} - \rho\frac{\partial\hat{\psi}}{\partial\chi_{\mu K}}\dot{\overline{\chi_{\mu K}}} + \rho\frac{\partial\hat{\psi}}{\partial\chi^\mu{}_{L:\alpha}}\chi^\lambda{}_L\chi_\lambda{}^K{}_{:\alpha}\dot{\overline{\chi^\mu{}_K}}. \tag{6.2}$$

In establishing (6.1) and (6.2) we have made use of (2.43) and of the relation

$$\chi^\beta{}_{K:\alpha}\chi_\beta{}^\nu = -\chi^\beta{}_K\chi_\beta{}^L{}_{:\alpha}, \tag{6.3}$$

which follows from (2.25). Note that, upon using (4.6d) and (2.15), we have

$$t^{\beta\alpha}{}_{;\beta}u_\alpha = -t^{\beta\alpha}u_{\alpha;\beta} = -\rho\frac{\partial\hat{\psi}}{\partial X^K{}_{,\beta}}\dot{\overline{X^K{}_{,\beta}}}. \tag{6.4}$$

On adding (6.2) to (6.1), we get

$$c^2\dot{\mathfrak{M}} + \dot{\psi} = (c^2 f - L_{\mu\beta}\nu^{\beta\mu}) + \dot{\sigma}_{\mu\beta}\nu^{\beta\mu} + \left(\frac{\partial\hat{\psi}}{\partial X^K{}_{,\beta}}\dot{\overline{X^K{}_{,\beta}}} + \frac{\partial\hat{\psi}}{\partial\chi_{\mu K}}\dot{\overline{\chi_{\mu K}}} + \frac{\partial\hat{\psi}}{\partial\chi^\mu{}_{K:\beta}}\dot{\overline{\chi^\mu{}_{K:\beta}}}\right). \tag{6.5}$$

We notice that the last term is nothing but $\dot{\hat{\psi}}$ and that

$$\dot{\sigma}_{\mu\beta}\nu^{\beta\mu} = \tfrac{1}{2}\dot{\overline{\sigma_{\mu\beta}\nu^{\beta\mu}}}. \tag{6.6}$$

Hence

$$c^2\dot{\mathfrak{M}} = \tfrac{1}{2}\dot{\overline{\sigma_{\mu\beta}\nu^{\beta\mu}}} + (c^2 f - L_{\mu\beta}\nu^{\beta\mu}). \tag{6.7}$$

If we assume that

$$f = c^{-2}L_{\mu\beta}\nu^{\beta\mu}, \tag{6.8}$$

then, by integration over proper time, (6.7) yields

$$c^2\mathfrak{M} = c^2 + \tfrac{1}{2}\sigma_{\mu\beta}\nu^{\beta\mu}, \tag{6.9}$$

where c^2 (the density of rest energy) is the constant of integration. We have thus the final form of the *total stress–energy–momentum tensor*,

$$T^{\alpha\beta} = \rho(1 + \psi c^{-2} + \tfrac{1}{2}c^{-2}\sigma_{\mu\beta}\nu^{\beta\mu})u^\alpha u^\beta - t^{\beta\alpha} - c^{-2}m^{\beta\gamma\mu}\nu_{\mu\gamma}u^\alpha, \tag{6.10}$$

while the 4-force reads

$$f^\alpha = \bar{f}^\alpha + c^{-2}L_{\mu\beta}\nu^{\beta\mu}. \tag{6.11}$$

Equations (3.3), (5.17), (5.19), (5.20), (5.24) subject to (5.26) with the definitions (6.10), (6.11), (5.8b), and (5.10) are in agreement with the results previously obtained by Kafadar and Eringen.[17] However, we must emphasize that the present work, in contrast to Ref. 17, is limited to nondissipative processes.

7. EQUATION OF BALANCE OF ENERGY

The equation of local energy balance has already been obtained in the process of determination of $\mathfrak{M}$. Indeed, with the known values of $\mathfrak{M}$ and f, (6.1) and (6.2) can be combined to yield

$$\rho\dot{\psi} - m^{\beta\mu\gamma}\nu_{\gamma\mu;\beta} + t^{\beta\alpha}{}_{;\beta}u_\alpha + \rho\left(\frac{\partial\hat{\psi}}{\partial\chi_{\gamma K}}\chi^\mu{}_K + \frac{\partial\hat{\psi}}{\partial\chi_{\lambda K:\gamma}}\chi_{\lambda K}{}^{:\mu} + \frac{\partial\hat{\psi}}{\partial\chi_{\gamma K:\lambda}}\right)\chi^\mu{}_{K:\lambda} \times \nu_{\gamma\mu} = 0. \tag{7.1}$$

The last term is none other than $-t^{[\beta\alpha]}\nu_{\beta\alpha}$ from (2.25) and (5.8b). Finally, upon use of (6.4), we obtain

$$\rho\dot{\psi} - m^{\beta\mu\gamma}\nu_{\gamma\mu;\beta} - t^{\beta\alpha}(u_{\alpha;\beta} + \nu_{\beta\alpha}) = 0, \tag{7.2}$$

which corresponds to the energy equation of Kafadar and Eringen (with $p^\alpha = q^\alpha = 0$ and the nonrealistic $\mathcal{K}^\alpha$ and $\mathcal{K}^{\alpha\beta}$ null). The relativistic Cauchy's equations are obtained by applying the projector $P^\alpha{}_\beta$ to the

equations (5.17a) and (5.24). We refer to Kafadar and Eringen[17] for these results.

8. RELATIVISTIC OBJECTIVITY

We use herein the results of Söderholm[42] already used in Maugin and Eringen.[40]

Starting with the relativistic strain energy function

$$\psi = \psi\,(X^K{}_{,\alpha}, \mathcal{K}^\mu{}_K, \mathcal{X}^\mu{}_{KL}), \tag{8.1}$$

we say that ψ satisfies the principle of relativistic objectivity in the sense of Söderholm if

$$\psi(\overset{-1}{\mathbf{F}}, \mathcal{K}, \mathcal{X}) = \tilde{\psi}(\mathcal{Q}\overset{-1}{\mathbf{F}}, \mathcal{Q}\,\mathcal{K}, \mathcal{Q}\,\mathcal{X}) \tag{8.2}$$

where $\mathcal{Q}$ is an orthogonal time-dependent transformation. For instance, let us select

$$\mathcal{Q}^\alpha{}_L = \chi^\alpha{}_L. \tag{8.3}$$

Then

$$\psi = \tilde{\psi}(\chi^{\alpha L} X^K{}_{,\alpha}, \chi^\mu{}_L \mathcal{K}_{\mu K}, \chi^\mu{}_M \mathcal{X}_{\mu KL}), \tag{8.4}$$

but

$$\chi^\mu{}_L \mathcal{K}_{\mu K} = \chi^\mu{}_L P_{\mu\alpha} \chi^\alpha{}_K = g_{KL}; \tag{8.5}$$

thus

$$\psi = \tilde{\psi}(\overset{-1}{\mathcal{C}}{}^{LK}, \Gamma_{KLM}), \tag{8.6}$$

where we have defined the *inverse relativistic Cosserat deformation* $\overset{-1}{\mathcal{C}}{}^{LK}$ and the *relativistic wryness tensor* Γ_{KLM} by

$$\overset{-1}{\mathcal{C}}{}^{LK} \equiv \chi^{\alpha L} X^K{}_{,\alpha}, \qquad \Gamma_{KLM} \equiv \chi^\mu{}_M \mathcal{X}_{\mu KL}. \tag{8.7}$$

Only the skew-symmetric part in K and M of Γ has to be considered. The 18 independent components of $\overset{-1}{\mathcal{C}}$ and Γ form a *minimal function basis* for ψ. It is straightforward to verify that (8.6) constitutes a solution for the partial differential equations (5.26). The requirement of relativistic objectivity and the Lorentz invariance requirement lead therefore, in the present case, to the same functional form for ψ. (This is to compare to the classical nonlinear theory of elasticity where coordinate frame invariance and objectivity yield the same result. Hence the study of objectivity is not worthwhile in the present case.)

With (8.6), the constitutive equations (5.8b) and (5.10) become

$$t^{\mu\lambda} = -\rho \frac{\partial\tilde{\psi}}{\partial \overset{-1}{\mathcal{C}}{}^{MN}} \chi^{\mu M} X^{N,\lambda},$$

$$m^{\lambda[\alpha\gamma]} = \rho \frac{\partial\tilde{\psi}}{\partial \Gamma_{KLM}} \chi^{[\alpha}{}_M \chi^{\gamma]}{}_K x^\lambda{}_L. \tag{8.8}$$

We have thus established a variational principle which provides the nondissipative counterpart of the general theory of relativistic polar media of Kafadar and Eringen.

9. PROSPECTS

The presentation of directors given in Sec. 2 is somewhat more general than the notion of micromotion of Kafadar and Eringen and, therefore, may help to solve, or at least to formulate, a set of equations for the *general theory* of *relativity* with *nonsymmetric* energy–momentum tensor. This attempt is not to be confused with the unified field theory ambition (i.e., we do not aim at describing all field phenomena in terms of geometrical objects). We just mention the fact that a generalization of the concept of *micromorphic media* (cf. Eringen and Suhubi[6]), more restrictly of *polar media* could find place in general relativity as was pointed out by Eringen[47] and discussed by Kafadar and Eringen.[17]

In general relativity [Einstein (1916)], we have the Einstein equations

$$G^{\alpha\beta} = \kappa T^{\alpha\beta} \tag{9.1}$$

and the balance law of energy–momentum

$$T^{\alpha\beta}{}_{;\beta} = 0. \tag{9.2}$$

The Einstein–Cartan tensor $G^{\alpha\beta}$ is defined as

$$G^{\alpha\beta} \equiv R^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta} R, \tag{9.3}$$

where $R^{\alpha\beta}$ is the Ricci curvature, R is the scalar curvature, and $g^{\alpha\beta}$ is the Riemannian normal hyperbolic symmetric metric of the universe manifold. κ is a constant proportional to Newton's constant of gravitation. In (9.1), the lhs has a pure geometrical significance, the rhs representing the source of energy–momentum (e.g., the electromagnetic stress–energy–momentum tensor in vacuum).

Equation (9.2) follows from (9.1) since there exists the demonstrable identity

$$G^{\alpha\beta}{}_{;\beta} \equiv 0. \tag{9.4}$$

It seems that a possible generalization could be to add to the motion x^α the set of directors $\mathbf{d}^{(\xi)}$ ($\xi = 1, 2, 3, 4$) (here we do not take $\mathbf{d}^{(4)} \propto \mathbf{u}$) and to add to (9.1), (9.2) the two equations

$$S^{\alpha\beta\mu}{}_{;\mu} - T^{[\alpha\beta]} = 0, \tag{9.5}$$

$$f^{\alpha\beta\mu}(\mathbf{K}) \propto S^{\alpha\beta\mu}. \tag{9.6}$$

By the foregoing statement, we mean that (9.5) and (9.6) are, respectively, the set of dynamical equations and the geometry-source relations that could supplement (9.2) and (9.1). $\mathbf{f}(\mathbf{K})$ is an "ad hoc" tensor-valued functional of the torsion tensor $K_\alpha{}^\beta{}_\gamma$ with

$$K_\alpha{}^\beta{}_\gamma = \Gamma_{[\alpha}{}^\beta{}_{\gamma]}, \tag{9.7}$$

where $\Gamma_\alpha{}^\beta{}_\gamma$ is the connection defined independently of the metric $g_{\alpha\beta}$. The latter is no longer symmetric. Since (9.5) is due to the consideration of directors $\mathbf{d}^{(\xi)}$, $\Gamma_\alpha{}^\beta{}_\gamma$ must be linked in some way to the $\mathbf{d}^{(\xi)}$. A suggestion provided for by the modern theories of dislocations is

$$\Gamma_{[\alpha}{}^\beta{}_{\gamma]} = d_{(\xi)}{}^\beta \nabla_{[\alpha} d^{(\xi)}{}_{\gamma]}, \tag{9.8}$$

where ∇_α denotes the covariant derivative with respect to $g_{\alpha\beta}$. This is, of course, a mere conjecture that could be considered as a starting point. Such an ambitious project is, however, beyond the scope of the present article and is left for further investigations.

* This work was supported by the National Science Foundation.
† Now at Université de Paris, Faculté des Sciences, Mécanique théorique, Tour 66-9, quai Saint-Bernard, 75005 Paris, France.

1 P. Duhem, Ann. Ecole Norm. Paris, **10**, 187 (1893).
2 W. Voigt, Abhandl. Wiss. Ges. Göttingen **34**, (1887).
3 E. Cosserat and F. Cosserat, *Théorie des corps déformables* (Hermann, Paris, 1909).
4 J. L. Eriksen and C. Truesdell, Arch. Ratl. Mech. Anal. **1**, 295 (1958).
5 R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).
6 A. C. Eringen and E. S. Suhubi, Intern. J. Eng. Sci. **2**, 189 (1964).
7 A. C. Eringen, Intern. J. Eng. Sci. **2**, 205 (1964).
8 G. Maugin, Compt. Rend. **271A**, 807 (1970).
9 A. C. Eringen, "Mechanics of Micromorphic Continua," in *Mechanics of Generalized Continua*, I.U.T.A.M. Symposium 1967, edited by E. Kröner (Springer, Berlin, 1968).
10 A. C. Eringen, Intern. J. Engng. Sci., **8**, 819-828 (1970).
11 A. C. Eringen, "Theory of Micropolar Continua" in Proceedings of the Midwestern Mech. Conference, Madison, Wisconsin, 1965.
12 A. C. Eringen, J. Math. Mech. **16**, 1 (1966).
13 A. C. Eringen, Intern. J. Eng. Sci. **5**, 191 (1967).
14 A. C. Eringen, Intern. J. Eng. Sci. **7**, 115 (1969).
15 A. C. Eringen, "Theory of Micropolar Elasticity", in *Fracture: An Advanced Treatise*, edited by H. Liebowitz (Academic, New York, 1968), Vol. II.
16 C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 271 (1971).
17 C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 307 (1971).
18 J. L. Eriksen, Trans. Soc. Rheol. **5**, 23 (1961).
19 A. E. Green and N. Laws, Arch. Ratl. Mech. Anal. **25**, 285 (1961).
20 A. E. Green and P. M. Naghdi, in Ref. 9.
21 J. B. Alblas, in Ref. 9.
22 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 143 (1972).
23 The literature on the subject of oriented media is extensive; for various summaries and review articles, see Eringen[9] and R. Stojanović, *Mechanics of Polar Continua*, Publ. C.I.S.M. No. 2 (C.I.S.M., Udine, Italy, 1969).
24 See for instance the striking analogy between, on one hand, the use of vierbeins in Einstein's (zero curvature – nonzero torsion) theory [S.B. Preuss. Akad. Wiss., 217 (1928)] the use of directors in 3-dimensional theories of dislocations with distant parallelism where the torsion is defined from the directors field (see Claus and Eringen,[25] Fox,[26] and Stojanović[27] and, on the other hand, the multipole-like theories of oriented media (for instance, Green and Rivlin[28]) and the multipole formalism of general relativistic theories (see Tulczyjew and Tulczyjew[29]).
25 W. D. Claus and A. C. Eringen, O.N.R. Technical Report No. 10 (Princeton University, Department of Aerospace and Mechanical Sciences, 1969).
26 N. Fox, Quart. J. Mech. Appl. Math. **19**, 343 (1966).
27 R. Stojanović, in Ref. 9.
28 A. E. Green and R. S. Rivlin, Arch. Ratl. Mech. Anal. **16**, 325 (1964).
29 B. Tulczyjew and W. Tulczyjew, in *Recent Developments in General Relativity* (Pergamon, Oxford and P. W. N., Warsaw, 1962).
30 In German "vierbein, tetrade"; in English "four-legs"; see E. Schmutzer, *Relativitische Physik* (Akademie-Verlag, Leipzig, 1968).
31 H. A. Kramers, *Quantentheorie des Elektrons und der Strahlung* (Akademie-Verlag, Leipzig, 1938).
32 A. Aymart, Compt. Rend. **243**, 885, 1100 (1956).
33 B. C. Unal and J. P. Vigier, Compt. Rend. **245**, 1781, 1891 (1957).
34 T. Takabayasi, Suppl. Prog. Theoret. Phys. (Kyoto), No. 4 (1957).
35 F. Halbwachs, *Théorie relativiste des fluides à spin* (Gauthiers-Villars, Paris, 1960).
36 F. Gürsey, Rev. Fac. Sci. Istanbul **21**, (1956).
37 L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1959), Sec. 125.
38 R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
39 R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 611 (1966).
40 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 1777 (1972).
41 Although the proper time that appears in the argument of (2.4) may in general differ by an additive function $f(\mathbf{X})$ (i.e., the origin of the proper time may be different for each world line), the theory developed can be shown to be independent of proper time. Notice, for example, that for material rates $\mathbf{X}$ is fixed and for strain measures (e.g., C_{KL}) the projection operator eliminates terms arising from the presence of $f(\mathbf{X})$. In regard to this invariance, see also Eqs. (3.5) and (3.11) of Ref. 42.
42 L. Söderholm, Arch. Ratl. Mech. Anal. **39**, 89 (1970).
43 Note that $\epsilon^{\mu\nu\rho}\epsilon_{\mu\nu\sigma\tau} = 3!\delta^{\rho}_{\tau}$, $\epsilon^{\mu\nu\sigma\tau}\epsilon_{\mu\nu\alpha\beta} = 2\delta^{\sigma\tau}_{\alpha\beta} = 2(\delta^{\sigma}_{\alpha}\delta^{\tau}_{\beta} - \delta^{\sigma}_{\beta}\delta^{\tau}_{\alpha})$.
44 There exist different approaches in the spirit of the modern theory of fields to represent angular velocities and their references: One may introduce fields of spinors $\Phi_{(\xi)}(x^{\alpha})$ or six relativistic Euler angles and the corresponding Cayley–Klein parameters. Our representation is equivalent to that given by a two-component spinor (Kramers[31]).
45 Another possible approach consists in generalizing the expression of the kinetic energy of rotation of classical rigid body mechanics $K = \frac{1}{2}I_{ij}\omega^i\omega^j$. Then, in lieu of (3.5), one can write $K = \frac{1}{2}I_{\alpha\beta\gamma\delta}\Omega^{\alpha\beta}\Omega^{\delta\gamma}$, where $I_{\alpha\beta\gamma\delta}$ is a fourth-order tensor. Upon use of certain symmetries, it can be shown that only ten components of $I_{\alpha\beta\gamma\delta}$ are independent.
46 G. A. Maugin, Ph.D. thesis (Princeton University, 1971), Footnote (3), p. 28.
47 A. C. Eringen, in *Proceedings of the Symposium on "A Critical Review of the Foundations of Relativistic and Classical Thermodynamics" (Pittsburgh, 1969)*, edited by E. B. Stuart, B. Gal Or, and A. J. Brainard (Mono, Baltimore, 1970).

Identities Related to Variational Principles

Edward Gerjuoy*
Physics Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

and

A. R. P. Rau† and Larry Spruch†
Physics Department, New York University, New York, New York 10003

(Received 15 May 1972)

The existence of a well-known identity associated with variational principles for any scattering parameter Q, and often serving as the starting point for the development of a variational bound on Q, strongly suggests that it might be useful to construct identities associated with variational principles for quantities other than scattering parameters. An identity associated with the variational principle for the determination of inner products of the linear form $g^{\dagger}\phi$, a generalization of the aforementioned identity, is presented. Here, g is a known function, and ϕ is an unknown function satisfying $M\phi = \omega$ and specified boundary conditions, where M is a known linear operator and ω is a known function. The generalized identity is obtained from a variational principle for $g^{\dagger}\phi$, this variational principle being itself a generalization of the usual Kohn variational principle for scattering amplitudes and phase shifts. An identity associated with a variational principle for the quadratic form $\phi^{\dagger}W\phi$, with ϕ as above and W a known linear operator, is also obtained. Finally, we obtain an identity for $\phi(\infty)$, where ϕ is defined as the solution of a nonlinear differential equation. The generalized identity related to $g^{\dagger}\phi$ is verified for a simple exactly solvable problem.

1. INTRODUCTION

There exists a variational principle for almost all scattering parameters Q, of the form

$$\langle Q\rangle_{\text{var}} = Q_t + (\phi_t, (H - E)\phi_t) = Q_t + \phi_t^{\dagger}[(H - E)\phi_t],$$

where ϕ_t is a trial scattering wavefunction whose asymptotic form determines the trial estimate Q_t of Q. Though the entire discussion is much more widely applicable, we will, for simplicity of discussion, restrict the analysis to the case of potential scattering, and, further, to a partial wave analysis. With η_l the exact phase shift for the lth partial wave, the Kohn variational principle[1] can be written as

* This work was supported by the National Science Foundation.
† Now at Université de Paris, Faculté des Sciences, Mécanique théorique, Tour 66-9, quai Saint-Bernard, 75005 Paris, France.

1 P. Duhem, Ann. Ecole Norm. Paris, **10**, 187 (1893).
2 W. Voigt, Abhandl. Wiss. Ges. Göttingen **34**, (1887).
3 E. Cosserat and F. Cosserat, *Théorie des corps déformables* (Hermann, Paris, 1909).
4 J. L. Eriksen and C. Truesdell, Arch. Ratl. Mech. Anal. **1**, 295 (1958).
5 R. A. Toupin, Arch. Ratl. Mech. Anal. **17**, 85 (1964).
6 A. C. Eringen and E. S. Suhubi, Intern. J. Eng. Sci. **2**, 189 (1964).
7 A. C. Eringen, Intern. J. Eng. Sci. **2**, 205 (1964).
8 G. Maugin, Compt. Rend. **271A**, 807 (1970).
9 A. C. Eringen, "Mechanics of Micromorphic Continua," in *Mechanics of Generalized Continua*, I.U.T.A.M. Symposium 1967, edited by E. Kröner (Springer, Berlin, 1968).
10 A. C. Eringen, Intern. J. Engng. Sci., **8**, 819-828 (1970).
11 A. C. Eringen, "Theory of Micropolar Continua" in Proceedings of the Midwestern Mech. Conference, Madison, Wisconsin, 1965.
12 A. C. Eringen, J. Math. Mech. **16**, 1 (1966).
13 A. C. Eringen, Intern. J. Eng. Sci. **5**, 191 (1967).
14 A. C. Eringen, Intern. J. Eng. Sci. **7**, 115 (1969).
15 A. C. Eringen, "Theory of Micropolar Elasticity", in *Fracture: An Advanced Treatise*, edited by H. Liebowitz (Academic, New York, 1968), Vol. II.
16 C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 271 (1971).
17 C. B. Kafadar and A. C. Eringen, Intern. J. Eng. Sci. **9**, 307 (1971).
18 J. L. Eriksen, Trans. Soc. Rheol. **5**, 23 (1961).
19 A. E. Green and N. Laws, Arch. Ratl. Mech. Anal. **25**, 285 (1961).
20 A. E. Green and P. M. Naghdi, in Ref. 9.
21 J. B. Alblas, in Ref. 9.
22 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 143 (1972).
23 The literature on the subject of oriented media is extensive; for various summaries and review articles, see Eringen[9] and R. Stojanović, *Mechanics of Polar Continua*, Publ. C.I.S.M. No. 2 (C.I.S.M., Udine, Italy, 1969).
24 See for instance the striking analogy between, on one hand, the use of vierbeins in Einstein's (zero curvature – nonzero torsion) theory [S.B. Preuss. Akad. Wiss., 217 (1928)] the use of directors in 3-dimensional theories of dislocations with distant parallelism where the torsion is defined from the directors field (see Claus and Eringen,[25] Fox,[26] and Stojanović[27] and, on the other hand, the multipole-like theories of oriented media (for instance, Green and Rivlin[28]) and the multipole formalism of general relativistic theories (see Tulczyjew and Tulczyjew[29]).
25 W. D. Claus and A. C. Eringen, O.N.R. Technical Report No. 10 (Princeton University, Department of Aerospace and Mechanical Sciences, 1969).
26 N. Fox, Quart. J. Mech. Appl. Math. **19**, 343 (1966).
27 R. Stojanović, in Ref. 9.
28 A. E. Green and R. S. Rivlin, Arch. Ratl. Mech. Anal. **16**, 325 (1964).
29 B. Tulczyjew and W. Tulczyjew, in *Recent Developments in General Relativity* (Pergamon, Oxford and P. W. N., Warsaw, 1962).
30 In German "vierbein, tetrade"; in English "four-legs"; see E. Schmutzer, *Relativitische Physik* (Akademie-Verlag, Leipzig, 1968).
31 H. A. Kramers, *Quantentheorie des Elektrons und der Strahlung* (Akademie-Verlag, Leipzig, 1938).
32 A. Aymart, Compt. Rend. **243**, 885, 1100 (1956).
33 B. C. Unal and J. P. Vigier, Compt. Rend. **245**, 1781, 1891 (1957).
34 T. Takabayasi, Suppl. Prog. Theoret. Phys. (Kyoto), No. 4 (1957).
35 F. Halbwachs, *Théorie relativiste des fluides à spin* (Gauthiers-Villars, Paris, 1960).
36 F. Gürsey, Rev. Fac. Sci. Istanbul **21**, (1956).
37 L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1959), Sec. 125.
38 R. Adler, M. Bazin, and M. Schiffer, *Introduction to General Relativity* (McGraw-Hill, New York, 1965).
39 R. A. Grot and A. C. Eringen, Intern. J. Eng. Sci. **4**, 611 (1966).
40 G. A. Maugin and A. C. Eringen, J. Math. Phys. **13**, 1777 (1972).
41 Although the proper time that appears in the argument of (2.4) may in general differ by an additive function $f(\mathbf{X})$ (i.e., the origin of the proper time may be different for each world line), the theory developed can be shown to be independent of proper time. Notice, for example, that for material rates $\mathbf{X}$ is fixed and for strain measures (e.g., C_{KL}) the projection operator eliminates terms arising from the presence of $f(\mathbf{X})$. In regard to this invariance, see also Eqs. (3.5) and (3.11) of Ref. 42.
42 L. Söderholm, Arch. Ratl. Mech. Anal. **39**, 89 (1970).
43 Note that $\epsilon^{\mu\nu\rho}\epsilon_{\mu\nu\sigma\tau} = 3!\delta^{\rho}_{\tau}$, $\epsilon^{\mu\nu\sigma\tau}\epsilon_{\mu\nu\alpha\beta} = 2\delta^{\sigma\tau}_{\alpha\beta} = 2(\delta^{\sigma}_{\alpha}\delta^{\tau}_{\beta} - \delta^{\sigma}_{\beta}\delta^{\tau}_{\alpha})$.
44 There exist different approaches in the spirit of the modern theory of fields to represent angular velocities and their references: One may introduce fields of spinors $\Phi_{(\xi)}(x^{\alpha})$ or six relativistic Euler angles and the corresponding Cayley–Klein parameters. Our representation is equivalent to that given by a two-component spinor (Kramers[31]).
45 Another possible approach consists in generalizing the expression of the kinetic energy of rotation of classical rigid body mechanics $K = \frac{1}{2}I_{ij}\omega^i\omega^j$. Then, in lieu of (3.5), one can write $K = \frac{1}{2}I_{\alpha\beta\gamma\delta}\Omega^{\alpha\beta}\Omega^{\delta\gamma}$, where $I_{\alpha\beta\gamma\delta}$ is a fourth-order tensor. Upon use of certain symmetries, it can be shown that only ten components of $I_{\alpha\beta\gamma\delta}$ are independent.
46 G. A. Maugin, Ph.D. thesis (Princeton University, 1971), Footnote (3), p. 28.
47 A. C. Eringen, in *Proceedings of the Symposium on "A Critical Review of the Foundations of Relativistic and Classical Thermodynamics" (Pittsburgh, 1969)*, edited by E. B. Stuart, B. Gal Or, and A. J. Brainard (Mono, Baltimore, 1970).

Identities Related to Variational Principles

Edward Gerjuoy*
Physics Department, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

and

A. R. P. Rau† and Larry Spruch†
Physics Department, New York University, New York, New York 10003

(Received 15 May 1972)

The existence of a well-known identity associated with variational principles for any scattering parameter Q, and often serving as the starting point for the development of a variational bound on Q, strongly suggests that it might be useful to construct identities associated with variational principles for quantities other than scattering parameters. An identity associated with the variational principle for the determination of inner products of the linear form $g^{\dagger}\phi$, a generalization of the aforementioned identity, is presented. Here, g is a known function, and ϕ is an unknown function satisfying $M\phi = \omega$ and specified boundary conditions, where M is a known linear operator and ω is a known function. The generalized identity is obtained from a variational principle for $g^{\dagger}\phi$, this variational principle being itself a generalization of the usual Kohn variational principle for scattering amplitudes and phase shifts. An identity associated with a variational principle for the quadratic form $\phi^{\dagger}W\phi$, with ϕ as above and W a known linear operator, is also obtained. Finally, we obtain an identity for $\phi(\infty)$, where ϕ is defined as the solution of a nonlinear differential equation. The generalized identity related to $g^{\dagger}\phi$ is verified for a simple exactly solvable problem.

1. INTRODUCTION

There exists a variational principle for almost all scattering parameters Q, of the form

$$\langle Q\rangle_{\text{var}} = Q_t + (\phi_t, (H - E)\phi_t) = Q_t + \phi_t^{\dagger}[(H - E)\phi_t],$$

where ϕ_t is a trial scattering wavefunction whose asymptotic form determines the trial estimate Q_t of Q. Though the entire discussion is much more widely applicable, we will, for simplicity of discussion, restrict the analysis to the case of potential scattering, and, further, to a partial wave analysis. With η_l the exact phase shift for the lth partial wave, the Kohn variational principle[1] can be written as

$$\langle \tan\eta_l \rangle_{\rm var} = \tan\eta_{lt} - (2m/\hbar^2 k) \int \phi_{lt}(H-E)\phi_{lt}\,dr, \quad (1.1)$$

where the integral runs from 0 to ∞. The exact wavefunction ϕ_l is defined by

$$(H-E)\phi_l(r) = 0, \quad H = H_0 + V,$$
$$H_0 = \frac{\hbar^2}{2m}\left(-\frac{d^2}{dr^2} + \frac{l(l+1)}{r^2}\right), \quad (1.2)$$

$$\phi_l(r) \sim \sin(kr - \tfrac{1}{2}l\pi) + \tan\eta_l \cos(kr - \tfrac{1}{2}l\pi), \ \phi_l(0) = 0. \quad (1.3)$$

The trial wavefunction $\phi_{lt}(r)$ satisfies $\phi_{lt}(0) = 0$, and specifies the trial phase shift η_{lt} through the requirement that

$$\phi_{lt}(r) \sim \sin(kr - \tfrac{1}{2}l\pi) + \tan\eta_{lt} \cos(kr - \tfrac{1}{2}l\pi). \quad (1.4)$$

There exists an identity[2] very similar in form to the variational principle (1.1), namely,

$$\tan\eta_l = \tan\eta_{lt} - (2m/\hbar^2 k) \int \phi_l(H-E)\phi_{lt}\,dr. \quad (1.5)$$

This identity is readily verified via integration by parts. Analogous identities exist[3] for other scattering parameters including those associated with breakup processes. Such identities will be referred to as "variational identities in scattering theory." The identity (1.5) leads immediately to the variational principle of Eq. (1.1); the recognition that $(H-E)\phi_{lt}$ is a first-order term makes it clear that the replacement of ϕ_l by ϕ_{lt} leads to an over-all second-order error. With minor modifications, Eq. (1.5) can also be used to generate the Hulthén[4] and Schwinger versions of the variational principle for $\tan\eta_l$.[5,6] More significantly, the identity is an excellent starting point for the development of variational bounds. Kato[2] showed that Eq. (1.5) can be employed to yield rigorous upper and lower bounds on $\tan\eta_l$, provided it is possible to estimate solutions to an associated eigenvalue equation, and he and others[7] did some work along these lines. Shimamura[8] used a somewhat extended version of this method of Kato's to obtain very good (though not wholly rigorous) upper and lower bounds on the singlet and triplet s-wave phase shifts in electron-hydrogen atom scattering. More recently, Miller[9] made another nonrigorous (but much simpler than Shimamura's) application of the "variational identity" to obtain with rather less effort correspondingly nonrigorous (but nevertheless also often very accurate) upper and lower bounds on $\tan\eta_l$. Furthermore, Bardsley, Gerjuoy, and Sukumar[10] have shown that the nonrigorous approximations in Miller's approach can be avoided in some circumstances, thereby providing truly rigorous upper and lower bounds on $\tan\eta_l$ for nondefinite $H-E$ without having to solve the associated eigenvalue equation of Kato[2] or Shimamura.[8] Spruch and Rosenberg and their collaborators showed some time ago that variational bounds can be obtained on a very wide range of scattering parameters[6,11,12] without introducing the generally cumbersome auxiliary eigenvalue problem. They recast the identity into a form in which the only unknown expression is the diagonal matrix element of a positive definite operator and is therefore of well-defined sign, and applied their formalism to a number of problems including the scattering of electrons and positrons by hydrogen atoms and the scattering of neutrons and protons by deuterons. The method is particularly simple for zero incident relative kinetic energy,[6,11] and, even for the nonzero energy case, is applicable to a wider class of problems than is the associated eigenvalue approach.

In view of the foregoing, it is clear that given any variational principle, it is interesting to attempt to obtain an identity associated as closely as possible with that principle. We will concern ourselves with three such variational principles, each involving an unknown function ϕ, defined by a differential equation. The first variational principle is an estimate of $B \equiv g^\dagger\phi$, involving ϕ linearly, with g a given function. The second is an estimate of $\langle W \rangle \equiv \phi^\dagger W \phi$, involving ϕ quadratically, with W a given Hermitian operator. The third example involves the value of a function ϕ at a point, where ϕ itself is defined by a nonlinear differential equation; a particular example is the phase amplitude method for determining the phase shift for potential scattering.

2. AN IDENTITY ASSOCIATED WITH THE VARIATIONAL PRINCIPLE FOR $B = g^\dagger\phi$

Some time ago Borowitz and Gerjuoy[13] showed that the Kohn variational principle for the scattering amplitude can be regarded as a special case of a more general variational principle for quantities of the form

$$B = g^\dagger\phi, \quad (2.1)$$

where the notation denotes the inner product of a known function g with an unknown function ϕ satisfying

$$M\phi = \omega \quad (2.2)$$

for some known (ϕ-independent) linear operator M. In the above, the dagger denotes the adjoint (complex conjugate transpose), and both g and ϕ may be column matrices, as they would be in the event ϕ were a wavefunction for a particle of nonvanishing spin. If ϕ has n discrete components, and depends on continuous variables collectively denoted by $\mathbf{r}$, then, of course,

$$B = \sum_{i=1}^{n} \int d\mathbf{r}\, g_i^*(\mathbf{r})\phi_i(\mathbf{r}), \quad (2.3)$$

where the asterisk denotes the complex conjugate. The Kohn variational principle (1.1) for $\tan\eta_l$, the Schwinger principles for $\tan\eta_l$ and the total scattering amplitude, and other less well-known variational principles for scattering amplitudes also are special cases of the general variational principle for quantities B of the form (2.1) as is demonstrated in a companion paper[14] to the present paper on generalized identities.

For these reasons, we have been impelled to seek—and then to find—a generalization of the identity (1.5) applicable to quantities of the form (2.1). Proving this generalized identity, and corresponding identities for somewhat different forms, is the main objective of this paper.

Whether this generalized identity can be employed to obtain useful bounds for arbitrary quantities B of the form (2.1), as it has proved possible to obtain bounds for the special $B = \tan\eta_l$ cases discussed earlier, and more generally for a much wider class of scattering

parameters, remains to be seen; no attempt to actually compute such bounds is made in this paper. Because the proof is both very formal and surprisingly simple, the final section of this paper illustrates and verifies the generalized identity for a trivially exactly solvable problem involving a somewhat unusual g, namely, the problem of the electric field in a spherical condenser.

To prove the generalized identity for quantities B of the form (2.1), it is convenient to start from the generalized variational principle for such quantities,[13,14] namely

$$\langle B\rangle_{\rm var} = g^\dagger\phi_t + f_t^\dagger([M\phi_t] - \omega), \tag{2.4}$$

where ϕ_t is a trial estimate of the exact ϕ appearing in (2.1) and where f_t is a trial estimate of the exact so-called auxiliary function f. The exact ϕ is specified by (2.2) together with appropriate boundary conditions; the trial ϕ_t normally will be restricted to the class of functions satisfying the same boundary conditions as ϕ. The procedure for specifying the exact auxiliary function f is as described in our companion paper,[14] namely, one specifies f via the requirement that the first variation of (2.4) must be zero if (2.4) really is to be a variational principle. In other words, we must have

$$\delta B = g^\dagger\delta\phi + f_t^\dagger([M\phi] - \omega) + f^\dagger[M\delta\phi] = 0, \tag{2.5}$$

where

$$\delta\phi = \phi_t - \phi, \quad \delta f = f_t - f. \tag{2.6}$$

In (2.5), it has been assumed that M is completely known, so that there is no term involving δM; this assumption is necessary for the derivation of the generalized identity given in this paper, but may be avoidable if merely a variational principle for $g^\dagger\phi$ is sought.[14] In effect, this last assumption rules out the circumstance that Eq. (2.2) determining ϕ is an eigenvalue equation, wherein the eigenvalue appearing in M cannot be exactly known unless ϕ itself is exactly known.

Via (2.2), the equality (2.5) can be rewritten in the form

$$g^\dagger\delta\phi + (M^\dagger f)^\dagger\delta\phi + f^\dagger[M\delta\phi] - (M^\dagger f)^\dagger\delta\phi = 0. \tag{2.7}$$

Therefore, Eqs. (2.7) and (2.5) will hold, i.e., (2.4) will be a variational principle, if[13,14]

$$g^\dagger + (M^\dagger f)^\dagger = 0, \tag{2.8}$$

subject to the condition

$$f^\dagger[M\delta\phi] - [M^\dagger f]^\dagger\delta\phi = 0. \tag{2.9}$$

Usually it is more convenient to replace (2.8) by its adjoint,

$$M^\dagger f + g = 0, \tag{2.10}$$

which presumably specifies f subject to the boundary condition implied by (2.9). In some circumstances, additional boundary conditions may be required[14] to uniquely specify the auxiliary function f. If so, it is assumed these boundary conditions have been imposed; however, these extra boundary conditions, though useful in practice, apparently are not needed for the proof of the generalized identity, as will be seen. It will be noted that $M^\dagger$ need not be identical with M, i.e., there has been no assumption that M is self-adjoint. It is assumed, of course, that the operator $M^\dagger$ satisfies (2.9) for a reasonably well-behaved class of functions f solving (2.10); otherwise, (2.4) is not a variational principle, and the corresponding identity cannot be demonstrated. In the application of the variational principle (2.4), one normally would try to choose a trial f_t satisfying the same boundary conditions as f; the choice of f_t is wholly irrelevant for the present paper, however.

It will be proved below that for ϕ_t satisfying the boundary and smoothness conditions required to make it an acceptable trial estimate of the exact ϕ in the variational principle (2.4), we have

$$B = g^\dagger\phi_t + f^\dagger([M\phi_t] - \omega), \tag{2.11}$$

where B is the exact $g^\dagger\phi$ of Eq. (2.1). Equation (2.11) is the generalized identity; the reason for so terming it is that, in potential scattering, Eq. (2.11) becomes identified with Eq. (1.5) when B is identified with $\tan\eta_l$. To be specific, it is known[5,6] that

$$\tan\eta_l = -(2m/\hbar^2)\int dr\, rj_l(kr)V(r)\phi_l(r), \tag{2.12}$$

where ϕ_l is the function defined by Eqs. (1.2) and (1.3) and j_l is the usual spherical Bessel function. Equation (2.12) expresses $\tan\eta_l$ in the standard form (2.1), with

$$g \equiv g(r) = -(2m/\hbar^2)rj_l(kr)V(r). \tag{2.13}$$

With $\omega = 0$ and $M = (2m/\hbar^2)(H - E)$, the variational principle (2.4) then becomes

$$\langle\tan\eta_l\rangle_{\rm var} = (2m/\hbar^2)\left(-\int dr\, rj_l(kr)V(r)\phi_{l\,t}(r) + \int dr f_t(r)[(H - E)\phi_{l\,t}(r)]\right), \tag{2.14}$$

with the requirements, from Eqs. (2.10) and (2.9), respectively,

$$(H - E)f = rV(r)j_l(kr), \tag{2.15}$$

$$(2m/\hbar^2)\int dr\{f(r)[H - E)\delta\phi] - [(H - E)f]\delta\phi\} = -\left(f\frac{d}{dr}\delta\phi - \delta\phi\frac{df}{dr}\right)\Big|_0^\infty = 0. \tag{2.16}$$

Because the exact ϕ_l vanishes at the origin, and because, as explained following Eq. (2.4), $\phi_{l\,t}$ normally will be required to satisfy the same boundary conditions as ϕ_l, the variation $\delta\phi$ defined by (2.6) vanishes at $r = 0$, so that (2.16) requires $f = 0$ at $r = 0$. Similarly, because Eqs. (1.3), (1.4), and (2.6) imply that $\delta\phi$ is proportional to $\cos(kr - \frac{1}{2}l\pi)$ as $r \to \infty$, Eq. (2.16) implies that the nonvanishing components of $f(r)$ as $r \to \infty$ also must be proportional to $\cos(kr - \frac{1}{2}l\pi)$ as $r \to \infty$. But if the function $\phi_l(r)$ uniquely specified by Eqs. (1.2) and (1.3) is written in the form

$$\phi_l(r) = krj_l(kr) + \chi_l(r), \tag{2.17}$$

the function $\chi_l(r)$ vanishes at $r = 0$, is proportional to $\cos(kr - \frac{1}{2}l\pi)$ at $r = \infty$, and, by using $(H_0 - E)rj_l(kr) = 0$, satisfies the equation

$$(H - E)\chi_l(r) = -V(r)krj_l(kr). \tag{2.18}$$

Therefore, the function f specified by Eq. (2.15) and the boundary conditions stated above must be

$$f = -(1/k)\chi_l = -(1/k)[\phi_l - krj_l(kr)]. \tag{2.19}$$

If, in accordance with (2.19), the trial estimate f_t of the exact f is chosen to be

$$f_t = -(1/k)[\phi_{lt} - krj_l(kr)], \tag{2.20}$$

with ϕ_{lt} satisfying (1.4), substitution in (2.14) yields

$$\langle \tan\eta_l \rangle_{\text{var}} = (2m/\hbar^2)\Big(\int dr\, rj_l(kr)[(H_0 - E)\phi_{lt}] - (1/k)\int dr\, \phi_{lt}[(H - E)\phi_{lt}]\Big). \tag{2.21}$$

By recalling (1.4), integration by parts then shows, by using $(H_0 - E)rj_l(kr) = 0$ once again, that

$$(2m/\hbar^2)\int_0^\infty dr\, rj_l(kr)[(H_0 - E)\phi_{lt}] = \tan\eta_{lt}, \tag{2.22}$$

i.e., shows that (2.21) is identical with the usual Kohn variational principle (1.1) for $\tan\eta_l$. Correspondingly, using the present values of f and g, Eqs. (2.19) and (2.13), respectively, in (2.11) yields, by recalling (2.12),

$$\tan\eta_l = (2m/\hbar^2)\Big(\int dr\, rj_l(kr)[(H_0 - E)\phi_{lt}] - (1/k)\int dr\, \phi_l[(H - E)\phi_{lt}]\Big); \tag{2.23}$$

by using (2.22), this is seen to be identical with the identity (1.5). Admittedly, the foregoing has not been the simplest means of deriving either the variational principle (1.1) or the identity (1.5), but the above derivations of (1.1) and (1.5) do serve to justify the terminology "generalized identity" for (2.11). Furthermore, we have deliberately chosen to go through the above derivation so that we can explicitly analyze the structure of the auxiliary function f. This is really not necessary for the establishment of the identity (2.23), and it is easy to verify that multiplying the defining equation for f, (2.15), by $\delta\phi$ and using (2.12) and (2.16) would lead immediately to (2.23). This feature that the identity can be established once the equation for f is known, without the need for solving this equation for f, will be demonstrated below in (2.31)–(2.33) when we establish the general identity (2.11) by such a procedure, and again in later sections for the case of analogous identities.

Before proving the general identity we will consider yet another derivation of the Kohn variational principle and its associated identity starting from an alternative to (2.12). This is done both because it is a shorter derivation and because the auxiliary function will take on a different structure from the one in (2.19), thereby demonstrating in a particular case the general result[14] that various alternative forms of the variational principle can be written down and these may involve different auxiliary functions. We begin then from the following definition of $\tan\eta_l$

$$\tan\eta_l = (2m/\hbar^2 k)[(H_0 - E)\phi_l]^\dagger\psi, \tag{2.24}$$

where ψ is the regular solution of $(H_0 - E)\psi = 0$, appropriately normalized, and is $krj_l(kr)$. That (2.24) defines the phase shift is seen perhaps most simply by looking at the combination $[(H_0 - E)\phi_l]^\dagger\psi - \phi_l^\dagger[(H_0 - E)\psi]$ and rearranging this in the form of a surface term. Evaluating this term on a surface at large r yields $\tan\eta_l$.

Exactly similarly, we have

$$\tan\eta_{lt} = (2m/\hbar^2 k)[(H_0 - E)\phi_{lt}]^\dagger\psi; \tag{2.25}$$

this is nothing but (2.22). The variational principle is written down in a routine fashion[14]

$$\langle \tan\eta_l \rangle_{\text{var}} = (2m/\hbar^2 k)\{[(H_0 - E)\phi_{lt}]^\dagger\psi - f_t^\dagger[(H - E)\phi_{lt}]\}. \tag{2.26}$$

The equation for the auxiliary function f is obtained by equating to zero the coefficient of $\delta\phi$ in (2.26); this gives

$$[(H_0 - E)\delta\phi]^\dagger\psi - f^\dagger[(H - E)\delta\phi] = 0.$$

Integration by parts transforms this to

$$-[(H - E)f]^\dagger\delta\phi + \frac{\hbar^2}{2m}\left[(f - \psi)\frac{d\delta\phi}{dr} + \left(\frac{d\psi}{dr} - \frac{df}{dr}\right)\delta\phi\right]\Bigg|_0^\infty = 0. \tag{2.27}$$

The expression in (2.27) will vanish if the surface terms vanish and if

$$(H - E)f = 0. \tag{2.28}$$

By recalling now from (1.3) and (1.4) that $\delta\phi$ and ψ vanish at $r = 0$ and that asymptotically $\delta\phi$ is proportional to $\cos(kr - \frac{1}{2}l\pi)$ and therefore $d\delta\phi/dr$ to $\sin(kr - \frac{1}{2}l\pi)$, it follows that the surface terms in (2.27) will vanish if f vanishes at $r = 0$ and if $f - \psi$ is asymptotically of the form $\cos(kr - \frac{1}{2}l\pi)$. From (2.28) and these boundary conditions, it is clear that

$$f = \phi_l. \tag{2.29}$$

Replacing $f_t^\dagger$ in (2.26) by $f^\dagger$, so that we have an identity rather than a variational principle, and using (2.25) and (2.29), we obtain

$$\tan\eta_l = \tan\eta_{lt} - (2m/\hbar^2 k)\phi_l^\dagger[(H - E)\phi_{lt}]. \tag{2.30}$$

This is, of course, identical to (2.23) and represents an alternative, somewhat shorter derivation of the identity. What is particularly worth noting is the difference between f in (2.29) and in (2.19).

Finally, we prove the identity (2.11) in its general form, without specializing as above to the case of the tangent of the phase shift. The proof is surprisingly simple. We start with the equations defining the auxiliary function, (2.8) and (2.9). Operating with (2.8) on $\delta\phi$, we obtain

$$g^\dagger\delta\phi + [M^\dagger f]^\dagger\delta\phi = 0. \tag{2.31}$$

By using the boundary condition (2.9) and (2.6), this becomes

$$g^\dagger(\phi_t - \phi) + f^\dagger[M(\phi_t - \phi)] = 0. \tag{2.32}$$

Since $M\phi = \omega$, this becomes

$$g^\dagger\phi = g^\dagger\phi_t + f^\dagger([M\phi_t] - \omega), \tag{2.33}$$

which is the sought for identity.

This procedure for finding the identity from the defining equations for the auxiliary function seems to be a generally valid one. It will be used again in Secs. 3 and 4. We also note that the variational principle (2.4) is retrieved from the identity (2.33) by the replacement of $f^\dagger$ by $f_t^\dagger$; such a replacement introduces only a second-order error because the term multiplying $f^\dagger$ in (2.33) is already of first order.

Finally we conclude this section with a special case of the identity (2.33) and its associated variational principle (2.4), namely the case when the value of the continuum wavefunction ϕ_l at some point r_0 (say) is desired. The function g is, therefore, $\delta(r_0 - r)$. The identity takes the form

$$\phi_l(r_0) = \phi_{lt}(r_0) + (2m/\hbar^2) f^\dagger[(H - E)\phi_{lt}(r)]. \quad (2.34)$$

The equations defining f, analogous to (2.9) and (2.10), follow from equating to zero the coefficient of $\delta\phi$ in (2.34). We have

$$(2m/\hbar^2)(H - E)f(r_0, r) = -\delta(r_0 - r), \quad (2.35)$$

and

$$\left(-f\frac{d}{dr}\delta\phi + \delta\phi\frac{df}{dr}\right)\Big|_0^\infty = 0. \quad (2.36)$$

The discussion after (2.28) applies to the analysis of (2.36) and leads to the boundary conditions on f:

$$f(0) = 0,\ f \sim \text{const} \times \cos(kr - \tfrac{1}{2}l\pi). \quad (2.37)$$

(2.35) and (2.37) together with the usual analysis of (2.35) as r passes the fixed point r_0 define completely the Green's function f which goes into the identity (2.34). Such identities and variational principles for the value of the wavefunction itself at a point may be useful because of their wide applicability. Once one has a wavefunction which is itself variational at every point, any matrix element evaluated with it will automatically be a variational estimate. As an identity valid for any ϕ_{lt}, (2.34) with the choice $\phi_{lt}(r) = krj_l(kr)$ becomes the usual integral equation for ϕ_l.

3. AN IDENTITY ASSOCIATED WITH THE VARIATIONAL PRINCIPLE FOR $\langle W\rangle = \phi^\dagger W\phi$

A variational principle for $\langle W\rangle$, for W an arbitrary linear self-adjoint operator, has been known for some time, work having been done by Dalgarno,[15] Schwartz,[16] Delves,[17] and many others. With E assumed known experimentally essentially "exactly", with ϕ defined by

$$H\phi = E\phi, \quad \phi^\dagger\phi = 1, \quad (3.1)$$

and, with an auxiliary function f defined by

$$(H - E)f = -[W - \langle W\rangle]\phi, \quad (3.2)$$

$$f^\dagger\phi = 0, \quad (3.3)$$

we have

$$\langle W\rangle_{\text{var}} = \phi_t^\dagger W\phi_t + f_t^\dagger(H - E)\phi_t + [(H - E)\phi_t]^\dagger f_t. \quad (3.4)$$

Here ϕ_t and f_t are approximations to ϕ and f, respectively, and $\phi_t^\dagger\phi_t = 1$. An identity is rather easily obtained.[18] It can be cast into the form

$$\langle W\rangle = \langle W\rangle_{\text{var}} - S,$$

where the second-order error term S is given explicitly, but it does not appear to be possible under most circumstances to determine the sign of S. Were it possible to determine the sign of S, one would have a *variational* bound on $\langle W\rangle$.

To obtain an identity of the form we are presently interested in, we take the inner product of $\phi_t^\dagger$ with (3.2) and of the adjoint version of (3.2) with ϕ_t, add, and use the hermiticity of $H - E$ to obtain

$$\tfrac{1}{2}(\phi_t^\dagger\phi + \phi^\dagger\phi_t)\langle W\rangle = [(H - E)\phi_t]^\dagger f + f^\dagger(H - E)\phi_t + \phi_t^\dagger W\phi + \phi^\dagger W\phi_t - \tfrac{1}{2}(\phi_t^\dagger\phi + \phi^\dagger\phi_t)\langle W\rangle. \quad (3.5)$$

We have decomposed the term proportional to $\langle W\rangle$ in order to be able to more readily recapture the variational principle from this identity. We begin by noting that the coefficient of $\langle W\rangle$, on each side of the equation, differs from unity by a term of second order. We note further that the replacement of f by f_t introduces a second-order error, as does the replacement of $f^\dagger$ by $f_t^\dagger$. Finally, we note, on replacing ϕ_t by $\phi + \delta\phi$, that $\phi_t^\dagger W\phi + \phi^\dagger W\phi_t - \langle W\rangle$ differs from $\phi_t^\dagger W\phi_t$ by a term of second order.

An alternate version of the identity has been obtained[19] which employs a Green's function rather than the auxiliary function f. Starting from this form, upper and lower bounds on $\langle W\rangle$ have been obtained which, however, are unfortunately not variational; the error is of first order. Applications have been made to the evaluation of $\langle r\rangle$ and $\langle r^2\rangle$ for the ground state of helium.

4. AN IDENTITY INVOLVING A FUNCTION DEFINED BY A NONLINEAR DIFFERENTIAL EQUATION

We consider now an identity and its associated variational principle involving a function $\phi(r)$ defined by a nonlinear differential equation of the form

$$\frac{d\phi}{dr} = [j(r) + n(r)\phi(r)]^2, \quad (4.1)$$

and by boundary conditions, where j and n are known functions of r. Such a differential equation arises, for instance, in the phase-amplitude method[20] for the determination of the phase shift for potential scattering. In this case ϕ is $\tan\delta(r)$, j and n are functions that involve the potential, the energy and standard functions like $j_l(kr)$ and $n_l(kr)$. With the boundary condition at the origin, $\phi(0) = 0$, the equation is integrated outward to infinity and $\phi(\infty)$ gives the value of $\tan\delta(\infty)$, that is, the "true" phase shift.

A variational principle for $\phi(\infty)$, with $\phi(r)$ defined by Eq. (4.1) (or by similar, more general, nonlinear equations) can be written down in routine fashion.[14] It is found to be

$$\phi_{\text{var}}(\infty) = \phi_t(\infty) - \int_0^\infty f_t\left(\frac{d\phi_t}{dr} - (j + n\phi_t)^2\right)dr, \quad (4.2)$$

where $\phi_t(r)$ is a trial function chosen to vanish at the origin and $f_t(r)$ is a trial estimate of the auxiliary function f which satisfies the equation

$$\frac{df}{dr} = -2n(j + n\phi)f \quad (4.3)$$

and the boundary condition $f(\infty) = 1$. We will demand

that $f_t(\infty) = 1$. It is not difficult to verify that such an f makes the estimate $\phi_{\text{var}}(\infty)$ stationary. As in Secs. 2 and 3, we derive an identity for $\phi(\infty)$, starting from the equation which defines the auxiliary function, which is now Eq. (4.3). Multiplying this equation by $\phi_t(r)$ and integrating, we have, after one integration by parts,

$$\phi_t(\infty) = \int_0^\infty f\left(\frac{d\phi_t}{dr} - 2n\phi_t(j + n\phi)\right)dr. \tag{4.4}$$

This equation is obviously also valid for the particular choice $\phi_t(r) = \phi(r)$. Subtracting Eq. (4.4) as it stands from (4.4) with ϕ_t replaced by ϕ and using (4.1), we have the identity

$$\phi(\infty) = \phi_t(\infty) - \int_0^\infty f\left\{\left(\frac{d\phi_t}{dr} - (j + n\phi)^2\right) - [2(\phi_t - \phi)n(j + n\phi)]\right\}dr. \tag{4.5}$$

The replacement of f by f_t obviously introduces only second-order error terms since each of the terms in curly brackets is of first order. Further, it is trivial to check that the replacement of ϕ by ϕ_t in the square brackets introduces a second-order error, proportional to $(\phi_t - \phi)^2$. We can thus readily retrieve the variational principle (4.2) from the identity (4.5).

5. THE GENERALIZED IDENTITY AND THE ELECTRIC FIELD IN A SPHERICAL CONDENSER

The validity of the identity (1.5) normally is verified via integration by parts, and this verification provides one illustration of the correctness of our simple formal proof [Eqs. (2.31)–(2.33)] of the generalized identity (2.11). However, because the functions and operations involved in the usual identity (1.5) are so well behaved, we have thought it advisable to verify (2.11) for a problem which—though trivially exactly solvable—involves rather less well-behaved quantities. In particular, we consider the problem of the determination of the electric field inside and at the spherical plates of a spherical condenser with spheres of radii 1 and 2. In this problem the potential ϕ obeys

$$M\phi = r^2\nabla^2\phi = \frac{d}{dr}\left(r^2\frac{d\phi}{dr}\right) = r^2\frac{d^2\phi}{dr^2} + 2r\frac{d\phi}{dr} = 0, \tag{5.1}$$

and the boundary conditions on ϕ (the potentials at the spherical plates) are taken to be

$$\phi(1) = 0, \quad \phi(2) = 1. \tag{5.2}$$

Our objective is to find $d\phi/dr$ in the domain $1 \le r \le 2$. Of course, it is a trivial exercise to determine that the desired exact ϕ satisfying (5.1), (5.2) is

$$\phi = 2[1 - (1/r)], \quad 1 \le r \le 2, \tag{5.3}$$

so that the desired $d\phi/dr$ is

$$\frac{d\phi}{dr} = \frac{2}{r^2}, \quad 1 \le r \le 2. \tag{5.4}$$

For our present purposes, however, we shall pretend that we have been unable to determine the exact ϕ and therefore are endeavoring to estimate $d\phi/dr$ from a variational principle of the form (2.4). The possible complication—and the reason we are examining this particular problem—is that $B = d\phi/dr$ can be put into the form $g^\dagger\phi$ only by the device of introducing a singular $g(r)$; the requirement that $d\phi/dr$ at $r = r_0$ be equal to $g^\dagger\phi$ requires that

$$g(r) = \delta(r - r_0)\frac{d}{dr}, \quad 1 \le r \le 2.$$

For r_0 different from 1 or 2, integration by parts leads to

$$g(r) = -\frac{d}{dr}\delta(r - r_0), \quad 1 \le r \le 2.$$

We do not actually need to know $g(r)$, since the variational principle involves g only in the form $g^\dagger\phi_t$, and the statement $g^\dagger\phi = d\phi/dr$ suggests the choice $g^\dagger\phi_t = d\phi_t/dr$. We have discussed the form of $g(r)$ to show that the values $r_0 = 1$ and $r_0 = 2$ might have to be treated differently than the values $1 < r_0 < 2$, and to exhibit the singular nature of $g(r)$, which suggests that there might conceivably be difficulty in applying the variational principle (2.4) to the present problem.

Nevertheless, let us attempt to construct the variational principle for $d\phi/dr$. Consider first the field at a spherical plate, at $r_0 = 1$, say. Then, according to (2.4)

$$\langle B\rangle_{\text{var}} \equiv \left\langle\left(\frac{d\phi}{dr}\right)_{r=1}\right\rangle_{\text{var}} = \left(\frac{d\phi_t}{dr}\right)_{r=1} + \int_1^2 dr\, f_t r^2\nabla^2\phi_t. \tag{5.5}$$

The trial function $\phi_t(r)$ is supposed to be well behaved and to obey the boundary conditions (5.2); f_t is a trial estimate of the exact auxiliary function f, whose defining conditions now must be determined from the requirement that the first variation of (5.5) be zero, i.e., from the requirement [as in (2.5)]

$$\delta B = \delta\left(\frac{d\phi}{dr}\right)_{r=1} + \int_1^2 dr\, fr^2\nabla^2\delta\phi + \int_1^2 dr(\delta f)r^2\nabla^2\phi = \left(\frac{d}{dr}\delta\phi\right)_{r=1} + \int_1^2 dr\, f\left(r^2\frac{d^2\delta\phi}{dr^2} + 2r\frac{d}{dr}\delta\phi\right) = 0, \tag{5.6}$$

where we have used (5.1) and have interchanged the operators δ and d/dr at $r = 1$. Integration by parts reduces (5.6) to

$$\left(\frac{d}{dr}\delta\phi\right)_{r=1} + r^2 f\frac{d}{dr}\delta\phi\Big|_1^2 - r^2\delta\phi\frac{df}{dr}\Big|_1^2 + \int_1^2 dr\,\delta\phi\frac{d}{dr}\left(r^2\frac{df}{dr}\right) = 0. \tag{5.7}$$

Because $\delta\phi = 0$ at $r = 1$ and $r = 2$, but is otherwise arbitrary (except for being well behaved), Eq. (5.7) implies

$$\frac{d}{dr}\left(r^2\frac{df}{dr}\right) = 0, \quad 1 \le r \le 2, \tag{5.8}$$

with the boundary conditions

$$f(1) = 1, \quad f(2) = 0. \tag{5.9}$$

In the problem under present consideration, the generalized identity (2.11) is [recalling (5.4)]

$$B \equiv \left(\frac{d\phi}{dr}\right)_{r=1} = 2 = \left(\frac{d\phi_t}{dr}\right)_{r=1} + \int_1^2 dr\, f\frac{d}{dr}\left(r^2\frac{d\phi_t}{dr}\right). \tag{5.10}$$

It is readily verified via integration by parts that

(5.10) indeed is true for any well-behaved ϕ_t satisfying (5.2) when f is the solution to (5.8) subject to (5.9), namely

$$f = (2/r) - 1, \quad 1 \le r \le 2. \tag{5.11}$$

The problem is somewhat more complicated when the field is desired at an interior point of the condenser. Variational principle (2.4) now is [to be compared with (5.5)]

$$\langle B\rangle_{\rm var} = \left\langle \left(\frac{d\phi}{dr}\right)_{r=r_0} \right\rangle_{\rm var} = \left(\frac{d\phi_t}{dr}\right)_{r=r_0} + \int_1^2 dr\, f_t r^2 \nabla^2 \phi_t, \quad 1 < r_0 < 2, \tag{5.12}$$

where, of course, f_t and its corresponding f in (5.12) need not be—and in general are not—identical with the f_t and corresponding f of (5.5) and (5.11). Varying (5.12) yields, as in (5.6),

$$\delta B = \frac{d}{dr}\delta\phi\Big)_{r=r_0} + \int_1^2 dr\, f\left(r^2 \frac{d^2}{dr}\delta\phi + 2r\frac{d}{dr}\delta\phi\right) = 0, \tag{5.13}$$

which is not identical with (5.6). In fact, it now is clear that integration by parts, as in (5.6), cannot possibly yield terms which will cancel $d(\delta\phi)/dr$ at $r = r_0$ unless $f(r)$ and/or df/dr are permitted to be discontinuous at $r = r_0$. In other words, before integration by parts (5.13) must be replaced by

$$\delta B = \frac{d}{dr}\delta\phi\Big)_{r=r_0} + \lim_{\epsilon\to 0}\left(\int_1^{r_0-\epsilon} dr + \int_{r_0+\epsilon}^2 dr\right) \times\left[f\left(r^2\frac{d^2}{dr^2}\delta\phi + 2r\frac{d}{dr}\delta\phi\right)\right] = 0. \tag{5.14}$$

Integrating (5.14) by parts yields, with $r_0 \pm \equiv \lim_{\epsilon\to 0}(r_0 \pm \epsilon)$,

$$\begin{aligned}&\frac{d}{dr}\delta\phi\Big)_{r=r_0} + r^2 f\frac{d}{dr}\delta\phi\Big|_1^{r_0^-} + r^2 f\frac{d}{dr}\delta\phi\Big|_{r_0^+}^2 \\ &- \left(r^2\delta\phi\frac{df}{dr}\Big|_1^{r_0^-} + r^2\delta\phi\frac{df}{dr}\Big|_{r_0^+}^2\right) \\ &+ \left(\int_1^{r_0^-} dr + \int_{r_0^+}^2 dr\right)\left[\delta\phi\left(\frac{d}{dr}\right)\left(r^2\frac{df}{dr}\right)\right] = 0,\end{aligned} \tag{5.15}$$

which is to be compared with (5.7). The presence of the integrals in (5.15) again requires that f satisfy the differential equation (5.8), but the boundary conditions on f now are quite different from (5.9). To eliminate the contributions to (5.15) at $r = 1$ and $r = 2$, we must have

$$f(1) = 0, \quad f(2) = 0. \tag{5.16}$$

Also, because $\delta\phi$ is presumed well behaved at $r = r_0$, and both $\delta\phi$ and $d(\delta\phi)/dr$ are essentially arbitrary at $r = r_0$, to eliminate from (5.15) the contributions proportional to $\delta\phi$ at $r = r_0$, it is necessary that

$$\left(\frac{df}{dr}\right)_{r_0^-} = \left(\frac{df}{dr}\right)_{r_0^+}. \tag{5.17}$$

Moreover, to eliminate from (5.15) the contributions proportional to $d(\delta\phi)/dr$ at $r = r_0$, it is further necessary that

$$-r_0^2[f(r_0+) - f(r_0-)] + 1 = 0. \tag{5.18}$$

Therefore, the exact $f \equiv f(r, r_0)$ must satisfy

$$\frac{d}{dr}\left(r^2\frac{df}{dr}\right) = 0, \quad \begin{cases} 1 \le r < r_0 \\ r_0 < r \le 2 \end{cases} \tag{5.19}$$

subject to the boundary conditions (5.16), (5.17), and (5.18). Evidently the auxiliary function f now under consideration is a Green's function of somewhat unusual type.

The solution to Eqs. (5.16)–(5.19) is

$$\begin{aligned} f(r, r_0) &= r_0^{-2}[(2/r) - 2], \quad 1 \le r < r_0, \\ f(r, r_0) &= r_0^{-2}[(2/r) - 1], \quad r_0 < r \le 2, \end{aligned} \tag{5.20}$$

as is readily verified. Again recalling (5.4), the generalized identity [to be compared with (5.10) and (5.12)] now is

$$\begin{aligned} B = \frac{d\phi}{dr}\Big)_{r=r_0} &= \frac{2}{r_0^2} = \frac{d\phi_t}{dr}\Big)_{r=r_0} + \frac{1}{r_0^2}\int_1^{r_0} dr\left(\frac{2}{r} - 2\right) \\ &\times \left(r^2\frac{d^2\phi_t}{dr^2} + 2r\frac{d\phi_t}{dr}\right) \\ &+ \frac{1}{r_0^2}\int_{r_0}^2 dr\left(\frac{2}{r} - 1\right)\left(r^2\frac{d^2\phi_t}{dr^2} + 2r\frac{d\phi_t}{dr}\right). \end{aligned} \tag{5.21}$$

The usual integration by parts demonstrates that (5.21) indeed is true for any well-behaved ϕ_t satisfying (5.2).

ACKNOWLEDGMENTS

We are indebted to Dr. J. Norman Bardsley and Dr. Leonard Rosenberg for a number of very useful conversations.

---

* This work was supported in part by the Advanced Research Project Agency under Contract No. DA-31-124-ARO-D-440 and by the National Aeronautics and Space Administration under Contract No. NGL-39-011-035.

† Work supported by the U.S. Army Research Office, Durham, Grant No. DA-ARO-D-31-124-72-G92 and by the Office of Naval Research, Grant No. N00014-67-A-0467-0007.

1 W. Kohn, Phys. Rev. **74**, 1763 (1948).

2 T. Kato, Progr. Theoret. Phys. (Kyoto) **6**, 295, 394 (1951); Phys. Rev. **80**, 475 (1950).

3 The variational identity for $\tan\eta_l$ given in Eq. (1.5) was obtained in Ref. 2 by evaluating $\phi_l^\dagger(H-E)\phi_{l\,t} - \phi_{l\,t}^\dagger(H-E)\phi_l$ in two ways, using $(H-E)\phi_l = 0$ on the one hand and Green's theorem on the other. By appropriate modifications, this same approach can be used to obtain variational identities for any scattering parameters involving multichannel scattering for energies and initial conditions such that there are two systems initially and two systems finally [L. Rosenberg and L. Spruch, Phys. Rev. **125**, 1407 (1962)], and for break-up processes [M. Lieber, L. Rosenberg and L. Spruch, Phys. Rev. D **5**, 1330, 1347 (1972)].

4 L. Hulthén, Kgl. Fysiograf. Sallskap. Lund Forh. **14**, No. 21 (1944); Arkiv Mat. Astron. Fys. **35A**, No. 25 (1948).

5 For material on variational principles in scattering theory, see, for example, N. F. Mott and H. S. W. Massey, *The Theory of Atomic Collisions* (Oxford U. P., New York, 1965); B. L. Moiseiwitsch, *Variational Principles* (Wiley, New York, 1966), Chap. 5; and Ref. 6.

6 L. Spruch, in *Lectures in Theoretical Physics, Boulder, 1961*, edited by W. E. Brittin, W. B. Downs, and J. Downs (Wiley, New York, 1962), Vol. 4, p. 161.

7 R. Bartram and L. Spruch, J. Math. Phys. **3**, 287 (1962), and refer-

ences therein.
[8] I. Shimamura, J. Phys. Soc. Japan **25**, 971 (1968); **28**, 1057 (1970).
[9] K. J. Miller, Phys. Rev. A **3**, 607 (1971).
[10] J. N. Bardsley, E. Gerjuoy, and C. V. Sukumar, to be published.
[11] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959).
[12] Y. Hahn, T. F. O'Malley, and L. Spruch, Phys. Rev. **134**, B911 (1964).
[13] S. Borowitz and E. Gerjuoy, University of Colorado, Boulder, Colorado, JILA Report 36 (May 14, 1965).
[14] E. Gerjuoy, A. R. P. Rau, and L. Spruch, to be published.
[15] A. Dalgarno and A. L. Stewart, Proc. Roy. Soc. **A257**, 534 (1960).
[16] Charles Schwartz, Ann. Phys. (N.Y.) **2**, 156, 170 (1959).
[17] L. M. Delves, Nucl. Phys. **41**, 497 (1963).
[18] L. Spruch, in *Proceedings of the Ninth Summer Meeting of Yugoslav Physicists, Hercegnovi, July, 1964*, edited by M. Cerineo (Federal Nuclear Energy Commision of Yugoslavia, Belgrade, 1965), Vol. 1, p. 191.
[19] R. Blau, A. R. P. Rau, and L. Spruch, submitted to Phys. Rev. A.
[20] See, for instance, F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).

On the Correlation of Field Quantities in a Random Linearly Elastic Solid

John J. McCoy
The Catholic University of America, Washington, D.C. 20017
(Received 19 June 1972)

The response of a random linearly elastic solid is considered. General formulisms are derived that govern the one- and two-point moments defined on the stress, strain, and displacement fields. The weakly inhomogeneous solid is discussed within the framework of the derived formulisms. Also, we consider the simplifications that are introduced for problems that involve two length scales, one defined by the randomly varying material parameters and one defined by the characteristic dimensions of the over-all geometry of the solid and of all forcing mechanisms.

INTRODUCTION

For an important class of engineering materials the homogeneous linearly elastic continuum represents an idealization that is valid only on a certain scale of observation, which may be termed the macroscale. A closer examination on a finer scale, which is still far above the atomic scale, reveals heterogeneity although the model of a linearly elastic continuum is still valid. This finer scale may be termed the microscale. Examples of such materials are numerous. The most important is possibly the polycrystal, which is an aggregate of a very large number of anisotropic crystals that are oriented in space in a random fashion. Each crystal is large enough to be idealized as a homogeneous linearly elastic continuum. Its mechanical properties are described by an elastic moduli tensor with components, referred to a space fixed system, that do not vary with position in the crystal. The components of the elastic moduli tensor for the polycrystal as a unit, again referred to a space fixed system, do vary with position in the polycrystal as one moves across the individual crystals. A second example of the class of materials of interest is the fiber reinforced composite. Here, again, the fibers or the regions of the matrix between fibers are large enough to be idealized by homogeneous linearly elastic continua. The parameters that define the mechanical properties do not vary with position in the fiber or in the matrix, but do vary with a position change from a fiber to the matrix.

A second feature that is common to the two examples cited is that the scalar fields needed to define the spatially varying material properties can only be described in statistical terms. That is, the scalar fields are given by stochastic processes. In this paper we observe the above described class of materials on the microscale and give an explicitly statistical interpretation to the problem.

A complete statistical formulation of the problem would be in terms of probability distribution functionals. The input to such a formulation would be a probability distribution functional defined on the stochastic processes needed to describe the spatially varying material parameters. The output would be a probability distribution functional defined on the stochastic processes needed to describe the response measures of the elastic continuum as well as those needed to describe the spatially varying material parameters. Using the theory of functionals and following the work of Hopf,[1] it seems clear that one could present such a complete statistical formulation, although to the author's knowledge this has not been done. It is to be expected that the functional formulation will be exceedingly complex.

The task we set for ourselves is much more limited. We intend to concentrate on the lower order statistical moments and develop a mathematical formulism that determines the ensemble, or statistical, averaged response measures [i.e., $\langle \tau_{ij}(\mathbf{x}) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}) \rangle$ and $\langle u_i(\mathbf{x}) \rangle$; the mean stress field, the mean strain field, and the mean displacement field] as well as one that determines the two-point moments defined by these response measures [i.e., $\langle \tau_{ij}(\mathbf{x}^1)\tau_{kl}(\mathbf{x}^2) \rangle$, $\langle \tau_{ij}(\mathbf{x}^1)\epsilon_{kl}(\mathbf{x}^2) \rangle$, $\langle \tau_{ij}(\mathbf{x}^1)u_k(\mathbf{x}^2) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}^1)\epsilon_{kl}(\mathbf{x}^2) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}^1)u_k(\mathbf{x}^2) \rangle$, and $\langle u_i(\mathbf{x}^1)u_j(\mathbf{x}^2) \rangle$]. The procedure to be used to accomplish the development is a slight extension of one that has been previously used in a variety of problems involving statistical continua. Beran and McCoy[2] have used it to develop the desired formulism on $\langle \tau_{ij}(\mathbf{x}) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}) \rangle$, and $\langle u_i(\mathbf{x}) \rangle$. A much abbreviated rederivation of this same formulism is reproduced in the present paper since it serves to clarify the development of the formulism that governs the two-point moments. Additional references on the procedure are cited in the text at appropriate places.

Some words of comment on the physical significance of the quantities we wish to determine are warranted. Inherent to the statistical interpretation of the problem is the idea that we are dealing not with a single solid, but with an assemblage (ensemble) of solids that are identical in some way (i.e., appear identical when observed on the macroscale), but differ in another way (i.e., appear different when observed on the microscale.) The meaning of $\langle u_i(\mathbf{x}) \rangle$, for example, is a weighted average of the displacements that one would measure at the same point in each of the solids of the assemblage. This interpretation of $\langle u_i(\mathbf{x}) \rangle$ is unambiguous and may always be applied. If, however, a problem exists for which variations in $\langle u_i(\mathbf{x}) \rangle$ with a change in position are only measurable for a posi-

ences therein.
[8] I. Shimamura, J. Phys. Soc. Japan **25**, 971 (1968); **28**, 1057 (1970).
[9] K. J. Miller, Phys. Rev. A **3**, 607 (1971).
[10] J. N. Bardsley, E. Gerjuoy, and C. V. Sukumar, to be published.
[11] L. Spruch and L. Rosenberg, Phys. Rev. **116**, 1034 (1959).
[12] Y. Hahn, T. F. O'Malley, and L. Spruch, Phys. Rev. **134**, B911 (1964).
[13] S. Borowitz and E. Gerjuoy, University of Colorado, Boulder, Colorado, JILA Report 36 (May 14, 1965).
[14] E. Gerjuoy, A. R. P. Rau, and L. Spruch, to be published.
[15] A. Dalgarno and A. L. Stewart, Proc. Roy. Soc. **A257**, 534 (1960).
[16] Charles Schwartz, Ann. Phys. (N.Y.) **2**, 156, 170 (1959).
[17] L. M. Delves, Nucl. Phys. **41**, 497 (1963).
[18] L. Spruch, in *Proceedings of the Ninth Summer Meeting of Yugoslav Physicists, Hercegnovi, July, 1964*, edited by M. Cerineo (Federal Nuclear Energy Commision of Yugoslavia, Belgrade, 1965), Vol. 1, p. 191.
[19] R. Blau, A. R. P. Rau, and L. Spruch, submitted to Phys. Rev. A.
[20] See, for instance, F. Calogero, *Variable Phase Approach to Potential Scattering* (Academic, New York, 1967).

On the Correlation of Field Quantities in a Random Linearly Elastic Solid

John J. McCoy
The Catholic University of America, Washington, D.C. 20017
(Received 19 June 1972)

The response of a random linearly elastic solid is considered. General formulisms are derived that govern the one- and two-point moments defined on the stress, strain, and displacement fields. The weakly inhomogeneous solid is discussed within the framework of the derived formulisms. Also, we consider the simplifications that are introduced for problems that involve two length scales, one defined by the randomly varying material parameters and one defined by the characteristic dimensions of the over-all geometry of the solid and of all forcing mechanisms.

INTRODUCTION

For an important class of engineering materials the homogeneous linearly elastic continuum represents an idealization that is valid only on a certain scale of observation, which may be termed the macroscale. A closer examination on a finer scale, which is still far above the atomic scale, reveals heterogeneity although the model of a linearly elastic continuum is still valid. This finer scale may be termed the microscale. Examples of such materials are numerous. The most important is possibly the polycrystal, which is an aggregate of a very large number of anisotropic crystals that are oriented in space in a random fashion. Each crystal is large enough to be idealized as a homogeneous linearly elastic continuum. Its mechanical properties are described by an elastic moduli tensor with components, referred to a space fixed system, that do not vary with position in the crystal. The components of the elastic moduli tensor for the polycrystal as a unit, again referred to a space fixed system, do vary with position in the polycrystal as one moves across the individual crystals. A second example of the class of materials of interest is the fiber reinforced composite. Here, again, the fibers or the regions of the matrix between fibers are large enough to be idealized by homogeneous linearly elastic continua. The parameters that define the mechanical properties do not vary with position in the fiber or in the matrix, but do vary with a position change from a fiber to the matrix.

A second feature that is common to the two examples cited is that the scalar fields needed to define the spatially varying material properties can only be described in statistical terms. That is, the scalar fields are given by stochastic processes. In this paper we observe the above described class of materials on the microscale and give an explicitly statistical interpretation to the problem.

A complete statistical formulation of the problem would be in terms of probability distribution functionals. The input to such a formulation would be a probability distribution functional defined on the stochastic processes needed to describe the spatially varying material parameters. The output would be a probability distribution functional defined on the stochastic processes needed to describe the response measures of the elastic continuum as well as those needed to describe the spatially varying material parameters. Using the theory of functionals and following the work of Hopf,[1] it seems clear that one could present such a complete statistical formulation, although to the author's knowledge this has not been done. It is to be expected that the functional formulation will be exceedingly complex.

The task we set for ourselves is much more limited. We intend to concentrate on the lower order statistical moments and develop a mathematical formulism that determines the ensemble, or statistical, averaged response measures [i.e., $\langle \tau_{ij}(\mathbf{x}) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}) \rangle$ and $\langle u_i(\mathbf{x}) \rangle$; the mean stress field, the mean strain field, and the mean displacement field] as well as one that determines the two-point moments defined by these response measures [i.e., $\langle \tau_{ij}(\mathbf{x}^1)\tau_{kl}(\mathbf{x}^2) \rangle$, $\langle \tau_{ij}(\mathbf{x}^1)\epsilon_{kl}(\mathbf{x}^2) \rangle$, $\langle \tau_{ij}(\mathbf{x}^1)u_k(\mathbf{x}^2) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}^1)\epsilon_{kl}(\mathbf{x}^2) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}^1)u_k(\mathbf{x}^2) \rangle$, and $\langle u_i(\mathbf{x}^1)u_j(\mathbf{x}^2) \rangle$]. The procedure to be used to accomplish the development is a slight extension of one that has been previously used in a variety of problems involving statistical continua. Beran and McCoy[2] have used it to develop the desired formulism on $\langle \tau_{ij}(\mathbf{x}) \rangle$, $\langle \epsilon_{ij}(\mathbf{x}) \rangle$, and $\langle u_i(\mathbf{x}) \rangle$. A much abbreviated rederivation of this same formulism is reproduced in the present paper since it serves to clarify the development of the formulism that governs the two-point moments. Additional references on the procedure are cited in the text at appropriate places.

Some words of comment on the physical significance of the quantities we wish to determine are warranted. Inherent to the statistical interpretation of the problem is the idea that we are dealing not with a single solid, but with an assemblage (ensemble) of solids that are identical in some way (i.e., appear identical when observed on the macroscale), but differ in another way (i.e., appear different when observed on the microscale.) The meaning of $\langle u_i(\mathbf{x}) \rangle$, for example, is a weighted average of the displacements that one would measure at the same point in each of the solids of the assemblage. This interpretation of $\langle u_i(\mathbf{x}) \rangle$ is unambiguous and may always be applied. If, however, a problem exists for which variations in $\langle u_i(\mathbf{x}) \rangle$ with a change in position are only measurable for a posi-

tion change that is observed on the macroscale, then an alternate interpretation of $\langle u_i(\mathbf{x})\rangle$ is possible. This alternate interpretation is that $\langle u_i(\mathbf{x})\rangle$ is a spatial average of the displacements that one would measure in each and every solid in the ensemble. The region over which the spatial average is to be taken appears very large when observed on the microscale at the same time that it appears to be very small when observed on the macroscale. In applying the classical elasticity theory to a polycrystal and in interpreting the predictions in the usual manner, we are implicitly assuming that the correct formulism on $\langle \tau_{ij}(\mathbf{x})\rangle$, $\langle \epsilon_{ij}(\mathbf{x})\rangle$, and $\langle u_i(\mathbf{x})\rangle$ reduces to the classical elasticity formulism under the conditions cited and are also assuming the validity of the ergodic hypothesis described above. The formulism developed by Beran and McCoy does reduce to the elasticity formulism under the conditions cited; but, by considering a specific example,[3] they also showed that the solutions of the general formulism do not uniformly agree with the solutions of the elasticity formulism in the appropriate limit. There will exist thin layers of all bounding surfaces and all forcing mechanisms within which the classical theory ceases to be valid and within which the only average that can be discussed is an ensemble average.

The direct physical significance of all of the two-point moments is not clear although the significance of some of the information contained therein is not to be disputed. For example, the limit of $\langle \tau_{ij}(\mathbf{x}^1)\tau_{ij}(\mathbf{x}^2)\rangle$ (no sum) as $\mathbf{x}^2$ approaches $\mathbf{x}^1$, under certain conditions, equals $\langle \tau_{ij}^2(\mathbf{x}^1)\rangle$. This latter term defines the variance of the stress tensor for the point located by $\mathbf{x}^1$. Also, the limit of $\langle \tau_{ij}(\mathbf{x}^1)\epsilon_{ij}(\mathbf{x}^2)\rangle$ (sum) as $\mathbf{x}^2$ approaches $\mathbf{x}^1$, under certain conditions, equals $2\langle V(\mathbf{x}^1)\rangle$ where $V(\mathbf{x})$ denotes the internal energy density. Further, if the variation of $\langle \tau_{ij}(\mathbf{x}^1)\tau_{ij}(\mathbf{x}^2)\rangle$ (no sum) with a change in absolute position can be observed only for a position change that is measured on the macroscale, then the Wiener–Khinchin theorem predicts that $\langle \tau_{ij}(\mathbf{x}^1)\tau_{ij}(\mathbf{x}^2)\rangle$ provides the amplitudes of the Fourier decomposition of the spatial variations of $\tau_{ij}(\mathbf{x})$ that one would observe on the macroscale. The reason that the formulism developed in this paper contains all of the 120 different correlation functions that can be defined by the stress, strain, and displacement fields, is not motivated by the physical significance that can be attached to all of them, but is rather a consequence of the derivation procedure used.

The general formulisms are developed in the next section although a number of intermediate details are carried out in an appendix. The results are given by Eqs. (16)–(19) and Eqs. (25) and (26). In the following section we consider the case of a weakly inhomogeneous solid for which it is possible to truncate the infinite series that appear in the general formulisms thereby achieving a completely defined formulation on the unknowns of interest. In this section we also consider the two length scale situation previously discussed and the simplifications that this introduces into the formulisms. The simplified formulisms are given by Eqs. (36) and (37).

GENERAL FORMULATION—SOME BASIC CONSIDERATIONS

Referred to a cartesian coordinate system, the equations governing the response of a linearly elastic solid in the absence of inertia effects may be written in the matrix form

$$AX = F. \tag{1}$$

Here, X denotes the response matrix, which is given by

$$X = \begin{Bmatrix} \tau(\mathbf{x}) \\ \epsilon(\mathbf{x}) \\ u(\mathbf{x}) \end{Bmatrix}, \tag{2}$$

where the submatrices τ, ϵ, and u contain the components of the stress, strain, and displacement fields, respectively. We write

$$\tau(\mathbf{x}) = \begin{Bmatrix} \tau_{11}(\mathbf{x}) \\ \tau_{22}(\mathbf{x}) \\ \tau_{33}(\mathbf{x}) \\ \tau_{23}(\mathbf{x}) = \tau_{32}(\mathbf{x}) \\ \tau_{13}(\mathbf{x}) = \tau_{31}(\mathbf{x}) \\ \tau_{12}(\mathbf{x}) = \tau_{21}(\mathbf{x}) \end{Bmatrix},$$

$$\epsilon(\mathbf{x}) = \begin{Bmatrix} \epsilon_{11}(\mathbf{x}) \\ \epsilon_{22}(\mathbf{x}) \\ \epsilon_{33}(\mathbf{x}) \\ 2\epsilon_{23}(\mathbf{x}) = 2\epsilon_{32}(\mathbf{x}) \\ 2\epsilon_{13}(\mathbf{x}) = 2\epsilon_{31}(\mathbf{x}) \\ 2\epsilon_{12}(\mathbf{x}) = 2\epsilon_{21}(\mathbf{x}) \end{Bmatrix},$$

$$u(\mathbf{x}) = \begin{Bmatrix} u_1(\mathbf{x}) \\ u_2(\mathbf{x}) \\ u_3(\mathbf{x}) \end{Bmatrix},$$

and note that the subscript associates the component with a like subscripted coordinate axis. The force field matrix F is given by

$$F = \begin{Bmatrix} f(\mathbf{x}) \\ 0 \\ 0 \end{Bmatrix}, \tag{4}$$

where the submatrix f contain the components of the body force per unit volume, i.e.,

$$f(\mathbf{x}) = \begin{Bmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ f_3(\mathbf{x}) \end{Bmatrix}. \tag{5}$$

The operator matrix A is given by

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ 0 & a_{32} & a_{33} \end{pmatrix}. \tag{6}$$

The submatrices a_{ij} are defined as follows:

$$a_{11} = \begin{pmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{pmatrix},$$

$$a_{21} = a_{32} = -I_6,$$

$$a_{33} = \begin{pmatrix} \partial_1 & 0 & 0 \\ 0 & \partial_2 & 0 \\ 0 & 0 & \partial_3 \\ 0 & \frac{1}{2}\partial_3 & \frac{1}{2}\partial_2 \\ \frac{1}{2}\partial_3 & 0 & \frac{1}{2}\partial_1 \\ \frac{1}{2}\partial_2 & \frac{1}{2}\partial_1 & 0 \end{pmatrix}, \tag{7}$$

and a_{22} is a symmetric 6×6 matrix with elements that are determined by the mechanical properties of the solid. In the above, I_6 denotes the 6×6 unit

matrix, and ∂_i denotes differentiation with respect to the x_i coordinate.

The random linearly elastic solid is defined as one that requires stochastic functions of position in the solid to describe the material properties matrix a_{22}.

In order to ensure uniqueness of solution, it is necessary to specify boundary conditions, which serve to restrict the domain of the operator A. At this point we shall simply require that the boundary conditions are conditions about which we have sure or deterministic information.

In this paper we intend to develop a mathematical formulism that is to be satisfied by the correlation matrix $\langle X(\mathbf{x}) \otimes X(\mathbf{y})\rangle$. Here $X(\mathbf{x})$ and $X(\mathbf{y})$ are the response matrices measured at two different points in the solid, $\otimes$ denotes the Kronecker product,[4] and the angular brackets denote an ensemble or statistical, average. In terms of the sumatrices τ, ϵ, and u, the correlation matrix is written

$$\langle X(\mathbf{x}) \otimes X(\mathbf{y})\rangle = \left\{ \begin{array}{c} \langle \tau(\mathbf{x}) \otimes \tau(\mathbf{y})\rangle \\ \langle \tau(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle \\ \langle \tau(\mathbf{x}) \otimes u(\mathbf{y})\rangle \\ \langle \epsilon(\mathbf{x}) \otimes \tau(\mathbf{y})\rangle \\ \langle \epsilon(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle \\ \langle \epsilon(\mathbf{x}) \otimes u(\mathbf{y})\rangle \\ \langle u(\mathbf{x}) \otimes \tau(\mathbf{y})\rangle \\ \langle u(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle \\ \langle u(\mathbf{x}) \otimes u(\mathbf{y})\rangle \end{array} \right\}. \tag{8}$$

A further expansion of the submatrices results in a correlation matrix that contains 225 scalar functions of $\mathbf{x}$ and $\mathbf{y}$. Only 120 of these scalar functions are distinct, however.

The procedures to be employed in developing the desired formulism is termed the method of smoothing. A survey article by Frisch[5] discusses the history of the method as it applies to developing a formulism to be satisfied by the mean response matrix. Following Frisch, we first average Eq. (1) to obtain

$$\langle A\rangle\langle X\rangle + \langle A'X'\rangle = F. \tag{9}$$

A prime has been introduced to denote the difference between a stochastic quantity and its mean value. That is, for example,

$$A' = A - \langle A\rangle. \tag{10}$$

We note that $\langle A'\rangle \equiv 0$ and that, for the problem of interest, A' is an algebraic matrix. Equation (9) is not the desired equation on the mean response matrix since, in addition to $\langle X\rangle$, it contains the unknown average $\langle A'X'\rangle$. It will become the desired equation once we obtain an expression for $\langle A'X'\rangle$ in terms of $\langle X\rangle$. To do this we subtract Eq. (9) from Eq. (1) to obtain

$$\langle A\rangle X' + (I - P)A'X' = -A'\langle X\rangle. \tag{11}$$

Here, I denotes the identity operator and P denotes the ensemble averaging operator, i.e.,

$$I\varphi = \varphi, \qquad P\varphi = \langle\varphi\rangle. \tag{12}$$

Equation (12) is viewed as a condition to be satisfied by X' in which $A'\langle X\rangle$ is taken to be a known forcing term. By direct substitution one can see that Eq. (12) is satisfied by the infinite series expression

$$X' = -\sum_{n=0} (-)^n[(I - P)\langle A\rangle^{-1}A']^n\langle A\rangle^{-1}A'\langle X\rangle \tag{13}$$

provided the series converges.[6] In this expression, $\langle A\rangle^{-1}$ denotes the inverse of the operator $\langle A\rangle$. If $\langle A\rangle$ is a differential operator, as in the present case, a uniquely specified inverse requires specification of boundary conditions. The appropriate boundary conditions are determined by the problem under consideration. Let us restrict attention to problem in which A and $\langle A\rangle$ require the same numbers and types of boundary conditions in order to uniquely determine their inverses. This is the case for the elasticity problem. Now, in the original problem statement we specify, in addition to the field equations, boundary conditions. The appropriate boundary conditions for $\langle A\rangle^{-1}$, as it occurs in Eq. (13), are homogeneous conditions of the same type as that contained in the original problem statement. In the elasticity example, the original problem statement will specify either the component of the traction vector or the component of the displacement vector in each of three noncoplanar directions at every boundary point. If the components of the traction vector are specified, for example, then, when we consider $\langle A\rangle$ in Eq. (13), the appropriate boundary condition would be that the traction vector is zero. Thus Eq. (13) represents an unambiguous prescription for A'. [No claim is made that Eq. (13) represents the only solution of Eq. (11). The statement is just that Eq. (13) represents an unambiguous prescription. The question as to whether the Dyson equation is sufficiently restrictive to uniquely determine the mean response matrix can be raised after the Dyson equation is obtained.] It is this prescription that we substitute into Eq. (10) to obtain the desired formulation on the mean response matrix. We write this equation as

$$D\langle X\rangle = (\langle A\rangle - M)\langle X\rangle = F, \tag{14}$$

where the operator M has the following infinite series prescription

$$M = -\sum_{n=0} (-)^n\langle A'[(I - P)\langle A\rangle^{-1}A']^n\langle A\rangle^{-1}A'\rangle. \tag{15}$$

This equation is often termed a Dyson equation, and the operator D is termed a Dyson operator. The operator M is often termed the mass operator. For the elasticity problem the Dyson equation leads to the following tensor field equations on the mean stress, mean strain, and mean displacement fields (see Appendix A):

$$\partial_j\langle\tau_{ij}\rangle = f_i, \tag{16}$$

$$\langle\tau_{ij}\rangle = \langle\tau_{ji}\rangle, \tag{17}$$

$$\langle\tau_{ij}(\mathbf{x})\rangle = \int \mathcal{C}_{ijkl}(\mathbf{x}, \mathbf{x}^1)\langle\epsilon_{kl}(\mathbf{x}^1)\rangle d\mathbf{x}^1 + \oint \mathcal{D}_{ijkl}(\mathbf{x}, \mathbf{x}_s^1)\langle\epsilon_{kl}(\mathbf{x}_s^1)\rangle\, d\mathbf{x}_s^1, \tag{18}$$

$$\langle\epsilon_{ij}\rangle = \tfrac{1}{2}(\partial_i\langle u_j\rangle + \partial_j\langle u_i\rangle). \tag{19}$$

In Eq. (18) the first integral is over the extent of the solid and the second integral is over the bounding surface. The infinite series of the Dyson equation appears in the prescriptions of the two-point tensor fields $\mathcal{C}_{ijkl}(\mathbf{x}, \mathbf{x}^1)$ and $\mathcal{D}_{ijkl}(\mathbf{x}, \mathbf{x}_s^1)$. As it applies to a

solid of infinite extent, this formulation was given in Ref. 2; some aspects of the solutions predicted by it were given in Ref. 3; a uniqueness theorem for the formulation was given in Ref. 7; and it was extended to incorporate inertia effects in Ref. 8. We note here the conclusion that when the randomly varying elastic moduli tensor defines one length scale, say l, and the variations in the mean strain field define a second length scale, say L, where $L \gg l$, then to zeroth order, Eqs. (16)–(19) are approximated over most of the solid by an effective moduli theory. (See next section.)

To apply the method of smoothing to obtain an equation on $\langle X(\mathbf{x}) \otimes X(\mathbf{y})\rangle$, we first form the Kronecker product of Eq. (1) as it applies to a field point $\mathbf{x}$ to the same equation as it applies to a second field point $\mathbf{y}$. The result is written

$$(A_x \otimes A_y)R = G, \tag{20}$$

where the subscript on A indicates the point at which it is to be applied:

$$\begin{aligned} R(\mathbf{x},\mathbf{y}) &= X(\mathbf{x}) \otimes X(\mathbf{y}), \\ G(\mathbf{x},\mathbf{y}) &= F(\mathbf{x}) \otimes F(\mathbf{y}). \end{aligned} \tag{21}$$

Next, we write the Dyson equation for Eq. (20). In the resulting prescription appears the inverse of

$$\langle A_x \otimes A_y\rangle = \langle A_x\rangle \otimes \langle A_y\rangle + \langle A'_x \otimes A'_y\rangle.$$

We can express this inverse in the form of an infinite series, each term of which requires only the inverses of $\langle A_x\rangle$ and $\langle A_y\rangle$. So doing and performing an extensive amount of rearranging of terms results in the desired equation on the correlation matrix. For details the reader is referred to McCoy.[9] The final equation is written

$$(D_x \otimes D_y - I_{xy})\langle R(\mathbf{x},\mathbf{y})\rangle = G(\mathbf{x},\mathbf{y}) \tag{22}$$

or

$$\begin{aligned} \langle R(\mathbf{x},\mathbf{y})\rangle &= (D_x^{-1} \otimes D_y^{-1})G(\mathbf{x},\mathbf{y}) \\ &\quad + (D_x^{-1} \otimes D_y^{-1})I_{xy}\langle R(\mathbf{x},\mathbf{y})\rangle \\ &= \langle X(\mathbf{x})\rangle \otimes \langle X(\mathbf{y})\rangle + (D_x^{-1} \otimes D_y^{-1})I_{xy}\langle R(\mathbf{x},\mathbf{y})\rangle. \end{aligned} \tag{23}$$

The operator I_{xy} is given in the form of an infinite series. Arranging this series in powers of A', we can write

$$\begin{aligned} I_{xy} = \langle A'_x \otimes A'_y\rangle &- \langle (A'_x\langle A_x\rangle^{-1}A'_x) \otimes A'_y\rangle \\ &- \langle A'_x \otimes (A'_y\langle A_y\rangle^{-1}A'_y)\rangle + o(A'^4), \end{aligned} \tag{24}$$

where $o(A'^4)$ contains all terms containing fourth and higher powers of A'. Equation (22) or (23) is often termed a Bethe–Salpeter equation. The operator I_{xy} is often termed the intensity operator.

Expansion of the Bethe–Salpeter equation for the elasticity example is an extremely tedious procedure, and the details have been relegated to the Appendix. We present here the following results. The correlation functions given by the displacement field, i.e., $\langle u(\mathbf{x}) \otimes u(\mathbf{y})\rangle$, satisfy the matrix operator equation[10]

$$\begin{aligned} &([a_{11}(\mathbf{x}) \otimes a_{11}(\mathbf{y})]\{[\langle a_{22}(\mathbf{x})\rangle - m_{22}(\mathbf{x})] \otimes [\langle a_{22}(\mathbf{y})\rangle \\ &- m_{22}(\mathbf{y})] - i_{55}(\mathbf{x},\mathbf{y})\}[a_{33}(\mathbf{x}) \otimes a_{33}(\mathbf{y})])\langle u(\mathbf{x}) \otimes u(\mathbf{y})\rangle \\ &= f(\mathbf{x}) \otimes f(\mathbf{y}). \end{aligned} \tag{25}$$

The matrix operators m_{22} and i_{55} are given by the infinite series in Eqs. (A12) and (A17). The remaining submatrices of $\langle X(\mathbf{x}) \otimes X(\mathbf{y})\rangle$ are directly calculable in terms of $\langle u(\mathbf{x}) \otimes u(\mathbf{y})\rangle$. We present here, as examples, the following equations on $\langle \epsilon(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle$ and $\langle \tau(\mathbf{x}) \otimes \tau(\mathbf{y})\rangle$:

$$\langle \epsilon(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle = [a_{33}(\mathbf{x}) \otimes a_{33}(\mathbf{y})]\langle u(\mathbf{x}) \otimes u(\mathbf{y})\rangle \tag{26a}$$

and

$$\begin{aligned} \langle \tau(\mathbf{x}) \otimes \tau(\mathbf{y})\rangle = \{&[\langle a_{22}(\mathbf{x})\rangle - m_{22}(\mathbf{x})] \otimes [\langle a_{22}(\mathbf{y})\rangle \\ &- m_{22}(\mathbf{y})] - i_{55}(\mathbf{x},\mathbf{y})\}\langle \epsilon(\mathbf{x}) \otimes \epsilon(\mathbf{y})\rangle. \end{aligned} \tag{26b}$$

These equations may be given in terms of a tensor notation. We do so for the case of a solid of unbounded extent.

Equation (25) leads to

$$\begin{aligned} &\partial_{i_2}\partial_{\alpha_2}\int\int[\mathcal{C}_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)\mathcal{C}_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y},\mathbf{y}^1) \\ &\quad - \mathcal{g}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1;\mathbf{y},\mathbf{y}^1)] \\ &\quad \times \partial^{(1)}_{i_3}\partial^{(1)}_{\alpha_3}\langle u_{i_4}(\mathbf{x}^1)u_{\alpha_4}(\mathbf{y}^1)\rangle d\mathbf{x}^1 d\mathbf{y}^1 = f_{i_1}(\mathbf{x})f_{\alpha_1}(\mathbf{y}), \end{aligned} \tag{27}$$

and Eqs. (26) become

$$\begin{aligned} \langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle = \tfrac{1}{4}(&\partial_{i_1}\delta_{i_2i_3} + \partial_{i_2}\delta_{i_1i_3})(\partial_{\alpha_1}\delta_{\alpha_2\alpha_3} \\ &+ \partial_{\alpha_2}\delta_{\alpha_1\alpha_3})\langle u_{i_3}(\mathbf{x})u_{\alpha_3}(\mathbf{y})\rangle \end{aligned}$$

and

$$\begin{aligned} \langle \tau_{i_1i_2}(\mathbf{x})\tau_{\alpha_1\alpha_2}(\mathbf{y})\rangle = &\int\int[\mathcal{C}_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)\mathcal{C}_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y},\mathbf{y}') \\ &- \mathcal{g}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1;\mathbf{y},\mathbf{y}^1)] \\ &\times \langle \epsilon_{i_3i_4}(\mathbf{x}^1)\epsilon_{\alpha_3\alpha_4}(\mathbf{y}^1)\rangle d\mathbf{x}^1 d\mathbf{y}^1. \end{aligned} \tag{28}$$

In these equations, an i subscript refers to the x coordinate and an α subscript refers to the y coordinate. The two-point tensor field denoted by a $\mathcal{C}$ is that appearing in Eq. (18), and the four-point tensor field denoted by $\mathcal{g}$ is given by the infinite series of Eq. (A20). Both of these fields depend on statistical moments of all orders of the elastic moduli tensor and on the nature of the boundary conditions of the problem to which the equations are to be applied. The equations for a bounded solid are complicated by the appearance of surface integrals. These equations for a bounded solid are readily obtained by making use of the expressions in the Appendix.

We consider some additional manipulations of the derived formulism in the next section, in which we investigate the weakly inhomogeneous solid and the two length scale problem discussed in the Introduction. We end this section by presenting the following discussion on the usefulness of the derived formulism.

The infinite series prescriptions for the various kernels prevent the derived formulisms from representing an end product that has computational value except for the special cases that enable our either truncating or summing the series. It would appear, therefore, that it should prove more fruitful to view the derived formulisms as intermediate steps that have been achieved without recourse to any physical approximations or restrictions.[11] To achieve formulisms that will have computational value we engage in some phenomenology and accept the derived formulisms, without the series prescriptions for the kernel

functions, as starting points. That is, we view the kernel functions as formulation parameters that are to be determined by physical experiments under laboratory conditions just as the elastic moduli tensor of the classical theory is to be determined. One distinction is to be noted between the kernel functions as formulation parameters and the elastic moduli tensor as a formulation parameter and that is that the former are not identified as material parameters. The nature of the boundary conditions of the problem to which the formulism is to be applied enters the infinite series definition of each of the kernel functions via the Green's function matrix G. (See Eq. (A1).) While it would be very desirable for our kernel functions to be dependent only on the material properties of the solid, this is not a requirement of a useful computational formulism. We do note that the dependence of the kernel functions on the boundary conditions is the same for a class of boundary conditions and that the class is identified by a canonical condition. We further note that it can be argued, for the class of materials to which we intend to apply the formulism (see the Introduction), that the dependence of the kernel functions on the boundary conditions will be significant only if the pair (or quartet) of points on which the functions are defined lie within a thin layer of the bounding surface. The reader is referred to Ref. 2 for these arguments. We might assume that any error that results from ignoring the boundary dependence of the kernel functions will similarly be limited to field points that fall within such boundary layers. Thus, we can argue that formulisms can be achieved for making predictions over much of the solid based on the kernel functions being viewed as material parameters. The regions in which the predictions made by such formulisms break down will be seen to be regions in which it is extremely difficult to make or interpret predictions.

The next step in the direction of obtaining a useful computational formulism is to place restrictions on the kernel functions.[12] This step is of obvious importance since it would take a nondenumerably infinite number of physical measurements to completely determine a kernel function. Some obvious restrictions on the kernel functions will be supplied by invariance requirements, causality requirements, etc. Care must be taken in this regard that one doesn't uncritically extrapolate some of the like requirements of the classical theory. For example, it is erroneous to conclude that the average increment of work done by the internal forces in a region is a region integral of $\langle \tau_{ij} \rangle \, d\langle \epsilon_{ij} \rangle$. After satisfying these general restrictions, we further limit the degrees of freedom in the kernel functions by delineating classes of problems to be analyzed. A specific class is to be defined by the nature of any additional restrictions we place on the kernel functions. (Again, we refer to Ref. 12.) The usefulness of a particular class will ultimately depend on the degree of coincidence of the predictions made by a given restricted formulism and experimental data.

We intend to consider the results of postulating some restrictions in a future work.

WEAKLY INHOMOGENEOUS SOLID

It proved convenient to carry out the derivation procedure leading to the Dyson and Bethe–Salpeter equations using a matrix notation. Once the procedure has been accomplished, however, it is just as convenient to switch to a tensor notation. The material properties of the solid are given in the tensor notation by C_{ijke}, which is termed the elastic moduli tensor. The relationship between the elements of this fourth rank tensor and those of the a_{22} matrix are readily obtained by comparing the equations given by (A5) and (A6) of the Appendix.

The infinite series prescriptions for the kernel functions that appear in the general formulisms are arranged in increasing powers of C'_{ijkl}, the randomly fluctuating part of the elastic moduli tensor. This suggests that, for a weakly inhomogeneous solid, we can obtain approximations to these kernel functions by truncating these series prescriptions. Retaining terms up to and including second order in powers of C'_{ijkl}, we have the following expressions:

$$
\begin{aligned}
&\mathcal{C}_{i_1i_2i_3i_4}(\mathbf{x}, \mathbf{x}^1) \\
&\quad = [\langle C_{i_1i_2i_3i_4} \rangle + A_{i_5i_6i_7i_8} \langle C'_{i_1i_2i_5i_6}(\mathbf{x}) C'_{i_7i_8i_3i_4}(\mathbf{x}) \rangle] \\
&\qquad \times \delta(\mathbf{x} - \mathbf{x}^1) + F_{i_5i_6i_7i_8}(\mathbf{x}, \mathbf{x}^1) \langle C'_{i_1i_2i_5i_6}(\mathbf{x}) C'_{i_7i_8i_3i_4}(\mathbf{x}^1) \rangle, \\
&\mathcal{D}_{i_1i_2i_3i_4}(\mathbf{x}, \mathbf{x}^1_s) = n_{i_l}(\mathbf{x}^1_s) E_{i_5i_6i_7}(\mathbf{x}, \mathbf{x}^1_s) \langle C'_{i_1i_2i_5i_6}(\mathbf{x}) C'_{i_7i_8i_3i_4}(\mathbf{x}^1_s) \rangle, \\
&\mathcal{G}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x}, \mathbf{x}^1; \mathbf{y}, \mathbf{y}^1) \\
&\quad = \langle C'_{i_1i_2i_3i_4}(\mathbf{x}) C_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y}) \rangle \delta(\mathbf{x} - \mathbf{x}^1) \delta(\mathbf{y} - \mathbf{y}^1), \\
&\mathcal{G}^{(1)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x}, \mathbf{x}^1_s; \mathbf{y}, \mathbf{y}^1_s) \\
&\quad \equiv \mathcal{G}^{(2)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x}, \mathbf{x}^1_s; \mathbf{y}, \mathbf{y}^1_s) \equiv 0. \qquad (29)
\end{aligned}
$$

Each of the terms are defined in the Appendix. Substitution of these into the general formulisms results in what might be termed the weakly inhomogeneous formulisms. It is to be assumed that the solutions of the weakly inhomogeneous formulisms converge to the solutions of the general formulisms in the indicated limit. However, we might suspect that the convergence will probably be an asymptotic convergence and that, further, it will not be uniform.

We might parenthetically remark here that although the case of a weakly inhomogeneous solid is of limited practical significance, it is of great theoretical significance since it is a case for which the general formulism is completely defined without resorting to any phenomenology. Thus, it provides us with a means to physically verify the derived formulism. This is of interest because of our present inability to rigorously justify all of the mathematical steps of the derivation procedure. In addition, the weakly inhomogeneous solid might provide forms for the kernel functions, which have a wider range of applicability.

Consider a weakly inhomogeneous solid of the type discussed in the Introduction. For such solids it is reasonable to assume that the statistics of a material property that is measured at two points in the solid are independent if the distance between the points exceeds some limiting value. We denote this limiting value by l and expect that, for a polycrystal, l is of the order of a linear dimension of a single crystal. For a composite, if the locations of the fibers are to be statistically independent from one another, then it is expected to be of the order of a characteristic fiber dimension. Statistical independence requires,

among other things, that $\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{i_5i_6i_7i_8}(\mathbf{x}^1)\rangle = 0$. Referring to Eq. (29), therefore, we immediately conclude that the kernels are zero if the pair (or quartet) of points on which they are defined fall outside a neighborhood, of linear dimension l, of one another. Further, upon making reference to Eq. (18), we note that the surface integral appearing therein offers a contribution only if the field point falls within a layer of linear dimension l of a bounding surface. In what follows, we shall not attempt to make predictions of the response measures for points within the boundary layers.

The smallest characteristic dimension of the overall geometry of the solid and of all forcing mechanisms is denoted by L, and we restrict attention to problems in which $l/L \ll 1$. We now assume that the variations in the average response fields with a change in absolute position is likewise measurable only on the L scale. This assumption is supported by a self-consistency argument between the assumption and the predictions of the approximate formulism that results from the assumption. We note that the self-consistency argument breaks down in the vicinity of boundary surfaces and forcing mechanisms. As a consequence of the assumption, the averaged strain field that appears in the integrals in Eq. (18) are taken to be constant over the region within which the integrands are nonzero. Thus, Eqs. (16)–(19) reduce to the classical elasticity formulism over much of the solid. Within the boundary layer discussed in the preceeding paragraph, there exist some differences. As already stated, we intend to ignore the boundary layers in this paper. For a further discussion see Refs. 2 and 3.

We now turn to the Bethe–Salpeter formulism and introduce both the weakly inhomogeneous prescriptions and the two length scale assumptions. In doing this we ignore all boundary layers of the type discussed.

By direct substitution of Eqs. (29) into Eqs. (27), we obtain a set of integro-partial differential equations that contain spatially varying coefficients, on the correlation of the displacement field. It is to be expected that this set of equations will present severe difficulties to even a numerical approach to the problem. It is to simplify this formulism that we introduce the presence of two length scales. Using the arguments of the preceeding paragraph, the integral operators that appear in the equations become differential operators. We write

$$\begin{aligned}&\partial_{i_2}\partial_{\alpha_2}[(C^*_{i_1i_2i_3i_4}C^*_{\alpha_1\alpha_2\alpha_3\alpha_4} - \Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y}))\\ &\quad\times \partial_{i_3}\partial_{\alpha_3}\langle u_{i_4}(\mathbf{x})u_{\alpha_4}(\mathbf{y})\rangle]\\ &\quad = f_{i_1}(\mathbf{x})f_{\alpha_1}(\mathbf{y}),\end{aligned} \tag{30}$$

where

$$\begin{aligned}&C^*_{i_1i_2i_3i_4} = \int \mathcal{C}_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}')d\mathbf{x}^1\\ &\text{and}\\ &\Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y}) = \langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle.\end{aligned} \tag{31}$$

In this equation, $C^*_{i_1i_2i_3i_4}$ is termed the effective elastic moduli tensor. For homogeneous statistics, $C^*_{i_1i_2i_3i_4}$ is a constant and $\Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}$ depends only on difference coordinates.

A further simplification of Eq. (30) is possible. To accomplish this it is convenient to introduce a series of integrations and differentiations of Eq. (30) to obtain an integral equation formulation on $\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle$. First we obtain from Eq. (30) the following integro-differential equations on the correlations of the displacement field:

$$\begin{aligned}&\langle u_{i_1}(\mathbf{x})u_{\alpha_1}(\mathbf{y})\rangle = \langle u_{i_1}(\mathbf{x})\rangle\langle u_{\alpha_1}(\mathbf{y})\rangle\\ &+ \int\int V^*_{i_1i_2}(\mathbf{x},\mathbf{x}^1)V^*_{\alpha_1\alpha_2}(\mathbf{y},\mathbf{y}^1)\partial^{(1)}_{i_3}\partial^{(1)}_{\alpha_3}\\ &\times[\Gamma_{i_2i_3i_4i_5\alpha_2\alpha_3\alpha_4\alpha_5}(\mathbf{x}^1,\mathbf{y}^1)\partial^{(1)}_{i_4}\partial^{(1)}_{\alpha_4}\langle u_{i_5}(\mathbf{x}^1)u_{\alpha_5}(\mathbf{y}^1)\rangle]d\mathbf{x}^1d\mathbf{y}^1,\end{aligned} \tag{32}$$

where $V^*_{i_1i_2}$ is the displacement Green's function tensor for the effective modulus formulation. It is the analog of $V_{i_1i_2}(\mathbf{x},\mathbf{x}^1)$, which is given in the Appendix, as it applies to the effective modulus formulation. Equation (32), in turn, leads to the following integro-differential equations on the correlations of the strain field. [See Eq. (28)].

$$\begin{aligned}&\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle = \langle \epsilon_{i_1i_2}(\mathbf{x})\rangle\langle \epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle\\ &\quad + \int\int E^*_{i_1i_2i_3}(\mathbf{x},\mathbf{x}^1)E^*_{\alpha_1\alpha_2\alpha_3}(\mathbf{y},\mathbf{y}^1)\partial^{(1)}_{i_4}\partial^{(1)}_{\alpha_4}\\ &\quad\times[\Gamma_{i_3i_4i_5i_6\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{x}^1,\mathbf{y}^1)\langle \epsilon_{i_5i_6}(\mathbf{x}^1)\epsilon_{\alpha_5\alpha_6}(\mathbf{y}^1)\rangle]d\mathbf{x}^1d\mathbf{y}^1,\end{aligned} \tag{33}$$

where $E^*_{i_1i_2i_3}(\mathbf{x},\mathbf{x}^1)$ is the strain Green's function tensor for the effective modulus formulation. [Equation (33) represents 81 equations on the 81 components of $\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle$. Of course, only 21 of these equations are different from the remaining, and there are only 21 distinct functions of $\mathbf{x}, \mathbf{y}$ that are defined by the components. A reference to an inversion or a partial inversion of Eq. (33) implies an inversion or partial inversion of the 21 different equations.] Finally, upon making use of Green's theorem as discussed in the Appendix, the following integral equations are obtained on the correlations of the strain field.[13]

$$\begin{aligned}&[\delta_{i_1i_3}\delta_{i_2i_4}\delta_{\alpha_1\alpha_3}\delta_{\alpha_2\alpha_4} - A^*_{i_1i_2i_5i_6}A^*_{\alpha_1\alpha_2\alpha_5\alpha_6}\Gamma_{i_5i_6i_3i_4\alpha_5\alpha_6\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})]\\ &\times\langle \epsilon_{i_3i_4}(\mathbf{x})\epsilon_{\alpha_3\alpha_4}(\mathbf{y})\rangle\\ &= \langle \epsilon_{i_1i_2}(\mathbf{x})\rangle\langle \epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle\\ &+ \int\int F^*_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)F^*_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y},\mathbf{y}^1)\Gamma_{i_3i_4i_5i_6\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{x}^1,\mathbf{y}^1)\\ &\times\langle \epsilon_{i_5i_6}(\mathbf{x}^1)\epsilon_{\alpha_5\alpha_6}(\mathbf{y}^1)\rangle\, d\mathbf{x}^1d\mathbf{y}^1,\end{aligned} \tag{34}$$

where $A^*_{i_1i_2i_3i_4}$ and $F^*_{i_1i_2i_3i_4}$ are the analog of $A_{i_1i_2i_3i_4}$ and $F_{i_1i_2i_3i_4}$ given in the Appendix, as they apply to the effective modulus formulation.

We now make use of the facts that $\Gamma_{i_3i_4i_5i_6\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{x}^1,\mathbf{y}^1)$ is nonzero only for $\mathbf{y}^1$ with a neighborhood of $\mathbf{x}^1$, of linear dimension l, and that $F^*_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y},\mathbf{y}^1)$ varies little with a change in $\mathbf{y}^1$ over a distance of order l to approximate the double integral in Eq. (34) by[14]

$$\begin{aligned}&\int F^*_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)F^*_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y},\mathbf{x}^1)[\int\Gamma_{i_3i_4i_5i_6\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{x}^1,\mathbf{y}^1)\\ &\qquad\times\langle \epsilon_{i_5i_6}(\mathbf{x}^1)\epsilon_{\alpha_5\alpha_6}(\mathbf{y}^1)\rangle\, d\mathbf{y}^1]d\mathbf{x}^1.\end{aligned}$$

As a consequence of this approximation we can solve

Eq. (34) in stages. First we solve for

$$\mathcal{E}_{i_1i_2\alpha_1\alpha_2}(\mathbf{x}) = \int \Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})\langle \epsilon_{i_3i_4}(\mathbf{x})\epsilon_{\alpha_3\alpha_4}(\mathbf{y})\rangle\, d\mathbf{y}, \tag{35}$$

and, then, for $\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle$. To obtain the equations that govern $\mathcal{E}_{i_1i_2\alpha_1\alpha_2}$ requires that we introduce the inverse of

$$[\delta_{i_1i_3}\delta_{i_2i_4}\delta_{\alpha_1\alpha_3}\delta_{\alpha_2\alpha_4} - A^*_{i_1i_2i_5i_6}A^*_{\alpha_1\alpha_2\alpha_5\alpha_6}\Gamma_{i_5i_6i_3i_4\alpha_5\alpha_6\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})],$$

which we denote by $B_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})$. We note that it will depend on difference coordinates alone for the case of homogeneous statistics and make reference to a previous comment for how it is to be interpreted. With this algebraic inverse and the above approximation of our two-fold integral, we obtain from Eq. (34)

$$\begin{aligned}\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle &= B_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})\langle \epsilon_{i_3i_4}(\mathbf{x})\rangle\langle \epsilon_{\alpha_3\alpha_4}(\mathbf{y})\rangle \\ &+ B_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})\int F^*_{i_3i_4i_5i_6}(\mathbf{x},\mathbf{x}^1) \\ &\times F^*_{\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{y},\mathbf{x}^1)\mathcal{E}_{i_5i_6\alpha_5\alpha_6}(\mathbf{x}^1)d\mathbf{x}^1. \end{aligned} \tag{36}$$

Finally, by multiplying Eq. (36) by $\Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})$ and integrating over $\mathbf{y}$ we obtain an integral equation formulation on $\mathcal{E}_{i_1i_2\alpha_1\alpha_2}(\mathbf{x})$:

$$\begin{aligned}\mathcal{E}_{i_1i_2\alpha_1\alpha_2}(\mathbf{x}) &= [\int D_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})\langle \epsilon_{\alpha_3\alpha_4}(\mathbf{y})\rangle\, d\mathbf{y}] \\ &\times \langle \epsilon_{i_3i_4}(\mathbf{x})\rangle + \int [D_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})F^*_{\alpha_3\alpha_4\alpha_5\alpha_6}(\mathbf{y},\mathbf{x}^1)d\mathbf{y}] \\ &\times F^*_{i_3i_4i_5i_6}(\mathbf{x},\mathbf{x}^1)\mathcal{E}_{i_5i_6\alpha_5\alpha_6}(\mathbf{x}^1)d\mathbf{x}^1, \end{aligned} \tag{37}$$

where

$$\begin{aligned}&D_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y}) \\ &\quad = \Gamma_{i_1i_2i_5i_6\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{x},\mathbf{y})B_{i_5i_6i_3i_4\alpha_5\alpha_6\alpha_3\alpha_4}(\mathbf{x},\mathbf{y}). \end{aligned} \tag{38}$$

For homogeneous statistics, $D_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}$ is a function of difference coordinate alone.

Equation (37) requires the solution of a system of one-fold integral equations to determine $\mathcal{E}_{i_1i_2\alpha_1\alpha_2}(\mathbf{x})$. Equation (36) is a system of integral formulas that allows the direct calculation of $\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle$ once $\mathcal{E}_{i_1i_2\alpha_1\alpha_2}(\mathbf{x})$ has been determined. In these equations, $B_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}$ and $D_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}$ are uniquely determined by the material properties of the solid. The tensor field $F^*_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)$ depends on the Green's function of the Dyson operator and, thus, depends on the boundary conditions of the problem.

The two-stage solution procedure offers two advantages over a direct attack on Eq. (34). The first advantage is that it only requires the construction of inverses of an algebraic operator and a one-fold integral operator. Equation (34) requires the inversion of a two-fold integral operator. The latter task is expected to be a good deal more complex. The second advantage is that the two-stage procedure is intuitively satisfying since it separates the two length scales. Equation (37) is defined on the L scale; Eq. (36) determines l scale variations in $\langle \epsilon_{i_1i_2}(\mathbf{x})\epsilon_{\alpha_1\alpha_2}(\mathbf{y})\rangle$.

Still, it is to be expected that obtaining solutions of Eqs. (35) and (36) will be a nontrivial task, probably requiring a numerical algorithm at some point in the analysis. We shall investigate some solutions of these equations in a future paper.

We conclude this section with two comments. The first applies to the strongly inhomogeneous solid in the presence of two length scales. For the strongly inhomogeneous solid, the analytic prescriptions given by Eqs. (29) cease to be valid. One can, however, present arguments that suggest that several important features of these prescriptions are still valid. The first feature is that the surface integrals that appear in the mass and intensity operators offer contributions only for field points that fall within a layer of the boundary surfaces. The second is that the remaining kernels $\mathcal{C}_{i_1i_2i_3i_4}(\mathbf{x},\mathbf{x}^1)$ and $\mathcal{G}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1;\mathbf{y},\mathbf{y}^1)$ are nonzero only if the pair (or quartet) of points on which they are defined fall within a neighborhood, of linear dimension l. Based on these arguments we still obtain Eq. (30) as the governing equation over most of the solid. The difference is that now we do not have useful prescriptions for $C^*_{i_1i_2i_3i_4}$ and $\Gamma_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{y})$, but, instead, view these quantities as formulation parameters to be ascertained by experimentation. This was discussed in the preceeding section.

The second comment is that the weakly inhomogeneous solid could be approached using a classical perturbation approach. That is, we could construct a Liousville–Neumann perturbation series for $u_i(\mathbf{x})$ using the ensembled averaged medium as the base solution. This series could then be truncated, and any statistical average desired could be formed using the resulting approximation. This was done by Molyneux and Beran.[15] It can be shown that the solution of the weakly inhomogeneous Dyson and Bethe–Salpeter formulisms contain selected subsets of infinite series expressions for the appropriate averages obtained by making use of the Liousville–Neumann series in calculating these averages. Thus, as a weakly inhomogeneous approximation, it can be expected to have different convergence properties than that obtained using a single term of the Liousville–Neumann series. In addition, there does not exist the possibility of extending the classical weakly inhomogeneous solution by accepting terms as formulation parameters as is present by first constructing the Dyson and Bethe–Salpeter formulisms.

SUMMARY

Using an iteration procedure termed the method of smoothing, mathematical formulisms were derived that govern the one- and two-point moments defined on the stress, strain, and displacement fields in a random linearly elastic solid. These formulisms are given by Eqs. (16)–(19) and by Eqs. (27) and (28). The kernel functions of the integral operators appearing in these equations are given in terms of infinite series. A truncation of these series is possible for the special case of a weakly inhomogeneous solid. The truncated expressions are given by Eq. (29).

In addition, we investigate the simplifications introduced in the formulisms for problems involving two distinct length scales. The weakly inhomogeneous version of Eqs. (16)–(19) reduce to an effective elastic moduli formulation. The weakly inhomogeneous

equations on the two point moments lead to Eqs. (36) and (37). It is speculated that, for a strongly inhomogeneous solid, one obtains similar equations. For strongly inhomogeneous solids, however, we do not have useful analytic prescriptions for the effective elastic moduli tensor or for the formulation parameters that appear in Eqs. (36) and (37).

ACKNOWLEDGMENTS

This work was supported by a grant, numbered DA-ARO-D-31-124-71-G-1713, of the Army Research Office, Durham, North Carolina. The author acknowledges, with gratitude, the discussions with R. Vaishnav on the work reported herein.

APPENDIX

In this appendix we carry out the manipulations required to obtain explicit expressions for the Dyson and Bethe–Salpeter equations as they apply to the elasticity example.

The operator $\langle A\rangle$ is a differential operator. Thus, its inverse is an integral operator, the kernel of which is termed the Green's function matrix. We denote the Green's function matrix by $G(\mathbf{x},\mathbf{x}^1)$ and formally write that $G(\mathbf{x},\mathbf{x}^1)$ satisfies the operator equation

$$\langle A\rangle^{+}G = \delta(\mathbf{x}-\mathbf{x}^1)I_{15}, \tag{A1}$$

where $\langle A\rangle^{+}$ denotes the adjoint to the $\langle A\rangle$ operator, $\delta(\mathbf{x})$ denotes the Dirac function, and I_{15} is the 15×15 unit matrix. It is convenient to express the Green's function matrix as

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \tag{A2}$$

The submatrix $g_{31}(\mathbf{x},\mathbf{x}^1)$ is the displacement Green's function matrix for the homogeneous, mean medium. In terms of our matrix notation, it satisfies the field equations given by

$$a_{11}\langle a_{22}\rangle a_{33}g_{31} = \delta(\mathbf{x}-\mathbf{x}^1)I_3, \tag{A3}$$

where I_3 is the 3×3 unit matrix. The elements of the $g_{31}(\mathbf{x},\mathbf{x}^1)$ matrix can be identified with the components of a second rank tensor referred to a cartesian coordinate system. We denote these tensor components by $V_{ij}(\mathbf{x},\mathbf{x}^1)$ and write the following set of equations as the tensor counterpart of Eq. (A3):

$$\langle C_{ijkl}\rangle \partial_j\partial_k V_{lm} = \delta(\mathbf{x}-\mathbf{x}^1)\delta_{im}. \tag{A4}$$

Here δ_{im} is the Kronecker delta and C_{ijkl} is the elastic moduli tensor, i.e., in tensor notation,

$$\tau_{ij} = C_{ijkl}\epsilon_{kl}. \tag{A5}$$

The relationship between the elements of the C_{ijkl} tensor and those of the a_{22} matrix are readily obtained by comparing the equations given by (A5) and those given by the matrix equation

$$\tau = a_{22}\epsilon. \tag{A6}$$

The appropriate boundary conditions required for a unique prescription of the $g_{31}(\mathbf{x},\mathbf{x}^1)$ submatrix are discussed in the main body of the paper.

The remaining submatrix elements of the Green's function matrix $G(\mathbf{x},\mathbf{x}^1)$ are given by the following formulas:

$$\begin{aligned}
g_{21}(\mathbf{x},\mathbf{x}^1) &= a_{33}g_{31}(\mathbf{x},\mathbf{x}^1),\\
g_{11}(\mathbf{x},\mathbf{x}^1) &= \langle a_{22}\rangle a_{33}g_{31}(\mathbf{x},\mathbf{x}^1),\\
g_{32}(\mathbf{x},\mathbf{x}^1) &= \int g_{31}(\mathbf{x},\mathbf{x}^2)a_{11}^{(2)}[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2,\\
g_{22}(\mathbf{x},\mathbf{x}^1) &= \int a_{33}g_{31}(\mathbf{x},\mathbf{x}^2)a_{11}^{(2)}[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2,\\
g_{12}(\mathbf{x},\mathbf{x}^1) &= \int\langle a_{22}\rangle a_{33}g_{31}(\mathbf{x},\mathbf{x}^2)a_{11}^{(2)}[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2,\\
g_{33}(\mathbf{x},\mathbf{x}^1) &= \int g_{31}(\mathbf{x},\mathbf{x}^2)a_{11}^{(2)}\langle a_{22}^{(2)}\rangle[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2,\\
g_{23}(\mathbf{x},\mathbf{x}^1) &= \int a_{33}g_{31}(\mathbf{x},\mathbf{x}^2)a_{11}^{(2)}\langle a_{22}^{(2)}\rangle[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2,\\
g_{13}(\mathbf{x},\mathbf{x}^1) &= \int\langle a_{22}\rangle a_{33}g_{31}(\mathbf{x},\mathbf{x}^1)a_{11}^{(2)}\langle a_{22}^{(2)}\rangle\\
&\quad\times[\delta(\mathbf{x}^2-\mathbf{x}^1)]d\mathbf{x}^2.
\end{aligned} \tag{A7}$$

The superscript (2) attached to an operator matrix indicates that the operation is to be applied at the point located by $\mathbf{x}^2$. It is seen for Eq. (A7) that some of the $g_{ij}(\mathbf{x},\mathbf{x}^1)$ are defined in terms of symbolic functions.

In terms of $G(\mathbf{x},\mathbf{x}^1)$, the inverse operation $\langle A\rangle^{-1}\varphi$, where φ is a 15 dimension column matrix with scalar fields as elements, is written

$$\int G(\mathbf{x},\mathbf{x}^1)\varphi(\mathbf{x}^1)d\mathbf{x}^1, \tag{A8}$$

where the integration is over the extent of the solid. We shall not use the integration symbol in what follows with the understanding that the appearance of a g_{ij} always requires an integration.

We are now ready to obtain expanded expressions for the individual terms of the mass operator matrix. Immediately we note that we may write

$$M = \begin{pmatrix} 0 & 0 & 0 \\ 0 & m_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{A9}$$

where the dimensions of m_{22} are that of g_{22}, i.e., a 6×6 matrix. We use

$$\varphi = \begin{Bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{Bmatrix} \tag{A10}$$

to denote the generic element of the space on which M is defined. The dimension of the φ_1 submatrices are 6×1, 6×1, and 3×1, respectively. Thus,

$$M\varphi = \begin{Bmatrix} 0 \\ m_{22}\varphi_2 \\ 0 \end{Bmatrix}, \tag{A11}$$

where $m_{22}\varphi_2$ is given by an infinite series. We write the first few terms as

$$\begin{aligned}
m_{22}\varphi_2 = [&\langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle\\
&-\langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^2)a'_{22}(\mathbf{x}^2)g_{22}(\mathbf{x}^2,\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle\\
&+\langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^2)a'_{22}(\mathbf{x}^2)g_{22}(\mathbf{x}^2,\mathbf{x}^3)\\
&\times a'_{22}(\mathbf{x}^3)g_{22}(\mathbf{x}^3,\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle\\
&-\langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^2)a'_{22}(\mathbf{x}^2)\rangle g_{22}(\mathbf{x}^2,\mathbf{x}^3)\\
&\times\langle a'_{22}(\mathbf{x}^3)g_{22}(\mathbf{x}^3,\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle+\cdots]\varphi_2(\mathbf{x}^1).
\end{aligned} \tag{A12}$$

The operations in this symbolic expression are to be carried out from right to left. We recall that the appearance of g_{22} implies an integration, and also note in its definition, Eq. (A7), the presence of the derivatives of the Dirac function. Thus, the above prescription for m_{22} is in the form of an infinite series of integro-differential operators of all orders.

The elements of the φ_2 matrix and the matrix given by $m_{22}\varphi_2$ can be identified with the components of two symmetric second rank tensors referred to a cartesian coordinate system. See Eq. (3). We introduce, here, $\psi_{ij}(\mathbf{x})$ to denote the components of the tensor defined by $\varphi_2(\mathbf{x})$ and present the following expression as the tensor counterpart of (A12):

$$
\begin{aligned}
&[E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^1)\partial^{(1)}_{i_8}\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}^1)\rangle \\
&- E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^2)\partial^{(2)}_{i_8}E_{i_9i_{10}i_{11}}(\mathbf{x}^2,\mathbf{x}^1)\partial^{(1)}_{i_{12}} \\
&\times \langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)C'_{i_{11}i_{12}i_3i_4}(\mathbf{x}^1)\rangle \\
&+ E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^2)\partial^{(2)}_{i_8}E_{i_9i_{10}i_{11}}(\mathbf{x}^2,\mathbf{x}^3)\partial^{(3)}_{i_{12}}E_{i_{13}i_{14}i_{15}}(\mathbf{x}^3,\mathbf{x}^1)\partial^{(1)}_{i_{16}} \\
&(\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)C'_{i_{11}i_{12}i_{13}i_{14}}(\mathbf{x}^3)C'_{i_{15}i_{16}i_3i_4}(\mathbf{x}^1)\rangle \\
&- \langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)\rangle\langle C'_{i_{11}i_{12}i_{13}i_{14}}(\mathbf{x}^3)C'_{i_{15}i_{16}i_3i_4}(\mathbf{x}^1)\rangle) \\
&+ \cdots]\psi_{i_3i_4}(\mathbf{x}^1). \qquad \text{(A13)}
\end{aligned}
$$

In this expression, $E_{ijk}(\mathbf{x},\mathbf{x}^1)$ is the tensor counterpart of the Green's function submatrix $g_{22}(\mathbf{x},\mathbf{x}^1)$. It is given in terms of the tensor $V_{ij}(\mathbf{x},\mathbf{x}^1)$ by the following formula

$$E_{ijk} = \tfrac{1}{2}(\partial_i V_{jk} + \partial_j V_{ik}). \qquad \text{(A14)}$$

Thus, $E_{ijk}(\mathbf{x},\mathbf{x}^1)$ may be termed the strain Green's function matrix for the homogeneous mean medium. Further, in (A13), a superscript attached to ∂ indicates that the derivative is to be applied at the point located by the like superscripted position vector. The series of integro-differential operations required by this prescription are to be carried out from right to left. We can obtain a more convenient form than (A13) by noting that the integro-differential operations occur in pairs as illustrated:

$$\int E_{ijk}(\mathbf{x},\mathbf{x}^1)\partial^{(1)}_l[\mathfrak{A}(\mathbf{x}^1)]d\mathbf{x}^1,$$

where $\mathfrak{A}(\mathbf{x}^1)$ denotes a general tensor function of $\mathbf{x}^1$. This suggests that we might apply Green's theorem and express this combination as

$$\oint n_2(\mathbf{x}^1_s)E_{ijk}(\mathbf{x},\mathbf{x}^1_s)\mathfrak{A}(\mathbf{x}^1_s)d\mathbf{x}^1_s - \int(\partial^{(1)}_l E_{ijk})\mathfrak{A}(\mathbf{x}^1)d\mathbf{x}^1.$$

Care must be exerted in interpreting this expression since the singularity that is introduced by $\partial^{(1)}_l E_{ijk}$ at the point located by $\mathbf{x}^1 = \mathbf{x}$ is too strong for the second integral to exist in an absolute sense. If, however, $\mathfrak{A}(\mathbf{x})$ is such as to approach a value as $\mathbf{x}^1$ approaches $\mathbf{x}$ that is independent of the direction of the approach, then, we can interpret the second integral in a "Cauchy principal value" sense. In this view the surface integral represented by $\oint$ is over the bounding surface of the solid and over a vanishingly small sphere surrounding the point located by $\mathbf{x}^1 = \mathbf{x}$. The contribution over the vanishingly small sphere can be calculated once an expression for $E_{ijk}(\mathbf{x},\mathbf{x}^1)$ has been obtained. We write

$$\mathfrak{A}(\mathbf{x})A_{ijkl} + \oint n_l(\mathbf{x}^1_s)E_{ijk}(\mathbf{x},\mathbf{x}^1_s)\mathfrak{A}(\mathbf{x}^1_s)d\mathbf{x}^1_s + \int F_{ijkl}(\mathbf{x},\mathbf{x}^1)\mathfrak{A}(\mathbf{x}^1)d\mathbf{x}^1,$$

where A_{ijkl} is obtained by the integration over the vanishingly small sphere, the second integral is over the bounding surface

$$F_{ijkl} = -\partial^{(1)}_l E_{ijk} = \partial_l E_{ijk},$$

and the last integral is interpreted in a "Cauchy principal value" sense.

Using the above for each of the integro-differential and assuming we can interchange the order of performing the integrations operator pairs, we can write, in place of (A13),

$$\oint \mathfrak{D}_{ijkl}(\mathbf{x},\mathbf{x}^1_s)\psi_{kl}(\mathbf{x}^1_s)d\mathbf{x}^1_s + \int \mathcal{G}_{ijkl}(\mathbf{x},\mathbf{x}^1)\psi_{kl}(\mathbf{x}^1)d\mathbf{x}^1, \qquad \text{(A15)}$$

where the first integration is over the bounding surface. We note that the infinite series of (A13) are collected in the two kernel functions and that the contributions such as $\mathfrak{A}(\mathbf{x})A_{ijkl}$ have been incorporated in $\mathcal{G}_{ijkl}$ by introducing the Dirac function. It is the form of the operator given by (A15) that is used in Eq. (14) to obtain the tensor Eqs. (16)–(19) on $\langle\tau_{ij}\rangle$, $\langle\epsilon_{ij}\rangle$, and $\langle u_i\rangle$. For an explicit expression for the first few terms of the infinite series definition of $\mathcal{G}_{ijkl}$ for the case of an infinite, locally, isotropic solid, the reader is referred to Ref. 2.

Turning to the Bethe–Salpeter equation, we wish to obtain an explicit expression for the intensity operator I_{xy} that is defined by Eq. (24). It is convenient to express I_{xy} as a 9×9 matrix with elements given by submatrices that have dimensions that conform with those defined by $a_{ij}(\mathbf{x}) \otimes a_{kl}(\mathbf{x})$. A complete expansion of I_{xy} would result in a 225×225 matrix of scalar operators. We immediately note that the only submatrix of I_{xy} that is not identically zero is the i_{55} matrix. This matrix contains 36×36 scalar operators. We write the following expression for the first few terms of i_{55} acting upon a generic 36-dimensional column matrix:

$$
\begin{aligned}
i_{55}\varphi_5 = \{&\langle a'_{22}(\mathbf{x}) \otimes a'_{22}(\mathbf{y})\rangle\delta(\mathbf{x}-\mathbf{x}^1)\delta(\mathbf{y}-\mathbf{y}^1) \\
&- \langle[a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^1)a'_{22}(\mathbf{x}^1)] \otimes a'_{22}(\mathbf{y})\rangle\delta(\mathbf{y}-\mathbf{y}^1) \\
&- \langle a'_{22}(\mathbf{x}) \otimes [a'_{22}(\mathbf{y})g_{22}(\mathbf{y},\mathbf{y}^1)a'_{22}(\mathbf{y}^1)]\rangle\delta(\mathbf{x}-\mathbf{x}^1) \\
&+ \langle[a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^1)a'_{22}(\mathbf{x}^1)] \\
&\otimes [a'_{22}(\mathbf{y})g_{22}(\mathbf{y},\mathbf{y}^1)a'_{22}(\mathbf{y}^1)]\rangle \\
&- \langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle \\
&\otimes \langle a'_{22}(\mathbf{y})g_{22}(\mathbf{y},\mathbf{y}^1)a'_{22}(\mathbf{y}^1)\rangle \\
&- \langle a'_{22}(\mathbf{x}) \otimes a'_{22}(\mathbf{y})\rangle[g_{22}(\mathbf{x},\mathbf{x}^1) \\
&\otimes g_{22}(\mathbf{y},\mathbf{y}^1)]\langle a'_{22}(\mathbf{x}^1) \otimes a'_{22}(\mathbf{y}^1)\rangle \\
&+ \langle[a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^2)a'_{22}(\mathbf{x}^2)g_{22}(\mathbf{x}^2,\mathbf{x}^1)a'_{22}(\mathbf{x}^1)] \\
&\otimes a'_{22}(\mathbf{y})\rangle\delta(\mathbf{y}-\mathbf{y}^1) \\
&- \langle a'_{22}(\mathbf{x}) \otimes a'_{22}(\mathbf{y}^1)\rangle([g_{22}(\mathbf{x},\mathbf{x}^2) \\
&\times \langle a'_{22}(\mathbf{x}^2)g_{22}(\mathbf{x}^2,\mathbf{x}^1)a'_{22}(\mathbf{x}^1)\rangle \otimes])\delta(\mathbf{y}-\mathbf{y}^1) \\
&- [\langle a'_{22}(\mathbf{x})g_{22}(\mathbf{x},\mathbf{x}^2)a'_{22}(\mathbf{x}^2)\rangle g_{22}(\mathbf{x}^2,\mathbf{x}^1) \\
&\otimes I]\langle a'_{22}(\mathbf{x}^1) \otimes a'_{22}(\mathbf{y})\rangle\,\delta(\mathbf{y}-\mathbf{y}^1) \\
&+ \langle a'_{22}(\mathbf{x}) \otimes [a'_{22}(\mathbf{y})g_{22}(\mathbf{y},\mathbf{y}^2)a'_{22}(\mathbf{y}^2)
\end{aligned}
$$

$$\times g_{22}(\mathbf{y}^2,\mathbf{y}^1)a'_{22}(\mathbf{y}^1)]\rangle\delta(\mathbf{x}-\mathbf{x}^1)$$
$$-\langle a'_{22}(\mathbf{x})\otimes a'_{22}(\mathbf{y})\rangle(I\otimes[g_{22}(\mathbf{y},\mathbf{y}^2)$$
$$\times\langle a'_{22}(\mathbf{y}^2)g_{22}(\mathbf{y}^2,\mathbf{y}^1)a'_{22}(\mathbf{y}^1)\rangle])\delta(\mathbf{x}-\mathbf{x}^1)$$
$$-(I\otimes[\langle a'_{22}(\mathbf{y})g_{22}(\mathbf{y},\mathbf{y}^2)a'_{22}(\mathbf{y}^2)\rangle$$
$$\times g_{22}(\mathbf{y}^2,\mathbf{y}^1)])\langle a'_{22}(\mathbf{x})\otimes a'_{22}(\mathbf{y}^1)\rangle\delta(\mathbf{x}-\mathbf{x}^1)$$
$$+o(a'^5_{22})\}\varphi_5(\mathbf{x}^1,\mathbf{y}^1). \tag{A16}$$

The operations in this symbolic expression are to be carried out from right to left. The appearance of $g_{22}(\mathbf{x},\mathbf{x}^1)$ or the Dirac function implies an integration. As was the case for the mass operator, the above prescription for the intensity operator is given by a series of integro-differential operators.

The elements of the φ_5 matrix as well as the matrix given by $i_{55}\varphi_5$ can be identified with the components of two fourth rank tensors that are obtained from the product of pairs of symmetric second rank tensors. We introduce, here, $\psi_{ij\alpha\beta}(\mathbf{x},\mathbf{y})$ to denote the components of the tensor defined by $\varphi_2(\mathbf{x},\mathbf{y})$ and present the following expression as the tensor counterpart of (A17):

$$\{C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle\delta(\mathbf{x}-\mathbf{x}^1)\delta(\mathbf{y}-\mathbf{y}^1)$$
$$-E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^1)\partial^{(1)}_{i_8}\delta(\mathbf{y}-\mathbf{y}^1)$$
$$\times\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}^1)C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$-E_{\alpha_5\alpha_6\alpha_7}(\mathbf{y},\mathbf{y}^1)\partial^{(1)}_{\alpha_8}\delta(\mathbf{x}-\mathbf{x}^1)$$
$$\times\langle C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y}^1)C'_{i_1i_2i_3i_4}(\mathbf{x})\rangle$$
$$+E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^1)\partial^{(1)}_{i_8}E_{\alpha_5\alpha_6\alpha_7}\partial^{(1)}_{\alpha_8}$$
$$[\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}^1)C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y}^1)\rangle$$
$$-\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}^1)\rangle\langle C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y}^1)\rangle$$
$$-\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})\rangle\langle C'_{i_7i_8i_3i_4}(\mathbf{x}^1)C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y}^1)\rangle]$$
$$+E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^2)\partial^{(2)}_{i_8}E_{i_9i_{10}i_{11}}(\mathbf{x}^2,\mathbf{x}^1)\partial^{(1)}_{i_{12}}\delta(\mathbf{y}-\mathbf{y}^1)$$
$$[\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)C'_{i_{11}i_{12}i_3i_4}(\mathbf{x}^1)C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$-\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle\langle C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)C'_{i_{11}i_{12}i_3i_4}(\mathbf{x}^1)\rangle$$
$$-\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_9i_{10}}(\mathbf{x}^2)\rangle\langle C'_{i_{11}i_{12}i_3i_4}(\mathbf{x}^1)C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle]$$
$$+E_{\alpha_5\alpha_6\alpha_7}(\mathbf{y},\mathbf{y}^2)\partial^{(2)}_{\alpha_8}E_{\alpha_9\alpha_{10}\alpha_{11}}(\mathbf{y}^2,\mathbf{y}^1)\partial^{(1)}_{\alpha_{12}}\delta(\mathbf{x}-\mathbf{x}^1)$$
$$[\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_9\alpha_{10}}(\mathbf{y}^2)C'_{\alpha_{11}\alpha_{12}\alpha_3\alpha_4}(\mathbf{y}^1)\rangle$$
$$-\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})\rangle\langle C_{\alpha_7\alpha_8\alpha_9\alpha_{10}}(\mathbf{y}^2)C'_{\alpha_{11}\alpha_{12}\alpha_3\alpha_4}(\mathbf{y}^1)\rangle$$
$$-\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_{11}\alpha_{12}\alpha_3\alpha_4}(\mathbf{y}^1)\rangle\langle C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_9\alpha_{10}}(\mathbf{y}^2)\rangle$$
$$+\cdots\}\psi_{i_3i_4\alpha_3\alpha_4}(\mathbf{x}^1,\mathbf{y}^1). \tag{A17}$$

The i indices go with the $\mathbf{x}$ coordinate, and the α indices go with the $\mathbf{y}$ coordinate. Thus, $\partial^{(1)}_{i_8}$, for example, is the partial derivative with respect to $x^1_{i_8}$. As in dealing with the mass operator, we note that the integro-differential operations occur in pairs suggesting the application of Green's theorem. Based on the reasoning already presented we can now write, in place of (A17),

$$\iint g_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1;\mathbf{y},\mathbf{y}^1)\psi_{i_3i_4\alpha_3\alpha_4}(\mathbf{x}^1,\mathbf{y}^1)d\mathbf{x}^1d\mathbf{y}^1$$
$$+\int\oint g^{(1)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1_s;\mathbf{y},\mathbf{y}^1)\psi_{i_3i_4\alpha_3\alpha_4}(\mathbf{x}^1_s,\mathbf{y}^1)d\mathbf{x}^1_sd\mathbf{y}^1$$
$$+\oint\int g^{(1)}_{\alpha_1\alpha_2\alpha_3\alpha_4i_1i_2i_3i_4}(\mathbf{y},\mathbf{y}^1_s;\mathbf{x},\mathbf{x}^1)\psi_{i_3i_4\alpha_3\alpha_4}(\mathbf{x}^1,\mathbf{y}^1_s)d\mathbf{x}^1d\mathbf{y}^1_s$$
$$+\oint\oint g^{(2)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1_s;\mathbf{y},\mathbf{y}^1_s)\psi_{i_3i_4\alpha_3\alpha_4}(\mathbf{x}^1_s,\mathbf{y}^1_s)d\mathbf{x}^1_sd\mathbf{y}^1_s. \tag{A18}$$

As before, $\oint$ indicates an integral over the bounding surface. We can write the following expressions for the kernel functions:

$$g_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1,\mathbf{y},\mathbf{y}^1)=\{[\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$-A_{i_5i_6i_7i_8}\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$-A_{\alpha_5\alpha_6\alpha_7\alpha_8}\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$+\ \ldots\ ]\delta(\mathbf{x}-\mathbf{x}')\delta(\mathbf{y}-\mathbf{y}')$$
$$-[\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}')C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle+\cdots]$$
$$\times F_{i_5i_6i_7i_8}(\mathbf{x},\mathbf{x}^1)\delta(\mathbf{y}-\mathbf{y}^1)$$
$$-[\langle C'_{i_1i_2i_3i_4}(\mathbf{x})C'_{\alpha_1\alpha_2\alpha_5\alpha_6}(\mathbf{y})C'_{\alpha_7\alpha_8\alpha_3\alpha_4}(\mathbf{y}^1)\rangle+\cdots]$$
$$\times F_{\alpha_5\alpha_6\alpha_7\alpha_8}(\mathbf{y},\mathbf{y}^1)\delta(\mathbf{x}-\mathbf{x}^1)+\cdots\},$$

$$g^{(1)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1_s;\mathbf{y},\mathbf{y}^1)$$
$$=-\{[\langle C'_{i_1i_2i_5i_6}(\mathbf{x})C'_{i_7i_8i_3i_4}(\mathbf{x}^1_s)C'_{\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{y})\rangle$$
$$+\cdots]n_{i_8}(\mathbf{x}^1_s)E_{i_5i_6i_7}(\mathbf{x},\mathbf{x}^1_s)\delta(\mathbf{y}-\mathbf{y}^1)+\cdots\},$$

$$g^{(2)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}(\mathbf{x},\mathbf{x}^1_s;\mathbf{y},\mathbf{y}^1_s)=\cdots. \tag{A19}$$

In these expressions, we have retained only terms up to and including third order in powers of the stochastically varying portion of the elastic moduli tensor. This is done in the interest of brevity. Incorporation of higher order terms would result in the appearance of integrals of correlation functions. We note that the lowest power of $C'_{i_1i_2i_3i_4}$ to enter the definition of $g^{(2)}_{i_1i_2i_3i_4\alpha_1\alpha_2\alpha_3\alpha_4}$ is the fourth.

By introducing the form of the mass operator given by Eq. (A9) and the 9×9 matrix, with all elements save the i_{55} element equal to zero, for the intensity operator into Eq. (22) results in a system of nine operator equations. Upon performing a sequence of operator manipulations, one can then conclude that these equations will be satisfied provided the correlation functions on the displacement field satisfies Eq. (25), and the correlation functions defined on the stress and strain fields are given by the correlation functions on the displacement field according to Eq. (26). We might note that Eq. (25) can be derived, perhaps more directly, by first obtaining a stochastic operator equation on the displacement field alone and, then, by forming the Bethe–Salpeter equation for this stochastic operator equation. If we choose to use the form of the Bethe–Salpeter equation given by Eq. (26), we need first construct the inverse of the Dyson operator. This can be done, at least symbolically, by introducing a Green's function matrix. This matrix is the complete analog of G given by Eq. (A2). In Eqs. (A3) and (A7), we simply replace $\langle a_{22}\rangle$ by $\langle a_{22}\rangle - m_{22}$. Equation (26) form of the Bethe–Salpeter equation leads directly to a system of integral equations on the correlation functions defined on the strain field. [See Eq. (34).] In addition, a set of formulas that allows the direct calculation of the remaining correlation functions are obtained, once $\langle\epsilon(\mathbf{x})\otimes\epsilon(\mathbf{y})\rangle$ has been determined.

[1] E. Hopf, J. Ratl. Mech. Anal. 1, 187 (1952).
[2] M. J. Beran and J. J. McCoy, Intern. J. Solids Struct. 6, 1035 (1970).
[3] M. J. Beran and J. J. McCoy, Intern. J. Solids Struct. 6, 1267 (1970).
[4] R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1960).
[5] V. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[6] Failure to provide a convergence proof necessitates that we view the method of smoothing as a formal mathematical procedure.
[7] J. J. McCoy, Proceedings of the First International Conference on Structural Mechanics in Reactor Technology (Berlin, 1971).
[8] J. J. McCoy, "On the Dynamic Response of Disordered Composites," J. Appl. Mech. (to be published).
[9] J. J. McCoy, J. Opt. Soc. Amer. **62**, 1, 30 (1972).
[10] This matrix operator equation represents nine scalar equations on the nine scalar fields in $\langle u(\mathbf{x}) \otimes u(\mathbf{y}) \rangle$. We note, however, that only six of the equations and six of the scalar fields are distinct.
[11] It might be expected that a rigorous justification of the derivation procedure will require the introduction of some restrictions.
[12] It is to be noted that any restriction that we place on the kernel functions must conform with the infinite series prescriptions of these functions if the phenomenological theory is to represent a valid restricted theory for the original stochastic model. It is to be expected that a check of this conformity will rarely be possible.
[13] The form of Eq. (34) is not dependent on the presence of two length scales provided we correctly interpret $A^*_{i_1i_2i_3i_4}$ and $F^*_{i_1i_2i_3i_4}$.
[14] The strong singularities of $F^*_{i_1i_2i_3i_4}(\mathbf{x}, \mathbf{x}^1)$ for $\mathbf{x} = \mathbf{x}^1$ leads us to suspect that there might be some convergence difficulties for this approximate form for $\mathbf{x} = \mathbf{y}$.
[15] J. E. Molyneux and M. J. Beran, J. Math. Mech. **14**, 3, 337 (1965).

Spin-$\frac{1}{2}$ Lattice System: Group Structure and Duality Relation

D. Merlini and C. Gruber
Laboratoire de Physique Théorique, Ecole Polytechnique Fédérale, Lausanne, Switzerland

(Received 16 May 1972)

Properties of general lattice systems are investigated using the group properties of such systems. The duality transformation is generalized and is expressed in terms of exact short sequence between the groups associated with one lattice and the groups associated with the dual lattice. A general method is given for constructing explicitly a dual lattice for any lattice system and is illustrated with some two- and three-dimensional models. In particular it is shown that the two-dimensional triangular lattice with three-body forces is self-dual.

1. INTRODUCTION

In a recent article F. Wegner[1] has established a duality relation for generalized Ising models. In particular he has shown that there exist dual Ising models to any generalized Ising models with positive interactions. However, his claim is based on an existence theorem—existence of solutions for a system of linear equations—and does not provide a simple method for explicitly constructing these duals.

In this article, we shall present another derivation of such a duality relation, which is not restricted to positive interactions and which gives a very simple method for constructing dual lattices. Moreover, the definition of duality which we shall introduce generalizes the definition usually adopted.

The duality relation is a relation between the "high temperature expansion" of the partition function (and correlation functions) and the "low temperature expansion" of some other system, called the dual of the first system. This duality reflects some symmetry of the system under consideration, and it was, therefore, expected that such a relation could be derived using only the group structure which can be associated with spin-$\frac{1}{2}$ lattice systems.[2] The purpose of this article is to show that general properties of lattice systems (spin-$\frac{1}{2}$), as well as the duality relation, can be discussed using only the properties of the groups defined on the lattice. The interest of such an approach, beside its simplicity, lies in its possible generalization to systems with higher spin.[3]

In Sec. 2 we shall discuss the group structure associated with any lattice system (spin-$\frac{1}{2}$) and derive from it some general properties of the system (Sec. 2B). This group formalism will then be used to express the high and low temperature expansion of the partition function in such a way that the definition of duality emerges very naturally. In particular, if we consider the usual definition of duality, then the duality transformation can be expressed as the existence of an exact short sequence (Sec. 2D). Those group properties will then give us a general method to construct dual lattices (Sec. 3), and we shall illustrate this technique with some simple models (Secs. 4–6). Let us mention that all known duality transformations can be obtained very simply using the general method of Sec. 3.

In particular, we shall show that the two-dimensional triangular lattice with three-body nearest neighbor interactions is self-dual, property which can be used to discuss the location of the zeros of the partition function and phase transition.[4] This model is very interesting since it is the only self-dual system, other than the Ising model, in two-dimensions and without external field.

2. GROUP STRUCTURE ON LATTICE SYSTEMS (SPIN-$\frac{1}{2}$)

The duality relation for general lattice systems will be derived using the group structure which can be associated with the systems; such a structure was introduced by Ginibre[2] in his derivation of Griffith's inequalities.

In this section we shall first recall some of the properties of the group associated with any finite set of points (Sec. 2A); we shall then describe the model and the general framework (Sec. 2B) from which we shall derive the duality transformation (Sec. 2C). Finally, we shall summarize the connection between the duality transformation and the group structure of the system (Sec. 2D).

Although some of the results which we shall derive—like the symmetry group of the system—have already been obtained by Wegner,[1] it is interesting to derive them using only the group properties of the system; it is indeed expected that a similar group structure can be introduced to study systems with higher spins,[3] and therefore the results obtained could be generalized to these models.

A. Group Associated with Set

With any finite set $\mathfrak{A}$ of order $|\mathfrak{A}|$, we can associate a group $\mathcal{P}(\mathfrak{A})$ defined by

[1] E. Hopf, J. Ratl. Mech. Anal. 1, 187 (1952).
[2] M. J. Beran and J. J. McCoy, Intern. J. Solids Struct. 6, 1035 (1970).
[3] M. J. Beran and J. J. McCoy, Intern. J. Solids Struct. 6, 1267 (1970).
[4] R. Bellman, *Introduction to Matrix Analysis* (McGraw-Hill, New York, 1960).
[5] V. Frisch, in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Bharucha-Reid (Academic, New York, 1968).
[6] Failure to provide a convergence proof necessitates that we view the method of smoothing as a formal mathematical procedure.
[7] J. J. McCoy, Proceedings of the First International Conference on Structural Mechanics in Reactor Technology (Berlin, 1971).
[8] J. J. McCoy, "On the Dynamic Response of Disordered Composites," J. Appl. Mech. (to be published).
[9] J. J. McCoy, J. Opt. Soc. Amer. **62**, 1, 30 (1972).
[10] This matrix operator equation represents nine scalar equations on the nine scalar fields in $\langle u(\mathbf{x}) \otimes u(\mathbf{y}) \rangle$. We note, however, that only six of the equations and six of the scalar fields are distinct.
[11] It might be expected that a rigorous justification of the derivation procedure will require the introduction of some restrictions.
[12] It is to be noted that any restriction that we place on the kernel functions must conform with the infinite series prescriptions of these functions if the phenomenological theory is to represent a valid restricted theory for the original stochastic model. It is to be expected that a check of this conformity will rarely be possible.
[13] The form of Eq. (34) is not dependent on the presence of two length scales provided we correctly interpret $A^*_{i_1i_2i_3i_4}$ and $F^*_{i_1i_2i_3i_4}$.
[14] The strong singularities of $F^*_{i_1i_2i_3i_4}(\mathbf{x}, \mathbf{x}^1)$ for $\mathbf{x} = \mathbf{x}^1$ leads us to suspect that there might be some convergence difficulties for this approximate form for $\mathbf{x} = \mathbf{y}$.
[15] J. E. Molyneux and M. J. Beran, J. Math. Mech. **14**, 3, 337 (1965).

Spin-$\frac{1}{2}$ Lattice System: Group Structure and Duality Relation

D. Merlini and C. Gruber
Laboratoire de Physique Théorique, Ecole Polytechnique Fédérale, Lausanne, Switzerland

(Received 16 May 1972)

Properties of general lattice systems are investigated using the group properties of such systems. The duality transformation is generalized and is expressed in terms of exact short sequence between the groups associated with one lattice and the groups associated with the dual lattice. A general method is given for constructing explicitly a dual lattice for any lattice system and is illustrated with some two- and three-dimensional models. In particular it is shown that the two-dimensional triangular lattice with three-body forces is self-dual.

1. INTRODUCTION

In a recent article F. Wegner[1] has established a duality relation for generalized Ising models. In particular he has shown that there exist dual Ising models to any generalized Ising models with positive interactions. However, his claim is based on an existence theorem—existence of solutions for a system of linear equations—and does not provide a simple method for explicitly constructing these duals.

In this article, we shall present another derivation of such a duality relation, which is not restricted to positive interactions and which gives a very simple method for constructing dual lattices. Moreover, the definition of duality which we shall introduce generalizes the definition usually adopted.

The duality relation is a relation between the "high temperature expansion" of the partition function (and correlation functions) and the "low temperature expansion" of some other system, called the dual of the first system. This duality reflects some symmetry of the system under consideration, and it was, therefore, expected that such a relation could be derived using only the group structure which can be associated with spin-$\frac{1}{2}$ lattice systems.[2] The purpose of this article is to show that general properties of lattice systems (spin-$\frac{1}{2}$), as well as the duality relation, can be discussed using only the properties of the groups defined on the lattice. The interest of such an approach, beside its simplicity, lies in its possible generalization to systems with higher spin.[3]

In Sec. 2 we shall discuss the group structure associated with any lattice system (spin-$\frac{1}{2}$) and derive from it some general properties of the system (Sec. 2B). This group formalism will then be used to express the high and low temperature expansion of the partition function in such a way that the definition of duality emerges very naturally. In particular, if we consider the usual definition of duality, then the duality transformation can be expressed as the existence of an exact short sequence (Sec. 2D). Those group properties will then give us a general method to construct dual lattices (Sec. 3), and we shall illustrate this technique with some simple models (Secs. 4–6). Let us mention that all known duality transformations can be obtained very simply using the general method of Sec. 3.

In particular, we shall show that the two-dimensional triangular lattice with three-body nearest neighbor interactions is self-dual, property which can be used to discuss the location of the zeros of the partition function and phase transition.[4] This model is very interesting since it is the only self-dual system, other than the Ising model, in two-dimensions and without external field.

2. GROUP STRUCTURE ON LATTICE SYSTEMS (SPIN-$\frac{1}{2}$)

The duality relation for general lattice systems will be derived using the group structure which can be associated with the systems; such a structure was introduced by Ginibre[2] in his derivation of Griffith's inequalities.

In this section we shall first recall some of the properties of the group associated with any finite set of points (Sec. 2A); we shall then describe the model and the general framework (Sec. 2B) from which we shall derive the duality transformation (Sec. 2C). Finally, we shall summarize the connection between the duality transformation and the group structure of the system (Sec. 2D).

Although some of the results which we shall derive—like the symmetry group of the system—have already been obtained by Wegner,[1] it is interesting to derive them using only the group properties of the system; it is indeed expected that a similar group structure can be introduced to study systems with higher spins,[3] and therefore the results obtained could be generalized to these models.

A. Group Associated with Set

With any finite set $\mathfrak{A}$ of order $|\mathfrak{A}|$, we can associate a group $\mathcal{P}(\mathfrak{A})$ defined by

$$\mathcal{P}(\mathfrak{A}) = \{A/A \subset \mathfrak{A}\},\ A_1A_2 = A_1 \cup A_2 - A_1 \cap A_2, \quad (2.1)$$

i.e., $\mathcal{P}(\mathfrak{A})$ is the set of subsets of $\mathfrak{A}$ together with the product defined by the symmetric difference.

In the following we shall denote by a capital letter $A, B, \cdots$ the subsets of $\mathfrak{A}$ and by $|A|$ the order of A.

This group is Abelian, of order $2^{|\mathfrak{A}|}$, and is generated by $|\mathfrak{A}|$ elements which are the subsets containing one point, i.e., the points of $\mathfrak{A}$ are the generators of $\mathcal{P}(\mathfrak{A})$.

Moreover, any element A is of order 2, and any subgroup of $\mathcal{P}(\mathfrak{A})$ is of order 2^N (N = nonnegative integer).

With any $A \in \mathcal{P}(\mathfrak{A})$, we associate the function σ_A defined on $\mathcal{P}(\mathfrak{A})$ by

$$\sigma_A(R) = (-1)^{|A \cap R|} \quad \forall\ R \in \mathcal{P}(\mathfrak{A}). \quad (2.2)$$

This function has the following properties:

$$\sigma_A(R) = \sigma_R(A), \quad (2.3)$$

$$\sigma_A(R)\sigma_A(S) = \sigma_A(R \cdot S), \quad (2.4)$$

$$\sigma_A(R)\sigma_B(R) = \sigma_{A \cdot B}(R). \quad (2.5)$$

Therefore, the functions σ_A are the characters of the group $\mathcal{P}(\mathfrak{A})$, and the mapping $A \to \sigma_A$ is the isomorphism between $\mathcal{P}(\mathfrak{A})$ and its character group $\widehat{\mathcal{P}(\mathfrak{A})}$.

Lemma 1: Let $\mathfrak{B}$ be a subset of $\mathcal{P}(\mathfrak{A})$ and

$$\mathcal{S} = \{S | \sigma_S(B) = 1\ \forall\ B \in \mathfrak{B}\}.$$

Then $\mathcal{S}$ is a subgroup of $\mathcal{P}(\mathfrak{A})$, which is isomorphic to $\mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}$, where $\overline{\mathfrak{B}}$ is the subgroup generated by the elements of $\mathfrak{B}$, i.e.,

$$\mathcal{S} \cong \mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}.$$

Moreover, $\overline{\mathfrak{B}} \equiv \{A | \sigma_A(S) = 1,\ \forall\ S \in \mathcal{S}\}$.

Proof[5]: (i) $\mathcal{S}$ is a subgroup of $\mathcal{P}(\Lambda)$ (trivial).

(ii) $\forall\ \overline{B} \in \overline{\mathfrak{B}}$ (= subgroup generated by $\mathfrak{B}$), $\sigma_S(\overline{B}) = 1$ $\forall\ S \in \mathcal{S}$ [follows from Eq. (2.4)]

(iii) With $(\widehat{\mathcal{P}(\mathfrak{A})/\overline{B}})$ the character group of $\mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}$, we consider the mapping

$$\underset{\cup}{S} \to (\widehat{\mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}})$$

$$S \twoheadrightarrow \chi_S,$$

where

$$\chi_S : \mathcal{P}(\underset{\cup}{\mathfrak{A}})/\overline{\mathfrak{B}} \to \underset{\cup}{\mathbb{R}}$$

$$[R] \qquad \sigma_S(R).$$

This mapping is well defined since $R' \in [R]$ implies

$$\sigma_S(R') = \sigma_S(R\overline{B}) = \sigma_S(R)\sigma_S(\overline{B}) = \sigma_S(R).$$

As one can checked easily this mapping is an isomorphism.

To conclude the proof it is sufficient to notice that $(\widehat{\mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}})$ is isomorphic to $\mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}$.

Therefore, $\mathcal{S} \cong \mathcal{P}(\mathfrak{A})/\overline{\mathfrak{B}}$.

(iv) Let $\mathcal{Y} = \{A | \sigma_A(S) = 1\ \forall\ S \in \mathcal{S}\}$, then $\mathcal{Y} \cong \mathcal{P}(\mathfrak{A})/\mathcal{S}$. Moreover since $\mathcal{Y} \supset \overline{\mathfrak{B}}$ and $|\mathcal{Y}| = |\overline{\mathfrak{B}}|$, we have $\mathcal{Y} = \overline{\mathfrak{B}}$.

B. General Framework and Properties of Lattice Systems

Let Λ be a finite set of points called sites and denoted by small letters $x, y, \cdots$; with each site $x \in \Lambda$ is associated a spin variable σ_x taking the values ± 1. A configuration of the system is then uniquely defined by the subset $X \subset \Lambda$ consisting of those sites x where $\sigma_x = -1$; therefore the configuration space is $\mathcal{P}(\Lambda)$.

The Hamiltonian of the system is then defined by

$$H = -\sum_{R \subset \Lambda} J(R)\sigma_R, \quad (2.6)$$

where $J(R)$ is a function on nonempty subset $R \subset \Lambda$ describing the many-body interactions, and σ_R was defined by Eq. (2.2).

To simplify the notation, we introduce

$$K(R) = \beta J(R), \quad \text{where } \beta = 1/kT. \quad (2.7)$$

A general lattice system is therefore defined by $\{\Lambda, K\}$, where Λ is a finite set of points describing the sites and K is a function on $\mathcal{P}(\Lambda)$ describing the interactions.

We introduce two subgroups of $\mathcal{P}(\Lambda)$ which will be the main concepts in the duality transformation.

Let $\mathfrak{B} \subset \mathcal{P}(\Lambda)$ be the support of the function $K(R)$:

$$\mathfrak{B} = \{B | K(B) \neq 0\}. \quad (2.8)$$

Since $\phi \notin \mathfrak{B}$, $\mathfrak{B}$ is a subset of $\mathcal{P}(\Lambda)$ and not a subgroup.

1. *Interaction Subgroup* $\overline{\mathfrak{B}}$

Let $\overline{\mathfrak{B}}$ be the subgroup of $\mathcal{P}(\Lambda)$ generated by $\mathfrak{B}$; it is a subgroup of order 2^{Ni} where Ni is the number of generators of $\overline{\mathfrak{B}}$.

We then have

$$\mathcal{P}(\Lambda) = \overset{p}{\underset{j=1}{\oplus}} R_j\overline{\mathfrak{B}}, \quad p = 2^{|\Lambda|-Ni}, \quad (2.9)$$

$$\mathcal{R} = \{R_j\} \cong \mathcal{P}(\Lambda)/\overline{\mathfrak{B}}. \quad (2.10)$$

2. *Symmetry Subgroup* $\mathcal{S}$

Let

$$\mathcal{S} = \{S | \sigma_S(B) = 1\ \forall\ B \in \mathfrak{B}\}. \quad (2.11)$$

$\mathcal{S}$ is a subgroup of $\mathcal{P}(\Lambda)$ of order 2^{N_S}, where N_S is the number of generators of $\mathcal{S}$. We have

$$\mathcal{P}(\Lambda) = \overset{q}{\underset{j=1}{\oplus}} T_j\mathcal{S}, \quad q = 2^{|\Lambda|-N_S}, \quad (2.12)$$

$$\mathcal{T} = \{T_j\} \cong \mathcal{P}(\Lambda)/\mathcal{S}. \quad (2.13)$$

Properties:

(1) $\overline{\mathfrak{B}} \equiv \{X | \sigma_x(S) = 1\ \forall\ S \in \mathcal{S}\}$. (2.14)

(2) $\mathcal{S} \cong \mathcal{P}(\Lambda)/\overline{\mathfrak{B}} \cong \mathcal{R}$, $\overline{\mathfrak{B}} \cong \mathcal{P}(\Lambda)/\mathcal{S} \cong \mathcal{T}$. (2.15)

(3) $N_S = |\Lambda| - N_i$. (2.16)

(4) If $K(B) > 0$ for all $B \in \mathcal{B}$, then any element $S \in \mathcal{S}$ defines a ground state: The ground state is 2^{N_S} degenerate.

Properties (1) and (2) are just those expressed by Lemma 1; Property (3) follows from Eq. (2.15), while the last property is a consequence of the definition of $\mathcal{S}$.

3. Unitary Group of Symmetry U

With all element $S \in \mathcal{S}$ we associate the unitary operator U_S defined on the configuration space by

$$U_S = \prod_{x \in S} U_x,$$

where

$$U_x \sigma_y U_x = \begin{cases} -\sigma_y & \text{if } y = x \\ \sigma_y & \text{if } y \neq x \end{cases}.$$

Properties:

(5) $U_S \sigma_x U_S^{-1} = \sigma_x(S)\sigma_x \quad \forall\, X \in \mathcal{P}(\Lambda).$ (2.17)

(6) $\{U_S\}_{S \in \mathcal{S}}$ is a symmetry group of the system

$$[U_S, H]_- = 0 \quad \forall\, S \in \mathcal{S}. \tag{2.18}$$

(7) With $\langle \sigma_y \rangle = \mathrm{Tr}(e^{-\beta H}\sigma y)/\mathrm{Tr}e^{-\beta H}$,

then for any finite system

$$\langle \sigma_y \rangle = 0 \quad \text{if} \quad y \notin \overline{\mathcal{B}}.$$

Proof:

(5) $U_S \sigma_x U_S^{-1} = (-1)^{|x \cap S|}\sigma_x = \sigma_x(S)\sigma_x,$

$\therefore U_S \sigma_x U_S^{-1} = \sigma_x(S)\sigma_x.$

(6) $U_S \sigma_B U_S^{-1} = \sigma_B$ (from the definition of $\mathcal{S}$),

$U_S H U_S^{-1} = H$ (from the definition of $\mathcal{B}$).

(7) If $y \notin \overline{\mathcal{B}}$, Eq. (2.14) implies that there exists $S_0 \in \mathcal{S}$ such that $\sigma y(S_0) = -1$,

$\therefore U_{S_0} \sigma_y U_{S_0}^{-1} = -\sigma_y$

and $[U_{S_0}, H]_- = 0$ yields $\langle \sigma_y \rangle = -\langle \sigma_y \rangle = 0$.

4. Character Table of $\mathcal{P}(\Lambda)$

The properties derived above are simply expressed with the help of the character table of the group $\mathcal{P}(\Lambda)$.

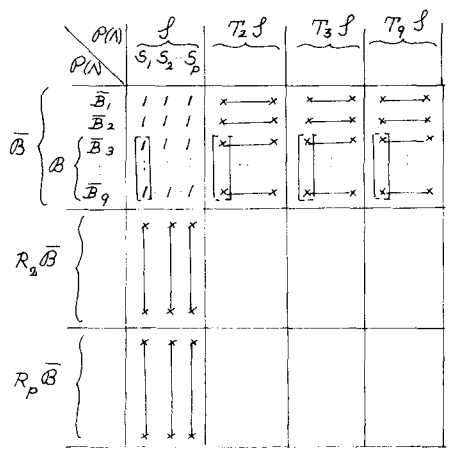


FIG. 1. Character table of $\mathcal{P}(\Lambda)$.

$g = 2^{N_i}$, $N_i = Nb$ of generators of $\overline{\mathcal{B}}$
$= Nb$ of generators of $\mathcal{P}(\Lambda)/\mathcal{S}$.
$p = 2^{N_S}$, $N_S = Nb$ of generators of $\mathcal{S}$
$= Nb$ of generators of $\mathcal{P}(\Lambda)/\overline{\mathcal{B}}$.
$|\Lambda| = N_i + N_S$, $\mathcal{T} = \{T_j\} \cong \mathcal{P}(\Lambda)/\mathcal{S} \cong \overline{\mathcal{B}}$, $\mathcal{R} = \{R_j\} \cong \mathcal{P}(\Lambda)/\overline{\mathcal{B}} \cong \mathcal{S}$.

Using the isomorphism $A \to \sigma_A$ and the relation

for any $\overline{B} \in \overline{\mathcal{B}}$, $\quad \sigma_{\overline{B}}(T_i S) = \sigma_{\overline{B}}(T_i) \quad \forall\, S \in \mathcal{S}$,

for any $S \in \mathcal{S}$, $\quad \sigma_{R_j \overline{B}}(S) = \sigma_{R_j}(S) \quad \forall\, \overline{B} \in \overline{\mathcal{B}}$.

The character table has then the form given in Fig. 1.

C. Duality Relation for the Partition Function

The partition function $Z\{\Lambda, K\}$ being defined by

$$Z\{\Lambda, K\} = \sum_{X \subset \Lambda} \exp \sum_{B \in \mathcal{B}} K(B)\sigma_B(X), \tag{2.19}$$

one obtains the *High Temperature Expansion*[2]:

$$Z\{\Lambda, K\} = 2^{|\Lambda|} \prod_{B \in \mathcal{B}} \cosh K(B) \times \sum_{\substack{\beta = (B_1 \ldots B_n) \subset \mathcal{B} \\ \Pi B_i = \phi}} \prod_{B \in \mathcal{B}} \tanh K(B). \tag{2.20}$$

On the other hand, we can obtain a *low temperature expansion* from

$$\begin{aligned} Z\{\Lambda, K\} &= \sum_{X \subset \Lambda} \prod_{B \in \mathcal{B}} \exp K(B)[\sigma_B(X) - 1 + 1] \\ &= \prod_{B \in \mathcal{B}} e^{K(B)} \sum_{X \subset \Lambda} \prod_{B \in \mathcal{B}} \times \exp - 2K(B)\tfrac{1}{2}[1 - \sigma_B(X)]. \\ &= \prod_{B \in \mathcal{B}} e^{K(B)} \sum_{T \in \mathcal{T}} 2^{N_S} \prod_{B \in \mathcal{B}} \times \exp - 2K(B)\tfrac{1}{2}[1 - \sigma_B(T)]. \end{aligned}$$

Let

$$\gamma(T) = \{B \mid \sigma_T(B) = -1\} \in \mathcal{P}(\mathcal{B}). \tag{2.21}$$

We thus obtain the *low temperature expansion*

$$Z\{\Lambda, K\} = 2^{N_S} \prod_{B \in \mathcal{B}} e^{K(B)} \sum_{T \in \mathcal{T}} \prod_{B \in \gamma(T)} e^{-2K(B)}. \tag{2.22}$$

A duality relation is a relation between the high temperature expansion of Z for a general lattice system (Λ, K) and the low temperature expansion of Z for another general lattice system (Λ^*, K^*) which will be called a dual lattice of (Λ, K).

By inspection of Eqs. (2.20) and (2.22), we are led to establish the following lemma.

Lemma 2: The mapping

$$\begin{array}{ccc} \gamma : \mathcal{P}(\Lambda) & \to & \mathcal{P}(\mathcal{B}) \\ \cup & & \cup \\ X & \nrightarrow & \gamma(X) = \{B \mid \sigma_X(B) = -1\} \end{array}$$

is an homomorphism whose kernel is the symmetry subgroup $\mathcal{S}$ and thus $\gamma(\tau) = \gamma(\mathcal{P}(\Lambda))$.

Proof: (i) homomorphism: $\gamma(X_1 X_2) = \{B \mid \sigma_{X_1 \dot{X}_2}(B) = -1\}$,

but $\sigma_{X_1 X_2}(B) = \sigma_{X_1}(B)\sigma_{X_2}(B) \Rightarrow B \in \gamma(X_1) \cup \gamma(X_2) - \gamma(X_1) \cap \gamma(X_2)$

$\therefore \gamma(X_1 X_2) = \gamma(X_1) \cdot \gamma(X_2)$ [product in $\mathcal{P}(\mathcal{B})$].

(ii) kernel of the mapping γ:

$$\ker\gamma = \{S \mid \sigma_B(S) = +1 \ \forall\, B \in \mathcal{B}\} = \mathcal{S}$$

since

$$\tau = \mathcal{P}(\Lambda)/\mathcal{S},$$
$$\gamma(\tau) = \gamma(\mathcal{P}(\Lambda)).$$

Corollary 1: The mapping γ defines an injective homomorphism of the subgroup $\tau \subset \mathcal{P}(\Lambda)$ into $\mathcal{P}(\mathfrak{B})$ and $|\tau| = |\gamma(\tau)|$.

Lemma 3: The mapping π

$$\begin{array}{ccc} \mathcal{P}(\mathfrak{B}) & \to \overline{\mathfrak{B}} \subset \mathcal{P}(\Lambda) \\ \cup & \cup \\ \beta = (B_1 \ldots B_n) & \twoheadrightarrow \overline{B} = \prod_i B_i \end{array}$$

is surjective homomorphism of the group $\mathcal{P}(\mathfrak{B})$ into the subgroup $\overline{\mathfrak{B}} \subset \mathcal{P}(\Lambda)$.

Proof: (i) surjective: by definition.

(ii) homomorphism: $\beta_1\beta_2 = \cup B_i - \cap B_i \twoheadrightarrow \overline{B}_{12} = \overline{B}_1 \cdot \overline{B}_2$ since $B_i^2 = \phi$.

Corollary 2: The *kernel* $\mathcal{K}$ of $\mathcal{P}(\mathfrak{B}) \to \overline{\mathfrak{B}}$ is a subgroup of $\mathcal{P}(\mathfrak{B})$ of order $2^{|\mathfrak{B}|N_i}$.

Using these mappings, we obtain the *high temperature expansion*

$$Z\{\Lambda, K\} = 2^{|\Lambda|} \prod_{B\in\mathfrak{B}} \cosh K(B) \sum_{\beta\in\mathcal{K}} \prod_{B\in\beta} \tanh K(B) \quad (2.23)$$

and the *low temperature expansion*

$$Z\{\Lambda, K\} = 2^{N_S} \prod_{B\in\mathfrak{B}} e^{K(B)} \sum_{\beta\in\gamma(\tau)} \prod_{B\in\beta} e^{-2K(B)}, \quad (2.24)$$

where $\mathcal{K}$ is the subgroup of $\mathcal{P}(\mathfrak{B})$ of order $2^{|\mathfrak{B}|-N_i}$ kernel of π and $\gamma(\tau)$ is the subgroup of $\mathcal{P}(\mathfrak{B})$ of order $2^{|\Lambda|-N_i} = 2^{N_i}$ image of γ:

$$\gamma(\tau) \cong \tau \cong \mathcal{P}(\Lambda)/\mathcal{S} \cong \overline{\mathfrak{B}}$$

From Eqs. (2.23) and (2.24), it is clear that we shall obtain a duality relation between two lattice systems $\{\Lambda, K\}$ and $\{\Lambda^* K^*\}$, if we can find a mapping between $\mathfrak{B} \to \mathfrak{B}^*$ such that the induced mapping $\mathcal{P}(\mathfrak{B}) \to \mathcal{P}(\mathfrak{B}^*)$ is a bijection between $\mathcal{K}$ and $\gamma^*(\tau^*)$.

The usual definition of duality assumes the existence of a bijection between $\mathfrak{B}$ and $\mathfrak{B}^*$ such that $\mathcal{K}$ is mapped onto $\gamma^*(\tau^*)$. However, this definition is not sufficiently general since it is possible to find dual transformation such that $\mathfrak{B} \nrightarrow \mathfrak{B}^*$ is not a bijection (Sec. 5C and Sec. 6).

Therefore before we define the duality transformation, we shall establish one more result.

Lemma 4: Let (Λ, K) and (Λ^*, K^*) be two lattice systems and $\mathcal{S}$ be a subgroup of $\mathcal{P}(\mathfrak{B})$. If there exists a surjective mapping

$$\begin{array}{ccc} d: \mathfrak{B} & \to \mathfrak{B}^* \\ \cup & \cup \\ B & \twoheadrightarrow B^* = dB \end{array}$$

such that the induced mapping

$$\begin{array}{ccc} D: \mathcal{P}(\mathfrak{B}) & \to \mathcal{P}(\mathfrak{B}^*) \\ \cup & \cup \\ \beta = \{\beta\} & \twoheadrightarrow \beta^* = \{B^*\} \end{array}$$

satisfies the condition Eq. (2.25) below, then D restricted to the subgroup $\mathcal{S}$ is an injective group homomorphism.

Condition: With $\beta_1, \ldots \beta_n$ a subset of $\mathcal{S}$ generating $\mathcal{S}$, then $\beta_i \equiv D^{-1}(D\beta_i)$, where $D^{-1}\beta^* = \{B \mid dB \in \beta^*\}$, i.e., $B \in \beta_i$ implies that β_i contains the whole fiber over B. (2.25)

Proof:

Let $\beta_i^* = D\beta_i = \{B_{ij}^*\}_{j=1\ldots n_i}$.

For all $\beta \in \mathcal{S}$, we have: $\beta = \prod_{i\in I} \beta_i$, where $I \subset \{1, \ldots, n\}$.

Equation (2.25) implies that

$$\beta = \prod_{i\in I} \{d^{-1}B_{ij}^*\}_{j=1\ldots n_i} = \{d^{-1}B_\alpha^*\},$$

where B_α^* are those B_{ij}^* which appear an odd number of times in $\{B_{ij}^*\}_{\substack{i\in I \\ j=1\cdots n_i}}$.

Moreover,

$$\beta^* = D\beta = \{B_\alpha^*\} \equiv \prod_{i\in I} \{B_{ij}\}_{j=1\cdots n_i},$$

therefore,

$$D(\prod_{i\in I} \beta_i) = \prod_{i\in I} (D\beta_i)$$

and thus $D \mid \mathcal{S}$ is a group homomorphism. It is injective since from Eq. (2.25) follows that

$$\beta_i \neq \beta_j \quad \text{implies} \quad \beta_i^* \neq \beta_j^*.$$

Corollary 3: The induced mapping D: $\mathcal{P}(\mathfrak{B}) \to \mathcal{P}(\mathfrak{B}^*)$ is a group isomorphism if and only if the mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ is a bijection.

Definition 1: With $\{\Lambda, K\}$ a general lattice system, then $\{\Lambda^*, K^*\}$ will be called a *dual lattice for* $\{\Lambda, K\}$ if there exists a surjective mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ such that the induced mapping D: $\mathcal{P}(\mathfrak{B}) \to \mathcal{P}(\mathfrak{B}^*)$ satisfies the following conditions:

(i) The subgroup $\mathcal{K} \subset \mathcal{P}(\mathfrak{B})$ satisfies the condition Eq. (2.25) above.

(ii) The image of $\mathcal{K}$ is $\gamma^*(\tau^*)$, i.e., $D/\mathcal{K}$ is a group isomorphism of $\mathcal{K}$ onto $\gamma^*(\tau^*)$.

(iii) $$e^{-2K^*(B^*)} = \prod_{B\in d^{-1}B^*} \tanh K(B). \quad (2.26)$$

The mapping $\{\Lambda, K\} \to \{\Lambda^*, K^*\}$ will be called a *duality transformation*. In general, the duality transformation is not unique; moreover, in most cases the mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ will not be injective.

Theorem 1:

If $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$, then

$$|\mathfrak{B}| - N_i = N_i^*. \quad (2.27)$$

Proof: This result follows immediately from the definition of duality and the properties

$$|\gamma^*(\tau^*)| = 2^{N_i}, \qquad |\mathcal{K}| = 2^{|\mathfrak{B}|-N_i}.$$

Let us derive the relation between the partition functions of a system $\{\Lambda, K\}$ and its dual $\{\Lambda^*, K^*\}$.

If $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$, then

$$\sum_{\beta\in\mathcal{K}} \prod_{B\in\beta} \tanh K(B) = \sum_{\beta^*\in\gamma^*(\ ^*)} \prod_{B^*\in\beta^*} \Big(\prod_{B\in d^{-1}B^*} \tanh K(B)\Big)$$
$$= \sum_{\beta^*\in\gamma^*(\tau^*)} \prod_{B^*\in\beta^*} e^{-2K^*(B^*)}.$$

We obtain therefore

$$Z\{\Lambda, K\} = 2^{|\Lambda|} \prod_{B\in\mathcal{B}} \cosh K(B) 2^{-N_s^*} \times \prod_{B^*\in\mathcal{B}^*} e^{-K^*(B^*)} Z\{\Lambda^*, K^*\}.$$

Via Eqs. (2.16), (2.26), and (2.27), we then have

Theorem 2: If $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$, then

$$Z\{\Lambda, K\} = 2^{(|\Lambda|+N_S)/2} 2^{-(|\Lambda|^*+N_S)/2} \times \prod_{B\in\mathcal{B}} \sqrt{\sinh 2K(B)}\, Z\{\Lambda^*, K^*\}. \quad (2.28)$$

Corollary: If $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$ and if the mapping d: $\mathcal{B} \to \mathcal{B}^*$ is a bijection, then

$$Z\{\Lambda, K\} = \frac{2^{(|\Lambda|+N_S)/2} \prod_{B\in\mathcal{B}} \sqrt{\cosh 2K(B)}}{2^{(|\Lambda^*|+N_S^*)/2} \prod_{B^*\in\mathcal{B}^*} \sqrt{\cosh 2K^*(B^*)}} Z\{\Lambda^*, K^*\}, \quad (2.29)$$

where $e^{-2K^*(B^*)} = \tanh K(B)$.

The relation (2.28) will be called the *duality relation for the partition function*. In the literature one usually defines the relation (2.29) as the duality relation; however, as we mentioned above, in most cases, the mapping d: $\mathcal{B} \to \mathcal{B}^*$ is not a bijection.

To construct explicitly dual lattices for $\{\Lambda, K\}$, we shall make use of the following result.

Theorem 3: $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$ if and only if:

(i) $|\mathcal{B}| - N_i = N_i^*$; (2.27)

(ii) There exists a surjective mapping d: $\mathcal{B} \to \mathcal{B}^*$ such that the subgroup $\mathcal{K}$ satisfies Eq. (2.25) and the image of generators of $\mathcal{K}$ are generators of $\gamma^*(\tau^*)$;

(iii) $K^*(B^*)$ is related to $K(B)$ by means of Eq. (2.26).

Proof: Lemma 4 implies that $D/\mathcal{K}$ is a group isomorphism between $\mathcal{K}$ and its image in $\mathcal{P}(\mathcal{B}^*)$; the number of generators of $\mathcal{K}$ being equal to the number of generators of $\gamma^*(\tau^*)$, Eq. (2.27), then the condition (ii) concludes the proof.

To conclude this section, we shall mention another definition of duality which is also useful to study properties of lattice systems.

Definition 2: With $\{\Lambda, K\}$ a general lattice system, then $\{\tilde{\Lambda}, \tilde{K}\}$ will be called an *inverse-dual lattice for* $\{\Lambda, K\}$ if there exists a surjective mapping d: $\mathcal{B} \to \tilde{\mathcal{B}}$ such that the induced mapping D: $\mathcal{P}(\mathcal{B}) \to \mathcal{P}(\tilde{\mathcal{B}})$ satisfies the following conditions:

(i) The subgroup $\gamma(\tau) \subset \mathcal{P}(\mathcal{B})$ satisfies the condition Eq. (2.25) above.

(ii) The image of $\gamma(\tau)$ is $\tilde{\mathcal{K}}$.

(iii) $$\tanh \tilde{K}(\tilde{B}) = \prod_{B\in d^{-1}\tilde{B}} e^{-2K(B)}. \quad (2.30)$$

Theorem 1': If $\{\tilde{\Lambda}, \tilde{K}\}$ is an inverse-dual lattice for $\{\Lambda, K\}$, then

$$N_i = |\tilde{\mathcal{B}}| - \tilde{N}_i.$$

Theorem 2': If $\{\tilde{\Lambda}, \tilde{K}\}$ is an inverse-dual lattice for $\{\Lambda, K\}$, then

$$Z\{\Lambda, K\} = 2^{-(|\Lambda|+N_S)/2}\, 2^{(|\tilde{\Lambda}|+\tilde{N}_S)/2} \times \prod_{\tilde{B}\in\tilde{\mathcal{B}}} \sqrt{(\sinh 2\tilde{K}(\tilde{B}))^{-1/2}}\, Z\{\tilde{\Lambda}, \tilde{K}\}.$$

Corollary 5: If the mapping d: $\mathcal{B} \to \tilde{\mathcal{B}}$ is a bijection, then the condition that $\{\tilde{\Lambda}, \tilde{K}\}$ is an inverse-dual lattice for $\{\Lambda, K\}$ is equivalent to the conditions:

(i) $0 \to \tau \xrightarrow{\gamma} \mathcal{P}(\mathcal{B}) \xrightarrow{\pi_{\tilde{\mathcal{O}}} D} \overline{\tilde{\mathcal{B}}} \to 0$ is an exact short sequence.

(ii) $\tanh \tilde{K}(\tilde{B}) = e^{-2K(B)}$.

Indeed d: $\mathcal{B} \to \tilde{\mathcal{B}}$ is a bijection implies that D: $\mathcal{P}(\mathcal{B}) \to \mathcal{P}(\tilde{\mathcal{B}})$ is a group homomorphism.

Moreover, the condition of inverse-duality is that the image of τ coincides with the kernel of the mapping $\pi^{\sim}$, which concludes the proof.

Remarks:

(1) If the mapping d: $\mathcal{B} \to \mathcal{B}^*$ is a bijection, then $\{\Lambda^*, K^*\}$ dual lattice for $\{\Lambda, K\}$, implies that $\{\Lambda, K\}$ is an inverse-dual lattice for $\{\Lambda^*, K^*\}$.

(2) We could also introduce a relation between the high (resp., low) temperature expansion of $\{\Lambda, K\}$ and the high (resp., low) temperature expansion of another system $\{\Lambda', K'\}$. Those relations are obtained by means of a bijection between the subgroups $\mathcal{K}$ and $\mathcal{K}'$ [resp., $\gamma(\tau)$ and $\gamma'(\tau')$] in the same manner as above and are useful to compute the partition functions of some systems.[4] Moreover such relations are also obtained directly as the product of a duality transformation followed by an inverse-duality transformation.

D. Summary of Group Structure and Duality Relation

A general lattice system (spin-$\frac{1}{2}$) was defined by $\{\Lambda, K\}$ where Λ is a finite set of points and K a function on $\mathcal{P}(\Lambda)$ with support $\mathcal{B}(\mathcal{B} \neq \phi)$ with $\mathcal{S} \subset \mathcal{P}(\Lambda)$, the symmetry subgroup of the system and $\tau \subset \mathcal{P}(\Lambda)$, the subgroup isomorphic to $\mathcal{P}(\Lambda)/\mathcal{S}$; we have obtained

$$\tau \cong \mathcal{P}(\Lambda)/\mathcal{S} \cong \overline{\mathcal{B}}$$

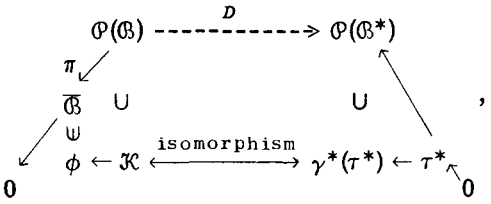
$$\begin{array}{ccccc} 0 & \to & \tau & \xrightarrow{\gamma} & \mathcal{P}(\mathcal{B}) \\ & & \cup & & \cup \\ & & T & \twoheadrightarrow & \gamma(T) = \{B \mid \sigma_T(B) = -1\} \end{array}$$

is an exact sequence

$$\begin{array}{ccccc} \mathcal{P}(\mathcal{B}) & & \xrightarrow{\pi} & \overline{\mathcal{B}} & \to 0 \\ \cup & & & \cup & \\ \beta = (B_1 \cdots B_n) & & \twoheadrightarrow & \Pi B_i & \end{array}$$

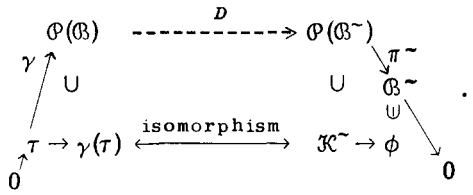
is an exact sequence.

$\{\Lambda^*, K^*\}$ will be a dual lattice for $\{\Lambda, K\}$ if there exists d: $\mathcal{B} \to \mathcal{B}^*$ such that Eq. (2.26) holds and such that



$\dashrightarrow$ is not a homomorphism.

$\{\Lambda^\sim, K^\sim\}$ will be an inverse-dual for $\{\Lambda, K\}$ if there exists $d: \mathcal{B} \to \mathcal{B}^*$ such that Eq. (2.30) holds and such that



If there exists a bijection $d: \mathcal{B} \to \mathcal{B}^\sim$ such that

$$\tanh K^\sim(B^\sim) = e^{-2K(B)},$$

then

$$0 \to \tau \xrightarrow{\gamma} \mathcal{P}(\mathcal{B}) \xrightarrow{\pi^\sim \circ D} \overline{\mathcal{B}}^\sim \to 0$$

is an exact short sequence if and only if $\{\Lambda^\sim, K^\sim\}$ is an inverse-dual for $\{\Lambda, K\}$. In this case $\{\Lambda, K\}$ is a dual for $\{\Lambda^\sim, K^\sim\}$.

3. GENERAL METHOD OF CONSTRUCTING DUAL LATTICES FOR $\{\Lambda, K\}$

In this section we shall give a method of constructing a dual lattice for $\{\Lambda, K\}$; repeating the argument we can then generate several dual lattices for the given lattice $\{\Lambda, K\}$.

To construct a dual lattice $\{\Lambda^*, K^*\}$ we shall make use of the following remarks.

Remarks:

(1) For any lattice $\{\Lambda^*, K^*\}$, the mapping

$$\gamma^*: \mathcal{P}(\Lambda^*) \to \mathcal{P}(\mathcal{B}^*)$$

is an homomorphism of $\mathcal{P}(\Lambda^*)$ into $\gamma^*(\tau^*)$ (Lemma 2). Therefore the image of a subset generating $\mathcal{P}(\Lambda^*)$ is a subset generating $\gamma^*(\tau^*)$ and, thus, the elements

$$\gamma^*(x^*) = \{B^* | B^* \ni x^*\}$$

generate $\gamma^*(\tau^*)$.

(2) If $\{\Lambda^*, K^*\}$ is a dual lattice for $\{\Lambda, K\}$, then Eq. (2.26) gives

$$|\Lambda^*| = |\mathcal{B}| - N_i + N_s^*. \tag{3.1}$$

Construction of the Lattice

Using the above remarks, with each lattice $\{\Lambda, K\}$, we associate another lattice $\{\Lambda^*, K^*\}$ defined in the following way.

Let $\beta_j = (B_{j1}, \ldots, B_{jnj})$, $j = 1 \ldots n$ be any subset of elements generating $\mathcal{K}$. With each element β_j, we associate a point $r^*_{\beta_j} \in \Lambda^*$, and we define the mapping

$$\begin{array}{ccc} d: & \mathcal{B} & \to \mathcal{B}^* \\ & \cup & \cup \\ & \mathcal{B} & \twoheadrightarrow B^* = \{r^*_{\beta_j} | \beta_j \ni B\}. \end{array} \tag{3.2}$$

Lemma 5: With $\{\Lambda^*, K^*\}$ and $d: \mathcal{B} \to \mathcal{B}^*$ defined above, then the induced mapping

$$D: \mathcal{P}(\mathcal{B}) \to \mathcal{P}(\mathcal{B}^*)$$

is such that the subgroup $\mathcal{K} \subset \mathcal{P}(\mathcal{B})$ satisfies condition (2.25), and therefore $D|\mathcal{K}$ is a group isomorphism between $\mathcal{K} \subset \mathcal{P}(\mathcal{B})$ and $D(\mathcal{K}) \subset \mathcal{P}(\mathcal{B}^*)$.

Proof:

$$\begin{array}{ccc} D: & \mathcal{P}(\mathcal{B}) & \to & \mathcal{P}(\mathcal{B}^*) \\ & \cup & & \cup \end{array}$$

$$\beta_j = (B_{j1}, \ldots, B_{jnj}) \twoheadrightarrow \beta_j^* = (B^*_{j1}, \ldots, B^*_{jmj}),$$

where $B^*_{j\alpha} = \{r^*_{\beta_i} | \beta_i \ni B_{j\alpha}\}$.

Since $\beta_j \ni B_{j\alpha}$, $B^*_{j\alpha} \ni r^*_{\beta_j}$, therefore,

$$B^* \in \beta_j^* \Rightarrow B^* \ni r^*_{\beta_j}. \tag{3.3}$$

Conversely if $B^* = dB \ni r^*_{\beta_j}$, then $\beta_j \ni B$ [(Eq. 3.2)] and $B^* \in \beta_j^*$.

Therefore $B^* \ni r^*_{\beta_j} \Leftrightarrow B^* \in \beta_j^*$, and

$$\beta_j^* = D\beta_j = \{B^* | B^* \ni r^*_{\beta_j}\}. \tag{3.4}$$

Finally the implications

$$B^* \in \beta_j^* \Rightarrow B^* \ni r^*_{\beta_j} \Rightarrow B \in \beta_j$$

gives $\beta_j \equiv D^{-1}\beta_j^*$ and Eq. (2.25) is satisfied.

Lemma 6: The mapping D induced by (3.2) is an isomorphism between $\mathcal{K}$ and $\gamma^*(\tau^*)$.

Proof: From Remark 1 above, $\gamma^*(\tau^*)$ is generated by the elements

$$\gamma^*(\tau^*) = \{B^* | B^* \ni r^*\}.$$

But from Eq. (3.4), those elements are the images of a subset generating $\mathcal{K}$. Finally, with $K^*(B^*) = -\frac{1}{2}\ln \prod_{B_j = D^{-1}B^*} \tanh K(B_j)$, we have

Theorem 4: The lattice $\{\Lambda^*, K^*\}$ defined above is a dual lattice for $\{\Lambda, K\}$. Moreover, $N_s^* = n - (|\mathcal{B}| - N_i)$.

Proof: Consequence of the construction, definition of duality, and Lemmas 5 and 6.

Remarks:

(1) From the high temperature expansion equation (2.23), it follows that without any restriction on the generality of the method, we can assume that for all $B \in \mathcal{B}$ there exists $\beta \in \mathcal{K}$ such that $\beta \ni \mathcal{B}$, and, therefore, $\mathcal{B}^*$ does not contain the empty set.

(2) It should be noted that choosing different subsets of $\mathcal{K}$ generating $\mathcal{K}$, one can obtain with the above method several dual lattice systems for the given lattice $\{\Lambda, \mathcal{K}\}$.

4. DUAL LATTICE FOR SYSTEM WITH EXTERNAL FIELD

In this section we consider the special case where the lattice system $\{\Lambda, K\}$ is such that

$$K(x) \neq 0 \quad \forall\, x \in \Lambda.$$

In this case $\mathfrak{B} \supset \Lambda$ and we have

$$\overline{\mathfrak{B}} = \mathcal{P}(\Lambda), \quad N_i = |\Lambda|, \quad N_s = 0,$$
$$\mathcal{S} = \{\phi\}, \quad \tau = \mathcal{P}(\Lambda).$$

Moreover $\mathcal{K}$ is generated by $|\mathfrak{B}| - |\Lambda|$ elements, and we can take as generators for $\mathcal{K}$ the elements

$$\beta = \{B, x_1, \ldots, x_n\}, \text{ where } \quad B \in \mathfrak{B}, |B| \geq 2$$
and
$$\{x_i\} = \{x | B \ni x\}.$$

Construction of $\{\Lambda^*, K^*\}$

The lattice $\{\Lambda^*, K^*\}$ is defined in the following way: With all $B \in \mathfrak{B}$ such that $|B| \geq 2$ we associate a point $r_B^* \in \Lambda^*$.

The mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ is then defined by

$$d: x \rightsquigarrow B_x^* = \{r_B^* | B \ni x\},$$
$$d: B \rightsquigarrow r_B^* \quad \text{if } |B| \geq 2.$$

Moreover,

$$K^*(B^*) = -\tfrac{1}{2} \ln \prod_{B = d^{-1} B^*} \tanh K(B).$$

This mapping associates with each generator $\beta = \{B, x_1, \ldots, x_n\}$ of $\mathcal{K}$ a point $r_\beta^* = r_B^* \in \Lambda^*$ and with each $B \in \mathfrak{B}$, the set $B^* = \{r_\beta^* | \beta \ni B\}$.

Using the result of the previous section, we conclude that the lattice $\{\Lambda^*, K^*\}$ so constructed, is a dual lattice for $\{\Lambda, K\}$.

Remark: This dual is easily constructed if we consider the lattice Λ^* defined by the barycenters of the interactions $B \in \mathfrak{B}$, $|B| \geq 2$; the interactions B^* are obtained by considering with each $x \in \Lambda$, the set of barycenters of the interaction B containing x.

5. APPLICATIONS—MODELS WITH EXTERNAL FIELD

In this section we shall illustrate the method discussed in Sec. 4 and construct a dual lattice for some specific lattice systems with an external field. In particular, these examples will show what kind of situations can occur: for some models the mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ is a bijection (Secs. 5A–5D), while for others it is not injective (Sec. 5E). Moreover, for some of the models, the dual lattice is identical with the original lattice (Secs. 5B and 5D), and we shall briefly discuss in Sec. 5F, the consequences one can draw from this self-duality for models with external field.

A. Square Lattice with Two-Body Interactions and External Field (Ising Model with External Field)

Let Λ be a square lattice with periodic boundary condition and K be defined by

$$K(r) = h \quad \forall\, r \in \Lambda,$$
$$K(r_1, r_2) = K \quad \text{if } r_1, r_2 \text{ are nearest neighbor},$$
$$K(r_1 \ldots r_n) = 0 \quad \text{otherwise}.$$

The barycenters of the two-body interaction defines the dual lattice Λ^*; the dual lattice is, therefore, a square lattice with periodic boundary condition; each point $x \in \Lambda$, belongs to four two-body interactions and, therefore, the dual lattice system is a square lattice with alternating four-body nearest neighbor interaction (i.e., the four barycenters of the interactions containing x) and external field. (See Fig. 2.)

B. Square Lattice with Four-Body Interaction and External Field

This model has been studied by Hintermann[6] using other techniques; however, within our formalism the dual lattice can be obtained directly. Let Λ be a square lattice with periodic boundary conditions and K be defined by

$$K(r) = h \quad \forall\, r \in \Lambda, \quad K(r_1, r_2, r_3, r_4) = K$$
if (r_1, r_2, r_3, r_4) are nearest neighbor,
$$K(r_1 \ldots r_n) = 0 \quad \text{otherwise}.$$

The barycenters of the four-body interactions define the lattice Λ^*; each point $x \in \Lambda$ belongs to four four-body interactions and, therefore, the dual lattice system is again a square lattice with periodic boundary conditions and four-body nearest neighbor interactions plus external field. (See Fig. 3.)

Conclusion: In this case the lattice system $\{\Lambda, K\}$ and the dual lattice $\{\Lambda^*, K^*\}$ are identical: We shall say that such a model is "self-dual" and discuss the properties of such system in Sec. 5F.

C. Triangular Lattice with Two-Body Forces and External Field

Let Λ be a triangular lattice with periodic boundary conditions and K be defined by

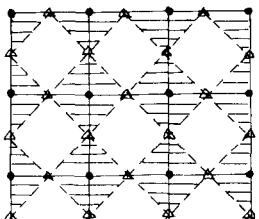


FIG. 2. Dual lattice for the square lattice with a two-body forces and external field.

· points of Λ; r_1 r_2 two-body interaction B,

Δ points of Λ^*; four-body interaction B_r^*,

$K^*(r^*) = -\tfrac{1}{2} \ln \tanh K \;\forall\, r^* \in \Lambda^*$,

$K^*(r_1^* \cdots r_4^*) = -\tfrac{1}{2} \ln \tanh h$
$\forall\, (r_1^* \cdots r_4^*) \in B_r^*$.

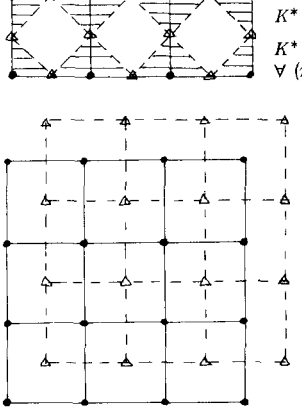


FIG. 3. Dual lattice for the square lattice with four-body forces and external field.

• points of Λ; [ ] four-body interaction B,

Δ points of Λ^*; Δ–Δ four-body interaction B_r^* Δ–Δ

$\tanh K^*(r^*) = -\tfrac{1}{2} \ln \tanh K$
$\forall\, r^* \in \Lambda^*$,

$\tanh K^*(r_1^*, r_2^*, r_3^*, r_4^*) =$
$-\tfrac{1}{2} \ln \tanh h \;\forall\, B_r^*$.

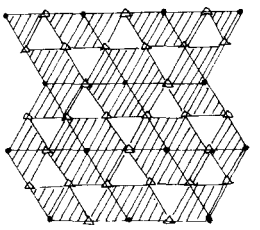


FIG. 4. Dual lattice for the triangular lattice with two-body forces and external field.

• point of Λ; r_1 r_2 two-body forces B,

Δ point of Λ^*; six-body forces B_r^*,

$\tanh K^*(r^*) = -\tfrac{1}{2} \ln \tanh K$,

$\tanh K^*(r_1^* \cdots r_6^*) = -\tfrac{1}{2} \ln \tanh h$.

$K(r) = h \quad \forall\, r \in \Lambda,$

$K(r_1, r_2) = K$ if (r_1, r_2) are nearest neighbor,

$K(r_1 \ldots r_n) = 0$ otherwise.

Repeating the construction, we find that the dual lattice is a "triangular" lattice with diluted six-body interactions and external field. (This is shown in Fig. 4.)

D. Triangular Lattice with Three-Body Alternating Forces and External Field

Let Λ be a triangular lattice with periodic boundary conditions and K be defined by

$K(r) = h \quad \forall\, r \in \Lambda,$

$K(r_1 r_2 r_3) = K$ if r_1, r_2, r_3 are the vertices of one of the shaded triangles (Fig. 5),

$K(r_1 \ldots r_n) = 0$ otherwise.

Repeating the same construction, it is easily seen (Fig. 6) that this lattice system is *self-dual*.

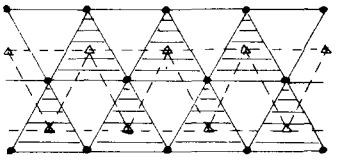


FIG. 5. Dual lattice for the triangular lattice with three-body alternating forces and external field.

• point of Λ;

three-body interactions B,

Δ point of Λ^*;

three-body interactions B_r^*.

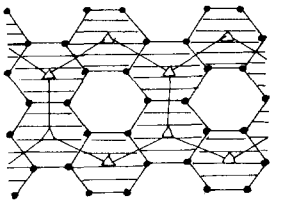


FIG. 6. Dual lattice for the hexagonal lattice with six-body diluted forces and external field.

• points of Λ;

six-body interactions B,

Δ points of Λ^*; Δ–Δ two-body interactions B_r^*.

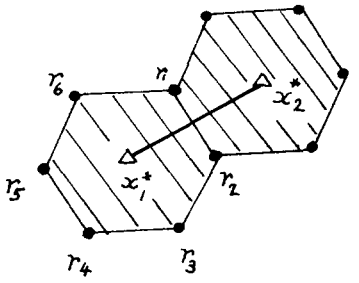


FIG. 7. mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ for the hexagonal lattice with six-body diluted force.
$(r_1 \cdots r_6) \to x_1^*$,
$r_1 \to (x_1^*, x_2^*)$,
$r_2 \to (x_1^*, x_2^*)$.

E. Hexagonal Lattice with Six-Body Diluted Forces and External Field

In all examples given so far the mapping d: $\mathfrak{B} \to \mathfrak{B}^*$ was a bijection. This model is intended to give an example of a situation where d: $\mathfrak{B} \to \mathfrak{B}^*$ is not injective.

Let Λ be a hexagonal lattice with periodic boundary conditions and K be defined by

$K(r) = h \quad \forall\, r \in \Lambda$

$K(r_1 \ldots r_6) = K$ if $(r_1 \ldots r_6)$ are the corners of one of the shaded hexagons of Fig. 6 below,

$K(r_1 \ldots r_n) = 0$ otherwise.

The dual lattice is, therefore, an hexagonal lattice with two-body nearest neighbor interactions and external field.

See Fig. 7 for more on the mapping d: $\mathfrak{B} \to \mathfrak{B}^*$.

The mapping is thus not a bijection, and the dual interaction is given by

$$\tanh K^*(x^*) = -\tfrac{1}{2} \ln \tanh K$$

$$\tanh K^*(x_1^*, x_2^*) = -\tfrac{1}{2} \ln [\tanh h]^2 = -\ln \tanh h$$

F. Self-Duality for Systems with External Field

The models discussed in Secs. 5B and 5D are self-dual, in the sense that

$$\{\Lambda, K\} \equiv \{\Lambda^*, K^*\}.$$

For such models the duality relation Eq. (2.28) becomes

$$Z_\Lambda(h, K) = (\sinh 2h)^{N/2} (\sinh 2K)^{N/2} \times Z_\Lambda(-\tfrac{1}{2} \ln \tanh K, -\tfrac{1}{2} \ln \tanh h),$$

where $N = |\Lambda|$.

Therefore introducing the variables

$$z = e^{-2h} \quad \text{and} \quad x = \tanh K,$$

we obtain

$$Z'_\Lambda(z, x) = [(1/z) - z]^{N/2} [-x + (1/x)]^{-N/2} Z'_\Lambda(x, z)$$

relation which can be used to discuss the locus of the zeros of the partition function, the singularities of the free energy, and the possible phase transition of the system.[4]

Moreover for those models the thermodynamic limit for the free energy is easily computed in the special case where $h \to 0$.

Indeed from the self-duality follows that

$$N_i = |\mathfrak{B}| - N_i.$$

Moreover, $N_s = 0$ implies then $|\Lambda| = N_i$ and

$$|\mathfrak{B}| = 2|\Lambda|.$$

If we consider the special case $h = 0$, we then have $|\mathfrak{B}^0| = \Lambda$ and

$$|\mathfrak{K}^0| = 2^{|\mathfrak{B}^0| - N_i} = 2^{N_s^0}$$

(where $\mathfrak{B}^0, \mathfrak{K}_1^0, \ldots$, corresponds to the $h = 0$ case).

Therefore if

$$\lim_{|\Lambda| \to \infty} (N_s^0 / |\Lambda|) = 0,$$

we will have

$$\lim_{|\Lambda| \to \infty} (\Lambda / |\Lambda|) \ln Z = \ln \cosh K \cdot 2, \qquad K = K(B)\ \forall\ B \in \mathfrak{B}^0.$$

In the particular case of the model Sec. 5B, we recover the expression computed explicitly by Hintermann[6] for the free energy.

6. APPLICATIONS—MODELS WITHOUT EXTERNAL FIELD

In this section, we shall apply the general method of Sec. 3 to construct the dual lattice for some specific models without external field. For some of those models (Secs. 6A, and 6B), we shall obtain a self-duality which is not as strong as the self-duality obtained for systems with external field; in fact those models will be self-dual only in the thermodynamic limit. Besides the well-known Ising model, we shall show that the two-dimensional triangular lattice with three-body nearest neighbor interactions, is self-dual in the thermodynamic limit.

Finally in one case (Sec. 6C) it will be possible to compute explicitly the partition function using the duality relation.

A. Two-Dimensional Ising Model

Let Λ_N be an $N \times N$ square lattice and K be defined by

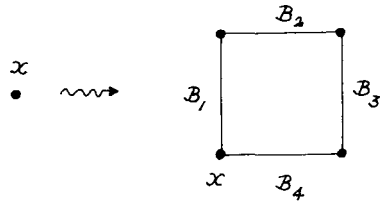


FIG. 8. Construction of the generators of $\mathcal{K}$ for the Ising model.

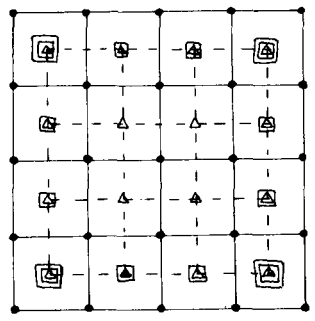


FIG. 9. Dual lattice for the Ising model.

Δ---Δ two-body interactions: $K^*(r_1^*, r_2^*) = -\frac{1}{2} \ln \tanh K$,

[Δ] one-body interaction: $K^*(r^*) = -\frac{1}{2} \ln \tanh K$,

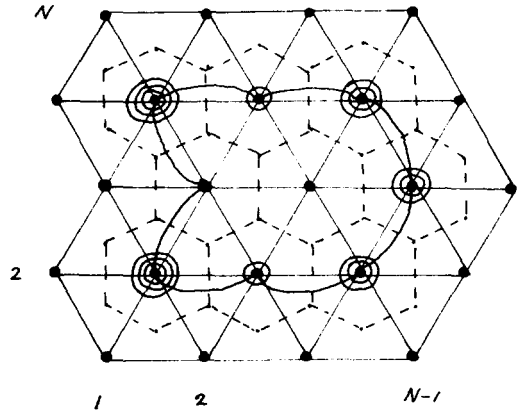


FIG. 10. Dual lattice for the triangular lattice with three body-forces. $K(r_1, r_2, r_3) = K(r_1, r_2, r_3)$ are nearest neighbor, $K(r_1, \ldots, r_n) = 0$ otherwise.
• point of Λ_N.

△, ▽; three-body interactions.

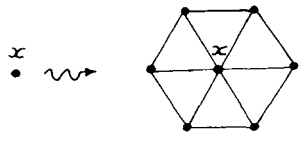


FIG. 11. Construction of the generators of $\mathcal{K}$.

$K(r_1, r_2) = K$ if (r_1, r_2) are nearest neighbor,

$K(r_1 \cdots r_n) = 0$ otherwise.

We shall construct a dual lattice $\{\Lambda^*, K^*\}$ choosing a set of N_K generators of $\mathcal{K}$. From

$$N_K = |\mathcal{B}| - N_i = |\mathcal{B}| - |\Lambda| + N_s$$

and

$$|\Lambda| = N^2; \quad N_s = 1; \quad |\mathcal{B}| = 2N(N-1),$$

we have $N_K = (N-1)^2$ = number of generators of $\mathcal{K}$.

With each point $x \in \Lambda_{N-1}$, we associate an element $\beta_x = (B_1, B_2, B_3, B_4) \in \mathcal{K}$ in the following manner (Fig. 8).

We obtain in this way $(N-1)^2$ element of $\mathcal{K}$, which are independent as one can easily see, i.e., there exists no subset $X \subset \Lambda_{N-1}$ such that $\beta y = \prod_{x \in X} \beta_x$ except $x = y$.

We then proceed with the general method of Sec. 3.

With each generator β_x we associate a point $r^*_{\beta_x} \in \Lambda^*$, which we will take to be the center of the square defined by β_x. (See Fig. 9.)

To define the dual interactions, to each $B \in \mathcal{B}$ we define B^* as the set of r^*, center of the squares containing B. The dual interactions are therefore two-body nearest neighbor interactions and an external field on the boundary of Λ^*. Moreover by construction $N_s^* = 0$.

As is well known[7] the Ising model with an external field on the boundary has the same thermodynamics (for the free energy) as the Ising model without field.

In conclusion the Ising model is self-dual in the thermodynamics, and we recover the result obtained by Wannier.[8]

B. Triangular Lattice with Three-Body Forces

Let Λ_N be the $N \times N$ triangular lattice and K defined by Fig. 10.

In this case

$$|\Lambda| = N^2 - 2, \quad |\mathcal{B}| = 2(N-1)^2 - 2, \quad N_s = 2.$$

Again to construct the dual we need a set of N_K generators for the subgroup $\mathcal{K}$ of $\mathcal{P}(\mathcal{B})$, from $N_K = |\mathcal{B}| - N_i$, we have $N_K = (N-2)^2$.

With each $x \in \Lambda_{N-2}$, we associate the elements $\beta_x = (B_1 \cdots B_6) \in \mathcal{K}$.

We obtain in this way $(N-2)^2$ element $\beta_x \in \mathcal{K}$ and as one can easily see those are independent: The β_x so defined can thus be taken as generators by means of hexagons constructed with the centers r'_B of each B (Fig. 11).

Following the general method we associate with each generator β_x a point $r^*_{\beta_x} \in \Lambda^*$, and we shall choose $r^*_{\beta_x} = x$: Λ^* is, therefore, a $(N-2) \times (N-2)$ triangular lattice.

To define the dual interaction, for each $B \in \mathcal{B}$ we define B^* as the set of r^* points associated with all generators β containing B, i.e., $B \to r'_B$ ($\in$ hexagonal lattice) $\to B^* = \{r^* |$ center of hexagon containing

$r'_B\}$. The dual interactions are thus three-body nearest neighbor interactions, together with a field and two-body forces on the boundary. Moreover, by construction $N_s^* = 0$. We therefore obtain the same result as in the case of the Ising model: In the thermodynamic limit this model is therefore self-dual.

C. Three-Dimensional Ising Model with 4-Spin Interaction (Suzuki[9])

Let Λ_N be a $N \times N \times N$ square lattice and K be defined by

$K(r_1r_2r_3r_4) = K$ if $(r_1r_2r_3r_4)$ are four-nearest neighbor in a plane $\parallel$ to (z,y) plane.

$K(r_1r_2r_3r_4) = K'$ if $(r_1r_2r_3r_4)$ are four-nearest neighbor in a plane $\parallel$ to (x,z) plane.

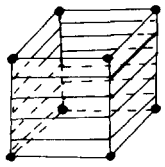


FIG. 12. Construction of the generators of $\mathcal{K}$.

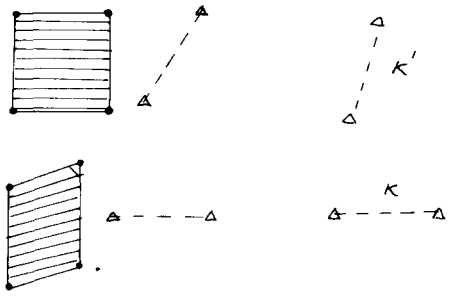


FIG. 13. Definition of the mapping d: $\mathcal{B} \to \mathcal{B}^*$.

This model has been studied by Suzuki[9] using a nonlinear σ-τ transformation of the spin variables. As we shall see the dual lattice associated with this model is the product of noninteracting two-dimensional Ising models, and therefore the partition function can be obtained without further computations.

In this case

$$|\Lambda| = N^3, \quad |\mathcal{B}| = 2N(N-1)^2, \quad N_s = N^2 + N - 1$$

and therefore $N_k = (N-1)^3$.

To construct the dual we need the N_k generators for $\mathcal{K}$; with each $x \in \Lambda_{N-1}$ we associate $\beta_x = (B_1, B_2, B_3, B_4) \in \mathcal{K}$ (as shown in Fig. 12).

We obtain in this way $(N-1)^3$ elements of $\mathcal{K}$, and as one can easily see the β_x so defined are independent and can be taken as generators of $\mathcal{K}$.

With each β_x we associate a point $r^*_{\beta_x} \in \Lambda^*$, and we shall take $r^*_{\beta_x}$ to be the center of the cube defined by β_x. Λ^* is, therefore, a $(N-1) \times (N-1) \times (N-1)$ cubic lattice.

To define the dual interactions for each $B \in \mathcal{B}$, we define B^* as the set of r^* associated with the β_x containing B. (See Fig. 13.)

In conclusion, the dual lattice is the product of $(N-1)$ two-dimensional Ising system with an external field on the boundary.

Finally, using the self-duality of the Ising model we conclude that

$$Z = [Z_{\text{Ising}}(K,K')]^{N-1}.$$

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. W. Greenberg and A. Hintermann for stimulating discussions.

1 F. J. Wegner, J. Math. Phys. **12**, 2259 (1971).
2 J. Ginibre, *Cargèse Lecture in Physics* (Gordon and Breach, New York, 1970), Vol. 4, p. 95.
3 W. Greenberg (private communication).
4 To appear in collaboration with A. Hintermann and F. Wegner.
5 The authors are indebted to W. Greenberg for the proof of the lemma.
6 A. Hintermann, Phys. Letters **39A**, 243 (1972).
7 D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
8 G. H. Wannier, Rev. Mod. Phys. **17**, 50 (1945).
9 M. Suzuki, Phys. Rev. Letters **28**, 507 (1972).

Remarkable Connection between the Multi-Veneziano Integrand and the Character of Special Unitary Groups

T. S. Santhanam*
Physikalisches Institüt der Universität, Bonn, Germany
(Received 10 May 1971)

The remarkable connection that exists between the multi-Veneziano integrand at a fixed point of the external momenta and the character of special unitary groups is exhibited explicitly. Some remarks are made on such a connection for arbitrary external momenta.

The remarkable connection that exists between the integrand of the multi-Veneziano amplitude (of scalar, equal mass particles) at a fixed point of the external momenta and the character of special unitary groups is shown to result from the evalution of partition functions, when one projects the positive root lattice on the simple (or primitive) root lattice. This connection is established using several different forms known for the characteristic in terms of the coordinates of the maximal toroid and in terms of the positive roots. Some remarks are made on such a connection for arbitrary external momenta and the problem of multiplicity in such a case.

The root system[1] of the group $SU(l+1) \sim A_l$ is given by $e_i - e_j, i,j = 1, \ldots, l+1$, where the e's are the unit vector in $(l+1)$-dimensional space and l is the rank of the group. The $\frac{1}{2}l(l+1)$ positive roots are $e_i - e_j, i < j$. It is well known that, for the group of rank l, there exists only l independent roots (called simple or primitive) which form a linearly independent basis in a l-dimensional space. For the group

$r'_B\}$. The dual interactions are thus three-body nearest neighbor interactions, together with a field and two-body forces on the boundary. Moreover, by construction $N_s^* = 0$. We therefore obtain the same result as in the case of the Ising model: In the thermodynamic limit this model is therefore self-dual.

C. Three-Dimensional Ising Model with 4-Spin Interaction (Suzuki[9])

Let Λ_N be a $N \times N \times N$ square lattice and K be defined by

$K(r_1r_2r_3r_4) = K$ if $(r_1r_2r_3r_4)$ are four-nearest neighbor in a plane $\parallel$ to (z,y) plane.

$K(r_1r_2r_3r_4) = K'$ if $(r_1r_2r_3r_4)$ are four-nearest neighbor in a plane $\parallel$ to (x,z) plane.

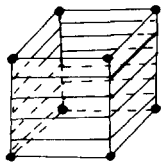


FIG. 12. Construction of the generators of $\mathcal{K}$.

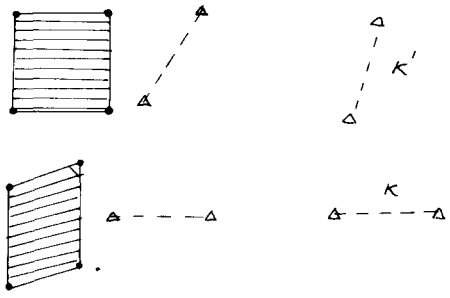


FIG. 13. Definition of the mapping d: $\mathcal{B} \to \mathcal{B}^*$.

This model has been studied by Suzuki[9] using a nonlinear σ-τ transformation of the spin variables. As we shall see the dual lattice associated with this model is the product of noninteracting two-dimensional Ising models, and therefore the partition function can be obtained without further computations.

In this case

$$|\Lambda| = N^3, \quad |\mathcal{B}| = 2N(N-1)^2, \quad N_s = N^2 + N - 1$$

and therefore $N_k = (N-1)^3$.

To construct the dual we need the N_k generators for $\mathcal{K}$; with each $x \in \Lambda_{N-1}$ we associate $\beta_x = (B_1, B_2, B_3, B_4) \in \mathcal{K}$ (as shown in Fig. 12).

We obtain in this way $(N-1)^3$ elements of $\mathcal{K}$, and as one can easily see the β_x so defined are independent and can be taken as generators of $\mathcal{K}$.

With each β_x we associate a point $r^*_{\beta_x} \in \Lambda^*$, and we shall take $r^*_{\beta_x}$ to be the center of the cube defined by β_x. Λ^* is, therefore, a $(N-1) \times (N-1) \times (N-1)$ cubic lattice.

To define the dual interactions for each $B \in \mathcal{B}$, we define B^* as the set of r^* associated with the β_x containing B. (See Fig. 13.)

In conclusion, the dual lattice is the product of $(N-1)$ two-dimensional Ising system with an external field on the boundary.

Finally, using the self-duality of the Ising model we conclude that

$$Z = [Z_{\text{Ising}}(K,K')]^{N-1}.$$

ACKNOWLEDGMENTS

It is a pleasure to thank Dr. W. Greenberg and A. Hintermann for stimulating discussions.

1 F. J. Wegner, J. Math. Phys. **12**, 2259 (1971).
2 J. Ginibre, *Cargèse Lecture in Physics* (Gordon and Breach, New York, 1970), Vol. 4, p. 95.
3 W. Greenberg (private communication).
4 To appear in collaboration with A. Hintermann and F. Wegner.
5 The authors are indebted to W. Greenberg for the proof of the lemma.
6 A. Hintermann, Phys. Letters **39A**, 243 (1972).
7 D. Ruelle, *Statistical Mechanics* (Benjamin, New York, 1969).
8 G. H. Wannier, Rev. Mod. Phys. **17**, 50 (1945).
9 M. Suzuki, Phys. Rev. Letters **28**, 507 (1972).

Remarkable Connection between the Multi-Veneziano Integrand and the Character of Special Unitary Groups

T. S. Santhanam*
Physikalisches Institüt der Universität, Bonn, Germany
(Received 10 May 1971)

The remarkable connection that exists between the multi-Veneziano integrand at a fixed point of the external momenta and the character of special unitary groups is exhibited explicitly. Some remarks are made on such a connection for arbitrary external momenta.

The remarkable connection that exists between the integrand of the multi-Veneziano amplitude (of scalar, equal mass particles) at a fixed point of the external momenta and the character of special unitary groups is shown to result from the evalution of partition functions, when one projects the positive root lattice on the simple (or primitive) root lattice. This connection is established using several different forms known for the characteristic in terms of the coordinates of the maximal toroid and in terms of the positive roots. Some remarks are made on such a connection for arbitrary external momenta and the problem of multiplicity in such a case.

The root system[1] of the group $SU(l+1) \sim A_l$ is given by $e_i - e_j, i,j = 1, \ldots, l+1$, where the e's are the unit vector in $(l+1)$-dimensional space and l is the rank of the group. The $\frac{1}{2}l(l+1)$ positive roots are $e_i - e_j, i < j$. It is well known that, for the group of rank l, there exists only l independent roots (called simple or primitive) which form a linearly independent basis in a l-dimensional space. For the group

$SU(l+1)$, these are given by $e_{i-1}-e_i$, $i=1,\ldots,l+1$. If we denote the system of positive roots by β and simple roots by α, then we have

$$\beta = \tilde{C}\alpha, \tag{1}$$

where the $\frac{1}{2}l(l+1)\times l$-dimensional rectangular matrix C is given by ($\sim$ denotes transposition)

$$C = \left[\begin{array}{c|cccc|cccc} & 1 & 0 & \ldots & 0 & 1 & 0 & \ldots & 1 \\ & 1 & 1 & & 0 & 1 & 1 & & 1 \\ & 0 & 1 & & 0 & 1 & 1 & & 1 \\ & 0 & 0 & & 0 & 0 & 1 & & 1 \\ I & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ & 0 & 0 & & 1 & 0 & 0 & & 1 \\ & 0 & 0 & & 1 & 0 & 0 & & 1 \end{array}\right] \begin{array}{l} \uparrow \\ l \\ \downarrow \end{array} \qquad \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ \\ \rightarrow \frac{1}{2}l(l+1) \end{array} \tag{2}$$

where I is the $(l\times l)$ unit matrix. This matrix C is typical of the group $SU(l+1)$. The multiplicity of a weight m belonging to the irreducible representation with highest weight Λ is given by Kostant's formula[2]

$$M^{\Lambda}(m) = \sum_{S\in W} \delta_s P[S(\Lambda+R_0)-(m+R_0)], \tag{3}$$

where the sum extends over the Weyl group which for the case of $SU(l+1)$ is simply the permutation group S_{l+1} of the components of the weight and is therefore of order $(l+1)!$. δ_s is the signature factor ± 1 depending on whether the permutation is even or odd, respectively. R_0 is half the sum of the positive roots. $P(M)$ is the ordered partition function which is the number of ways M can be expressed as

$$M = \sum_{i=1}^{l(l+1)/2} a_i\beta_i, \tag{4}$$

where the a_i are nonnegative integers. From the definition, it follows that $P(0)=1$ and $P(M)=0$ for $M<0$. Since M is defined in an l-dimensional space, we also have

$$M = \sum_{i=1}^{l} K_i\alpha_i, \tag{5}$$

where the K's are again nonnegative integers. Hence, for a given s, the value of the partition function is simply the number of solutions of the equation

$$\sum_{i=1}^{l} K_i\alpha_i = \sum_{j=1}^{l(l+1)/2} a_j\beta_j. \tag{6}$$

Since the α's are related to the β's by Eq. (1) and since the α's form a basis, it follows that Eq. (6) reduces to the diophantine equation[3]

$$K = Ca. \tag{7}$$

The problem is therefore finding the number of solutions of Eq. (7) for a given K. Equation (7) can be realized as simply the projection of $\frac{1}{2}l(l+1)$-dimensional positive-root lattice on the l-dimensional simple root lattice. The number of solutions of Eq. (7) can be easily realized to be given by the coefficient of $x_1^{K_1}\ldots x_l^{K_l}$ of the generating function[4] defined by

$$f(x_1,\ldots,x_l) = \prod_{i=1}^{l(l+1)/2} (1-x_1^{c_{1i}}\cdots x_l^{c_{li}})^{-1}, \qquad |x|<|,$$

$$= \prod_{j=1}^{l}\prod_{i=j}^{l}\Big(1-\prod_{k=1}^{j} x_{i-k+1}\Big)^{-1}. \tag{8}$$

In the following, $f(x_1,\ldots,x_l)$ will be shown to be just $1/\Delta$ where Δ is the characteristic of the unit representation. The character χ of an irreducible representation with highest weight Λ of a compact group is given by the trace of the representation matrix in its generic form. It is given by Weyl's formula[5]:

$$\chi = \xi/\Delta, \tag{9}$$

where ξ is the alternating elementary sum (characteristic)

$$\xi = \sum_{S\in W} \delta_s \exp\{2\pi i[S(\Lambda+R_0),\phi]\}, \tag{10}$$

where the ϕ's are the coordinates of the maximal toroid. The eigenvalues of the representation (in the diagonal form which can always be achieved for a compact group) are given by $\epsilon_i = \exp(2\pi i\phi_i)$, $0\le\phi\le 1$. Δ is simply the characteristic of the unit representation. There exist several equivalent formulas for $1/\Delta$. The first is in terms of the toroid parameters

$$1/\Delta = \prod_{i<j}(\epsilon_i-\epsilon_j)^{-1}, \qquad i,j=1,\ldots,l+1,$$

$$= \det(\epsilon^l,\ldots,\epsilon^1,1). \tag{11}$$

The second is Weyl's formula

$$1/\Delta = \Big(\sum_{S\in W}\delta_s \exp\{2\pi i[SR_0,\phi]\}\Big)^{-1}$$

$$= \Big(\sum_{\{p\}}\delta_p P(\rho_1,\ldots,\rho_{l+1})\epsilon_1^{\rho_1}\cdots\epsilon_{l+1}^{\rho_{l+1}}\Big)^{-1} \tag{12}$$

P is the permutation group S_{l+1}. The ρ's are the components of R_0. The third is in terms of positive roots[6]

$$\frac{1}{\Delta} = \exp[-2\pi i(R_0,\phi)]\prod_{i=1}^{l(l+1)/2}\{1-\exp[-2\pi i(\beta_i,\phi)]\}^{-1}.$$

If we set

$$x_i = \exp[-2\pi i(\alpha_i,\phi)], \qquad i=1,\ldots,l, \tag{13}$$

then Eq. (13) reads as

$$1/\Delta = x_1\cdots x_l f(x_1,\ldots,x_l), \tag{14}$$

where f is the generating function defined in Eq. (8). Thus the generating function f is simply $1/\Delta$ shifted by $\exp[-2\pi i(R_0,\phi)]$. The geometrical meaning of $1/\Delta$ as the sum over all the points in the positive domain of the infinite lattice formed by the positive roots is already well known.[7] Thus finding the coefficient of $x_1^{K_1}\ldots x_l^{K_l}$ of $f(x_1,\ldots,x_l)$ simply means projecting the lattice (of positive roots) E_m, $m=\frac{1}{2}l(l+1)$ on the lattice (of simple roots) E_l. The origin of the multiplicity structure of the weights of irreducible representations directly comes from this projection.

The remarkable point now is that the multi-Veneziano integrand for the $(l+3)$ external particles is identical to $1/\Delta$ at the unphysical point when the exponents

are equal to -1. The x acquire the meaning of the integration variables corresponding to the nondual channels of the dual diagram (nonintersecting diagonals). Equation (11) is the Koba–Nielson form[8] and Eq. (14) is the Bardakci–Ruegg–Virasoro[9] form. Equation (12) is yet another equivalent form which comes out of our analysis. This remarkable connection need not surprise one since the residue of the multi-Veneziano $(l+3)$-point amplitude expressed in the obvious notation[10]

$$\int_0^1 \prod_{i=1}^{l} dx_i x_i^{-\alpha(1,\ldots,i+1)-1} \times \left[\prod_{j=1}^{l}\prod_{i=j}^{l}\left(1-\prod_{K=1}^{j} x_{i-K+1}\right)^{-y_{ij}}\right] \tag{15}$$

where

$$\alpha(1,\ldots i+1) = a_{i+1} + \mathcal{s}(1,\ldots,i+1),$$
$$\mathcal{s}(1,\ldots,i+1) = (P_1 + \ldots + P_{i+1})^2,$$
$$y_{i,j} = -2P_{i-j+2}\cdot P_{i+2} - C_j,$$
$$C_j = a_j - 2a_{j-1} + a_{j-2},\ j \geq 4,$$
$$C_2 = a_2 + 2m^2 + 1,$$
$$C_3 = a_3 - 2a_2 - m^2, \quad \text{etc.},$$
$$m = \text{common mass of external particles},$$

when a chain of resonances occur simultaneously for $\alpha(12) = K_1, \ldots, \alpha(1,\ldots,l+1) = K_l$ is given by the coefficient of $x_1^{K_1}\ldots x_l^{K_l}$ in the product of factors given in the paranthesis, which can be realized to be simply the positive-root lattice of $SU(l+1)$, and the fixed point where the connection is made is when $y_{i,j} = 1$.

Encouraged by this remarkable connection, we make the following remarks for arbitrary $y_{i,j}$. In this case also, the paranthesis can still be envisaged as the sum of points on the positive domain of the lattice but the points having multiplicities given by the respective binomial coefficient depending on the trajectory functions in the dual channels. The projection physically means that we allow chain of resonances only in the nondual channels.

The $SU(l+1)$ structure can be seen to result from the invariance of the Hamiltonian of the system under the permutation of variables $x_1, \ldots, x_l$. The situation is quite analogous to the $SU(3N)$ invariance of the N-particle system in a three-dimensional oscillator potential.[11] The chain of subgroup considered in this case is $SU(N) \times O(3)$. The only change in our present case is that the $O(3)$ group is replaced by its relativistic counterpart. We also believe that the generalization of the multi-Veneziano amplitude when the external particles are nucleons can come from such a group analysis (for instance, we know that $SU(l+1)$ has spinor representations only for $l = 1$ and 3 when there exist isomorphisms with orthogonal groups, but orthogonal groups have always spin representations). We have been able to show that at this fixed point, the factorization of the multi-Veneziano amplitude[12] simply amounts to expressing $1/\Delta$ of $SU(l+1)$ in terms of its normal subgroups. The details of this calculation will be presented elsewhere.

ACKNOWLEDGMENTS

The author has benefitted from the discussions he had with the members of the theoretical groups at CERN and Torino. It is a pleasure to thank Professors H. Rollnik and K. Dietz for their hospitality at the Physikalisches Institute, Bonn.

* On leave of absence from MATSCIENCE, The Institute of Mathematical Sciences, Madras, India.

1 See for example E. B. Dynkin, Am. Math. Soc. Transl. **6**, 353 (1957).
2 B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959); P. Cartier, Bull. Am. Math. Soc. **67** 228 (1961).
3 P. A. MacMahon, *Combinatory Analysis* (Chelsea, New York, 1960), Vol. II, Sec VIII.
4 T. S. Santhanam, J. Math. Phys. **10**, 1704 (1969).
5 H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J.), p. 201.
6 N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 250.
7 J. P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 (1964).
8 Z. Koba and H. B. Nielson, Z. Physik **229**, 243 (1969).
9 K. Bardakci and H. Ruegg, Phys. Letters **28B**, 342 (1968); M. A. Virasoro, Phys. Rev. Letters **22**, 37 (1969).
10 We follow Y. Nambu, Chicago University, Report EFI 69-64 (unpublished).
11 See, for instance, V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960).
12 S. Fubini and G. Veneziano, Nuovo Cimento **64A**, 811 (1969); K. Bardakci and S. Mandelstam, Phys. Rev. **184**, 1640 (1969).

Nonnegativity of the Yukawa Hamiltonian*

John Piepenbrink
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 12 June 1972)

In a recent paper [J. Math. Phys. 8, 423 (1967)], Dyson and Lenard gave a lower bound of the N-particle Hamiltonian, Coulomb interaction, which was quadratic in N. An essential step in their proof is the fact that the one-particle Hamiltonian with Yukawa potential $-r^{-1}e^{-\mu r}$ is nonnegative if μ is not less than some value. The aim of this paper is to improve this positivity result and thereby improve the lower bound of Dyson and Lenard.

1. INTRODUCTION

In a recent paper,[1] Dyson and Lenard investigated the lower bound for the spectrum of the N-particle Hamiltonian with Coulombic interaction. The charges of the particles were of equal magnitude but of unspecified signs. In particular the authors were interested in obtaining the best asymptotic behavior of the lower bound as $N \to \infty$. As a simple preliminary result they proved the following.

Theorem (Dyson–Lenard): The minimum energy $E_{\min}$ for the N-particle Hamiltonian satisfies the inequality

$$E_{\min} > -[N(N-1)/\sqrt{2}\,]\text{Ry}. \tag{1.1}$$

Here Ry is Rydberg's constant $\text{Ry} = me^4/2h^2$, m is the mass of each particle, e is the magnitude of charge, and h is Planck's constant.

are equal to -1. The x acquire the meaning of the integration variables corresponding to the nondual channels of the dual diagram (nonintersecting diagonals). Equation (11) is the Koba–Nielson form[8] and Eq. (14) is the Bardakci–Ruegg–Virasoro[9] form. Equation (12) is yet another equivalent form which comes out of our analysis. This remarkable connection need not surprise one since the residue of the multi-Veneziano $(l+3)$-point amplitude expressed in the obvious notation[10]

$$\int_0^1 \prod_{i=1}^{l} dx_i x_i^{-\alpha(1,\ldots,i+1)-1} \times \left[\prod_{j=1}^{l}\prod_{i=j}^{l}\left(1-\prod_{K=1}^{j} x_{i-K+1}\right)^{-y_{ij}}\right] \tag{15}$$

where

$$\alpha(1,\ldots i+1) = a_{i+1} + \mathcal{s}(1,\ldots,i+1),$$
$$\mathcal{s}(1,\ldots,i+1) = (P_1 + \ldots + P_{i+1})^2,$$
$$y_{i,j} = -2P_{i-j+2}\cdot P_{i+2} - C_j,$$
$$C_j = a_j - 2a_{j-1} + a_{j-2},\ j \geq 4,$$
$$C_2 = a_2 + 2m^2 + 1,$$
$$C_3 = a_3 - 2a_2 - m^2, \quad \text{etc.},$$
$$m = \text{common mass of external particles},$$

when a chain of resonances occur simultaneously for $\alpha(12) = K_1, \ldots, \alpha(1,\ldots,l+1) = K_l$ is given by the coefficient of $x_1^{K_1}\ldots x_l^{K_l}$ in the product of factors given in the paranthesis, which can be realized to be simply the positive-root lattice of $SU(l+1)$, and the fixed point where the connection is made is when $y_{i,j} = 1$.

Encouraged by this remarkable connection, we make the following remarks for arbitrary $y_{i,j}$. In this case also, the paranthesis can still be envisaged as the sum of points on the positive domain of the lattice but the points having multiplicities given by the respective binomial coefficient depending on the trajectory functions in the dual channels. The projection physically means that we allow chain of resonances only in the nondual channels.

The $SU(l+1)$ structure can be seen to result from the invariance of the Hamiltonian of the system under the permutation of variables $x_1, \ldots, x_l$. The situation is quite analogous to the $SU(3N)$ invariance of the N-particle system in a three-dimensional oscillator potential.[11] The chain of subgroup considered in this case is $SU(N) \times O(3)$. The only change in our present case is that the $O(3)$ group is replaced by its relativistic counterpart. We also believe that the generalization of the multi-Veneziano amplitude when the external particles are nucleons can come from such a group analysis (for instance, we know that $SU(l+1)$ has spinor representations only for $l = 1$ and 3 when there exist isomorphisms with orthogonal groups, but orthogonal groups have always spin representations). We have been able to show that at this fixed point, the factorization of the multi-Veneziano amplitude[12] simply amounts to expressing $1/\Delta$ of $SU(l+1)$ in terms of its normal subgroups. The details of this calculation will be presented elsewhere.

ACKNOWLEDGMENTS

The author has benefitted from the discussions he had with the members of the theoretical groups at CERN and Torino. It is a pleasure to thank Professors H. Rollnik and K. Dietz for their hospitality at the Physikalisches Institute, Bonn.

* On leave of absence from MATSCIENCE, The Institute of Mathematical Sciences, Madras, India.

1 See for example E. B. Dynkin, Am. Math. Soc. Transl. **6**, 353 (1957).
2 B. Kostant, Trans. Am. Math. Soc. **93**, 53 (1959); P. Cartier, Bull. Am. Math. Soc. **67** 228 (1961).
3 P. A. MacMahon, *Combinatory Analysis* (Chelsea, New York, 1960), Vol. II, Sec VIII.
4 T. S. Santhanam, J. Math. Phys. **10**, 1704 (1969).
5 H. Weyl, *The Classical Groups* (Princeton U.P., Princeton, N.J.), p. 201.
6 N. Jacobson, *Lie Algebras* (Interscience, New York, 1962), p. 250.
7 J. P. Antoine and D. Speiser, J. Math. Phys. **5**, 1226 (1964).
8 Z. Koba and H. B. Nielson, Z. Physik **229**, 243 (1969).
9 K. Bardakci and H. Ruegg, Phys. Letters **28B**, 342 (1968); M. A. Virasoro, Phys. Rev. Letters **22**, 37 (1969).
10 We follow Y. Nambu, Chicago University, Report EFI 69-64 (unpublished).
11 See, for instance, V. Bargmann and M. Moshinsky, Nucl. Phys. **18**, 697 (1960).
12 S. Fubini and G. Veneziano, Nuovo Cimento **64A**, 811 (1969); K. Bardakci and S. Mandelstam, Phys. Rev. **184**, 1640 (1969).

Nonnegativity of the Yukawa Hamiltonian*

John Piepenbrink
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
(Received 12 June 1972)

In a recent paper [J. Math. Phys. 8, 423 (1967)], Dyson and Lenard gave a lower bound of the N-particle Hamiltonian, Coulomb interaction, which was quadratic in N. An essential step in their proof is the fact that the one-particle Hamiltonian with Yukawa potential $-r^{-1}e^{-\mu r}$ is nonnegative if μ is not less than some value. The aim of this paper is to improve this positivity result and thereby improve the lower bound of Dyson and Lenard.

1. INTRODUCTION

In a recent paper,[1] Dyson and Lenard investigated the lower bound for the spectrum of the N-particle Hamiltonian with Coulombic interaction. The charges of the particles were of equal magnitude but of unspecified signs. In particular the authors were interested in obtaining the best asymptotic behavior of the lower bound as $N \to \infty$. As a simple preliminary result they proved the following.

Theorem (Dyson–Lenard): The minimum energy $E_{\min}$ for the N-particle Hamiltonian satisfies the inequality

$$E_{\min} > -[N(N-1)/\sqrt{2}\,]\text{Ry}. \tag{1.1}$$

Here Ry is Rydberg's constant $\text{Ry} = me^4/2h^2$, m is the mass of each particle, e is the magnitude of charge, and h is Planck's constant.

Inequality (1.1) is far from ideal in its asymptotic behavior. A later and much deeper theorem in Ref. 1 shows that $E_{\min} > -AN^{5/3}$ Ry. But since the constant A is on the order of 50, it is clear that if N is not too large, the lower bound (1.1) is actually better. One goal of the present paper is to improve (1.1).

Dyson and Lenard's proof of (1.1) is based on the following Lemma concerning the one-particle Yukawa Hamiltonian. We normalize units so that $e = 1$ and $h^2/2m = 1$.

Lemma (Dyson–Lenard): The one-particle Hamiltonian

$$H = -\Delta - r^{-1}e^{-\mu r} \tag{1.2}$$

is nonnegative if

$$\mu \geq 1/\sqrt{2}\,. \tag{1.3}$$

Now for those values of μ which make (1.2) nonnegative it is proved in Ref. 1 that

$$E_{\min} > -\tfrac{1}{4}\mu N(N-1). \tag{1.4}$$

For $\mu = 1/\sqrt{2}$, inequality (1.1) is found. It is clear that any sharpening of the condition (1.3) leads to a corresponding sharpening of (1.4).

This improvement in the lemma is the goal of the present paper. We state the result as a theorem.

Theorem: The one-particle Hamiltonian

$$H = -\Delta - r^{-1}e^{-\mu r} \tag{1.2}$$

is nonnegative if

$$\mu \geq (1.64)^{-1}.$$

In Sec. 2 the problem of the positivity of (1.2) is reduced to the problem of determining a lower bound for the spectrum of an ordinary differential operator in a weighted L^2 space. In Sec. 3 this determination is made using a method of Weinberger.[2] A Rayleigh–Ritz computation then is made to determine the error.

2. THE EIGENVALUE PROBLEM

By making a scale change $r \to \mu r$ it is clear that (1.2) is nonnegative if and only if the operator

$$-\Delta - (1/\mu)(e^{-r}/r) \tag{2.1}$$

is nonnegative. Operator (2.1) is nonnegative if and only if

$$\frac{1}{\mu} \leq \frac{\int |\nabla u|^2 dx}{\int (e^{-r}/r)|u|^2 dx}, \tag{2.2}$$

for all u that are C^∞ and have compact supports in E^3. Here $x = (x_1, x_2, x_3)$ is in E^3, $r^2 = x_1^2 + x_2^2 + x_3^2$. But finally (2.2) is equivalent to requiring that $1/\mu$ be a lower bound for the spectrum of the operator

$$L = -\frac{\Delta}{(e^{-r}/r)} \tag{2.3}$$

in the space $\mathcal{H} = \{u \mid \int (e^{-r}/r)|u(x)|^2 dx < \infty\}$.

It is known that the spectrum of L is totally discrete since the equation

$$\Delta u + (e^{-r}/r)u = 0$$

is unconditionally nonoscillatory (see Refs. 3 and 4). Also since e^{-r}/r is spherically symmetric, the eigenfunction corresponding to the least eigenvalue is also spherically symmetric. When this eigenfunction is multiplied by r^{-1}, it becomes an eigenfunction for the ordinary differential operator

$$Au = -\frac{u''}{(e^{-r}/r)}, \quad r > 0, \tag{2.4}$$

in the space $\mathcal{H}_1 = \{u(r) \mid \int_0^\infty (e^{-r}/r)|u(r)|^2 dr < \infty\}$. By reversing the argument it is clear that the smallest eigenvalues for (2.3) and (2.4) are the same. Thus the problem is reduced to finding a lower bound for the first eigenvalue of the problem

$$(*)\quad \begin{aligned} &u'' + \lambda(e^{-t}/t)u = 0, \quad t > 0, \\ &u(0) = 0,\ \int_0^\infty (e^{-t}/t)|u(t)|^2 dt < \infty. \end{aligned}$$

3. THE SECULAR EQUATION

We apply a method of Weinberger to find a lower bound. Let the inner product (u, v) be defined by

$$(u, v) = \int_0^\infty e^t/t\, u(t)\overline{v(t)}dt\,.$$

The first step is to find a function $p(t)$ in $\mathcal{H}_1$ and a positive constant ρ such that

$$(u, p) = 0 \quad \text{implies that} \quad (Au, u) \geq \rho(u, u). \tag{3.1}$$

Let $w(t)$ be a positive continuously differentiable function with $w(t) \leq 1$. Define the operator A' by

$$A'u = -(w(t)u')'/t^{-1}e^{-t}. \tag{3.2}$$

Now if u is in the domain of A,

$$(A'u, u) = \int_0^\infty w(t)|u'(t)|^2 dt \leq \int_0^\infty |u'(t)|^2 dt = (Au, u).$$

Thus

$$(A'u, u) \leq (Au, u). \tag{3.3}$$

Now we pick $w(t)$ so that the spectrum of A' can be computed. In fact, if $w(t) = e^{-t}$, the eigenvalue problem for (3.2) is equivalent to

$$(+)\quad \begin{aligned} &t^2u'' - t^2u' + \lambda tu = 0, \quad t > 0, \\ &u(0) = 0, \quad u \in \mathcal{H}_1. \end{aligned}$$

By standard power series methods the eigenvalues of (+) are $\lambda = n$, $n = 1, 2, 3, \ldots$, and the corresponding eigenfunctions are polynomials $p_n(t)$ with degree $p_n = n$ and $p_n(0) = 0$. We will only need the first eigenfunction $p(t) = p_1(t) = t$. Thus because of (3.3) we find that (3.1) holds with $p(t) = t$ and $\rho = 2$.

For the second step we derive the secular equation for our lower bound. Let $q(t)$ be a real function in

the domain of A with $(p,q) \neq 0$. Then each function u in the domain of A can be represented uniquely as

$$u = \alpha q + v, \tag{3.4}$$

where $(v,p) = 0$. Because of (3.1), where $\rho = 2$, we have

$$\frac{(Au,u)}{(u,u)} \geq \frac{|\alpha|^2(Aq,q) + 2\,\mathrm{Re}\alpha(Aq,v) + 2(v,v)}{(u,u)} \equiv \frac{Q(u)}{(u,u)}.$$

Here the quadratic form Q is defined as

$$Q(u_1,u_2) = \alpha_1\bar{\alpha}_2(Aq,q) + \alpha_1(Aq,v_2) + \bar{\alpha}_2(v,Aq) + 2(v_1,v_2),$$

where

$$u_j = \alpha_j q + v_j, \quad j = 1,2.$$

Thus the smallest eigenvalue of A is bounded below by the constant

$$\mu = \inf_u [Q(u)/(u,u)]$$

A compactness argument will show that the infimum in this problem is attained at a function $\tilde{u} = \tilde{\alpha}q + \tilde{v}$. The eigenvalue equation for Q,

$$Q(u,\tilde{u}) = \mu(u,\tilde{u}), \tag{3.5}$$

is then satisfied for all u of the form (3.4). If we choose first $v = 0$ in (3.4), then $\alpha = 0$, and (3.5) implies the equations

$$(q,Aq - \mu q)\tilde{\alpha} + (\tilde{v},Aq - \mu q) = 0, \tag{3.6}$$

$$(Aq - \mu q)\tilde{\alpha} + (2 - \mu)\tilde{v} + zp = 0, \tag{3.7}$$

where z is some scalar.

We can eliminate $\tilde{v}$ from this pair of equations. First multiply (3.6) by $2 - \mu$ and substitute from (3.7). Then take the inner product of (3.7) with p. The equations that result are

$$(Aq - 2q, Aq - \mu q)\tilde{\alpha} + (p,Aq - \mu q)z = 0,$$
$$(p,Aq - \mu q)\tilde{\alpha} + (p,p)z = 0.$$

Since this system has a nonzero solution $\tilde{\alpha}$ and z, we must have

$$\det\begin{pmatrix}(Aq - 2q, Aq - \mu q) & (p,Aq - \mu q)\\ (p,Aq - \mu q) & (p,p)\end{pmatrix} = 0. \tag{3.8}$$

This is called the *secular equation*. It is easy to verify that this quadratic equation has two real roots. The smaller of these roots will be our lower bound.

4. DETERMINATION OF A LOWER BOUND

The secular equation (3.8) takes the form

$$A\mu^2 + B\mu + C = 0, \tag{4.1}$$

where

$$A = (p,q)^2, \quad B = (Aq - 2q,q) - 2(p,Aq)(p,q),$$
$$C = (p,Aq)^2 - (Aq - 2q,Aq).$$

Use is made of the fact that $(p,p) = 1$. If q is real valued, then A, B and C are real valued and $B^2 \geq 4AC$. So the lower bound we compute is

$$\mu = (-B - \sqrt{B^2 - 4AC})/2A. \tag{4.2}$$

It is clear from (4.2) that $\mu \geq b$ if and only if

$$2Ab + B \leq 0 \tag{4.3}$$

and

$$Ab^2 + Bb + C \geq 0. \tag{4.4}$$

Now if

$$q(t) = 1 - e^{-\alpha t}$$

and

$$\alpha = 1.5,$$

we find that

$$A = 0.36, \quad -1.342\,58 \leq B \leq -1.342\,56,$$
$$C = 1.234\,375.$$

When these values are substituted into (4.3) and (4.4), it is found that

$$b = 1.64$$

satisfies inequalities (4.3) and (4.4). So if λ_1 is the smallest eigenvalue of A, then

$$\lambda_1 \geq 1.64. \tag{4.5}$$

To give an idea of how good this bound is, we try a Rayleigh–Ritz calculation, using $q(t) = 1 - e^{-\alpha t}$ as test functions. It is found that $\alpha = 1.5$ is approximately a stationary value for the function

$$f(\alpha) = (q,Aq)/(q,q)$$

and that

$$\lambda_1 \leq f(1.5) \leq 1.68.$$

Thus we have the following bounds for λ_1:

$$1.64 \leq \lambda_1 \leq 1.68. \tag{4.6}$$

Returning to the original problem, we see that the conclusion of our theorem in Sec. 1 follows. Also, by (1.4), a sharper quadratic lower bound for $E_{\min}$ is obtained.

---

* This work was supported in part by contracts AFOSR 883-67 and AFOSR 71-2098.

[1] F. J. Dyson and A. Lenard, J. Math. Phys. 8, 423 (1967).
[2] Hans F. Weinberger, "Lower Bounds for Eigenvalues" (unpublished).
[3] John Piepenbrink, "Rellich Densities and an Application to Unconditionally Nonoscillatory Elliptic Equations," Pacific J. Math. **42**, No. 3 (1972).
[4] John Piepenbrink, "The Oscillation Constant for Second Order Elliptic Equations" (unpublished).

Relativistic Excitation of Cavity Modes

Richard L. Liboff*
Faculté des Sciences,† Université Libre de Bruxelles, Bruxelles, Belgium
(Received 28 April 1972)

Fields and energies due to a charged particle and other planar charged configurations, which move with relativistic speed along the axis of finite and infinite cylinders with perfectly conducting walls, are obtained. Explicit criteria are derived for modes of zero and maximum energy deposited in a finite closed can by a point particle. Infinities of wake fields are discussed vis a vis field energies of particles at rest, enclosed in cavities. The wake energy deposited in a semi-infinite can with forward target wall is found to include an infinite self-energy term modulated by the causality factor γ^{-1}.

1. INTRODUCTION AND DISCUSSION OF RESULTS

The technique introduced in a previous paper[1] for obtaining electromagnetic fields and energies excited by a charged particle traversing a closed finite cylinder with infinitely conducting walls, is extended to other planar charged configurations. These include the charged ring and charged disk. For all but the latter, total energy in the wake field excited by such traversal, suffer divergence in the high transverse wavenumbers.

Energy of the wake field partitions naturally into normal modes. The corresponding modal spectrum is believed relevant to more realistic configurations in which the charged aggregate enters and leaves the cavity through finite holes. Such conditions impose a natural cut-off on transverse wavenumber summation rendering total wake energy finite. Another cut-off (in modal frequency) obtains if the finite plasma frequency of the walls is brought into play.

The analysis begins with a short review of the results relevant to the finite closed can. In the ensuing subsections these results are carried to various limiting configurations by letting either the length or radius of the can pass to infinity.

The wake energies excited in these various configurations are studied in Sec. 3. The energy accompanying a particle moving in an infinitely long tube is consistent with the relativistic form of Poynting's theorem.[2] This is possible since such fields are obtained from the corresponding static ones through a single Lorentz transformation. If a target wall is placed in the infinitely long tube, one obtains the generic form of the transition radiation problem of a particle incident on an infinite plane.[3] This latter geometry is realized if the radius of the semi-infinite tube goes to infinity. The self-energy singularity contained in the transition radiation field is found to be modulated by the causality factor γ^{-1}. In the limit $v \to c$, the source particle does not *see* its image, and this causality factor goes to zero.

The energy spectrum of the point particle traversing a finite-closed tube is obtained explicitly. It is found that modal energy varies harmonically with particle velocity. It goes to zero at certain critical values of mode index numbers, radius/length, and particle velocity. Such properties of the spectrum should be easily subject to experimental observation.

An interesting result, representative to this study, obtains for the case of a charged particle which has traversed the space between two infinite condenser plates (Secs. 2D, 3D). Total wake energy is found to suffer a larger order divergence than that of the static field energy connected to a particle fixed between plates. This static energy is simply renormalized via the self-energy of the particle. Owing to the larger order singularity of the wake energy, a similar renormalization does not carry over.

2. THE POINT PARTICLE

A. Finite Cylindrical Box

In a previous analysis,[1] the exact fields due to a point-charged particle moving with relativistic speed in a closed, perfectly conducting, grounded, cylindrical box of finite length are obtained. The only constraint is that the particle moves with constant velocity along the axis of the can. The particle enters and leaves the cavity through point holes.

The technique of solution involves Lorentz transforming the static solution appropriate to a particle in a grounded cylindrical box. This yields the "inhomogeneous" fields. They are singular at the particle's position and satisfy the boundary conditions at the wall, $E_{\|} = 0$. To satisfy the initial condition that there are no fields in the box at $t = 0$, a superposition of cavity modes are adjoined to the inhomogeneous fields. The resulting cylindrical components of electric and magnetic fields are given by the expressions

$$E_z = \frac{4q}{r_0^2} \sum_{j=1}^{\infty} \frac{J_0(k_j r)}{J_1^2(\alpha_j)} \left[\frac{1}{\sinh K_j} \begin{pmatrix} \sinh \Omega_j t \cosh K_j[1 - (z/L)] \\ - \cosh(K_j z/L) \sinh(K_j - \Omega_j t) \end{pmatrix} - \sum_{p=-\infty}^{\infty} \frac{K_j \Omega_j \sin\omega t \cos\pi pz/L}{\omega[(\pi p)^2 + K_j^2]} \right]$$

$$E_r = \frac{4q\gamma}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} \left[\frac{1}{\sinh K_j} \begin{pmatrix} \sinh\Omega_j t \sinh K_j(1 - \frac{z}{L}) \\ \sinh(K_j z/L) \sinh(K_j - \Omega_j t) \end{pmatrix} - \sum_p \frac{\pi p \Omega_j \sin\omega t \sin\pi pz/L}{\omega[(\pi p)^2 + K_j^2]} \right] \tag{1}$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} \left[\frac{1}{\sinh K_j} \begin{pmatrix} \cosh\Omega_j t \cosh K_j(1 - \frac{z}{L}) \\ \cosh(K_j z/L) \cosh(K_j - \Omega_j t) \end{pmatrix} - \sum_p \frac{K_j \cos\omega t \cos\pi pz/L}{(\pi p)^2 + K_j^2} \right].$$

In these formulas, we have written ω for ω_{pj}, i.e.,

$$\omega^2 \equiv \omega_{pj}^2 \equiv c^2[k_j^2 + (\pi p/L)^2].$$

The zeros of the Bessel function J_0 are α_j:

$$J_0(k_j r_0) \equiv J_0(\alpha_j) = 0.$$

Other identities are

$$\Omega_j \equiv \gamma k_j v, \quad K_j \equiv \gamma k_j L, \quad \gamma^2 \equiv (1-\beta^2)^{-1}, \quad \beta \equiv v/c.$$

The length of the can is L, while the speed of the particle is v. The column vector notation is such that, for example,

$$E_r = \begin{bmatrix} E_r(z > vt) \\ E_r(z < vt) \end{bmatrix} = \begin{bmatrix} E_r(\text{"ahead" of particle}) \\ E_r(\text{"behind" particle}) \end{bmatrix} = \begin{bmatrix} E_r^> \\ E_r^< \end{bmatrix}.$$

These enter the inhomogeneous contribution to the fields, represented by this first j summation over the zeros of J_0. The hyperbolic factors insure exponential convergence of these summations except at $z = vt$, where these factors go to unity (as $j \to \infty$). Nevertheless the following continuity conditions maintain in the particle plane

$$E_z^> - E_z^< = 4\pi q\delta(r), \quad E_r^> - E_r^< = 0, \quad B_\phi^> - B_\phi^< = 0.$$

The asymmetric structure of E_z^i (where i denotes *inhomogeneous*) is due to the fact that the image contribution to E_z^i is continuous across the particle plane while the source contribution jumps across this plane.

The fields given by Eq. (1) are those seen by an observer at (r, z, t) and are due to a point charge q, moving with speed v which was at $z = 0$, $r = 0$ at $t = 0$. These fields are causal so that $\mathbf{E} = \mathbf{B} = 0$ ahead of the light front $z = ct$. Furthermore they serve as the Green's function for a finite line pulse of charge.[1] We wish now to exhibit certain asymptotic properties of these fields.

B. Semi-Infinite Can with After (Ejection) Wall

The first limit we consider is that in which the far wall at $z = L$ is moved to $z = \infty$. In the homogeneous p summations, the discrete index p becomes the continuous variable ξ, through the transformation

$$\pi p/L \to \xi, \quad \pi/L \to d\xi$$

so that

$$\omega^2 \to c^2(k_j^2 + \xi^2).$$

Ahead of the light front the fields vanish. For $z < ct$ and with Ω written for Ω_j, the fields are

$$E_z = \frac{4q}{r_1^2} \sum_j \frac{J_0(k_j r)}{J_0^2(\alpha_j)} \left[\begin{pmatrix} e^{-\gamma k_j z} \sinh\Omega t \\ -e^{-\Omega t} \cosh\gamma k_j z \end{pmatrix} - U \right],$$

$$E_r = \frac{4q\gamma}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} \left[\begin{pmatrix} e^{-\gamma k_j z} \sinh\Omega t \\ e^{-\Omega t} \sinh\gamma k_j z \end{pmatrix} - V \right], \tag{2}$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} \left[\begin{pmatrix} e^{-\gamma k_j z} \cosh\Omega t \\ e^{-\Omega t} \cosh\gamma k_j z \end{pmatrix} - W \right].$$

The homogeneous contributions are

$$U = \frac{\Omega^2}{\pi v} \int_{-\infty}^{\infty} \frac{\sin\omega t \cos\xi z}{\omega(\xi^2 + \gamma^2 k^2)} d\xi,$$

$$W = \frac{\Omega}{\pi v} \int_{-\infty}^{\infty} \frac{\cos\omega t \cos\xi z}{\xi^2 + \gamma^2 k^2} d\xi,$$

$$V = \frac{\Omega}{\pi} \int_{-\infty}^{\infty} \frac{\xi \sin\omega t \sin\xi z}{\omega(\xi^2 + \gamma^2 k^2)} d\xi,$$

Let us check that at $t = 0$ all fields $(z > vt)$ vanish. This is trivially true for $\mathbf{E}$. For $\mathbf{B}$ the relevant factors become

$$B_\phi(t = 0) = \cdots \left(e^{-\gamma k z} - \frac{\gamma k}{\pi} \int_{-\infty}^{\infty} \frac{\cos\xi z \, d\xi}{\xi^2 + \gamma^2 k^2} \right) = 0.$$

C. Infinitely Long Can and Free Particle Solutions

To obtain the fields appropriate to a charged particle moving in a can with no beginning nor end, one takes the limit $\Omega t \gg 1$ in the above expressions, Eq. (2). The ξ integrations (U, V, W) all vanish, and one obtains

$$E_z = \frac{2q}{r_0^2} \sum_j \frac{J_0(k_j r)}{J_1^2(\alpha_j)} e^{-\gamma k_j |z - vt|} \operatorname{sgn}(z - vt),$$

$$E_r = \frac{2q\gamma}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} e^{-\gamma k_j |z - vt|}, \tag{3}$$

$$B_\phi = \frac{2q\gamma\beta}{r_0^2} \sum_j \frac{J_1(k_j r)}{J_1^2(\alpha_j)} e^{-\gamma k_j |z - vt|}.$$

These fields may be simply obtained through a single Lorentz transformation of the static field due to a charged particle in an infinitely long tube. The free-particle fields emerge in the limit $r_0 \to \infty$. In this limit the k_j spectrum goes over to a continuum through the transformation (see Appendix A)

$$\lim_{r_0 \to \infty} \frac{2}{r_0^2} \sum_{j=1}^{\infty} \frac{J_0(k_j r)}{J_1^2(\alpha_j)} = \int_0^\infty k J_0(kr)\, dk = 2\pi\delta(r). \tag{4}$$

The delta function is represented by both the summation (with finite r_0) and integral and obeys the normalization[4]

$$\int_0^{r_0} \delta(r) 2\pi r \, dr = 1.$$

There results

$$E_z = q\int_0^\infty k J_0(kr) e^{-\gamma k|z - vt|} \operatorname{sgn}(z - vt) dk = \frac{q\gamma |z - vt|}{[\gamma^2 |z - vt|^2 + r^2]^{3/2}} \operatorname{sgn}(z - vt), \tag{5}$$

$$E_r = q\gamma\int_0^\infty k J_1(kr) e^{-\gamma k|z - vt|} dk = \frac{q\gamma r}{[\gamma^2 |z - vt|^2 + r^2]^{3/2}}, \quad B_\phi = \beta E_r.$$

These are the fields[5] seen by an observer at (r, z) due to a particle at $z = vt$, $r = 0$.

D. Particle between Condenser Plates

Let us return to the solution, Eq (1), for the finite can and again consider the limit $r_0 \to \infty$. This generates the geometry of two plane plates of infinite extent separated by the distance L. Using the transformation (4) together with the change to continuum,

$$K_j \to K, \quad \Omega_j \to \Omega, \quad \omega^2 \to c^2[k^2 + (\pi p/L)^2],$$

there results

$$E_z = 2q\int_0^\infty dk\, kJ_0(kr) \left[\frac{1}{\sinh K}\begin{pmatrix}\sinh\Omega t\, \cosh K(1-z/L)\\ -\cosh(Kz/L)\, \sinh(K-\Omega t)\end{pmatrix} - \sum_p\right],$$

$$E_r = 2q\gamma\int_0^\infty dk\, kJ_1(kr) \left[\frac{1}{\sinh K}\begin{pmatrix}\sinh\Omega t\, \sinh K(1-z/L)\\ \sinh(Kz/L)\, \sinh(K-\Omega t)\end{pmatrix} - \sum_p\right], \tag{6}$$

$$B_\phi = 2q\gamma\beta\int_0^\infty dk\, kJ_1(kr) \left[\frac{1}{\sinh K}\begin{pmatrix}\cosh\Omega t\, \cosh K(1-z/L)\\ \cosh(Kz/L)\, \cosh(K-\Omega t)\end{pmatrix} - \sum_p\right].$$

The homogeneous p summations are the same as in Eq. (1) with the modification that $k_j \to k$ in ω, Ω, K. In the limit that the forward plate goes to infinity, $L \to \infty$ and the homogeneous p summations go over the U, V, W integrals listed in Eq. (2). One obtains

$$E_z = 2q\int_0^\infty dk\, kJ_0(kr)\left[\begin{pmatrix}e^{-\gamma kz}\sinh\Omega t\\ -e^{-\Omega t}\cosh\gamma kz\end{pmatrix} - U\right],$$

$$E_r = 2q\gamma\int_0^\infty dk\, kJ_1(kr)\left[\begin{pmatrix}e^{-\gamma kz}\sinh\Omega t\\ e^{-\Omega t}\sinh\gamma kz\end{pmatrix} - V\right], \tag{7}$$

$$B_\phi = 2q\gamma\beta\int_0^\infty dk\, kJ_1(kr)\left[\begin{pmatrix}e^{-\gamma kz}\cosh\Omega t\\ e^{-\Omega t}\cosh\gamma kz\end{pmatrix} - W\right].$$

Again, the long-time behavior of these fields is free-particlelike (far removed from the plate), and we conclude that no radiation accompanies a charged particle ejected from a conducting plane. In the domain near the plate, the fields of the image are felt and the hyperbolic forms must be maintained (for the "behind" solution) in Eq. (7).

E. Semi-Infinite Can with Forward (Target) Wall

The geometry of this configuration is shown in Fig. 1. A charged particle moves with constant speed v along the axis of a semi-infinitely long cylinder. It begins its flight at $z = +\infty$, $t = -\infty$ and is absorbed by the target wall at $z = 0$, $t = 0$:

$$z = -vt, \quad r = 0, \quad v > 0, \quad t \leq 0.$$

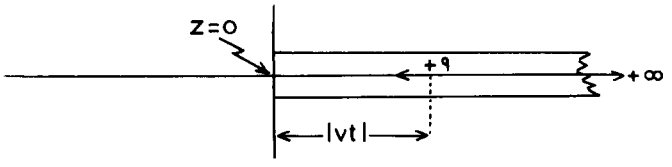


FIG. 1. Configuration of the semi-infinite tube with forward target wall. Collision occurs at $t = 0$.

The fields which accompany the particle are directly obtained from the totally infinite can solutions, Eq. (3), by superposing source and image solutions. When the source is at $z = |vt|$, a negative image is at $z = -|vt|$. The resulting fields appear as (with $z \geq 0$ and $t \leq 0$)

$$E_z = -\frac{4q}{r_0^2}\sum\frac{J_0(kr)}{J_1^2(\alpha_j)}\begin{pmatrix}e^{\Omega_j t}\cosh\gamma k_j z\\ e^{-\gamma kz}\sinh\Omega_j t\end{pmatrix},$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum\frac{J_1(kr)}{J_1^2(\alpha_j)}\begin{pmatrix}e^{\Omega_j t}\sinh\gamma k_j z\\ -e^{-\gamma k_j z}\sinh\Omega_j t\end{pmatrix}, \tag{8}$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum\frac{J_1(kr)}{J_1^2(\alpha_j)}\begin{pmatrix}e^{\Omega_j t}\cosh\gamma k_j z\\ e^{-\gamma k_j z}\cosh\Omega_j t\end{pmatrix}.$$

At the instant the particle is absorbed by the forward plate, only B_ϕ survives:

$$B_\phi(t=0) = -\frac{4q\gamma\beta}{r_0^2}\frac{\partial}{\partial r}\sum_j\frac{J_0(k_j r)}{k_j J_1^2(\alpha_j)}e^{-\gamma k_j z}. \tag{9}$$

Along the absorbing wall ($z = 0$), B_ϕ is singularly zero at $r = 0$. In the two-dimensional analysis of Ott and Shmoys,[3] it is found that this zero in the field expands back into the forward region with the speed c with the singularity in the field carried along in the wavefront.

3. ENERGY SPECTRUM OF THE WAKE FIELDS

A. The Finite Cylindrical Box

Consider again the first case (2a) of the point particle. We wish now to obtain the fields in the can after the particle leaves, that is, for times $vt > L$. Let $\mathbf{G}$ represent that total field when the particle is at the far wall. Then,

$$\mathbf{G} = \mathbf{G}^i + \mathbf{G}^h,$$

where i denotes inhomogeneous and h homogeneous, so that

$$\Box^2\mathbf{G}^h = 0,$$

we seek a solution to

$$\Box^2\mathbf{G}_{\text{WAKE}} = 0, \quad vt > L$$

which reduces to $\mathbf{G}$ at $vt = L$. Let

$$\Box^2\tilde{\mathbf{G}} = \mathbf{G}^i, \quad vt = L, \quad \Box^2\tilde{\mathbf{G}} = 0, \quad vt > L,$$

Then the desired solution is

$$\mathbf{G}_{\text{WAKE}} = \tilde{\mathbf{G}} + \mathbf{G}^h.$$

In this manner we find that, for $vt \geq L$,

$$E_z = \frac{4q}{r_0^2}\sum_j\sum_p\frac{K_j\Omega_j\cos\pi pz/L}{\omega_{pj}[(\pi p)^2 + K_j^2]}\frac{J_0(k_j r)}{J_1^2(\alpha_j)}S_{pj}(t),$$

$$E_r = \frac{4q\gamma}{r_0^2}\sum_j\sum_p\frac{\pi p\Omega_j\sin\pi pz/L}{\omega_{pj}[(\pi p)^2 + K_j^2]}\frac{J_1(k_j r)}{J_1^2(\alpha_j)}S_{pj}(t), \tag{10}$$

$$B_\phi = \frac{4q\gamma\beta}{r_0^2}\sum_j\sum_p\frac{K_j\cos\pi pz/L}{[(\pi p)^2 + K_j^2]}\frac{J_1(k_j r)}{J_1^2(\alpha_j)}C_{pj}(t).$$

The time factors S, C and dimensionless frequencies $\bar{\omega}, \bar{\omega}_\pm$ are given by the following; writing ω for ω_{pj},

$$S \equiv \sin(\omega t - \bar{\omega}_-) - \sin\omega t, \qquad C \equiv \cos(\omega t - \bar{\omega}_-) - \cos\omega t$$
$$\bar{\omega}_\pm \equiv \bar{\omega} \pm \pi p, \qquad \bar{\omega} \equiv \omega L/\beta c.$$

They satisfy the relation

$$S^2 + C^2 = 2[1 - \cos\pi p \cos\bar{\omega}] = 4 \sin^2\tfrac{1}{2}\bar{\omega}_+$$

which is independent of time.

The energy $\mathcal{E}$ in the wake field can be obtained in one of two ways, owing to Poynting's theorem. Thus,

$$\mathcal{E} = \int_0^{r_0} dr\ 2\pi r \int_0^L dt \left(\frac{E_z^2 + E_r^2 + B_\phi^2}{8\pi}\right)_{\text{WAKE}}$$
$$= -\int_0^{r_0} dr\ 2\pi r \int_0^L dz \int_0^{L/v} dt\ \mathbf{J}\cdot\mathbf{E}. \quad (11)$$

The second integral involves the fields in the box before the particle leaves. It represents the work done by the particle *on* the fields. The current is given by

$$\mathbf{J} = \mathbf{e}_z q v \delta(z - vt)\delta(r),$$

where $\mathbf{e}_z$ is a unit vector in the z direction. The right equality in Eq. (11) will be used to calculate $\mathcal{E}$ for the condenser problem treated in Sec. 3B. For the problem at hand we use the left equality involving field densities. Toward these ends recall the normalization

$$\int_0^{r_0} J_0^2(k_j r) r\, dr = \int_0^{r_0} J_1^2(k_j r) r\, dr = \tfrac{1}{2} r_0^2 J_1^2(\alpha_j).$$

Furthermore, a somewhat subtle entry of a factor of 2 appears in the summations

$$\int_0^L \left(\sum_{-\infty}^{\infty} a_p \cos\frac{\pi p z}{L}\right)^2 dz = 2\sum_{-\infty}^{\infty} a_p^2 \int_0^L \cos^2\frac{\pi p z}{L}\, dz$$
$$= L\sum_{-\infty}^{\infty} a_p^2,$$

$$\int_0^L \left(\sum_{-\infty}^{\infty} b_p \sin\frac{\pi p z}{L}\right)^2 dz = 2\sum_{-\infty}^{\infty} b_p^2 \int_0^L \sin^2\frac{\pi p z}{L}\, dz$$
$$= L\sum_{-\infty}^{\infty} b_p^2,$$

where a_p is even in p and b_p is odd.

We also note the relation

$$(\pi p)^2 + K^2 = (\gamma\beta)^2[\bar{\omega}^2 - (\pi p)^2] = (\gamma\beta)^2 \bar{\omega}_+ \bar{\omega}_- .$$

Performing the first integral in Eq. (11), using the wake fields Eq. (10) together with the above relations, one finds that the energy partitions into time-independent orthogonal modes as follows:

$$\mathcal{E} = \sum_j \sum_p \mathcal{E}_{pj},$$
$$\mathcal{E}_{pj} = \left(\frac{2\pi}{g}\right)^3 \frac{q^2}{r_0} \frac{\alpha_j^2}{J_1^2(\alpha_j)} \frac{\sin^2(\bar{\omega}_+/2)}{\beta^2 \bar{\omega}_+^2 \bar{\omega}_-^2}. \quad (12)$$

The dimensionless frequencies $\bar{\omega}_\pm$ are defined above, while the dimensionless frequency $\bar{\omega}$ obeys the inequality

$$\bar{\omega} = (\pi/\beta)[(\alpha_j/g)^2 + p^2]^{1/2} > \pi p. \quad (13)$$

The geometric ratio g is given by

$$2g \equiv 2\pi r_0/L.$$

The velocity and p dependence of $\mathcal{E}_{pj}$ is entirely contained in the factor

$$\Lambda_{pj}(\beta) = \sin^2(\bar{\omega}_+/2)/\beta^2 \bar{\omega}_+^2 \bar{\omega}_-^2, \quad (14)$$

and we may set

$$\mathcal{E}_{pj} = \tilde{\mathcal{E}}_j \Lambda_{pj}(\beta). \quad (15)$$

Owing to the following properties of $\bar{\omega}_\pm$,

$$\bar{\omega}_+ > 2\pi p, \qquad \bar{\omega}_- = \bar{\omega}_+ - 2\pi p > 0,$$
$$(\beta\bar{\omega}_-/\pi)^2 = (\beta\bar{\omega}_+/\pi)^2 = (\alpha/g)^2 + p^2, \qquad \text{for } \beta = 0, \quad (16)$$
$$\beta\bar{\omega}_- > 0, \qquad \text{for } \beta < 1,\ p < \infty.$$

Λ is bounded. The only question of a singularity in the spectrum occurs at large p and $\beta \simeq 1$. This can be seen from Eq. (13). In this limit $\bar{\omega}_- \to 0$. To examine this case in detail we look at

$$\bar{\omega}_\pm\big|_{p\to\infty} \sim \pi p[(1/\beta) \pm 1],$$

which gives

$$\beta\bar{\omega}_- \sim \pi p\epsilon, \qquad \epsilon \equiv 1 - \beta$$

while

$$\tfrac{1}{2}\bar{\omega}_+ \sim \tfrac{1}{2}\pi p[1 + (1/\beta)] \sim \pi p(1 + \tfrac{1}{2}\epsilon).$$

Thus,

$$\sin^2(\tfrac{1}{2}\bar{\omega}_+) \sim \sin^2\pi p(1 + \tfrac{1}{2}\epsilon) \sim (\tfrac{1}{2}\pi p\epsilon)^2 = (\tfrac{1}{4}\pi p)^2(1 - \beta)^2.$$

It follows that in the limit $\beta \to 1$, $p \to \infty$

$$\Lambda \sim (8\pi p)^{-2} \to 0.$$

Thus we may conclude that $\mathcal{E}_{pj}$ is bounded for all p, j, and β. It goes to zero faster than p^{-2} with increasing p. Owing to the asymptotic property of J_n,

$$J_n(x) \sim (2/\pi x)^{1/2} \cos(x - \tfrac{1}{2}n\pi - \tfrac{1}{4}\pi),$$

we see that $\mathcal{E}_{pj}$ goes to zero as j^{-1} with increasing j. Thus, while the sum $\sum_p \mathcal{E}_{pj}$ over p converges, the sum over j (for fixed p) diverges logarithmically.

A sketch of Λ_{pj} vs $\bar{\omega}_+$ is shown in Fig. 2, while $\bar{\omega}_+, \bar{\omega}_-$ vs β is sketched in Fig. 3. The Λ curve begins at $\bar{\omega}_+ = \pi p + \epsilon$ and decreases harmonically with increasing $\bar{\omega}_+$, or equivalently, with decreasing β.

From Eq. (12) we see that the zeros of $\mathcal{E}_{pj}$ occur when $(\bar{\omega}_+/\pi)$ is an even integer while its maxima occur when this same parameter is an odd integer. To guarantee the first inequality in Eq. (16), we set $\bar{\omega}_+/\pi = 2p + l$. This gives the following criteria. No

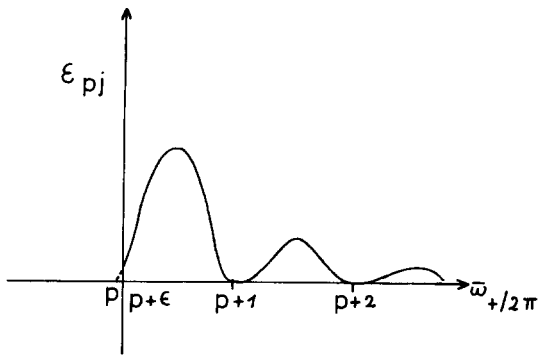


FIG. 2. Velocity dependence of energy in the $p-j$ mode.

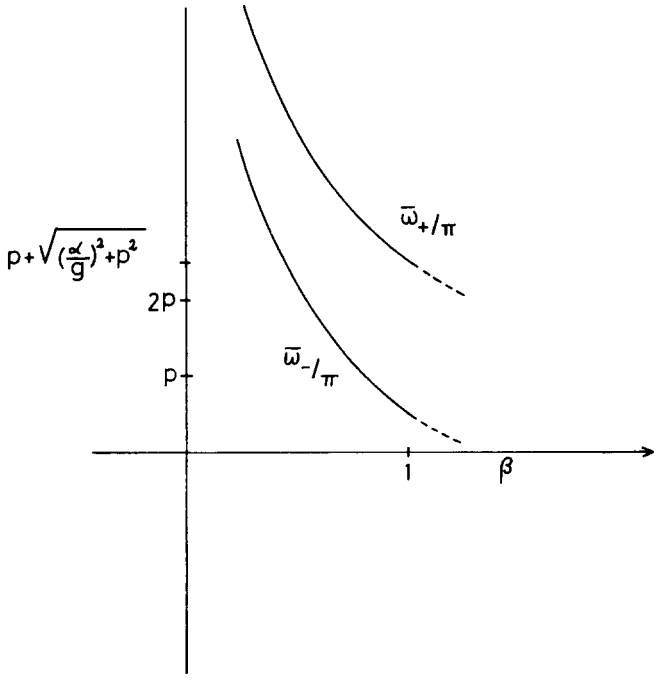


FIG. 3. Dependence of $\bar{\omega}_\pm$ on particle velocity.

energy is excited in the $\mathcal{E}_{pj}$ mode if there is an even integer l which fits the relation

$$p^2 + (\alpha/g)^2 = \beta^2(p + l)^2. \tag{17}$$

If there is an odd integer which fits this relation, then $\mathcal{E}_{pj}$ is maximum. These criteria are graphically exhibited by the right triangle shown in Fig. 4. As a simple case in point, consider a can with

$$g = \tfrac{1}{4}\alpha_j$$

and $\beta \simeq 1$. For such a configuration, (almost) no energy is deposited in the $(j, 3)$ modes.

The maxima of $\mathcal{E}_{pj}$ are

$$\mathcal{E}_{pj}^{\mathrm{MAX}} = \frac{8}{\pi g^3}\,\frac{q^2}{r_0}\,\frac{\alpha_j^2}{J_1^2(\alpha_j)}\,\frac{1}{\beta^2 l^2(2p + l)^2}, \tag{18}$$

where l is an odd integer as in Eq. (17).

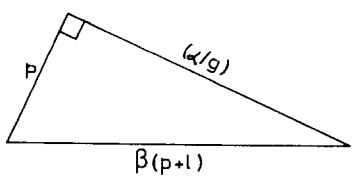


FIG. 4. For a cylinder with geometric ratio g, (no, maxi.) energy is excited in the α_j, p mode by a particle with speed βc if triangle is completed by an (even, odd) integer l.

The $p = 0$ modes (z-independent) are of particular interest. These modes also reveal the large α_j dependence, as may be seen from the following.

$$\bar{\omega}_\pm\big|_{j\to\infty} \to \pi\alpha/g\beta = \bar{\omega}_\pm(p = 0) \equiv y,\ y \equiv kL/\beta = K/\beta\gamma.$$

In this limit, $\mathcal{E}_{pj}$ becomes

$$\mathcal{E}_{0j} = \frac{8\pi}{g}\,\frac{q^2}{r_0}\,\frac{1}{J_1^2(\alpha_j)}\,\frac{\sin^2 y/2}{y^2},\quad y \geq \pi\alpha/g. \tag{19}$$

For given k_j, these represent the modes of lowest frequency [see Eq. (12)] and should persist longest after excitation. In the ideal problem with perfectly conducting walls, modes may still escape, owing to (a) the presence of entrance and exit holes in the can and (b) escape of modes above the wall–plasma frequency.

In either case one should expect to detect experimentally the component of the $\mathcal{E}_{pj}$ spectrum [Eq. (12)] to which such perturbation is minor. This would apply in the first case to modes with wavelength (k_j^{-1}) large compared to hole diameter[6] (which in turn is small compared to can radius r_0). In the second case, energy modes following the $\mathcal{E}_{pj}$ spectrum [Eq. (12)] should be present for frequences (ω_{pj}) small compared to the wall plasma frequency.

We now proceed to execute the p summation in Eq. (12). First we rewrite

$$\mathcal{E}_{pj} = \frac{8\pi}{g}\,\frac{q^2}{r_0}\,\frac{1}{J_1^2(\alpha_j)}\,\frac{y^2\sin^2\tfrac{1}{2}\bar{\omega}_+}{\bar{\omega}_+^2\bar{\omega}_-^2},$$
$$\bar{\omega}_\pm = [y^2 + (\pi p/\beta)^2]^{1/2} \pm \pi p, \tag{20}$$

where y is as defined in Eq. (19). Then (see Appendix B),

$$\sum_{p=-\infty}^{\infty} \frac{\sin^2\tfrac{1}{2}\bar{\omega}_+}{\omega_+^2\omega_-^2} = \frac{1}{4y^2},$$

and we obtain

$$\mathcal{E}_j \equiv \sum_{p=-\infty}^{\infty} \mathcal{E}_{pj} = \frac{2\pi}{g}\,\frac{q^2}{r_0}\,\frac{1}{J_1^2(\alpha_j)}. \tag{21}$$

This energy is velocity independent and extremely divergent. For large j it grows as j. The velocity independence is believed related to the imposed constraint that the particle move with constant speed while in the cavity. As noted by Jackson,[7] the time integral of $\mathbf{v}\cdot\mathbf{E}$ for a relativistic particle moving with constant speed is independent of v. On the other hand, velocity independence of field energy is also found to be a property of the closed, finite cavity. If either or both walls are brought to infinity, related energies become dependent on γ, as shown below.

In passing, we note the close relation between $\mathcal{E}_j$ and the energy contained in the single $p = 0$ mode:

$$\mathcal{E}_{j0} = \mathcal{E}_j\,\frac{\sin^2 y/2}{y^2} \leq \frac{\mathcal{E}_j}{4} = \frac{2}{\pi}\int_{-\infty}^{\infty}\mathcal{E}_{0j}\,dy.$$

Although the energy $\mathcal{E}_{j0}$ goes to zero as j^{-1}, for large j, its sum over all j suffers a logarithmic divergence.

The divergence of the summation of $\mathcal{E}_j$, on the other hand, is the same as that of the cylindrical delta function

$$\sum_{j=1}^{\infty} \mathcal{E}_j = 2\pi L\delta(0)$$

[see Eq. (4)]

B. Wake Fields Excited between Condenser Plates

To obtain the wake fields and energy excited by a particle which has traversed the space between two plates of infinite extent, we employ the transformation, Eq. (4), on the wake fields appropriate to a point particle in a finite box, as given by Eq. (10).

There results

$$E_z = 2q\sum_p \cos\left(\frac{\pi pz}{L}\right)\int_0^\infty \frac{kJ_0(kr)K\Omega S_p(t)}{\omega_p[(\pi p)^2 + K^2]}\,dk,$$

$$E_r = 2q\gamma\sum_p \pi p\,\sin\left(\frac{\pi pz}{L}\right)\int_0^\infty \frac{kJ_1(kr)S_p(t)}{\omega_p[(\pi p)^2 + K^2]}\,dk, \qquad (22)$$

$$B_\phi = 2q\gamma\beta\sum_p \cos\left(\frac{\pi pz}{L}\right)\int_0^\infty \frac{kKJ_1(kr)C_p(t)}{(\pi p)^2 + K}\,dk.$$

To calculate the energy spectrum we will use the right equality in Eq. (11), together with E_z as given in Eq. (6). Owing to the symmetry of the inhomogeneous fields, only the homogeneous component contributes to $\mathcal{E}$. With the current given by the expression directly beneath Eq. (11), one obtains

$$\mathcal{E}_p = -\int_0^L dz\int_0^\infty dr\; 2\pi r\int_0^{L/v} dt\; E_z^p\, qv\delta(z - vt)\delta(r)$$
$$= \frac{8q^2L^3}{\beta^2}\int_0^\infty dk\, k^3\,\frac{\sin^2(\bar{\omega}_+/2)}{\bar{\omega}_+^2\bar{\omega}_-^2}. \qquad (23)$$

We note that this expression may have been directly obtained from the spectrum relevant to the finite box, Eq. (11), with the aid of the transformation, Eq. (4). The divergence of $\mathcal{E}_p$ in the high k domain is again logarithmic. Summing over all p gives the result

$$\bar{\mathcal{E}} = \sum_p \mathcal{E}_p = \frac{2q^2}{L}\int_0^\infty d\tilde{k}\,\tilde{k}, \qquad (24)$$

where $\tilde{k} \equiv kL$ is nondimensional wavenumber.

It is interesting to compare this energy expression with that of a point particle fixed at midplane between two infinite parallel plates. The potential due to a particle at $r = 0$, $z = \bar{z}$ is

$$\phi(r,z) = \begin{pmatrix}\phi(z > \bar{z})\\ \phi(z < \bar{z})\end{pmatrix}$$
$$= 2q\int_0^\infty \frac{dk\,J_0(kr)}{\sinh kL}\begin{pmatrix}\sinh k\bar{z}\,\sinh k(L - z)\\ \sinh k(L - \bar{z})\,\sinh kz\end{pmatrix}.$$

It follows that the field energy due to a particle at midplane $z = L/2$ is

$$\mathcal{E}_L = \tfrac{1}{2}\int \rho(\mathbf{x})\phi(\mathbf{x})d\mathbf{x} = \frac{q}{2}\int_0^\infty\int_0^L \delta(r)\delta(z - \tfrac{1}{2}L)\,\phi(r,z)dz$$
$$\times\, 2\pi r\,dr = \frac{q^2}{L}\int_0^\infty d\tilde{k}\,\frac{\sinh^2\tilde{k}/2}{\sinh\tilde{k}}. \qquad (25)$$

As $L \to \infty$, one obtains the self-energy of a particle in free space:

$$\mathcal{E}_L \to \mathcal{E}_0 = \frac{q^2}{2}\iint \frac{dz\,2\pi r\,dr\,\delta(r)\delta(z)}{\sqrt{r^2 + z^2}}$$
$$= \frac{q^2}{2}\iiint e^{-kz}J_0(kr)\delta(r)\delta(z)dz\;2\pi r\;drdk$$
$$= \frac{q^2}{2}\int_0^\infty dk. \qquad (26)$$

Thus, for a fixed particle between condenser plates, it is possible to subtract the self-energy $\mathcal{E}_0$ and obtain the finite result

$$\tilde{\mathcal{E}}_L = \mathcal{E}_0 - \mathcal{E}_L = \frac{q^2}{L}\int_0^\infty d\tilde{k}\left(\frac{1}{2} - \frac{\sinh^2\tilde{k}/2}{\sinh\tilde{k}}\right) = \frac{q^2}{L}\ln 2. \qquad (27)$$

On the other hand, the total wake energy $\bar{\mathcal{E}}$ is not the same order infinity as $\mathcal{E}_0$ nor $\mathcal{E}_L$ and is not renormalizable via these terms. A finite $\mathcal{E}_p$ spectrum, appropriate to a particle which exits the target plate through a hole of diameter d, is obtained by cutting off the integral in Eq. (23) at $2\pi/d$.

C. Field Energy for the Infinitely Long Can

Here we consider the energy contained in a can of infinite length, in which a charged particle is moving with fixed speed v. From Eq. (3) one obtains

$$\mathcal{E} = \frac{1}{8\pi}\iint (E_z^2 + E_r^2 + B_\phi^2)2\pi r\,dr\,dz$$
$$= \frac{q^2}{2r_0{}^2}\sum_j \frac{1}{J_1^2(\alpha_j)}(1 + \gamma^2 + \gamma^2\beta^2)\int_{-\infty}^\infty dz e^{-2\gamma k_j|z - vt|}$$
$$= \gamma\frac{q^2}{r_0}\sum_j \frac{1}{\alpha_j J_1^2(\alpha_j)} = \gamma\mathcal{E}_\infty. \qquad (28)$$

The energy $\mathcal{E}_\infty$ is that due to a fixed charge on the axis of an infinitely long tube. This result is an example of the relativistic Poynting expression[2]

$$\mathcal{E} = \gamma\int u d\sigma - [(\gamma^2 - 1)^{1/2}/c]\int \mathbf{S}\cdot\hat{\mathbf{n}}\,d\sigma.$$

Energy density is u while $d\sigma$ is an invariant spacelike surface element, and $\hat{n}$ is a unit 3-vector. In the rest frame, $d\sigma = d^3x$, and the Poynting vector $\mathbf{S}$ vanishes (for the problem at hand).

D. Wake Energy for the Semi-Infinite Can with Target Wall

Poynting's theorem affords a technique for estimating wake energy without knowledge of the wake fields in as much as the rhs of Eq. (11) involves the electric field while the particle is still in the cavity. This term represents work done by the particle against fields induced by interaction between the boundaries and the particle. It is equal to the energy deposited in the fields (bounded or radiant). As a simple example, a freely moving particle, in the absence of boundries, gives $\int \mathbf{E}\cdot\mathbf{J}\,d^4x = 0$. No work is done nor is any radiation emitted.

We now apply this technique to calculate the energy deposited in the semi-infinite can after the particle

is absorbed by the target wall. Using the fields given in Eq. (8), one obtains (see Appendix C)

$$\mathcal{E} = qv\int_{-\infty}^{0} dt\int_{0}^{r_0} dr\, 2\pi r\left[\int_{\infty}^{|vt|} dz E_z(z > |vt|) + \int_{|vt|}^{0} \times dz E_z(z < |vt|)\right]\delta(r)\delta(z + vt)$$

$$= \frac{q^2}{\gamma r_0}\sum \frac{1}{\alpha_j J_1^{\,2}(\alpha_j)} = \frac{1}{\gamma}\mathcal{E}_\infty. \tag{29}$$

In the limit that $r_0 \to \infty$, Eq. (29) gives the total energy in the transition radiation due to a particle absorbed in a perfectly conducting grounded plate. It is simply related to the self-energy of a stationary particle $\mathcal{E}_0$:

$$\mathcal{E} = \frac{q^2}{\gamma}\int_0^\infty dk = \frac{2}{\gamma}\mathcal{E}_0. \tag{30}$$

The inverse γ dependence has a simple physical interpretation. In the limit that $v \to c$, the source–particle does not *see* its image and, therefore, does no work against it. This result is somewhat idealized and again depends on the point quality of the source as well as other classical properties of the configuration. If the particle passes through a hole of diameter d, one obtains

$$\mathcal{E} \simeq \gamma^{-1}(2\pi q^2/d).$$

While the self-energy $\mathcal{E}_0$ is rendered finite, the modulation γ^{-1} persists. Thus we may conclude that, to within the validity of the stated assumptions, the energy decrement accompanying a high energy particle which passes through a hole in a plate (whose diameter is small compared to plate dimension and initial particle displacement) decreases as the inverse kinetic energy of the particle.

4. OTHER CHARGED CONFIGURATIONS

A. The Charged Ring

To effect solutions relevant to other planar charge configurations, we first recall the form[8] of the static potential for a particle in a grounded cylindrical box of length L, due to a point charge q, at $\bar{z}, \bar{r}, \bar{\theta}$. For $z < \bar{z}$,

$$\phi = \frac{4q}{r_0}\sum_j\sum_m (2 - \delta_{m,0})\frac{\sinh k_j(L - \bar{z})\sinh k_j z}{\sinh k_j L} \times \frac{J_m(k_j\bar{r})J_m(k_j r)}{\alpha_j[J_{m+1}(\alpha_j)]^2}\cos m(\theta - \bar{\theta}).$$

To obtain the potential for a ring of charge q and radius a, centered on axis within a finite closed cylindrical box, one makes the replacement

$$q \to \int dq = q\int_0^{2\pi} d\bar{\theta}\int_0^{r_0} dr\ \delta(a - r).$$

There results (for $z < \bar{z}$)

$$\phi^R = \frac{4q}{r_0}\sum_j \frac{\sinh k_j(L - \bar{z})\sinh(k_j z)}{\sinh(k_j L)}\frac{J_0(k_j a)J_0(k_j r)}{\alpha_j J_1^{\,2}(\alpha_j)}.$$

The only difference between this potential and that due to a point charge q, on axis, is the presence of the coefficient $J_0(k_j a)$. The z dependence of both forms is identical. Image locations are the same. The Lorentz transformations carried out in Ref. 1 are appropriate with the modification that summations over j are replaced as follows.

The Ring

$$\sum_j \to \sum_j J_0(k_j a) \quad \text{for fields}$$

$$\sum_j \to \sum_j J_0^2(k_j a) \quad \text{for energies}$$

These replacements are made in the equations above relevant to a point particle in the various configurations analyzed. Thus, for example, E_z ahead of a charged ring of radius a propagating down a finite closed cylindrical box, is

$$E^R{}_z = \frac{4q}{r_0^2}\sum_j \frac{J_0(k_j a)J_0(k_j r)}{J_1^2(\alpha_j)}\left(\frac{\sinh\Omega_j t\,\cosh K_j\left(1 - \frac{z}{L}\right)}{\sinh K_j} - \sum_p \frac{K_j\Omega_j \sin\omega t\,\cos(\pi p z/L)}{\omega[(\pi p)^2 + K_j^2]}\right).$$

The energy–spectrum in the wake field is given by [see Eq. (12)]

$$\mathcal{E}^R_{pj} = \left(\frac{2\pi}{g}\right)^3 \frac{q^2}{r_0}\frac{\alpha_j^2 J_0^2(\alpha_j a/r_0)}{J_1^2(\alpha_j)}\frac{\sin^2(\bar{\omega}_+/2)}{\beta^2\bar{\omega}_+^2\bar{\omega}_-^2}. \tag{31}$$

Although the energy in all the p modes

$$\sum_p \mathcal{E}^R_{pj} = \mathcal{E}^R_j = \frac{2\pi}{g}\frac{q^2}{r_0}\frac{J_0^2(\alpha_j a/r_0)}{J_1^2(\alpha_j)}$$

still gives a divergent sum, the energy in the $p = 0$ modes

$$\mathcal{E}^R_{0j} = \frac{8\pi}{g}\frac{q^2}{r_0}\frac{J_0^2(\alpha_j a/r_0)}{J_1^2(\alpha_j)}\frac{\sin^2\frac{1}{2}y}{y^2} \tag{32}$$

gives a finite sum (the large j components go as j^{-2}).

B. The Charged Disk

For a uniformly charged disk of total charge q and radius $b < r_0$ at $z = \bar{z}$ centered on axis within a finite, closed cylindrical box one sets

$$q \to \int dq = q\iint[(\bar{r}\,d\bar{r}\,d\bar{\theta})/\pi b^2].$$

There results

$$\phi = \frac{8q}{b}\sum_j \frac{\sinh k_j(L - \bar{z})\sinh k_j z}{\sinh k_j L}\frac{J_1(k_j b)J_0(k_j r)}{\alpha_j^2 J_1^2(\alpha_j)}.$$

Again, the z dependence is the same as for the point particle on axis. To obtain the relativistic fields relevant to the disk in motion, one makes the following replacements in the appropriate equations above.

The Disk

$$\sum_j \to (2r_0/b)\sum[J_1(k_j b)/\alpha_j] \quad \text{For Fields}$$

$$\sum_j \to [2r_0/b]^2\sum[J_1^2(k_j b)/\alpha_j^2] \quad \text{For Energies} \tag{33}$$

For example, B_ϕ behind a charged disk propagating down a semi-infinite can (with an after wall) is given by [from Eq.(2)]

$$B_\phi^D = \frac{8q\gamma\beta}{br_0}\sum_j \frac{J_1(k_j b)J_1(k_j r)}{\alpha_j J_1^2(\alpha_j)}(e^{-\Omega_j t}\cosh\gamma k_j z - W).$$

The spectrum of energies excited in the finite closed box is [from Eq.(12)]

$$\mathcal{E}_{pj}^D = \left(\frac{2r_0}{b}\right)^2\left(\frac{2\pi}{g}\right)^3\frac{q^2}{r_0}\frac{J_1^2(k_j b)}{J_1^2(k_j r_0)}\frac{\sin^2(\frac{1}{2}\bar{\omega}_+)}{\beta^2\bar{\omega}_+^2\bar{\omega}_-^2}. \quad (34)$$

This spectrum gives totally finite results. The energy in all of the p modes, at fixed j, is

$$\mathcal{E}_j^D = \sum_p \mathcal{E}_{pj} = (8Lq^2/b)\{J_1^2(k_j b)/[\alpha_j^2 J_1^2(\alpha_j)]\}. \quad (35)$$

This energy goes like j^{-2} at large j and gives a finite sum

$$\mathcal{E}^D = \frac{8Lq^2}{b^2}\sum_{j=1}^{\infty}\frac{J_1^2(k_j b)}{\alpha_j^2 J_1^2(\alpha_j)} < \infty. \quad (36)$$

This covergence is due to the fact that the fields surrounding a stationary, charged disk are everywhere finite. Furthermore, a charged disk has finite self-energy:

$$\mathcal{E}_0^D = (8/3\pi)(q^2/b),$$
$$\mathcal{E}_j^D = \left(\frac{3\pi L}{b}\right)\mathcal{E}_0^D\frac{J_1^2(\alpha_j b/r_0)}{\alpha_j^2 J_1^2(\alpha_j)}. \quad (37)$$

In the limit that the radius of the disk shrinks to zero,

$$J_1^2(\alpha_j b/r_0) \sim \left(\frac{\alpha_j b}{2r_0}\right)^2\left(1 + \frac{\alpha_j b}{2r_0} + \cdots\right),$$

and

$$\mathcal{E}_j^D \sim \frac{2Lq^2}{r_0^2}\frac{1}{J_1^2(\alpha_j)} = \mathcal{E}_j,$$

the divergent wake energy associated with a point particle which has passed through a finite, closed cavity.

In conclusion, we note that of the results presented above, the following two are readily subject to experimental corroboration: (a) The wake energy spectrum $\mathcal{E}_{pj}$ appropriate to a finite, closed, cylindrical box (Sec 3A); (b) the γ^{-1} modulation of transition radiation excited by a particle passing through a hole in a large plate (Sec. 3D). In applying this result, measurement should find the relative decrement of particle energy

$$\frac{E_{\text{TRANS}}}{E_{\text{INC}}} \simeq 1 - \frac{r_F/d}{\gamma(\gamma-1)},$$

where r_F is the Fermi distance, E_{INC} is the particle's incident kinetic energy, $mc^2(\gamma-1)$, and E_{TRANS} is the particle's final kinetic energy.

APPENDIX A: ASYMPTOTIC INTEGRAL REPRESENTATION OF mth ORDER FOURIER–BESSEL EXPANSIONS

In this appendix we derive a generalization of the transformation given by Eq.(4).

The mth order Fourier–Bessel expansion has the generic form

$$S_m(r, r_0) \equiv \frac{2}{r_0^2}\sum_j \frac{a(k_j, r_0)J_m(k_j r)}{J_{m+1}^2(k_j r_0)} \equiv \sum_j A_j. \quad (A1)$$

The jth zero of $J_m(x)$ is

$$\alpha_j = k_j r_0, \quad J_m(\alpha_j) = 0. \quad (A2)$$

We seek a representation of the above series in the limit $r_0 \to \infty$. In this limit the spectrum of k_j values becomes a continuum over the interval $(0, \infty)$

$$k_j = \alpha_j/r_0 \to k. \quad (A3)$$

Thus, only large α_j contribute to finite k_j values. Using the asymptotic expansion for α_j gives

$$k_j \to \pi j/r_0 = k. \quad (A4)$$

The related asymptotic structure of J_{m+1} gives

$$J_{m+1}^2(\alpha_j) \to (2/\pi\alpha_j)\cos^2\alpha_j \to (2/\pi r_0 k). \quad (A5)$$

For the special case that $a(k_j, r_0)$ is a function only of k_j, one may write

$$\lim_{r_0\to\infty}\sum_j A_j \to \int_0^\infty dj\, A_j = \frac{r_0}{\pi}\int_0^\infty dk\, A_k$$
$$= \int_0^\infty a(k)J_m(kr)k\, dk \equiv S_m(r). \quad (A6)$$

A few applications of this result follow.

Case a: $a(k) = k^{-1}$.

$$\lim_{r_0\to\infty}\frac{2}{r_0^2}\sum_j\frac{J_m(k_j r)}{k_j J_{m+1}^2(k_j r_0)} = \frac{1}{r} \quad m > -1. \quad (A7)$$

Case b: $a(k) = k^{-2}$.

$$\lim_{r_0\to\infty}\frac{2}{r_0^2}\sum_j\frac{J_m(k_j r)}{k_j^2 J_{m+1}^2(k_j r_0)} = \frac{1}{m}, \quad m \geq 1. \quad (A8)$$

Case c: $a(k) = e^{-kd}/k$.

$$\lim_{r_0\to\infty}\frac{2}{r_0^2}\sum_j\frac{J_m(k_j r)e^{-k_j d}}{k_j J_{m+1}^2(k_j r_0)} = \frac{[(d^2+r^2)^{1/2} - d]^m}{r^m\sqrt{d^2+r^2}},$$
$$m > -1. \quad (A9)$$

Note that Case a follows from the latter if one sets $d = 0$.

Example A is also valid for $m = 1$ *and* $J_0(k_j r_0) = 0$: That is,

$$\lim_{r_0\to\infty}\frac{2}{r_0^2}\sum_j\frac{J_1(k_j r)}{k_j J_1^2(k_j r_0)} = \frac{1}{r}. \quad (A10)$$

The validity of this result is related to the fact that both $\{J_0\}$ and $\{J_1\}$, respectively, are orthogonal sequences over the fundamental interval of J_0.

A corollary of the above (A6) is the following. Let

$$a(k_j, r_0) = (r_0)^l a(k_j). \quad (A11)$$

Then, if $l > 0$,

$$\lim_{r_0 \to \infty} S_m(r, r_0) = \infty. \tag{A12}$$

If $l < 0$,

$$\lim_{r_0 \to \infty} S_m(r, r_0) = 0. \tag{A13}$$

These results are valid providing $S_m(r)$ (see A6) is finite.

APPENDIX B: THE SUMMATION OF $\mathcal{E}_{pj}$

In this appendix we wish to perform the summation of $\mathcal{E}_{pj}$ given by Eq. (20), *et seq.* Namely, we wish to show that

$$M \equiv \sum_{p=-\infty}^{\infty} \frac{\sin^2 \bar{\omega}_+/2}{\bar{\omega}_+^2 \bar{\omega}_-^2} = \frac{1}{4y^2}, \qquad y \equiv K/\beta\gamma. \tag{B1}$$

First, rewrite

$$M = \frac{K^4}{2y^4} \sum_p \frac{(1 - \cos\pi p \cos\bar{\omega})}{[(\pi p)^2 + K^2]^2} \equiv \frac{K^4 \bar{M}}{2y^4}, \qquad \beta^2\gamma^2\bar{\omega}^2 = K^2 + (\pi p)^2\gamma^2. \tag{B2}$$

If C denotes a contour of infinitesimal width which encircles the Rez axis, then

$$\bar{M} = \frac{1}{2\pi i} \int_C \frac{dz}{\sin z} \frac{(\cos z - \cos\bar{\omega})}{(z^2 + K^2)^2}. \tag{B3}$$

This integral may be evaluated by separating C at $\pm\infty$ into two parallel segments and then completing these segments, respectively, with infinite semicircles. Two closed paths are thus effected,[9] C_1 and C_2,

$$\int_C dz = \int_{C_1} dz + \int_{C_2} dz.$$

The integrand of $\bar{M}$ is analytic in C_1 and C_2 except at $\pm iK$, respectively, where it has second-order poles. To evaluate $\bar{M}$ we take the derivative of Cauchy's formula, viz.,

$$f'(a) = (1/2\pi i) \oint [dz f(z)/(z - a)^2],$$

where a prime denotes differentiation. If we define

$$\zeta(z) \equiv \cos z - \cos\bar{\omega}, \tag{B4}$$

then

$$-\bar{M} = [\zeta(z)/\sin z(z + iK)^2]'_{+iK} + [\zeta(z)/\sin z(z - iK)^2]'_{-iK}. \tag{B5}$$

Since $\bar{\omega}(\pm iK) = \pm iK$,

$$\zeta(\pm iK) = 0. \tag{B6}$$

One need only differentiate $\zeta(z)$ to obtain the desired result. From the equality

$$\zeta'(z)/\sin z = -1 + (z/\sin z)(\sin\bar{\omega}/\bar{\omega}\beta^2), \tag{B7}$$

it follows that this form is even in z, and

$$\zeta'(z)/\sin z\,|_{\pm iK} = 1/\gamma^2\beta^2. \tag{B8a}$$

Also,

$$[z \pm iK]^2\,|_{\pm iK} = -4K^2. \tag{B8b}$$

Combining these expressions gives

$$\bar{M} = \tfrac{1}{2}\beta^2\gamma^2K^2, \tag{B9}$$

whence

$$M = (K^4/2y^4)\bar{M} = 1/4y^2 \tag{B10}$$

which is the desired result.

APPENDIX C: THE EVALUATION OF $\int \mathbf{E}\cdot\mathbf{J}\,dz$

In this appendix we establish the integration, Eq. (29):

$$\mathcal{E}_j = \frac{2q^2}{r_0^2 J_1^2(\alpha_j)} \int_0^{r_0} dr\, 2\pi r \delta(r) J_0(k_j r) \bar{\mathcal{E}} = \frac{q^2}{\gamma r_0 \alpha_j J_1^2(\alpha_j)} \tag{C1}$$

The relevant factor is

$$\bar{\mathcal{E}} = \int_{-\infty}^0 d\xi \int_{+\infty}^0 dz\, e^{\kappa(\xi - z)}\, \delta(\xi + z) + \int_{-\infty}^0 d\xi \int_{-\infty}^{\xi} dz e^{\kappa(\xi + z)} \times \delta(\xi + z) - \int_{-\infty}^0 d\xi \int_{\xi}^0 dz\, e^{-\kappa(\xi + z)}\, \delta(\xi + z) \equiv \bar{\mathcal{E}}_1 + \bar{\mathcal{E}}_2 + \bar{\mathcal{E}}_3, \tag{C2}$$

where we have set

$$\kappa = \gamma k, \qquad \xi \equiv vt \tag{C3}$$

and $\bar{\mathcal{E}}_{1,2,3}$ are, respectively, defined. The transformation of variables $(\xi, z) \to (y, x)$

$$y = \xi - z, \quad 2\xi = x + y, \qquad x = \xi + z, \quad 2z = x - y \tag{C4}$$

carries a Jacobian of 1/2. The transformation of domains is as shown in Fig. 5. The numbers refer to

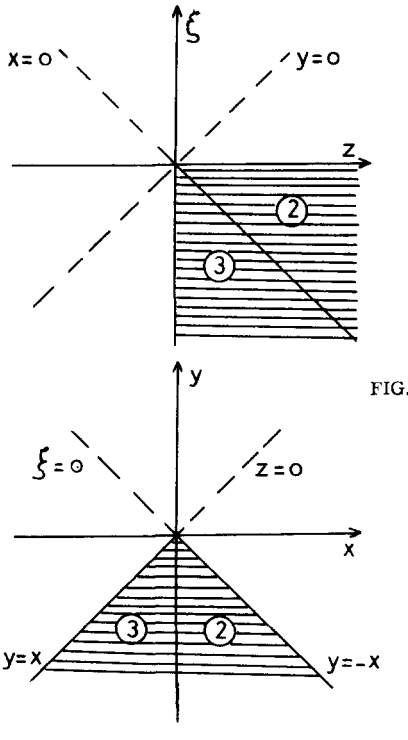


FIG. 5. Domains of integration in (ξ, z) and (y, x) planes.

the domains relevant to the integrals $\bar{\mathcal{E}}_2, \bar{\mathcal{E}}_3$. There results

$$\bar{\mathcal{E}} = \frac{1}{2}\int_{-\infty}^{0} dy\, e^{\kappa y} \int_{-y}^{y} dx\, \delta(x) + \tfrac{1}{2}\int_{-\infty}^{0} dy \left(\int_{-y}^{0} dx\, e^{\kappa x}\delta(x) - \int_{0}^{y} dx\, e^{-\kappa x}\delta(x)\right). \quad \text{(C5)}$$

Only the first integral survives (work done against the image field), and one obtains

$$\bar{\mathcal{E}} = \tfrac{1}{2}(\gamma k)^{-1}. \quad \text{(C6)}$$

(The Jacobian $\frac{1}{2}$ goes to unity for finite problems, e.g., the cylinder of finite length.)

ACKNOWLEDGMENTS

Fruitful discussions with G. Vasu, M. de Haan, P. Mandel, H. Lekkerkerker, and S. Nagarajan are gratefully acknowledged. This work was supported in part under contract N00014-67-A-0077-0015 with the Physics Branch of the Office of Naval Research, and in part by a Solvay Fellowship from the Free University of Brussels.

* Permanent address: College of Engineering, Cornell University, Ithaca, N.Y., 14850.

† Euratom–Etat Belge.

1 R. L. Liboff, J. Math. Phys. **11**, 1295 (1970); **13**, 272 (1972).

2 R. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965).

3 G. M. Garibian, Zh. Eksp. Teor. Fiz. **33**, 1079 (1958); E. Ott and Shmoys, Quart. J. Appl. Math. **25**, 378 (1968), **26**, 188 (1968); F. G. Bass and V. M. Yakovenko, Usp. Fiz. Nauk **86**, 189 (1965) [Sov. Phys. Usp. **8**, 420 (1965)].

4 The normalization of the cylindrical delta function used in this paper also may be found in W. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley, Reading, Mass., 1962), 2nd ed. It differs from that which is found in J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1963). If Jackson's delta function is written $\Delta(r)$, then its relation to $\delta(r)$, which appears above, is $\Delta(r) = 2\pi r\delta(r)$.

5 Equation (5) reduces to the free particle fields listed in J. D. Jackson (Ref. 4, Eq. 11.118) if one sets $z = 0$, $r = b$.

6 Y. N. Dnestrovskii and D. P. Kostomarov, Radioteknika Electron. **4**, 1644 (1959); R. W. King and T. R. Wu, *The Scattering and Diffraction of Waves* (Harvard U. P., Cambridge, Mass., 1959).

7 J. D. Jackson, Ref. 4, p. 382.

8 W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill, New York, 1968), 3rd ed.

9 See Ref. 1, Fig. 4.

Structure of the Bondi-Metzner-Sachs Group

Patrick J. McCarthy

*Department of Physics, University of British Columbia, Vancouver 8, Canada**

(Received 6 March 1972)

A simple example of an aymptotic symmetry group in two dimensions is described. The structure of the corresponding group for asymptotically flat (four-dimensional) space-times, the BMS group, is given explicitly. The recent result that all induced representations of the BMS group have discrete spins is explained in terms of the relationship between this group and the Poincaré group. In fact, it is shown that the BMS group is, in a sense, the smallest generalization of the Poincaré group which eliminates the (physically embarrassing) continuous spin representations of the Poincaré group.

1. INTRODUCTION

The purpose of this paper is twofold. Firstly, I wish to explain to nonspecialists in relativity, by means of a simple analogy, how the BMS group arose and to describe the structure of the group in some detail. Secondly, I wish to explain the "reason" that the induced representations of the BMS group do *not* contain, as a subclass, the (physically embarrassing) "continuous spin" representations of the Poincaré group.[1–4] What is new in this paper is rather trivial (the 2D example, the results of Secs. 4 and 5, and the criterion for distinguishing translations from supertranslations); but I feel that an easily available reference "explaining" the BMS group and giving its structure in detail, is needed, especially since the only available reference giving the structure of the group[5] is very brief and contains misprints (see below). In Sec. 2, the idea of an "asymptotic symmetry group" is illustrated by deriving such a group in two dimensions. In Sec. 3, a simple description of the BMS group is given in convenient coordinates, and this description is related to the original "spherical polar coordinate" description. An "explanation" is given in Sec. 4 of the "elimination of continuous spins," and it is shown in Sec. 5 that the BMS group is the "smallest" group which performs this elimination.

2. THE BONDI GROUP OF THE PLANE

The isometries of a Riemannian manifold are those diffeomorphisms which preserve the metric tensor. In coordinate language, they are those coordinate transformations $x^i \to \bar{x}^i(x)$ which preserve the functional form of the metric tensor, $\bar{g}^{ij}(x) = g^{ij}(x)$ (coordinate indices i, j run over the dimension of the manifold). To compute them, it is simplest to find the "infinitesimal generators" $\xi^i(x)$, given by coordinate transformations $x^i \to \bar{x}^i = x^i + \epsilon\xi^i(x)$ for which the "Killing derivatives" $\mathcal{L}_\xi g^{ij}$ vanish, namely

$$\mathcal{L}_\xi g^{ij} \equiv \lim_{\epsilon\to 0}\{[\bar{g}^{ij}(x) - g^{ij}(x)]/\epsilon\} = g^{ik}\xi^j_{,k} + g^{jk}\xi^i_{,k} - g^{ij}_{,k}\xi^k = 0$$

(Killing equation), where commas denote partial derivatives, and the summation convention is used. With an eye to generalization to the asymptotically flat case, I now compute these generators for the plane, with positive definite metric, in polar coordinates $x^i = (x^1, x^2) = (r, \theta)$, $0 < r < \infty$, $0 \leq \theta < 2\pi$. Thus $g^{ij}(r, \theta) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & r^{-2} \end{smallmatrix}\right]$. Solving $\mathcal{L}_\xi g^{11} = 0$ and $\mathcal{L}_\xi g^{12} = \mathcal{L}_\xi g^{21} = 0$ gives

$$\xi^1 = \zeta(\theta), \qquad \xi^2 = \pi(\theta) + r^{-1}\zeta_{,2}(\theta),$$

the domains relevant to the integrals $\bar{\mathcal{E}}_2, \bar{\mathcal{E}}_3$. There results

$$\bar{\mathcal{E}} = \frac{1}{2}\int_{-\infty}^{0} dy\, e^{\kappa y} \int_{-y}^{y} dx\, \delta(x) + \tfrac{1}{2}\int_{-\infty}^{0} dy \left(\int_{-y}^{0} dx\, e^{\kappa x}\delta(x) - \int_{0}^{y} dx\, e^{-\kappa x}\delta(x)\right). \quad \text{(C5)}$$

Only the first integral survives (work done against the image field), and one obtains

$$\bar{\mathcal{E}} = \tfrac{1}{2}(\gamma k)^{-1}. \quad \text{(C6)}$$

(The Jacobian $\frac{1}{2}$ goes to unity for finite problems, e.g., the cylinder of finite length.)

ACKNOWLEDGMENTS

Fruitful discussions with G. Vasu, M. de Haan, P. Mandel, H. Lekkerkerker, and S. Nagarajan are gratefully acknowledged. This work was supported in part under contract N00014-67-A-0077-0015 with the Physics Branch of the Office of Naval Research, and in part by a Solvay Fellowship from the Free University of Brussels.

* Permanent address: College of Engineering, Cornell University, Ithaca, N.Y., 14850.

† Euratom–Etat Belge.

1 R. L. Liboff, J. Math. Phys. **11**, 1295 (1970); **13**, 272 (1972).

2 R. Rohrlich, *Classical Charged Particles* (Addison-Wesley, Reading, Mass., 1965).

3 G. M. Garibian, Zh. Eksp. Teor. Fiz. **33**, 1079 (1958); E. Ott and Shmoys, Quart. J. Appl. Math. **25**, 378 (1968), **26**, 188 (1968); F. G. Bass and V. M. Yakovenko, Usp. Fiz. Nauk **86**, 189 (1965) [Sov. Phys. Usp. **8**, 420 (1965)].

4 The normalization of the cylindrical delta function used in this paper also may be found in W. Panofsky and M. Phillips, *Classical Electricity and Magnetism* (Addison-Wesley, Reading, Mass., 1962), 2nd ed. It differs from that which is found in J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1963). If Jackson's delta function is written $\Delta(r)$, then its relation to $\delta(r)$, which appears above, is $\Delta(r) = 2\pi r\delta(r)$.

5 Equation (5) reduces to the free particle fields listed in J. D. Jackson (Ref. 4, Eq. 11.118) if one sets $z = 0$, $r = b$.

6 Y. N. Dnestrovskii and D. P. Kostomarov, Radioteknika Electron. **4**, 1644 (1959); R. W. King and T. R. Wu, *The Scattering and Diffraction of Waves* (Harvard U. P., Cambridge, Mass., 1959).

7 J. D. Jackson, Ref. 4, p. 382.

8 W. R. Smythe, *Static and Dynamic Electricity* (McGraw-Hill, New York, 1968), 3rd ed.

9 See Ref. 1, Fig. 4.

Structure of the Bondi-Metzner-Sachs Group

Patrick J. McCarthy

*Department of Physics, University of British Columbia, Vancouver 8, Canada**

(Received 6 March 1972)

A simple example of an aymptotic symmetry group in two dimensions is described. The structure of the corresponding group for asymptotically flat (four-dimensional) space-times, the BMS group, is given explicitly. The recent result that all induced representations of the BMS group have discrete spins is explained in terms of the relationship between this group and the Poincaré group. In fact, it is shown that the BMS group is, in a sense, the smallest generalization of the Poincaré group which eliminates the (physically embarrassing) continuous spin representations of the Poincaré group.

1. INTRODUCTION

The purpose of this paper is twofold. Firstly, I wish to explain to nonspecialists in relativity, by means of a simple analogy, how the BMS group arose and to describe the structure of the group in some detail. Secondly, I wish to explain the "reason" that the induced representations of the BMS group do *not* contain, as a subclass, the (physically embarrassing) "continuous spin" representations of the Poincaré group.[1–4] What is new in this paper is rather trivial (the 2D example, the results of Secs. 4 and 5, and the criterion for distinguishing translations from supertranslations); but I feel that an easily available reference "explaining" the BMS group and giving its structure in detail, is needed, especially since the only available reference giving the structure of the group[5] is very brief and contains misprints (see below). In Sec. 2, the idea of an "asymptotic symmetry group" is illustrated by deriving such a group in two dimensions. In Sec. 3, a simple description of the BMS group is given in convenient coordinates, and this description is related to the original "spherical polar coordinate" description. An "explanation" is given in Sec. 4 of the "elimination of continuous spins," and it is shown in Sec. 5 that the BMS group is the "smallest" group which performs this elimination.

2. THE BONDI GROUP OF THE PLANE

The isometries of a Riemannian manifold are those diffeomorphisms which preserve the metric tensor. In coordinate language, they are those coordinate transformations $x^i \to \bar{x}^i(x)$ which preserve the functional form of the metric tensor, $\bar{g}^{ij}(x) = g^{ij}(x)$ (coordinate indices i, j run over the dimension of the manifold). To compute them, it is simplest to find the "infinitesimal generators" $\xi^i(x)$, given by coordinate transformations $x^i \to \bar{x}^i = x^i + \epsilon\xi^i(x)$ for which the "Killing derivatives" $\mathcal{L}_\xi g^{ij}$ vanish, namely

$$\mathcal{L}_\xi g^{ij} \equiv \lim_{\epsilon\to 0}\{[\bar{g}^{ij}(x) - g^{ij}(x)]/\epsilon\} = g^{ik}\xi^j_{,k} + g^{jk}\xi^i_{,k} - g^{ij}_{,k}\xi^k = 0$$

(Killing equation), where commas denote partial derivatives, and the summation convention is used. With an eye to generalization to the asymptotically flat case, I now compute these generators for the plane, with positive definite metric, in polar coordinates $x^i = (x^1, x^2) = (r, \theta)$, $0 < r < \infty$, $0 \leq \theta < 2\pi$. Thus $g^{ij}(r, \theta) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & r^{-2} \end{smallmatrix}\right]$. Solving $\mathcal{L}_\xi g^{11} = 0$ and $\mathcal{L}_\xi g^{12} = \mathcal{L}_\xi g^{21} = 0$ gives

$$\xi^1 = \zeta(\theta), \qquad \xi^2 = \pi(\theta) + r^{-1}\zeta_{,2}(\theta),$$

where $\zeta(\theta)$ and $\pi(\theta)$ are arbitrary (differentiable) functions. $\mathcal{L}_\xi g^{22} = 0$ gives $2r^{-2}\pi_{,2}(\theta) + 2r^{-3} \times [\zeta_{,22}(\theta) + \zeta(\theta)] = 0$. This can only be true for all r if $\pi_{,2}(\theta) = 0$ and $\zeta_{,22}(\theta) + \zeta(\theta) = 0$. So the solutions are

$$\xi^1 = \zeta^{(a,b)}(\theta) \equiv a\cos\theta + b\sin\theta,$$

$$\xi^2 = K + r^{-1}\zeta^{(a,b)}_{,2}(\theta),$$

where a, b, K are constants. So there are three linearly independent generators ξ^i, and, as is well known, they describe the infinitesimal group of motions of the plane, a and b corresponding to translations, K to rotations. The full isometry group (excluding reflections) $E(2)$ is well known to have the structure of a semidirect product

$$E(2) = R^2 \circledS_T SO(2),$$

R^2 and $SO(2)$ being the translation and rotation groups in two dimensions. Thus, $E(2)$ has underlying space $R^2 \times SO(2)$, the remaining structure of the semidirect product being specified by the action T of $SO(2)$ on R^2 given by rotation of R^2 vectors; so the product law is

$$(v_1, R_1)(v_2, R_2) = (v_1 + T(R_1)v_2, R_1R_2),$$

where

$$T(R)v \equiv Rv.$$

Now consider any positive definite two-dimensional differentiable Riemannian manifold, with metric given by $g^{ij}(x)$ in some coordinate patch $\{x^i\}$, $i = 1, 2$. Define, in this patch, two differentiable scalar functions $r(x)$ and $\theta(x)$, assumed to satisfy conditions appropriate to "polar" coordinates, namely

$$g^{ij}r_{,i}r_{,j} = 1, \quad g^{ij}r_{,i}\theta_{,j} = 0.$$

Use r and θ as new coordinates so that, in the patch $\{x^i\}$, the metric tensor takes the form $g^{ij}(r, \theta) = \left[\begin{smallmatrix} 1 & 0 \\ 0 & f(r,\theta) \end{smallmatrix}\right]$, where $f(r, \theta)$ is an arbitrary (differentiable) function. Now impose conditions that the manifold be "asymptotically flat." A polar form for the coordinates has been chosen to facilitate this. The conditions are

(a) that the patch $\{x^i\}$ cover the whole manifold outside some bounded region, the coordinate ranges outside the region being $r_0 \leq r < \infty$ and $0 \leq \theta < 2\pi$ for some $r_0 > 0$;

(b) that, in these ranges, $f(r, \theta) = r^{-2} + O(r^{-\lambda})$ with $2 < \lambda \leq 3$;

(c) that $(\partial/\partial r)[f(r, \theta) - r^{-2}] = O(r^{-q})$ with $q \geq \lambda$ and that the derivative with respect to θ satisfy the same condition.

[The order symbol has its usual meaning, $h(r, \theta) = O(r^{-\alpha})$ if $\lim r^\beta |h(r, \theta)| = 0$ as $r \to \infty$ for all θ and any $\beta < \alpha$.] By transforming to "cartesians," one easily shows that $g^{ij} \to \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ in cartesians as $r \to \infty$. So the metric describing a wide class of "asymptotically flat" manifolds is $g^{ij} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & r^{-2+O(r^{-\lambda})} \end{smallmatrix}\right]$. Unless the $O(r^{-\lambda})$ term has certain very simple forms, there are no nontrivial isometries, and this is reflected in the fact that the Killing equations $\mathcal{L}_\xi g^{ij} = 0$ have no non-zero solutions. However, if one requires only *asymptotic* isometries, namely, transformations that preserve the metric asymptotically,

$$\mathcal{L}_\xi g^{11} = 0, \quad \mathcal{L}_\xi g^{12} = \mathcal{L}_\xi g^{21} = 0,$$

and

$$\mathcal{L}_\xi g^{22} = O(r^{-\lambda}),$$

there *are* solutions, which, for large r, are independent of the $O(r^{-\lambda})$ term. Indeed, solving the first two equations, one finds

$$\xi^1 = \zeta(\theta),$$

$$\xi^2 = \pi(\theta) + r^{-1}\zeta_{,2}(\theta) + O(r^{-\mu}), \quad 1 < \mu \leq 2,$$

$\zeta(\theta)$ and $\pi(\theta)$ being arbitrary differentiable functions. The last equation gives

$$2r^{-2}\pi_{,2}(\theta) + 2r^{-3}[\zeta_{,22}(\theta) + \zeta(\theta)] + O(r^{-s}) = O(r^{-\lambda}), \quad s \geq \lambda.$$

Multiplying this by r^2 and making $r \to \infty$, one finds that $\pi_{,2}(\theta) = 0$, so that $\pi(\theta) = K$, a constant. However, since $2 < \lambda \leq 3$, one gets *no condition on* $\zeta(\theta)$. So the final solutions are

$$\xi^1 = \zeta(\theta), \quad \xi^2 = K + r^{-1}\zeta_{,2}(\theta) + O(r^{-\mu}), \quad 1 < \mu \leq 2.$$

One sees that the solutions are, for large r, independent of the $O(r^{-\lambda})$ term. Comparing these solutions with the solutions for the flat space case, one sees that the *two*-parameter family $\zeta^{(a,b)}(\theta)$ have become enlarged to an *infinite* parameter family of *arbitrary* functions $\zeta(\theta)$. This is analogous to what happens in the space-time case. Note that, if the calculation had been performed with $\lambda > 3$, the solutions would have been

$$\xi^1 = \zeta^{(a,b)}(\theta) = a\cos\theta + b\sin\theta,$$
$$\xi^2 = K + r^{-1}\zeta^{(a,b)}_{,2}(\theta) + O(r^{-\mu}), \quad \mu > 2.$$

Thus, if the metric approaches flatness too rapidly as $r \to \infty$, the $\zeta^{(a,b)}(\theta)$ functions do *not* become enlarged to an infinite parameter family.

The finite coordinate transformations which preserve the metric asymptotically can be computed by a method analogous to the original method for deriving the BMS group (see references cited in Ref. 1). One gets

$$r \to \bar{r} = r + \zeta(\theta) + \tfrac{1}{2}r^{-1}[\zeta_{,2}(\theta)]^2 + O(r^{-2}),$$
$$\theta \to \bar{\theta} = \theta + K + r^{-1}\zeta_{,2}(\theta) + O(r^{-2}).$$

By performing successive transformations, one verifies that the asymptotic symmetry group $E^B(2)$ has the structure of a semidirect product $\{\zeta\} \circledS_T SO(2)$, $\{\zeta\}$ being the Abelian group of differentiable functions $\zeta(\theta)$ under pointwise addition and the action T of $SO(2)$ on $\{\zeta\}$ being given by (K is the rotation angle) $(T(K)\zeta)(\theta) = \zeta(\theta - K)$. Restricting $\{\zeta\}$ to the subgroup of functions of the form $\zeta^{(a,b)}(\theta) = a\cos\theta + b\sin\theta$ restricts $E^B(2)$ to $E(2)$. I have found $E^B(2)$ a most useful paradigm for the BMS group.

3. THE STRUCTURE OF B

The BMS group B arises from a similar calculation to the above one in the case of asymptotically flat (four-dimensional) space-times (see references cited in Ref. 1). It was originally derived in "null polar" coordinates. Here I first describe it in more convenient coordinates. First, recall that the Poincaré group P is a semidirect product of the Abelian translation group V in four dimensions with the (connected component of) the Lorentz group L, $P = V\circledS_T L$, the action T of L on V being given by multiplication of the translations a^μ by 4×4 Lorentz matrices Λ^μ_ν, $\mu, \nu = 0, 1, 2, 3$; thus $(T(\Lambda)a)^\mu \equiv (\Lambda a)^\mu = \Lambda^\mu_\nu a^\nu$ (summation convention). Associate with V the four-parameter family of functions $\{f^a\}$ defined on the future pointing Minkowski null cone N as follows:

$$f^a(n) = n \cdot a \text{ (Lorentz scalar product)}, \qquad n \in N, \quad a \in V.$$

Addition is defined in $\{f^a\}$ in the usual way, $(f^a + f^b)(n) = f^a(n) + f^b(n) = f^{a+b}(n)$. Note that these functions are homogeneous of degree 1 in n:

$$f^a(tn) = tf^a(n) \quad \text{all } t > 0.$$

The action T of L on V induces an action on $\{f^a\}$ as follows:

$$(T(\Lambda)f^a)(n) \equiv n \cdot T(\Lambda)a = n \cdot \Lambda a = \Lambda^{-1}n \cdot a = f^a(\Lambda^{-1}n).$$

Thus one sees that P may be written

$$P = \{f^a\}\circledS_T L \quad \text{where} \quad (T(\Lambda)f^a)(n) = f^a(\Lambda^{-1}n).$$

The BMS group B is obtained from P by replacing the four-parameter family $\{f^a\}$ of functions with the above homogeneity property by the set $\{f\}$ of *all* (differentiable) functions with the same homogeneity property, the action of L being analogous to that in P;

$$B = \{f\}\circledS_T L, \quad f(tn) = tf(n) \quad \text{all } t > 0,$$
$$\text{and} \qquad (T(\Lambda)f)(n) = f(\Lambda^{-1}n).$$

[Addition is defined in $\{f\}$ by $(f_1 + f_2)(n) = f_1(n) + f_2(n)$]. Thus B differs from P in that the four-parameter family of functions $\{f^a\}$ ("translations") become replaced by the infinite parameter family $\{f\}$ ("supertranslations"). (In the Appendix, a criterion for distinguishing homogeneous functions of the form f^a from other homogeneous functions is given.) Obviously, $\{f^a\}$ is an invariant subspace of $\{f\}$, and restricting $\{f\}$ to $\{f^a\}$ in B gives P. In the calculation of B, conditions analogous to (a), (b), and (c) of Sec. 2 are put on a space-time. The analog of (a) is the use of null polar coordinates, of (b) are the "asymptotic flatness" conditions, and of (c) are the so-called "asymptotic smoothness" conditions (see references cited in Ref. 1). Again, if the metric of the space-time approaches flatness too quickly as $r \to \infty$, $\{f^a\}$ is not enlarged to $\{f\}$. So B only arises when gravitational radiation is present. To obtain the realization of B in the coordinates in which it was originally obtained, one may proceed as follows. Associate with each line generator of the null cone N the point of intersection of this line with the 3-space $n^0 = 1$ (n^0 is the zero component of n^μ). This set of points is the unit sphere S in 3-space, and, since the Lorentz group L takes line generators into line generators, this gives an action of L on S, the "conformal" action. Defining the null vector $l^\mu = (1, m^1, m^2, m^3)$, where $(m^1, m^2, m^3) = m^i$ is a unit 3-vector, and associating this unit 3-vector with the point $p \in S$, one finds that the conformal action $p \to \Lambda p$ is given by $m^i \to (\Lambda^0_\mu l^\mu)^{-1}(\Lambda^i_\mu l^\mu)$. The "conformal factor" associated with the action is defined as $K_\Lambda(p) = \Lambda^0_\mu l^\mu$. In view of the homogeneity condition on the functions $\{f\}$, every $f \in \{f\}$ may be associated with a differentiable function α on S, and conversely, for every differentiable α on S, an $f \in \{f\}$ may be found. The action of L on $\{f\}$ thus gives an action on the corresponding set $\{\alpha\}$, which is easily computed as

$$(T(\Lambda)\alpha)(p) = K_{\Lambda^{-1}}(p)\alpha(\Lambda^{-1}p).$$

This finally gives the original realization of B, as a semidirect product of the Abelian group $\{\alpha\}$ of differentiable functions on S with L, $B = \{\alpha\}\circledS_T L$, where the action T of L on $\{\alpha\}$ is the action just given. Historically, this semidirect product structure of B was realized by Cantoni (see third reference of Ref. 1). As printed, the formula given in Geroch and Newman[5] for the action is incorrect (top of second column). In the present notations, it would read $(T(\Lambda)\alpha)(p) = K_\Lambda(p)\alpha(\Lambda p)$, which gives $T(\Lambda_1)T(\Lambda_2) = T(\Lambda_2\Lambda_1)$. The correct formula is the one given above.

Henceforth, the Abelian group of supertranslations $\{f\}$ or $\{\alpha\}$ will be written as A and the translation subgroup as V. It has been noted that V is an invariant subspace of A under the action T of L on A (so that V is a normal subgroup of A) and that restricting A to V restricts B to P. As noted by Sachs,[6] the complementary subspace Σ to V in A is *not* invariant under the action. This fact plays a crucial role in the representation theory of B.

4. WHY B ELIMINATES CONTINUOUS SPINS

In view of the fact that B *contains* P, some people find it surprising that the representations of B exclude the continuous spin representations of P.[1-4] (In this section, "representation" always means "unitary representation.") In order to understand this fact, it is necessary to recall the ingredients required for the construction of induced representations of semidirect products.[7] Let W be a vector space regarded as an Abelian group, H a group, and $G = W\circledS_J H$ be a semidirect product specified by an action J of H on W. To find the induced representations, one first finds the set of all irreducible representations of W. This set may be identified with the dual vector space W' ("momentum" space). The action J of H on W determines a dual action J' of H on W' in the usual way. *It is the action J' of H on W' which determines the structure of the irreducible representations. In particular, this action determines the "orbits" and "little groups."*[7] For example, in the case of $P = V\circledS_T L$, the action T' of L on V' gives rise to the well-known orbits and little groups for P. The little groups of timelike, lightlike, and spacelike vectors are the 3D rotation group, the group of null rotations [identifiable with the motion group $E(2)$ of the plane] and the

(2 + 1) Lorentz group. Of these little groups, the first is compact; but the other two are noncompact, containing "boosts" arbitrarily close to "speed of light boosts." The compact little group gives rise to discrete spins; the noncompact ones to continuous spins.

In the case of $B = A \circledS_T L$, the action T of L on A is, as remarked above, such that V is invariant, but the complement Σ is not. *On passing to the dual action T' of L on A', the subspace Σ' dual to Σ becomes invariant, and the complementary subspace V', which is dual to V, becomes noninvariant.*[1] (If the complement *were* invariant, V' would be invariant in A'.) Since V' is not invariantly contained in A' under T', the usual action of L on Poincaré momentum space, discussed in the preceding paragraph, is not "contained" in the action of L in BMS "supermomentum" space A'. In particular, there is no reason that the BMS little groups should include Poincaré little groups. In fact, the little groups all turn out to be compact for B.[1-4] One may understand this intuitively as follows. It turns out that the action T' of L on A' may be realized in terms of functions ϕ on S as[1]

$$(T'(\Lambda)\phi)(p) = K^{-3}_{\Lambda^{-1}}(p)\phi(\Lambda^{-1}p).$$

This action is similar to the transformation law for intensity distributions on distant ("celestial") spheres under L. (The latter action appears with a $K^{-2}_{\Lambda^{-1}}$ rather than $K^{-3}_{\Lambda^{-1}}$ factor; but the analogy works nevertheless). The K factor corresponds physically to a redshift factor and the conformal action $p \to \Lambda p$ to "aberration." A subgroup Q of L is a little group if it leaves a (nonzero) ϕ fixed under the action T'. Now one can show[2] that Q must be closed, and if it is noncompact, it must contain a sequence of pure boosts whose velocity parameter becomes arbitrarily close to the speed of light. For an intensity distribution ϕ to be fixed under Q, then, it must be such that, when subjected to arbitrarily large "blueshifts" over a region of the sphere with arbitrarily small complement (the complement is redshifted), it remains the same. It is not surprising that there are no such nonzero ϕ's, so that there cannot be any noncompact little groups. So B spins are always discrete.

5. UNIQUENESS OF B

I have indicated that the essential reason that the little groups of B do not contain those of P is that, though V is invariant under the action T defining B, the complement Σ is not. In other words, the representation T of L on A is reducible (to V), but not decomposable. One may well ask—is it possible to generalize the Poincaré group by "enlarging" V to a *finite*-dimensional vector space Y in such a way that the action of L on Y was reducible (to V) but *not* decomposable? One could then hope that such a *finite*-dimensional enlargement of P would eliminate the continuous spins. The answer is no, for the following reason. Any finite-dimensional representation of a connected semi-simple Lie group is necessarily decomposable.[8] Since L is connected and simple, the result follows.

Formulating things a little more precisely, I will call any semidirect product of L with a vector space Z, in which the action T of L on Z is reducible (to V) a *superextension* of P. The above result then reads—*"there are no finite-dimensional superextensions of P which eliminate the continuous spins."* I will now show that there is a sense in which B is the *smallest* superextension of P which does eliminate the continuous spins. For this purpose, let the (now, necessarily infinite-dimensional) vector space defining the superextension be Y, so that one is given an action T of L on Y, reducible to V. If the superextension is to eliminate continuous spins, the complement of V in Y cannot be invariant. Now, there is the following possibility. Y may decompose into two invariant subspaces Y_1 and Y_2, $Y = Y_1 \oplus Y_2$, one of which contains the invariant subspace V (this one having the complement of V in it noninvariant, so necessarily being infinite-dimensional). To exclude these situations, I require that Y be irreducible in the usual *infinite*-dimensional sense of the word, that is, operator irreducible.[9] In other words, every (closed) operator which commutes with all the representations of L on Y must necessarily be a multiple of the unit operator. In particular, this excludes the splitting $Y_1 \oplus Y_2$ mentioned above; but it does not exclude the possibility that Y contains V with a noninvariant complement. But it is shown in the Appendix that there is one and only one real, operator irreducible representation of L which contains the invariant subspace V, and that is the supertranslation space A.

Hence B is the smallest superextension of P which eliminates the continuous spins.

APPENDIX

First, I shall identify the representation T of L on A with one of the operator irreducible representations of L defined by Gel'fand *et al.*[9] Consider the complex space of functions of two complex variables z_1, z_2 satisfying

(i) $F(z_1, z_2; \bar{z}_1, \bar{z}_2)$ is (infinitely) differentiable in its arguments everywhere except at (0, 0);

(ii) $F(az_1, az_2; \bar{a}\bar{z}_1, \bar{a}\bar{z}_2) = a^{n_1-1}\bar{a}^{n_2-1}F(z_1, z_2; \bar{z}_1, \bar{z}_2)$ for all complex $a \neq 0$,

where n_1 and n_2 are any complex numbers differing by an integer. Define an action of the group G of unimodular 2×2 complex matrices in the space $D_{(n_1,n_2)}$ of functions defined by a fixed pair (n_1, n_2) above as follows:

$$(T(g)F)(z_1, z_2; \bar{z}_1, \bar{z}_2) = F(\alpha z_1 + \gamma z_2, \beta z_1 + \delta z_2; \quad = \overline{\alpha z_1 + \gamma z_2}, \overline{\beta z_1 + \delta z_2}),$$

where $g = \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right] \in G$. Then[9] every operator irreducible representation of G is equivalent to one of the above form (with a "dense subspace" definition of equivalence.[9]) [The operator irreducible representations of G which have a single nontrivial invariant subspace are those in which $(n_1 n_2)$ are integers of the same sign. The subspace is finite-dimensional, consisting of homogeneous polynomials, if $(n_1 n_2)$ are both positive integers and infinite-dimensional if they are negative integers. All others are irreducible in the "subspace" sense.]

Next note that if the well-known homomorphism of G onto L, whose kernel is the subgroup Z_2 (consisting of plus and minus the unit matrix) of G, is employed, every statement about L and its representations

throughout this paper can be written as a statement about G—the representation defining B is unfaithful on Z_2. (In Ref 1, B is *defined* as $A \circledS G$.) I shall now show that the representation of L on A corresponds to the representation of G on $D_{(2,2)}$. Write $z_A = (z_1, z_2)$ and $\bar{z}_{A'} = (\bar{z}_1, \bar{z}_2)$ for convenience. The well-known correspondence between these "spinors" and future pointing null vectors is given by $n_\mu = \sigma_\mu^{AA'} z_A \bar{z}_{A'}$ (summation convention), n_μ being the *covariant* null vector, $\sigma_\mu^{AA'}$ being the van der Waerden matrices.[10] Thus, $z_A, \bar{z}_{A'}$ uniquely determine n_μ, and n_μ determines $z_A, \bar{z}_{A'}$ up to a phase factor $z_A \to e^{i\theta} z_A$, $\bar{z}_{A'} \to e^{-i\theta} \bar{z}_{A'}$. In terms of this correspondence, the homomorphism $G \to L$ mentioned above is given by $g = \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right] \to \Lambda^\mu_\nu \in L$ with

$$n_\nu \Lambda^\nu_\mu = \sigma_\mu^{AA'} (zg)_A \overline{(zg)}_{A'}, (zg)_A = (\alpha z_1 + \gamma z_2, \beta z_1 + \delta z_2).$$

Now, consider the action $(T(\Lambda)f)(n) = f(\Lambda^{-1}n)$ on the homogeneous functions f which describes the action of L on A in B. By an obvious reformulation of Sec. 3, this action may be redefined in terms of an action of L on N by *right* translation, $(T(\Lambda)f)(n) = f(n\Lambda)$, the n's now understood as covariant. Associate with $f \in \{f\}$ the function $F(z_A, \bar{z}_{A'})$ by writing $F(z_A, \bar{z}_{A'}) = f(n_\mu)$ $= f(\sigma_\mu^{AA'} z_A \bar{z}_{A'})$. Since f has the homogeneity property and n_μ only determines $z_A, \bar{z}_{A'}$ up to a phase, one finds a one-to-one correspondence between the f's and those differentiable functions $F(z_A, \bar{z}_{A'})$ satisfying

$$F(te^{i\theta} z_A, te^{-i\theta} \bar{z}_{A'}) = t^2 F(z_A, \bar{z}_{A'}) \text{ for real } t, \theta \text{ with } t > 0$$

or

$$F(az_A, \overline{az}_{A'}) = a\bar{a} F(z_A, \bar{z}_{A'}) \text{ for all complex } a \neq 0.$$

Further, the action $(T(\Lambda)f)(n) = f(n\Lambda)$ corresponds to $(T(g)F)(z_A, \bar{z}_{A'}) = F((zg)_A, \overline{(zg)}_{A'})$ in this correspondence. Comparing with (i) and (ii), one sees, finally, the action of L on A is equivalent to the $D_{(2,2)}$ representation (except for the fact that A is a space of real functions, so that the $D_{(2,2)}$ here must be understood as the real subspace of Gel'fand's complex $D_{(2,2)}$[9]). The penultimate sentence of Sec. 5 is thus proved. Using the homogeneity property of the F's, one may associate, with each such F, an arbitrary (differentiable) function α of *one* complex variable z (including $z = \infty$) according to the scheme

$$\begin{aligned} F(z_1, z_2; \bar{z}_1, \bar{z}_2) &= r^2 F(z_1/r, z_2/r; \bar{z}_1/r, \bar{z}_2/r) \\ &= r^2 F(z/p(z), 1/p(z); \bar{z}/p(z), 1/p(z)) \equiv r^2 \alpha(z, \bar{z}) \\ &\text{where } r \equiv (|z_1|^2 + |z_2|^2)^{1/2}, z \equiv z_1/z_2 \text{ and} \\ &p(z) = (1 + |z|^2)^{1/2} \end{aligned}$$

and conversely. This gives the realization of the action T of G on A used in Ref. 1. (I feel that the considerable detail of the preceding discussion is necessary because this important identification $A \approx D_{(2,2)}$ is nowhere proved in the literature. See, however, the final remark following Ref. 5. below.)

Finally, I wish to arrive at a criterion which distinguishes the homogeneous functions $\{f^a\}$ corresponding to *translations* from the remaining ones (see Sec. 3). First, I shall prove the following.

Theorem: Let h: $R^n \to R$ be a real-valued function on $n - D$ real space, satisfying $h(tx) = t^q h(x)$ for some real q and all real $t > 0$. Then, if h is everywhere infinitely differentiable, $h(x)$ is a homogeneous polynomial in x if q is a nonnegative integer; otherwise it is zero.

Proof: Suppose first that q is a nonnegative integer. Differentiate the homogeneity relation q times with respect to t to give

$$x^{i_1} x^{i_2} \cdots x^{i_q} h_{,i_1 i_2 \cdots i_q}(tx) = q! h(x),$$

where x^i is the ith component of $x \in R^n$, the summation convention is used, and commas denote partials. Taking the limit $t \to 0$ and using the continuity of the qth mixed partials at $0 \in R^n$ gives the result. Next, if $q < 0$, multiplying both sides of the homogeneity relation by t^{-q} and taking the limit $t \to 0$, using the continuity of h at $0 \in R^n$, gives $h(x) \equiv 0$. Finally, if q is positive but nonintegral, write $q = n - r$, where n is a positive integer and $0 < r < 1$, differentiate n times with respect to t, multiply by t^r, and take the limit $t \to 0$ using the continuity of the nth mixed partials to give $h(x) \equiv 0$. QED

[This easy theorem, which was conjectured by Mr. J. Pulham in the case $q = 1$, with the stronger requirement that $h(tx) = t^q h(x)$ for all real $t \geq 0$ and $h(x)$ everywhere differentiable, must be well known to some people; but I have not seen it anywhere.] Obvious generalizations are requirements that h be only k times differentiable, for k chosen appropriately for a given q, or that there exist a neighborhood of $0 \in R^n$ for which h is differentiable. Functions satisfying the conditions of the theorem differ from homogeneous functions in that they are required to be differentiable everywhere *including* the origin $0 \in R^n$.

To distinguish functions f: $N \to R$ (notation as Sec. 3) of the form f^a, with $f^a(n) = a \cdot n$ from other homogeneous functions, differentiate each side of $f(tn^\mu) = tf(n^\mu)$ with respect to t to give $n^\mu f_{,\mu}(tn) = f(n^\mu)$. If f is differentiable everywhere on N *including the vertex*, take the limit $t \to 0$ to give $f(n^\mu) = k_0 n^0 + k_1 n^1 + k_2 n^2 + k_3 n^3$, where k_μ are constants $f_{,\mu}(0)$. These are clearly of the form f^a. Thus the translations are those f's in $\{f\}$ which are differentiable everywhere *including the vertex*. This gives the promised criterion.

It may be of interest to note that the theorem is easily adapted, by differentiating (ii) with respect to a and $\bar{a}$ and taking the limit $|a| \to 0$, to prove that *the finite-dimensional representations of G (given by homogeneous polynomials, so that n_1 and n_2 are both nonnegative integers)*[9] *are distinguished from the infinite-dimensional ones in that they are given by homogeneous functions differentiable in the z's everywhere including the origin (0, 0).*

---

* All results below appear in the author's Ph. D. thesis.[3]

1 P. J. McCarthy, "Representations of the BMS Group. I," to appear in Proc. Roy. Soc. (London) (1972).

2 P. J. McCarthy, "Representations of the BMS Group. II," submitted to Proc. Roy. Soc. (London) (1972).

3 P. J. McCarthy, Ph.D. thesis (University of London, King's College 1971).

4 P. J. McCarthy, Report to Trieste conference, Preprint No.

1C/71/144 July (1971). [The sentence in Ref. 4 which starts on line 9, p. 26 should read: "The others, which have a fixed value of the mass, contain, for $m^2 > 0$ and $m^2 \leq 0$, respectively, direct sums or integrals of different appropriate Poincaré spins."] See Ref. 2 above.

5 R. J. Geroch and E. T. J. Newman, J. Math. Phys. **12**, 314 (1971). [As printed in Ref. 5, the definition of the invariance group of electrodynamics is also incorrect, since it is given there as a direct product. The sentence beginning on line 11, second column, of Ref. 5 should end "$Pf(q) = f(P^{-1}q)$, $q \in M$." However, the identification in this reference of the representation of L which specifies the BMS semidirect product, proved explicitly above, is correct.]

6 R. K. Sachs, Phys. Rev. **128**, 2851 (1962), commutation relations (IV. 19) on p. 2859.

7 G. W. Mackey, "Theory of Group Representations," Chicago notes (1955) and references cited there.

8 R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

9 I. M. Gel'fand, M. I. Graev, N. Ya Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.

10 Brandeis Summer Institute, *General Relativity*, chapter by F.A.E. Pirani (Prentice Hall, Englewood Cliffs, N.J., 1964).

Erratum: Exchange Interaction Model of Ferromagnetism

[J. Math. Phys. **13**, 725 (1972)]

H. H. Chen and R. I. Joseph

Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 9 June 1972)

(1) The correct name of the second author of this paper is R. I. Joseph.

(2) The numbers given for the coefficients a_7 for the plane square lattice in Appendix B are incorrect and should be replaced by

$$-0.321\ 825, -0.700\ 033, -1.090\ 470, -1.205\ 625, \text{and} -1.188\ 391$$

for $S = \frac{1}{2} - \frac{5}{2}$, respectively

(3) The right-hand side of Eq. (15) should be multiplied by the factor $n!$.

1C/71/144 July (1971). [The sentence in Ref. 4 which starts on line 9, p. 26 should read: "The others, which have a fixed value of the mass, contain, for $m^2 > 0$ and $m^2 \leq 0$, respectively, direct sums or integrals of different appropriate Poincaré spins."] See Ref. 2 above.

5 R. J. Geroch and E. T. J. Newman, J. Math. Phys. **12**, 314 (1971). [As printed in Ref. 5, the definition of the invariance group of electrodynamics is also incorrect, since it is given there as a direct product. The sentence beginning on line 11, second column, of Ref. 5 should end "$Pf(q) = f(P^{-1}q)$, $q \in M$." However, the identification in this reference of the representation of L which specifies the BMS semidirect product, proved explicitly above, is correct.]

6 R. K. Sachs, Phys. Rev. **128**, 2851 (1962), commutation relations (IV. 19) on p. 2859.

7 G. W. Mackey, "Theory of Group Representations," Chicago notes (1955) and references cited there.

8 R. Hermann, *Lie Groups for Physicists* (Benjamin, New York, 1966).

9 I. M. Gel'fand, M. I. Graev, N. Ya Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.

10 Brandeis Summer Institute, *General Relativity*, chapter by F.A.E. Pirani (Prentice Hall, Englewood Cliffs, N.J., 1964).

Erratum: Exchange Interaction Model of Ferromagnetism

[J. Math. Phys. **13**, 725 (1972)]

H. H. Chen and R. I. Joseph

Department of Electrical Engineering, The Johns Hopkins University, Baltimore, Maryland 21218

(Received 9 June 1972)

(1) The correct name of the second author of this paper is R. I. Joseph.

(2) The numbers given for the coefficients a_7 for the plane square lattice in Appendix B are incorrect and should be replaced by

$$-0.321\ 825, -0.700\ 033, -1.090\ 470, -1.205\ 625, \text{and} -1.188\ 391$$

for $S = \frac{1}{2} - \frac{5}{2}$, respectively

(3) The right-hand side of Eq. (15) should be multiplied by the factor $n!$.