JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 5, NUMBER 3 MARCH 1964

A No-Interaction Theorem in Classical Relativistic Hamiltonian Particle Dynamics*

JOHN T. CANNON† AND THOMAS F. JORDAN‡
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received 31 May 1963)

It is shown that a relativistically invariant classical mechanical Hamiltonian description of a system of three (spinless) particles admits no interaction between the particles. If a set of ten functions of the canonical variables of the three-particle system satisfies the Poisson bracket relations characteristic of the ten generators of the inhomogeneous Lorentz group, and—with the canonical position variables of the particles—satisfies the Poisson bracket equations which express the familiar transformation properties of the (time-dependent) particle positions under space translation, space rotation, and Lorentz transformation, then this set of ten functions can only describe a system of three free particles. A significant part of the proof is valid for a system containing any fixed number of particles. In this general case, a simplified form is established for the Hamiltonian and generators of Lorentz transformations, and it is shown that the generators of space translations and space rotations can be put in the standard form characteristic of free-particle theories. The proof of the latter involves a generalization from one to many three-vector variables of the angular momentum Helmholtz theorem of Lomont and Moses.

I. INTRODUCTION

WE shall continue here the investigation begun by Currie, Jordan, and Sudarshan[1,2] of classical mechanical relativistically invariant Hamiltonian theories of systems of spinless particles. Our results lend further support to the conjecture that only systems of free particles can be described by the theories under consideration.

As in Ref. 1, a classical mechanical relativistically invariant Hamiltonian theory of particle dynamics is defined by ten functions $H, P_i, J_i, K_i, i = 1, 2, 3$ of the canonical variables $q_i^n, p_i^n, n = 1, 2, \cdots N$; $i = 1, 2, 3$ for the N (spinless) particles. These ten functions are, respectively, the generators of the time translations, space translations, space rotations, and (pure) Lorentz transformations which comprise the inhomogeneous Lorentz group. Symmetry of the theory under this group is attained by postulating that these functions satisfy the Poisson bracket equations

$$[P_i, P_j] = 0, \quad [P_i, H] = 0, \quad [J_i, H] = 0,$$
$$[J_i, J_j] = \epsilon_{ijk}J_k, \quad [J_i, P_j] = \epsilon_{ijk}P_k,$$
$$[J_i, K_j] = \epsilon_{ijk}K_k, \quad [K_i, H] = P_i,$$
$$[K_i, K_j] = -\epsilon_{ijk}J_k, \quad [K_i, P_j] = \delta_{ij}H, \qquad \text{(A)}$$

for $i, j, k = 1, 2, 3$. (The summation convention is used for these indices.) These are the Lie bracket

* Supported in part by the U. S. Atomic Energy Commission.
† Alfred P. Sloan Foundation Scholar visiting as research assistant under the cooperative program from Antioch College, Yellow Springs, Ohio. Present address: Fine Hall, Princeton University, Princeton, New Jersey.
‡ Present address: Institut für theoretische Physik, Universität Bern, Bern, Switzerland.
[1] D. G. Currie, T. F. Jordan, and E. C. G. Sudarshan, Rev. Modern Phys. 35, 350 (1963).
[2] D. G. Currie, "Interaction *contra* Classical Relativistic Hamiltonian Particle Mechanics", University of Rochester Rept. NYO-10242 (to be published), and "The Hamiltonian Description of Interaction for Classical Relativistic Particles", Thesis, University of Rochester (1962).

equations characteristic of the inhomogeneous Lorentz group.[1]

We consider the motion of classical mechanical particles to be described by the time dependence of their positions in space. In keeping with this interpretation, we postulate, as part of our requirement of relativistic invariance, that the time-dependent particle positions—or points on the world lines of the particles—transform in the familiar manner under space translations, space rotations, and Lorentz transformations of the reference frame.[1] Within our Hamiltonian formalism this is equivalent to postulating that the Poisson bracket equations

$$[Q_i^n, P_k] = \delta_{ik}, \qquad [J_i, Q_j^n] = \epsilon_{ijk}Q_k^n, \qquad [Q_i^n, K_i] = Q_i^n[Q_i^n, H], \tag{B}$$

be satisfied by the ten functions H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ together with the functions Q_i^n which represent the positions of the particles.[1] We assume that the Q_i^n are those functions whose values at a point of phase space are the q coordinates of the point,

$$Q_i^n(q, p) = q_i^n.$$

We will also need to use the canonical functions

$$P_i^n(q, p) = p_i^n.$$

We make no assumptions about the transformation properties of the canonical "momenta" P_i^n. In fact we could avoid making any physical interpretation of these functions. This would leave us free to make canonical transformations in the variables P_i^n without changing the physical content of the theory.[3] In any case, canonical transformations that change the functions Q_i^n cannot be considered as giving equivalent theories since the role of the Q_i^n is fixed by their interpretation as particle positions. This point of view is explained in detail in Ref. 1.

In Ref. 1 the following theorem was proved for the case of two particles.

Theorem. Let H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ be ten functions of the canonical particle variables q_i^n and p_i^n. If H, $\mathbf{P}$, $\mathbf{J}$ and $\mathbf{K}$ satisfy Eqs. (A) and (B) then they are equivalent, via canonical transformations of the P_i^n, to the standard functions for a free-particle theory.

This theorem has an immediate corollary that the Hamiltonian function H yields zero acceleration for each of the particles. It should be emphasized that the validity of this result does not depend in any way on any decision as to whether the P_i^n should be given a physical interpretation or whether canonical transformations of the P_i^n should be regarded as changing the physical content of the theory.

Corollary. Let H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ be ten functions of the canonical particle variables q_i^n and p_i^n. If H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ satisfy Eqs. (A) and (B), then

$$[[Q_i^n, H], H] = 0.$$

In this paper we shall generalize this result to the case of three particles and, in part, to the case of any fixed number of particles.

The results that are valid for any number of particles are contained in Sec. II. We show there that the functions $\mathbf{P}$ and $\mathbf{J}$ can be put in the standard form of free-particle theories and that H and $\mathbf{K}$ can be put in a form which separates the dependence on the variables p_i^n for different n and exhibits the standard relation between the terms of $\mathbf{K}$ and the terms of H. Our treatment of $\mathbf{J}$ depends on a generalization, which we prove in the Appendix, of the angular momentum Helmholtz theorem of Lomont and Moses.[4,5] In Sec. III we prove the above-stated no-interaction theorem for the case of three particles. The proof does not generalize for more than three particles.

II. STANDARD FORMS IN THE CASE OF N PARTICLES

In this section we consider functions H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ of the canonical particle variables q_i^n and p_i^n, $i = 1, 2, 3$; $n = 1, 2, \cdots N$, with N any positive integer. It will always be assumed that H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ satisfy Eqs. (A) and (B). Our purpose is to show that $\mathbf{P}$ and $\mathbf{J}$ can be put in the standard forms characteristic of a free-particle theory, and that H and $\mathbf{K}$ can also be put into simple forms which will be the starting point for our proof of the no-interaction theorem.

We shall make repeated use of a change of variables from q_i^n to r_i^n, $n = 1, 2, \cdots N$; $i = 1, 2, 3$,

[3] If the functions P_i^n are given a physical interpretation, then their transformation properties also might be specified. If they are required to transform in the manner that is familiar for momenta, we get another set of equations to be satisfied by the functions H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$. This extra set of equations should make it possible to establish our theorems without doing canonical transformations. This was demonstrated for the case of two particles in Ref. 1.

[4] J. S. Lomont and H. E. Moses, Commun. Pure Appl. Math. **14**, 69 (1961). This theorem is used by J. S. Lomont and H. E. Moses, Nuovo Cimento **16**, 96 (1960), to establish the standard form of the single-particle quantum mechanical angular momentum operator.

[5] J. B. Keller, Commun. Pure Appl. Math. **14**, 77 (1961), gives an alternative proof of the theorem of Lomont and Moses (Ref. 4).

defined by

$$r_i^m = \sum_{n=1}^{N} g_n^m q_i^n, \tag{2.1}$$

with

$$g_n^m = 1 \quad \text{if} \quad n \quad \text{or} \quad m = 1,$$
$$g_n^m = -1 \quad \text{if} \quad n = m \neq 1,$$
$$g_n^m = 0 \quad \text{otherwise},$$

from which follows the useful equations

$$\frac{\partial}{\partial r_i^1} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial}{\partial q_i^n}, \tag{2.2}$$

$$\frac{\partial}{\partial q_i^1} = \sum_{n=1}^{N} \frac{\partial}{\partial r_i^n}, \tag{2.3}$$

$$\frac{\partial}{\partial q_i^n} = \frac{\partial}{\partial r_i^1} - \frac{\partial}{\partial r_i^n}, \quad \text{for} \quad n \neq 1. \tag{2.4}$$

Proposition 1. The functions **P** can be put into the standard form

$$P_i = \sum_{n=1}^{N} P_i^n \tag{2.5}$$

by a canonical transformation which leaves the functions Q_i^n unchanged.

Proof: Let

$$P_i = \sum_{n=1}^{N} P_i^n + F_i,$$

with F_i a function of the canonical variables $\mathbf{q}^n$ and $\mathbf{p}^n$. From the first of Eqs. (B), it follows that

$$\partial F_i / \partial p_k^n = [F_i, Q_k^n] = 0,$$

so F_i is in fact a function of the variables $\mathbf{q}^n$ only. The transformation

$$P_i^1 \to P_i^1 - (1/N)F_i - (1/N) \sum_{n=2}^{N} W_i^n,$$

$$P_i^n \to P_i^n - (1/N)F_i + (1/N)W_i^n \quad \text{for} \quad n \neq 1,$$

with the Q_i^n remaining unchanged, where

$$W_i^n = \tfrac{1}{3} \sum_k \int \frac{\partial F_k}{\partial r_i^n} \, dr_k^1,$$

puts P_i into the desired form (2.5). To see that this is a canonical transformation one must compute the Poisson brackets of the transformed canonical variables. This can be easily accomplished and yields the desired results if one works in terms of the variables (2.1) and uses the condition (A) that

$$N[(\partial F_i / \partial r_k^1) - (\partial F_k / \partial r_i^1)] = [P_i, P_k] = 0.$$

Proposition 2. Let **P** have the standard form (2.5). Then the functions **J** can be put into the standard form

$$J_i = \sum_{n=1}^{N} \epsilon_{ijk} Q_j^n P_k^n \tag{2.6}$$

by a canonical transformation which leaves the functions Q_i^n and the standard form (2.5) for P_i unchanged.

Proof: We shall need the following lemma which follows as a corollary to the theorem proved in the Appendix.

Lemma. Let **A** be a real three-vector function of the M three-vector variables $\mathbf{r}^m$, $m = 1, 2, \cdots M$. If

$$\sum_{m=1}^{M} (\mathbf{r}^m \times \nabla_m) \times \mathbf{A} = -A,$$

then there exists a function W of the variables $\mathbf{r}^m$ such that

$$\mathbf{A} = \sum_{m=1}^{M} r^m \times \nabla_m W.$$

To prove Proposition 2 let

$$J_i = \sum_{n=1}^{N} \epsilon_{ijk} Q_j^n P_k^n + F_i.$$

From the second of Eqs. (B) it follows that

$$(\partial F_i / \partial p_k^n) = [F_i, Q_k^n] = 0,$$

so F_i is a function of the variables $\mathbf{q}^n$ only. Now using the standard form (2.5) for **P**, the fifth of Eqs. (A), and Eq. (2.2), we find that

$$N(\partial F_i / \partial r_k^1) = [F_i, P_k] = 0,$$

so F_i is independent also of the variables $\mathbf{r}^1$. Now if we substitute our form for **J** into the fourth of Eqs. (A), we obtain the equation

$$\sum_{n=2}^{N} (\mathbf{r}^n \times \nabla_n) \times \mathbf{F} = -F.$$

Thus, by the lemma stated above, there exists a function W of the variables $\mathbf{r}^n$, $n = 2, \cdots N$ such that

$$\mathbf{F} = \sum_{n=2}^{N} \mathbf{r}^n \times \nabla_n W,$$

with ∇_n the gradient with respect to the $\mathbf{r}^n$. The transformation

$$P_i^1 \to P_i^1 - \sum_{n=2}^{N} \frac{\partial W}{\partial r_i^n}, \quad P_i^n \to P_i^n + \frac{\partial W}{\partial r_i^n} \quad \text{for} \quad n \neq 1,$$

with Q_i^n remaining unchanged, leaves the standard

form (2.5) for $\mathbf{P}$ unchanged and puts $\mathbf{J}$ into the standard form (2.6),

$$J \to \mathbf{Q}^1 \times \left(\mathbf{P}^1 - \sum_{n=2}^{N} \nabla_n W\right) + \sum_{n=2}^{N} \mathbf{Q}^n \times (\mathbf{P}^n + \nabla_n W) + \mathbf{F} = \sum_{n=1}^{N} \mathbf{Q}^n \times \mathbf{P}^n$$
$$- \sum_{n=2}^{N} (\mathbf{Q}^1 - \mathbf{Q}^n) \times \nabla_n W + \mathbf{F} = \sum_{n=1}^{N} \mathbf{Q}^n \times \mathbf{P}^n.$$

That this is a canonical transformation can be checked by a straightforward computation of the Poisson brackets of the transformed canonical variables.

Proposition 3. Let $\mathbf{P}$ and $\mathbf{J}$ have the standard forms (2.5) and (2.6). The functions H and $\mathbf{K}$ can be written in the forms

$$H = \sum_{n=1}^{N} H_n, \tag{2.7}$$

and

$$K_i = \sum_{n=1}^{N} Q_i^n H_n, \tag{2.8}$$

with H_n a function only of the scalar variables $(\mathbf{p}^n)^2$, $(\mathbf{p}^n \cdot \mathbf{r}^k)$, and $(\mathbf{r}^k \cdot \mathbf{r}^m)$, for $k, m = 2, \cdots N$.

Proof: The last of Eqs. (B) is

$$(\partial K_k/\partial p_i^n) = Q_k^n(\partial H/\partial p_i^n).$$

From this it follows that

$$(Q_k^n - Q_k^m)(\partial^2 H/\partial p_i^n\, \partial p_l^m) = 0,$$

or, for $m \neq n$, that[6]

$$(\partial^2 H/\partial p_i^n\, \partial p_l^m) = 0.$$

Hence

$$H = \sum_{n=1}^{N} H_n,$$

where H_n is a function only of the variables $\mathbf{p}^n$ and $\mathbf{q}^k$, $k = 1, 2, \cdots N$.

We can choose the H_n to be independent of $\mathbf{r}^1$ as follows. We compute

$$\frac{\partial H}{\partial r_i^1} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial H}{\partial q_i^n} = \left(\frac{1}{N}\right)[H, P_i] = 0$$

by using the second of Eqs. (A). From this it follows that

$$\frac{\partial H_1}{\partial r_i^1} = -\sum_{n=2}^{N} \frac{\partial H_n}{\partial r_i^1}.$$

Since the right-hand side is independent of $\mathbf{p}^1$, and the left-hand side is independent of all of the other $\mathbf{p}$ variables, the above quantity depends only on the $\mathbf{q}$ variables. Further, since

$$\epsilon_{ijk}(\partial^2 H_1/\partial r_j^1\, \partial r_k^1) = 0,$$

there exists a function G of the variables $\mathbf{r}^n$, $n = 1, 2, \cdots N$, such that

$$\frac{\partial H_1}{\partial r_i^1} = -\sum_{n=2}^{N} \frac{\partial H_n}{\partial r_i^1} = \frac{\partial G}{\partial r_i^1}.$$

Hence we can redefine H_1 and H_2 by subtracting G from H_1 and adding G to H_2 so as to obtain a new division of H in the form (2.7) in which H_1 and $\sum_{n=2}^{N} H_n$ are both independent of $\mathbf{r}^1$. This process can be continued to make each H_n independent of $\mathbf{r}^1$.

To show that each H_n can be chosen to be a scalar function we use the third of Eqs. (A) to write

$$A_i \equiv [H_1, J_i] = \left[J_i, \sum_{n=2}^{N} H_n\right].$$

By actually evaluating the brackets one can check that A_i is independent of the $\mathbf{p}$ variables and of $\mathbf{r}^1$. From the fourth of Eqs. (A), it follows that

$$[H_1, [J_i, J_j]] = \epsilon_{ijk}[H_1, J_k],$$

which, by use of the Jacobi identity, can be written as

$$[J_i, [H_1, J_j]] - [J_j, [H_1, J_i]] = \epsilon_{ijk}[H_1, J_k],$$

or

$$[J_i, A_j] - [J_j, A_i] = \epsilon_{ijk}\,.$$

Evaluation of the brackets then gives

$$\sum_{n=2}^{N} (\mathbf{r}^n \times \nabla_n) \times \mathbf{A} = -\mathbf{A}.$$

According to the lemma used above, there exists a function W of the variables $\mathbf{r}^n$, $n = 2, \cdots N$, such that

$$\mathbf{A} = \sum_{n=2}^{N} \mathbf{r}^n \times \nabla_n W.$$

From this it follows that

$$[H_1, \mathbf{J}] = \left[\mathbf{J}, \sum_{n=2}^{N} H_n\right] = \sum_{n=2}^{N} \mathbf{r}^n \times \nabla_n W = [W, \mathbf{J}],$$

as can be seen easily by evaluating the bracket on the right. Hence we can redefine H_1 and H_2 by subtracting W from H_1 and adding W to H_2 so as to obtain a new division of H in the form (2.7) in which H_1 and $\sum_{n=2}^{N} H_n$ are both scalars, i.e., have vanishing Poisson bracket with $\mathbf{J}$. This

[6] As in Ref. 1, we assume that all of the functions are sufficiently differentiable to make the Poisson brackets and other necessary derivatives meaningful, and we use the resulting continuity in our proofs.

process can be continued to make each H_n a scalar. Therefore each H_n can be chosen without loss of generality to be a function only of the scalar variables $(\mathbf{p}^n)^2$, $(\mathbf{p}^n \cdot \mathbf{r}^k)$, and $(\mathbf{r}^k \cdot \mathbf{r}^m)$ for $k, m = 2, \cdots N$.

Finally we must show that $\mathbf{K}$ can be assumed, without loss of generality, to have the form (2.8). Let

$$K_i = \sum_{n=1}^{N} Q_i^n H_n + F_i.$$

If we substitute this in the last of Eqs. (B), we find that

$$\partial F_i/\partial p_i^1 = [Q_i^n, F_i] = 0,$$

so F_i must be a function only of the $\mathbf{q}$ variables. From the last of Eqs. (A), the first of Eqs. (B), and the fact that each H_n is independent of $\mathbf{r}^1$, it follows that

$$N(\partial F_i/\partial r_k^1) = [F_i, P_k] = 0,$$

while from the sixth of Eqs. (A), the second of Eqs. (B), and the fact that each H_n is a scalar, it follows that

$$[J_i, F_j] = \epsilon_{ijk} F_k.$$

Hence $\mathbf{F}$ must have the form

$$F_i = \sum_{n=2}^{N} r_i^n F_n = \sum_{n=2}^{N} (Q_i^1 - Q_i^n) F_n,$$

with each F_n a function only of the scalar variables $\mathbf{r}^k \cdot \mathbf{r}^m$, $k, m = 2, \cdots N$. If we add $\sum_{n=2}^{N} F_n$ to H_1 and subtract F_n from H_n for $n = 2, \cdots N$, we obtain a new division of H in the form (2.7) with each H_n still a function only of the variables $(\mathbf{p}^n)^2$, $(\mathbf{p}^n \cdot \mathbf{r}^k)$, and $(\mathbf{r}^k \cdot \mathbf{r}^m)$, $k, m = 2, \cdots N$, and we also attain the form (2.8) for $\mathbf{K}$.

III. THE NO-INTERACTION THEOREM FOR THREE PARTICLES

Theorem. Let H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ be ten functions of the canonical variables q_i^n and p_i^n, $n, i = 1, 2, 3$ for three particles. If H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ satisfy Eqs. (A) and (B), then they are equivalent, via canonical transformations of the P_i^n, to the standard functions for a free-particle theory.

In proving this theorem we may use the results of the preceding section, taking $\mathbf{P}$, $\mathbf{J}$, H, and $\mathbf{K}$ to have the forms (2.5), (2.6), (2.7), and (2.8), respectively. The theorem will be established when we find a canonical transformation that leaves these forms as well as each Q_i^n unchanged but puts each H_n into the standard free-particle form

$$H_n = ((\mathbf{P}^n)^2 + m_n^2)^{\frac{1}{2}}, \tag{3.0}$$

with m_n some nonnegative number. To prove the theorem it is therefore sufficient to prove the following.

Proposition 4. Let H, $\mathbf{P}$, $\mathbf{J}$ and $\mathbf{K}$ be ten functions of the canonical particle variables q_i^n, p_i^n, $n, i = 1, 2, 3$. If H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ have the forms (2.7), (2.5), (2.6), and (2.8), respectively, and satisfy Eqs. (A), there is a canonical transformation which leaves each Q_i^n unchanged, leaves the forms (2.5), (2.6), (2.7), and (2.8) unchanged, and puts each H_n into the form (3.0).

Proof: Since we begin with functions H, $\mathbf{P}$, $\mathbf{J}$, and $\mathbf{K}$ in the forms (2.7), (2.5), (2.6), and (2.8), respectively, Eqs. (B) and most of Eqs. (A) will be satisfied. To gain more information about H and $\mathbf{K}$, we need to use only the seventh of Eqs. (A). Our main task is to extract enough information from this equation to determine a form for H_n that is equivalent to the free-particle form (3.0).

In terms of the variables (2.1), the seventh of Eqs. (A) is

$$\sum_{n=1}^{3} \left\{ \frac{\partial H_n^2}{\partial (\mathbf{p}^n)^2} - 1 \right\} \mathbf{P}^n + \sum_{n=2}^{3} \left\{ \sum_{m=1}^{3} \frac{1}{2} \frac{\partial H_m^2}{\partial (\mathbf{p}^m \cdot \mathbf{r}^n)} + [H, H_n] \right\} \mathbf{r}^n = 0. \tag{3.1}$$

For any n let

$$\mathbf{p}^k = 0 \quad \text{if} \quad k \neq n. \tag{3.2}$$

From Eq. (3.1) we obtain an equation in three linearly independent vectors $\mathbf{p}^n$, $\mathbf{r}^2$, and $\mathbf{r}^3$, which is valid under the restriction (3.2). From the linear independence we can conclude in particular that the coefficient of $\mathbf{P}^n$ must vanish,

$$\partial H_n^2/\partial (\mathbf{p}^n)^2 = 1. \tag{3.3}$$

But the restriction (3.2) is of no consequence for Eq. (3.3) since H_n is independent of $\mathbf{p}^k$ for $k \neq n$. Hence Eq. (3.3) is valid without restriction.

As a result of Eq. (3.3), our original Eq. (3.1) involves only the two linearly independent vectors $\mathbf{r}^2$ and $\mathbf{r}^3$, and we can conclude that the coefficient of each of these must vanish:

$$\sum_{m=1}^{3} \frac{1}{2} \frac{\partial H_m^2}{\partial (\mathbf{p}^m \cdot \mathbf{r}^n)} + [H, H_n] = 0 \tag{3.4}$$

for $n = 2, 3$. If we evaluate the bracket in Eq. (3.4) we get some terms linear in $(\mathbf{p}^1 \cdot \mathbf{p}^n)$ and some terms linear in $(\mathbf{p}^s \cdot \mathbf{p}^n)$ where s is the index 2 or 3 which is not equal to n. These terms and the other terms occurring in the bracket depend on $(\mathbf{p}^m)^2$, $(\mathbf{p}^m \cdot \mathbf{r}^k)$, and $(\mathbf{r}^k \cdot \mathbf{r}^h)$ for $m = 1, 2, 3$; $k, h = 2, 3$. But for fixed values of the latter variables, the quantities

$(\mathbf{p}^1\cdot\mathbf{p}^n)$ and $(\mathbf{p}^s\cdot\mathbf{p}^n)$ are free to take two different values. For example, $\mathbf{p}^1$ has fixed length and fixed components in the plane spanned by $\mathbf{r}^2$ and $\mathbf{r}^3$, but may still be reflected through that plane to produce two values for $(\mathbf{p}^1\cdot\mathbf{p}^n)$. It follows that the coefficients of $(\mathbf{p}^1\cdot\mathbf{p}^n)$ and $(\mathbf{p}^s\cdot\mathbf{p}^n)$ must vanish. This yields the equations

$$\frac{\partial H_1}{\partial(\mathbf{p}^1\cdot\mathbf{r}^n)}\frac{\partial H_n}{\partial(\mathbf{p}^n)^2}+\sum_{m=2}^{3}\frac{\partial H_n}{\partial(\mathbf{p}^n\cdot\mathbf{r}^m)}\frac{\partial H_1}{\partial(\mathbf{p}^1)^2}=0, \tag{3.5}$$

$$\frac{\partial H_s}{\partial(\mathbf{p}^s\cdot\mathbf{r}^n)}\frac{\partial H_n}{\partial(\mathbf{p}^n)^2}-\frac{\partial H_n}{\partial(\mathbf{p}^n\cdot\mathbf{r}^s)}\frac{\partial H_s}{\partial(\mathbf{p}^s)^2}=0. \tag{3.6}$$

Let

$$F_{mn}=\partial H_m^2/\partial(\mathbf{p}^m\cdot\mathbf{r}^n)$$

for $m=1, 2, 3$; $n=2, 3$. If we multiply Eqs. (3.5) and (3.6) by $4H_1H_n$ and $4H_sH_n$, respectively, and use Eq. (3.3), we find that

$$\sum_{m=1}^{3}F_{mn}=0 \tag{3.7}$$

for $n=2, 3$, and that

$$F_{23}=F_{32}. \tag{3.8}$$

From Eqs. (3.7) and (3.8) and the fact that H_n is independent of $\mathbf{p}^m$ for $m\neq n$, it follows that each F_{mn} is independent of all $\mathbf{p}$ variables. We have thus found that

$$H_n^2=(\mathbf{p}^n)^2+\sum_{m=2}^{3}F_{nm}(\mathbf{p}^n\cdot\mathbf{r}^m)+G_n \tag{3.9}$$

for $n=1, 2, 3$ where F_{nm} and G_n are functions only of $(\mathbf{r}^2)^2$, $(\mathbf{r}^3)^2$, and $(\mathbf{r}^2\cdot\mathbf{r}^3)$, with F_{nm} satisfying Eqs. (3.7) and (3.8).

As a consequence of Eq. (3.7), Eq. (3.4) becomes

$$[H, H_n]=0 \tag{3.10}$$

for $n=2, 3$. If we evaluate this bracket for the case $n=3$, we get some terms linear in $(\mathbf{p}^1\cdot\mathbf{p}^3)$ and $(\mathbf{p}^2\cdot\mathbf{p}^3)$ which sum to zero because of Eq. (3.7). The sum of the remaining terms must vanish to satisfy Eq. (3.10). If we consider these under the restriction that

$$(\mathbf{p}^2\cdot\mathbf{r}^3)=0, \tag{3.11}$$

we find several that are independent of $\mathbf{p}^2$, several that are linear in $1/H_2$ but otherwise independent of $\mathbf{p}^2$, and some that depend on $\mathbf{p}^2$ only through the linear factor $(\mathbf{p}^2\cdot\mathbf{r}^2)/H_2$. If the sum of these terms is to vanish, the sum of each of the three sets of terms must vanish independently, for, even under the restriction (3.11) that $\mathbf{p}^2$ be perpendicular to $\mathbf{r}^3$, the variations of $\mathbf{p}^2$ will cause different variations of $1/H_2$ and $(\mathbf{p}^2\cdot\mathbf{r}^2)/H_2$. We make use of the fact that the coefficient of $(\mathbf{p}^2\cdot\mathbf{r}^2)/H_2$ must vanish. Among the terms of this coefficient are some linear in $(\mathbf{p}^3\cdot\mathbf{r}^2)$ and some linear in $(\mathbf{p}^3\cdot\mathbf{r}^3)$. There is no other dependence on the $\mathbf{p}$ variables. It follows that the coefficients of $(\mathbf{p}^3\cdot\mathbf{r}^2)$ and of $(\mathbf{p}^3\cdot\mathbf{r}^3)$ must vanish separately. This yields the equations

$$2[\partial F_{32}/\partial(\mathbf{r}^2)^2]=\partial F_{22}/\partial(\mathbf{r}^2\cdot\mathbf{r}^3), \tag{3.12}$$

$$\partial F_{33}/\partial(\mathbf{r}^2)^2=\partial F_{22}/\partial(\mathbf{r}^3)^2. \tag{3.13}$$

Since these equations have no dependence on any $\mathbf{p}$ variables, the restriction (3.11) is of no consequence for them; they are valid without restriction. We could apply the same procedure to Eq. (3.10) for $n=2$. The equations that would result are the same as Eqs. (3.12) and (3.13) with the indices 2 and 3 interchanged. Making this interchange on Eq. (3.12) gives us the new equation

$$2[\partial F_{23}/\partial(\mathbf{r}^3)^2]=\partial F_{33}/\partial(\mathbf{r}^2\cdot\mathbf{r}^3). \tag{3.14}$$

Equation (3.13) remains unchanged.

The transformation

$$\mathbf{P}^n\rightarrow\mathbf{P}^n-\tfrac{1}{2}\sum_{m=2}^{3}F_{nm}\mathbf{r}^m, \tag{3.15}$$

with the $\mathbf{Q}^n$ remaining unchanged, is a canonical transformation. To see this one must compute the Poisson brackets of the transformed canonical variables. This can be easily accomplished, yielding the desired results, if Eqs. (3.7), (3.8), and (3.12)–(3.14) are used.

That the standard form (2.5) of $\mathbf{P}$ is preserved by the transformation (3.15) is ensured by Eq. (3.7). One can check that Eqs. (3.7) and (3.8) also guarantee that the standard form (2.6) of $\mathbf{J}$ is preserved. The forms (2.7) and (2.8) of H and $\mathbf{K}$ are also preserved and each H_n is taken by the transformation (2.15) from the form (3.9) to the desired form (3.0) with

$$m_n^2=G_n-(\tfrac{1}{4})F_{n2}^2(\mathbf{r}^2)^2-(\tfrac{1}{2})F_{n2}F_{n3}(\mathbf{r}^2\cdot\mathbf{r}^3)-(\tfrac{1}{4})F_{n3}^2(\mathbf{r}^3)^2.$$

It remains to show that m_n^2 is in fact a constant, i.e., that it is independent of the variables $(\mathbf{r}^2)^2$, $(\mathbf{r}^2\cdot\mathbf{r}^3)$, and $(\mathbf{r}^3)^2$, on which the terms in the above expression depend. This is most easily accomplished as follows. Suppose that each m_n^2 is a function of the variables $(\mathbf{r}^2)^2$, $(\mathbf{r}^2\cdot\mathbf{r}^3)$, and $(\mathbf{r}^3)^2$, and substitute the form (3.0) for H_n into Eq. (3.10). One finds that for this equation to be satisfied it is necessary that the partial derivatives of each of the functions m_n^2 with respect to each of the variables $(\mathbf{r}^2)^2$, $(\mathbf{r}^2\cdot\mathbf{r}^3)$,

and $(\mathbf{r}^3)^2$ vanish. Since H must be a real function, it follows that m_n^2 is a nonnegative number. This completes the proof of Proposition 4, and thus completes the proof of the no-interaction theorem for three particles.

APPENDIX: GENERALIZATION OF THE ANGULAR MOMENTUM HELMHOLTZ THEOREM OF LOMONT AND MOSES

Let $\mathbf{L} = \sum_{n=1}^{N} \mathbf{x}^n \times \nabla_n$ where ∇_n is the gradient operator with respect to the three-vector variable $\mathbf{x}_n$, and N is some positive integer.

Theorem. Every real three-vector function $\mathbf{f}(\mathbf{x}^1, \cdots \mathbf{x}^N)$ of N three-vector variables $\mathbf{x}^1, \cdots \mathbf{x}^N$ can be expressed uniquely as the sum of real terms,

$$\mathbf{f}(\mathbf{x}^1, \cdots \mathbf{x}^N) = \mathbf{g}(\mathbf{x}^1, \cdots \mathbf{x}^N) + \mathbf{h}(\mathbf{x}^1, \cdots \mathbf{x}^N),$$

with

$$\mathbf{L} \times \mathbf{g} = -\mathbf{g}, \qquad \mathbf{g} = \mathbf{L}\phi(\mathbf{x}^1, \cdots \mathbf{x}^N),$$

$$\mathbf{L} \cdot \mathbf{h} = 0, \quad \mathbf{h} = \mathbf{L} \times \psi(\mathbf{x}^1, \cdots \mathbf{x}^N) + \psi(\mathbf{x}^1, \cdots \mathbf{x}^N).$$

Proof: Lomont and Moses[4] and also Keller[5] have proved this theorem for the case $N = 1$ without emphasizing the uniqueness. We shall outline the essential results of the proof by Lomont and Moses, pointing out where the uniqueness can be seen. We shall then set up an analogous general proof which reduces to the proof for the case $N = 1$.

The spherical harmonics $Y_{km}(\theta, \phi)$ for fixed k form a basis of a $(2k + 1)$-dimensional irreducible representation Γ_k of the three-dimensional rotation group. In this representation $\mathbf{L}$ is given by a triplet of $(2k + 1) \times (2k + 1)$ matrices $L^{(k)}_{imm'}$, $j = 1, 2, 3$. For example, with

$$\mathbf{f}(\mathbf{x}) = \sum_{km} \varrho^{km}(r) Y_{km}(\theta, \phi),$$

the operation of the third component L_3 of $\mathbf{L}$ on the second component $f_2(\mathbf{x})$ of $\mathbf{f}(\mathbf{x})$ is given by

$$L_3 f_2(\mathbf{x}) = \sum_{km} \sum_{m'} L^{(k)}_{3mm'} \rho_2^{km'}(r) Y_{km}(\theta, \phi).$$

We think of ρ_2^{km} as the components of a $(2k + 1)$-dimensional vector $\rho_2^{(k)}$, and of $L^{(k)}_{3mm'}$ as the matrix elements of an operator $L_3^{(k)}$, and write

$$L_3 f_2 = \sum_{km} (L_3^{(k)} \rho_2^{(k)})_{km} Y_{km}.$$

Having set the tone of the notation to follow that of Lomont and Moses, we now outline their results. They showed that a triplet of three arbitrary vectors $\rho_j^{(k)}$, $j = 1, 2, 3$ (or $\varrho^{(k)}$) in $(2k + 1)$-dimensional space can be written as

$$\varrho^{(k)} = \xi^{(k)} + \mathbf{n}^{(k)},$$

with

$$\mathbf{L}^{(k)} \times \xi^{(k)} = -\xi^{(k)}, \tag{A1}$$

$$\mathbf{L}^{(k)} \cdot \mathbf{n}^{(k)} = 0, \tag{A2}$$

and that the equations

$$\mathbf{L}^{(k)} \alpha^{(k)} = \xi^{(k)}, \tag{A3}$$

$$\mathbf{L}^{(k)} \times \beta^{(k)} + \beta^{(k)} = \mathbf{n}^{(k)} \tag{A4}$$

have solutions $\alpha^{(k)}$ and $\beta^{(k)}$. The theorem for the case $N = 1$ follows after expanding

$$\mathbf{f}(\mathbf{x}) = \sum_{km} \varrho^{km}(r) Y_{km}(\theta, \phi),$$

and setting

$$\mathbf{g}(\mathbf{x}) = \sum_{km} \xi^{km}(r) Y_{km}(\theta, \phi),$$

$$\mathbf{h}(\mathbf{x}) = \sum_{km} \mathbf{n}^{km}(r) Y_{km}(\theta, \phi),$$

$$\phi(\mathbf{x}) = \sum_{km} \alpha^{km}(r) Y_{km}(\theta, \phi),$$

$$\psi(\mathbf{x}) = \sum_{km} \beta^{km}(r) Y_{km}(\theta, \phi),$$

where ξ^{km}, $\mathbf{n}^{km}$, α^{km}, and β^{km} correspond, respectively, to $\xi^{(k)}$, $\mathbf{n}^{(k)}$, $\alpha^{(k)}$, and $\beta^{(k)}$.

To see the uniqueness, consider the matrices

$$M^{(k)} = \begin{bmatrix} 0 & L_3^{(k)} & -L_2^{(k)} \\ -L_3^{(k)} & 0 & L_1^{(k)} \\ L_2^{(k)} & -L_1^{(k)} & 0 \end{bmatrix},$$

$$\Lambda^{(k)} = \begin{bmatrix} L_1^{(k)} & L_1^{(k)} & L_1^{(k)} \\ L_2^{(k)} & L_2^{(k)} & L_2^{(k)} \\ L_3^{(k)} & L_3^{(k)} & L_3^{(k)} \end{bmatrix},$$

which operate on triplets of $(2k + 1)$-dimensional vectors or on $3(2k + 1)$-dimensional vectors. In terms of these matrices, Eqs. (A1) and (A2) can be written as

$$M^{(k)} \xi^{(k)} = \xi^{(k)}, \qquad \Lambda^{(k)\mathrm{t}} \mathbf{n}^{(k)} = 0,$$

where t denotes the transpose of a matrix. Lomont and Moses showed that the subspace of eigenvectors of $M^{(k)}$ having eigenvalue one is the orthogonal compliment of the subspace of eigenvectors of $\Lambda^{(k)}$ having eigenvalue zero. Thus, since $\xi^{(k)}$ and $\mathbf{n}^{(k)}$ are the components of $\varrho^{(k)}$ in these orthogonal subspaces, they are determined uniquely. But any $\mathbf{g}(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ satsifying

$$\mathbf{f} = \mathbf{g} + \mathbf{h},$$

$$\mathbf{L} \times \mathbf{g} = -\mathbf{g}, \qquad \mathbf{L} \cdot \mathbf{h} = 0$$

yield, when expressed in terms of spherical harmonics, the equations

$$\varrho^{(k)} = \xi^{(k)} + \mathbf{n}^{(k)},$$

with $\xi^{(k)}$ and $\mathbf{n}^{(k)}$ satisfying Eqs. (A1) and (A2). Thus $\mathbf{g}$ and $\mathbf{h}$ are unique.

We now turn to the proof of the generalization. The function $\mathbf{f}(\mathbf{x}^1, \cdots \mathbf{x}^N)$ can be expanded in terms of spherical harmonics as

$$\mathbf{f}(\mathbf{x}^1, \cdots \mathbf{x}^N) = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \varrho^{k_1,\cdots k_N, m_1,\cdots m_N}(r_1, \cdots r_N) \times Y_{k_1 m_1}(\theta_1, \phi_1) \cdots Y_{k_N m_N}(\theta_N, \phi_N). \quad \text{(A5)}$$

For fixed $k_1 \cdots k_N$, the $Y_{k_1 m_1} \ldots Y_{k_N m_N}$ form a basis in the representation space of the direct product $\Gamma_{k1} \otimes \Gamma_{k2} \cdots \otimes \Gamma_{kN}$, which is a reducible representation of three-dimensional rotation group. The operation of L_i can be expressed in terms of the matrix

$$L^{(k_1,\cdots k_N)}_{i m_1,\cdots m_N, m_1',\cdots m_N'} = L^{(k_1)}_{i m_1 m_1'} \otimes I^{(2)} \cdots \otimes I^{(N)} + \cdots + I^{(1)} \otimes I^{(2)} \cdots \otimes L^{(k_N)}_{i m_N m_N'},$$

or in terms of the corresponding operator $L_i^{(k_1,\cdots k_N)}$.

The direct-product representation can be completely reduced to a direct sum of irreducible representations

$$\Gamma_{k1} \otimes \Gamma_{k2} \cdots \otimes \Gamma_{kN} = \sum_k \oplus \Gamma_k,$$

with the summation over a set of possibly repeated values of k depending on $k_1, \cdots k_N$. The representation space of the direct product is thus broken up into orthogonal subspaces which are the representation spaces of the irreducible representations Γ_k. The operator $L_i^{(k_1,\cdots k_N)}$ on the representation space of the direct product is the direct sum of operators on the irreducible representation spaces of the Γ_k,

$$L_i^{(k_1,\cdots k_N)} = \sum_k \oplus L_i^{(k)},$$

where the $L_i^{(k)}$ are the same as those appearing above. In other words, on each irreducible subspace the infinitesimal generators $L_i^{(k_1,\cdots k_N)}$ of the direct-product representation act as the infinitesimal generators $L_i^{(k)}$ of the irreducible representation Γ_k.

A vector $\rho_i^{(k_1,\cdots k_N)}$ of $(2k_1 + 1) \cdots (2k_N + 1)$ components can be considered as an element of the direct-product representation space, and written as the sum,

$$\rho_i^{(k_1,\cdots k_N)} = \sum_k \rho_i(k),$$

of its uniquely determined components $\rho_i^{(k)}$ in the mutually orthogonal irreducible subspaces. Such a vector results from the coefficients of the spherical harmonic expansion (A5) for each component of the vector function $\mathbf{f}$ and for each set of fixed values for $r_1, \cdots r_N$ and for $k_1, \cdots k_N$.

From the proof of Lomont and Moses, as outlined above, we have a unique decomposition

$$\varrho^{(k)} = \xi^{(k)} + \mathbf{n}^{(k)},$$

with $\xi^{(k)}$ and $\mathbf{n}^{(k)}$ satisfying Eqs. (A1) and (A2). Let

$$\xi_i^{(k_1,\cdots k_N)} = \sum_k \xi_i^{(k)}$$

and

$$\eta_i^{(k_1,\cdots k_N)} = \sum_k \eta_i^{(k)}$$

be the unique vectors whose components in each irreducible subspace are, respectively, $\xi_i^{(k)}$ and $\eta_i^{(k)}$. Then

$$\varrho^{(k_1,\cdots k_N)} = \xi^{(k_1,\cdots k_N)} + \mathbf{n}^{(k_1,\cdots k_N)}.$$

We define functions $\mathbf{g}(\mathbf{x}_1, \cdots \mathbf{x}_N)$ and $\mathbf{h}(\mathbf{x}_1, \cdots \mathbf{x}_N)$ by spherical harmonic expansions of the form (A5) with ρ replaced by ξ and η, respectively. It follows that

$$\mathbf{f}(\mathbf{x}_1, \cdots \mathbf{x}_N) = \mathbf{g}(\mathbf{x}_1, \cdots \mathbf{x}_N) + \mathbf{h}(\mathbf{x}_1, \cdots \mathbf{x}_N).$$

It also follows that

$$\mathbf{L} \times \mathbf{g} = -\mathbf{g}, \qquad \mathbf{L} \cdot \mathbf{h} = 0.$$

For example, the equation

$$\begin{aligned}
\mathbf{L} \times \mathbf{g} &= \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \sum_{m_1',\cdots m_N'} \mathbf{L}^{(k_1,\cdots k_N)}_{m_1,\cdots m_N, m_1',\cdots m_N'} \\
&\times \xi^{k_1,\cdots k_N, m_1',\cdots m_N'} \times Y_{k_1 m_1} \cdots Y_{k_N m_N} \\
&= \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} [\mathbf{L}^{(k_1,\cdots k_N)} \times \xi^{(k_1,\cdots k_N)}]_{m_1,\cdots m_N} \\
&\times Y_{k_1 m_1} \cdots Y_{k_N m_N} \\
&= \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} [\sum_k \mathbf{L}^{(k)} \times \xi^{(k)}]_{m_1,\cdots m_N} Y_{k_1 m_1} \cdots Y_{k_N m_N} \\
&= -\sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} [\sum_k \xi^{(k)}]_{m_1,\cdots m_N} Y_{k_1 m_1} \cdots Y_{k_N m_N} \\
&= -\sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \xi^{k_1,\cdots k_N, m_1,\cdots m_N} Y_{k_1 m_1} \cdots Y_{k_N m_N} \\
&= -\mathbf{g}
\end{aligned}$$

is seen to be a consequence of Eq. (A1). Similarly, the equation

$$\mathbf{L} \cdot \mathbf{h} = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} [\sum_k \mathbf{L}^{(k)} \cdot \mathbf{n}^{(k)}]_{m_1,\cdots m_N} \times Y_{k_1 m_1} \cdots Y_{k_N m_N} = 0$$

results from Eq. (A2). Furthermore, that $\mathbf{g}$ and $\mathbf{h}$ are the unique decomposition of $\mathbf{f}$ having these properties is clear from uniqueness of the decomposition of each $\boldsymbol{\varrho}^{(k)}$ into $\boldsymbol{\xi}^{(k)}$ and $\mathbf{n}^{(k)}$ satisfying Eqs. (A1) and (A2).

Finally, we shall show that there exist functions ϕ and ψ such that

$$\mathbf{g} = \mathbf{L}\phi, \qquad \mathbf{h} = \mathbf{L}\times\boldsymbol{\psi} + \boldsymbol{\psi}.$$

Let

$$\phi = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \alpha^{k_1,\cdots k_N, m_1,\cdots m_N}(r_1, \cdots r_N) \times Y_{k_1 m_1}(\theta_1, \phi_1) \cdots Y_{k_N m_N}(\theta_N, \phi_N),$$

$$\boldsymbol{\psi} = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \boldsymbol{\beta}^{k_1,\cdots k_N, m_1,\cdots m_N}(r_1, \cdots r_N) \times Y_{k_1 m_1}(\theta_1, \phi_1) \cdots Y_{k_N m_N}(\theta_N, \phi_N).$$

In accordance with the notation used above, we consider the expansion coefficients $\alpha^{k_1,\cdots k_N, m_1\cdots m_N}$ and $\beta_i^{k_1,\cdots k_N, m_1,\cdots m_N}$ for each set of fixed values of $r_1, \cdots r_N$ and of $k_1, \cdots k_N$ to be the components, respectively, of vectors $\alpha^{(k_1,\cdots k_N)}$ and $\beta_i^{(k_1,\cdots k_N)}$ in the $[(2k_1 + 1) \cdots (2k_N + 1)]$-dimensional representation space of the direct-product representation of the three-dimensional rotation group. Let

$$\alpha^{(k_1,\cdots k_N)} = \sum_k \alpha^{(k)}, \qquad \beta_i^{(k_1,\cdots k_N)} = \sum_k \beta_i^{(k)}$$

be the decomposition of these vectors in terms of their components $\alpha^{(k)}$ and $\beta_i^{(k)}$ in each of the irreducible subspaces. Then

$$\mathbf{L}\phi = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \Big(\sum_k \mathbf{L}^{(k)}\alpha^{(k)}\Big)_{m_1,\cdots m_N} Y_{k_1 m_1} \cdots Y_{k_N m_N},$$

$$\mathbf{L}\times\boldsymbol{\psi} + \boldsymbol{\psi} = \sum_{\substack{k_1,\cdots k_N \\ m_1,\cdots m_N}} \Big[\sum_k (\mathbf{L}^{(k)}\times\boldsymbol{\beta}^{(k)} + \boldsymbol{\beta}^{(k)}\Big]_{m_1,\cdots m_N} \times Y_{k_1 m_1} \cdots Y_{k_N m_N}.$$

The existence of functions ϕ and $\boldsymbol{\psi}$ having the desired properties is thus ensured by the existence of solutions $\alpha^{(k)}$ and $\boldsymbol{\beta}^{(k)}$ of Eqs. (A3) and (A4). These were established by Lomont and Moses. This completes the proof of the theorem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Quantum Electrodynamics of the Stimulated Emission of Radiation. I

R. M. BEVENSEE
Electrical Engineering Department, University of California, Berkeley, California
(Received 20 September 1963)

The quantum electrodynamic formalism of Senitzky is developed so as to yield *long-time* differential equations for the time-average expectation values of electric field energy and excess molecular population in a system composed of one resonant cavity mode, two-level molecules, and a dissipation mechanism. The cavity-mode resonant frequency is essentially the mean molecular transition frequency of either a Gaussian or Lorentzian distribution of transition frequencies. The intermolecular and lattice T_1 and T_2 time constants are assumed to be much longer than the stimulated emission period. The field is coupled to either the electric or the magnetic dipole moments of the molecules.

Senitzky's papers are summarized and the relevant expressions for the Heisenberg field and molecular operators are listed. Modified time-average expressions are obtained for the presence of a transient coherent driving field and perhaps an "off-resonance" molecular distribution of transition frequencies. Time averages of the derivatives of various terms in the expected field energy are compared to derivatives of time averages of those terms. Well-behaved differential equations are obtained for the time-average expectation values of field energy and molecular excess population, justified on a long-time basis by the application of an intermittent similarity transformation. A differential equation for the time-average dispersion (second moment) of electric field energy is obtained, which indicates that the relative dispersion tends to decrease during the pumping interval and increase back to the thermal value during the emission interval. Energy transfer from pumping field to molecules to the resonant field during the pumping interval is described qualitatively. The direct-product form for the density matrix $\rho = \rho(\text{field}) \times \rho(\text{molecules}) \times \rho(\text{dissipation mechanism})$ is justified by maximum-entropy inference. In conclusion, the equations of motion for the time-average expectation values of field energy and molecular population are interpreted so as to explain the envelope modulation of a solid-state laser beam during the emission period.

I. INTRODUCTION

IN this paper we present a detailed description of the interaction between one quantized resonant cavity mode of the electromagnetic field and an assemblage of molecules, each of which has effectively only two energy levels. The cavity mode may be of either microwave or optical frequency ω, and is coupled not only to the molecules via either their electric or their magnetic dipole moments, but also to a stochastic dissipation mechanism in the walls. With the exception of Sec. 7, the pumping field is absent after the establishment of the initial nonequilibrium distribution of the molecules among their energy states. The molecular transition frequency between the two energy levels is assumed to be distributed in either a Gaussian or Lorentzian fashion about a mean frequency $\omega_M \approx \omega$. The wavefunction for the molecular system is taken to be a direct product of the individual wavefunctions with a random distribution of their phases. Since the T_2 or intermolecular coupling merely causes a reshuffling of energy among the molecules all of which are radiating into the field, this coupling is not examined. The T_1-coupling to the lattice is neglected entirely in this first paper.

Perhaps the most important derivation in this paper is a pair of second-order differential equations, one for the time-average field energy at the natural frequencies of the cavity as determined in a nonlinear way by itself and the time-average excess molecular population, and one for the time-average excess molecular population. This pair is valid for arbitrarily long time intervals and implies conservation of energy between the molecules and the field. "Time-average" means an average over many periods of the field frequency ω but during an interval short compared to the time constant of the empty cavity and the stimulated emission period. Our equations are not analogous to the usual rate equations for the amplitudes of the molecular wavefunction, so often assumed in semiclassical analysis which neglects the perturbation of the field by the molecules.[1] Semiclassical analysis is often satisfactory for gas masers and lasers but this paper is primarily devoted to a description of the solid-state phenomena. Nor are our differential equations the result of evaluating commutators according to the Heisenberg prescription with operator representations consisting of matrices unchanged by the interaction. Rather, our equations are obtained by solving the Heisenberg equations of motion for the field and molecular operators by means of second-order perturbation theory (in which terms quadratic in the field-molecular coupling are retained and the

[1] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., New York, 1955) 2nd ed., Sec. 35.

second-order field energy terms grow as t^2). The resultant equations, valid for an arbitary short-time interval (long compared to the interval of time averages) are then interpreted over the entire interval during which stimulated emission occurs. Philosophically our analysis is similar to that of Wangsness and Bloch,[2] who obtained a set of first-order differential equations for the elements of the reduced density matrix for a statistical ensemble of nuclei.

Our differential equations apply to time-average field energy and molecular excess population as averaged over an ensemble of similar systems. In order to show how well they quantatively describe an individual member of the ensemble we also examine a differential equation for the dispersion of electric field energy $\mathfrak{IC}_e$, defined as $\langle \mathfrak{IC}_e^2 \rangle_{ta} - (\langle \mathfrak{IC}_e \rangle_{ta})^2$, where subscript ta denotes a time average. It turns out that the dispersion which is $2(\langle \mathfrak{IC}_e \rangle_{ta})^2$ in thermal equilibrium tends to decrease during the pumping interval and increase back to the equilibrium value during the subsequent emission interval. (We can analyze the dispersion at the natural frequencies of the cavity independently of the dispersion of the pumping fields, according to subsequent remarks.)

One might think that the application of second-order perturbation theory successively over a large number of short-time intervals would unavoidably introduce an increasing cumulative error in the time-average expectation values of field energy and molecular population. That this, in fact, does not occur is shown by employing an "intermittent similarity transformation" to the molecular operators at the end of each short-time interval. This similarity transformation restores the dipole moment and molecular Hamiltonians to their initial off-diagonal and diagonal forms, respectively, so that the perturbation analysis can be repeated for the next interval. Of course the transformation must also be applied to the density matrix for the system—which is assumed to be in the direct-product form of ρ(field) $\times$ ρ(molecules) $\times$ ρ(loss mechanism) according to maximum entropy inference[3]—thereby obliterating the details of ρ(molecules) even though it does not change the expectation values of molecular operators. A further transformation must be applied to ρ(field) in order to establish continuity of the expectation values of field operators, consisting of zeroth-, first-, and second-order portions, at the end of one short-time interval with the corresponding expectation values at the beginning of the next interval, of zero-order only. But the individual elements of ρ(field) and ρ(molecules) are not important because the equations for the time-average expectation values of field energy and (fortunately) of electric energy dispersion do not involve them.

Our analysis is an elaboration upon the work of Senitzky, whose many papers on the subject of induced and spontaneous emission in a coherent field will be summarized in the next section.

2. SENITZKY'S ANALYSIS OF THE FIELD-MOLECULAR INTERACTION

In a series of advanced papers on the subject of induced and spontaneous emission in a coherent field, Senitzky has furnished a quantum electrodynamic description of the phenomena which is a logical and relatively thorough extension of Einstein's derivation[4] of the spontaneous and induced "A" and "B" coefficients for a single molecule in a thermal field. An important assumption in his perturbation treatment is that of resonance of the molecular centerfrequency with the cavity-mode frequency: $\omega_M = \omega$. Later we shall liberalize this condition somewhat.

In Part I[5] of his series he considers the interaction of one coherent, resonant, lossless-cavity mode with a number of molecules all with the same transition frequency between two energy levels equal to the mode resonant frequency. The molecules are coupled to the field (but not among themselves or to the lattice in his series of papers) via their electric dipole moments and both field and molecules are treated quantum mechanically in the Heisenberg framework. Expressions for the coherent and incoherent portions of the field strength and field energy are derived by means of second-order perturbation theory. The coherent portion of the field has predictable phase; the coherent portion of the field energy arises from the square of the expectation values of the field amplitudes. Terms in the expressions for field amplitudes and energies are identified as representing either induced (proportional to a field strength or its square) or spontaneous emission. The field energy is found to consist of the following portions [Eq. (18) of Ref. 5]: an energy due to the initial field; induced coherent energy growing as Nt^2 for $\omega t \gg 1$, N being the number of molecules and t the time; another portion of induced coherent energy depending upon the

[2] R. K. Wangsness and F. Bloch, Phys. Rev. **89**, 728 (1953).
[3] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[4] A. Einstein, Z. Physik. **18**, 121 (1917).
[5] I. R. Senitzky, Phys. Rev. **111**, 3 (1958).

initial phases of the various molecules; a spontaneous coherent portion growing as t^2 and dependent upon the number of molecules in the upper energy state; a spontaneous coherent portion proportional to t^2 but dependent upon the phase distribution of the molecules; and, finally, a spontaneous incoherent energy growing as t^2 and dependent upon the number of molecules in the upper energy state. For the situation in which all the molecules are originally in the upper energy state the ratio of induced (coherent) energy to spontaneous (incoherent) energy remains constant, for $\omega t \gg 1$, at $(E_0^2/8\pi)/\hbar\omega$, the equivalent number of photons in the cavity $E_0^2/8\pi$ being just the classical driving-field energy. Lastly, correlated vs uncorrelated states are discussed for the molecules with respect to spontaneous emission.

In part II[6] of the series Senitzky extends the analysis to allow for interaction of the fields of the one cavity mode in resonance with the molecules with an ensemble of loss dipoles in the walls. Initially, the molecules are absent and the interaction of the electric field operator amplitude P (Q for the magnetic field) with the loss dipoles is discussed. The integral equation is obtained for P (averaged over the loss-mechanism states) as driven by the coupling operator, where the latter is driven by the loss-dipole Hamiltonian which, in turn, is driven by the electric field coupled to it. It is shown that the effect of the loss mechanism is not only a linear dissipation term in the equation of motion for P and Q, but also a stochastic "drive" of these operators by the incoherent coupling operator $(F^{(0)}(t))$. As a result, the expectation values of the field operators with respect to the loss mechanism are not affected by $F^{(0)}$ but the expectation value of the field energy contains an incoherent (induced) thermal portion which is just $\hbar\omega(e^{\hbar\omega/kT} - 1)^{-1}$ after the transient has died away. The commutator $[P, Q] = -i\hbar$ through second-order terms in the perturbation theory, and the zero-point energy remains $\frac{1}{2}\hbar\omega$ for all time as it must.

Senitzky's paper II continues with the introduction of the molecules into a driven field, the transient cavity field and transient portion of the driving field having decayed away. The complete Hamiltonian now consists of field energy, field-dissipation and field-molecular coupling energy, loss-mechanism Hamiltonian, and molecular Hamiltonian. The Heisenberg integral equations of motion for $P(t)$ and $Q(t)$ and the other operators are developed as a power series in the field-molecular coupling coefficient. Expressions for the zeroth-, first-, and second-order portions of P and Q (the second-order portions being integral expressions of the zero-order cavity electric field, including the thermal field) are developed for time $t \geq 0$ and, from them, expressions are obtained for the electric and magnetic field energies through second-order terms. The result is a zero-order field energy containing driving field, zero-point, and thermal portions, a first-order field energy which has no expectation over an assemblage of molecules with random initial phases, and a second-order energy consisting of two parts. The first part is spontaneous and incoherent, becoming proportional to the number of molecules in the upper energy state as the transient $e^{-\frac{1}{2}\beta t}$ term goes to zero; the second part consists of an induced coherent portion and an induced incoherent portion due to the thermal field. For $\beta t \gg 1$ the spontaneous-emission energy approaches a constant, the coherent induced-emission energy tends to grow linearly with the time, and the incoherent thermal energy approaches another constant. The molecular Hamiltonian is also studied through second-order terms and various statements are made about the induced and spontaneous emission.

In paper III,[7] the results of which appear below for reference in our paper, Senitzky extends the analysis of the first two papers to include a molecular frequency spread. Each molecule, uncoupled to others and to the lattice, has two well-defined energy levels E_{m2} and E_{m1} and the transition frequency $(E_{m2} - E_{m1})/\hbar$ has a Gaussian distribution centered about the cavity resonant frequency. The phase of the molecular wavefunctions is random. The cavity driving field remains a coherent classical electric dipole of frequency equal to the cavity resonant frequency and transient fields have died away. Those portions of the field energy resulting from induced and spontaneous emission initially increase as t^2, but approach steady-state values after transient periods (usually different) determined by two time constants: cavity relaxation time and inverse molecular-frequency spread. Both the induced and spontaneous emission *powers* radiated by the molecules initially increase linearly with the time and approach steady-state values after transient periods. The ratio of coherent induced to spontaneous field energy approaches a steady-state value which depends upon the ratio of cavity bandwidth to molecular frequency spread. [See our Eq. (21).]

[6] I. R. Senitzky, Phys. Rev. **115**, 227 (1959).

[7] I. R. Senitzky, Phys. Rev. **119**, 1807 (1960).

The important expressions from Senitzky's paper III[7] are now listed. The Hamiltonian for the system is

$$\mathfrak{H} = \left(2\pi c^2 P^2 + \frac{\omega^2}{8\pi c^2} Q^2\right) + \sum_j \mathfrak{H}_j + \sum_m \mathfrak{H}_m + 4\pi c P[U \sum_j F_j + u \sum_m \gamma_m + D u_D] \quad (1)$$

in which the first two terms are the field energy, $\mathfrak{H}_j$ and $\mathfrak{H}_m$ are the Hamiltonians of the jth loss dipole and mth molecule, U measures the strength of the electric-field–loss-dipole coupling, u measures the strength of the electric-field–molecule coupling ($E = -4\pi c P u$), and D is the prescribed classical drive

$$D = D_0 \sin(\omega t + \theta). \quad (2)$$

The electric and magnetic fields are

$$E = -4\pi c P(t) u(r), \qquad H = Q(t) \nabla \times u, \quad (3a)$$

with the patterns normalized as

$$\int_v u^2(r)\, dv = 1, \quad \int_v (\nabla \times u)^2\, dv = k^2, \quad k = \omega/c. \quad (3b)$$

For electric-dipole coupling, it is convenient to let u be dimensionless; for magnetic-dipole coupling it is convenient to specify the patterns differently according to subsequent remarks. The zero-order uncoupled portion of F_j, denoted $F_j^{[0]}$, is represented by an off-diagonal matrix which need not be quoted here[8]; the zero-order molecular operators are represented by

$$\gamma_m^{[0]}(t) = \tilde{\gamma} \begin{bmatrix} 0 & e^{-i\omega_m t} \\ e^{i\omega_m t} & 0 \end{bmatrix}, \quad \mathfrak{H}_m^{[0]} = \begin{bmatrix} E_{m_1} & 0 \\ 0 & E_{m2} \end{bmatrix}, \quad (4)$$

for molecule m. The wavefunction for the total system is taken to be of direct-product form,

$$\Psi = \psi(\text{dissipation})\psi(\text{field}) \sum \psi_m, \qquad \psi_m = a_{m1}\varphi_{m1} + a_{m2}\varphi_{m2}, \quad (5)$$

with a random phase distribution among the a_{mi}.

After obtaining the Heisenberg equation of motion for the operators, forming integral equations for $P(t)$ and $Q(t)$, averaging them over the states of the loss mechanism, and forming second-order differential equations for the averaged P and Q, the solutions are (see the end of this section for the expressions for operators P and Q appropriate for magnetic-dipole coupling between field and molecules)

$$P_D(t) = (\omega u_D/c\beta) D_0 \cos(\omega t + \theta) \equiv P_{D_0} \cos(\omega t + \theta), \quad (6a)$$

$$P_F(t) = -\frac{\omega}{c} U \sum_j \int_{-\infty}^{t} dt_1 F_j^{[0]}(t_1) e^{-\frac{1}{2}\beta(t-t_1)} \times \sin\omega(t - t_1), \quad (6b)$$

$$P_\gamma(t) = -\frac{\omega}{c} u \sum_m \int_0^t dt_1 \gamma_m(t_1) e^{-\frac{1}{2}\beta(t-t_1)} \times \sin\omega(t - t_1), \quad (6c)$$

for the driven, loss-mechanism, and molecular-induced portions of P, respectively. Q can be obtained from the equation of motion

$$\dot{P} = -(\omega^2/4\pi c^2) Q. \quad (7)$$

$\beta = \omega/Q_e$ is the dissipation constant, Q_e being the cavity quality factor or the inverse of the empty cavity bandwidth.

P (as well as Q and γ_m) is expanded in ascending orders,

$$P(t) = P^{[0]} + P^{[1]} + P^{[2]} + \cdots, \quad (8)$$

where $\langle P^{[0]} \rangle$, as averaged[9] over the direct-product wavefunction of Eq. (5), is

$$\langle P^{[0]}(t) \rangle = \langle P_D + P_F \rangle = P_D, \quad (9)$$

because of the random phase of the loss dipole operator $F_j^{[0]}$ in Eq. (6b). $P^{[1]} = P_\gamma^{[1]} \sim \int \gamma_m^{[0]}(t_1)\, dt_1$, according to Eq. (6c) and $P^{[2]} = P_\gamma^{[2]} \sim \int \gamma_m^{[1]}(t_1) dt_1$, where $\gamma_m^{[1]}$ as a solution to the equations of motion for itself and $\mathfrak{H}_m(t)$ is

$$\gamma_m^{[1]}(t) = \frac{8\pi c}{\hbar} u \tilde{\gamma}^2 I_m \int_0^t dt_1 P^{[0]}(t_1) \sin\omega_m(t - t_1) \quad (10a)$$

$$I_m = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \langle I_m \rangle = \psi_m^* I_m \psi_m = -|a_{m1}|^2 + |a_{m2}|^2. \quad (10b)$$

Evaluation of these operators yields, for the time-average ensemble average over states of the loss mechanism and the N molecules,

$$\langle P^{[1]}(t) \rangle = 0, \quad (11a)$$

$$\langle P^{[2]}(t) \rangle_{ta} = (4\pi/\beta\hbar)\omega u^2 \tilde{\gamma}^2 N \langle I_m \rangle_{av} P_D(t) F_1(t), \quad (11b)$$

with

$$F_1(t) = \int_0^t dt_1 e^{-t_1^2/4\alpha^2} - e^{-\frac{1}{2}\beta t} \int_0^t dt_1 e^{-t_1^2/4\alpha^2 + \frac{1}{2}\beta t_1}. \quad (11c)$$

[7] I. R. Senitzky, Phys. Rev. **119**, 670 (1960).

[9] The $\langle\ \rangle$ denote an ensemble average at a given time with respect to the total wavefunction Ψ of Eq. (5). Since Ψ remains in direct-product form in spite of intermittent similarity transformations, operators which are integral expressions of the zero-order molecular operators are averaged with respect to both $\sum\psi_m$ and ψ (dissipation) while operators which are integral expressions of zero-order field operators are averaged with respect to ψ (field) and ψ (dissipation).

$\langle I_m \rangle_{av}$ is the time-average population difference of Eq. (10b) for those molecules of transition frequencies $\omega_m \simeq \omega$. Parameter α measures the spectral width of the Gaussian distribution in transition frequency ω_m,

$$\rho(\omega_m) = N(\alpha/\pi^{\frac{1}{2}}) \exp[-\alpha^2(\omega_m - \omega)^2], \quad (12a)$$

which is normalized as

$$\int_0^\infty \rho(\omega_m)\, d\omega_m \simeq \int_{-\infty}^{\infty} \rho(\omega_m)\, d\omega_m = N,$$

N is total number of molecules. (12b)

For the expectation value of field energy, we require the following ensemble average:

$$\langle P^2 \rangle = \langle P^{[0]2} \rangle + \langle P^{[1]2} \rangle + \langle [P^{[0]}, P^{[1]}]_+ \rangle + \langle [P^{[0]}, P^{[2]}]_+ \rangle, \quad (13)$$

where $[A, B]_+ \equiv AB + BA$. The expansion for Q^2 is similar. The result is

$$\langle \mathfrak{IC}_{field} \rangle = \langle \mathfrak{IC}_f^{[0]} \rangle + \langle \mathfrak{IC}_f^{[2]} \rangle, \quad (14a)$$

$$\langle \mathfrak{IC}_f^{[0]} \rangle_{ta} = E_{D_0}^2/8\pi + \tfrac{1}{2}\hbar\omega + \hbar\omega(e^{\hbar\omega/kT} - 1)^{-1}, \quad (14b)$$

$$\langle \mathfrak{IC}_f^{[2]}(t) \rangle_{ta} = \frac{\omega u^2 \tilde{\gamma}^2}{\beta\hbar} NE_{D_0}^2 \langle I_m \rangle_{av} F_1(t) + \frac{2\pi}{\beta} \omega u^2 \tilde{\gamma}^2 N[1 + \langle I_m \rangle_{av} f(T)] F_2(t), \quad (14c)$$

with

$$f(T) = 1 + 2(e^{\hbar\omega/kT} - 1)^{-1}, \quad (15a)$$

$$F_2(t) = \int_0^t dt_1 e^{-t_1^2/4\alpha^2 - \frac{1}{2}\beta t_1} - e^{-\beta t} \int_0^t dt_1 e^{-t_1^2/4\alpha^2 + \frac{1}{2}\beta t_1}. \quad (15b)$$

The first portion of $\langle \mathfrak{IC}_f^{[2]} \rangle$ is induced and coherent; the F_2 portion can be decomposed into an induced thermal (incoherent) portion proportional to $(e^{\hbar\omega/kT} - 1)^{-1}$ and a remainder which is spontaneous and incoherent.

The basic equations for the molecules are

$$\sum_m \langle \dot{\mathfrak{IC}}_m \rangle = -4\pi c u \sum_m \langle P\dot{\gamma}_m \rangle \quad \text{(the dot is } d/dt), \quad (16)$$

the first-order portion of which is

$$\sum_m \langle \dot{\mathfrak{IC}}_m^{[1]} \rangle = -4\pi c u \sum_m \langle P^{[0]} \dot{\gamma}_m^{[0]} \rangle, \quad (17)$$

which has no expectation over the assembly of molecules with random phases among their wavefunctions. The second-order portion evaluates conveniently as

$$-\sum_m \langle \dot{\mathfrak{IC}}_m^{[2]} \rangle_{ta} = -2\pi c u \sum_m \{\langle [P^{[0]}, \dot{\gamma}_m^{[1]}]_+ \rangle + \langle [P^{[1]}, \dot{\gamma}_m^{[0]}]_+ \rangle\}_{ta} = \frac{\omega u^2 \tilde{\gamma}^2}{2\hbar} NE_{D_0}^2 \langle I_m \rangle_{av} f_1(t) + 2\pi(\omega u \tilde{\gamma})^2 N[1 + \langle I_m \rangle_{av} f(T)] f_2(t), \quad (18)$$

with

$$f_1(t) = \int_0^t dt_1 e^{-t_1^2/4\alpha^2}, \quad (19a)$$

$$f_2(t) = \int_0^t dt_1 e^{-t_1^2/4\alpha^2 - \frac{1}{2}\beta t_1}. \quad (19b)$$

The f_1 portion is induced; the f_2 portion can be decomposed into thermally induced and spontaneous parts.

Examination of the *double*-integral expressions for the various portions of the field energy indicates, after differentiation with respect to time, that the induced coherent power flow into the total field is exactly that from the molecules less the amount

$$\frac{\beta}{4\hbar} \omega u^2 \tilde{\gamma}^2 N \langle I_m \rangle_{av} E_{D_0}^2 e^{-\frac{1}{2}\beta t} \times \int_0^t dt_1 \int_0^{t_1} dt_2 e^{\frac{1}{2}\beta t_1} e^{-(t_1 - t_2)^2/4\alpha^2}; \quad (20a)$$

the induced incoherent power flow due to the thermal field and that part of the spontaneous power flow due to molecular dipole moment fluctuations is less into the total field than from the molecules by an amount

$$2\pi\beta(\omega u \tilde{\gamma})^2 N \langle I_m \rangle_{av} f(T) e^{-\beta t} \times \int_0^t dt_1 \int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1 + t_2)} e^{-(t_1 - t_2)^2/4\alpha^2}; \quad (20b)$$

and the remainder of the spontaneous power flow due to vacuum fluctuations is less into the field than from the molecules by

$$\pi\beta(\omega u \tilde{\gamma})^2 N e^{-\beta t} \times \int_0^t dt_1 \int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1 + t_2)} e^{-(t_1 - t_2)^2/4\alpha^2}. \quad (20c)$$

Expressions (20) represent the time-average power flowing into the dissipation mechanism.

The important conclusions of Senitzky's paper III may be summarized as follows. The ratio of induced (nonthermal) to spontaneous energy, with the molecules initially in the upper energy state, approaches the steady-state value of

$$\frac{2n}{\exp(r^2)(1 - \operatorname{erf} r)}, \qquad n = \frac{E_{D_0}^2/8\pi}{\hbar\omega},$$

photons in cavity, (21)

with

$$r = \frac{\text{cavity bandwidth}}{\text{molecular linewidth}} = \tfrac{1}{2}\alpha\beta.$$

"erf" stands for the error function. The narrower the frequency spread of the molecules the larger is the induced steady-state power emitted but the longer the initial transient period becomes, with the result that the steady-state expression for induced field energy or induced transition probability may not be applicable.

In Senitzky's paper IV,[10] to which we do not refer further in this paper, the perturbation restriction upon the field is removed and each molecule is assumed to radiate only slightly during the interaction time as in a molecular beam maser. Two different molecular distributions in transition frequency are considered, as well as several types of driving field. The molecules are presumed to be initially in an emissive state.

The foregoing equations are appropriate for electric-dipole coupling with the field. The modifications for magnetic-dipole coupling are rather trivial *provided* no dc magnetic field is present. If this is so, $4\pi cPu\sum\gamma_m$ of Eq. (1) is replaced by $Q\sum\gamma_m\nabla\times u$, γ_m now being the magnetic-dipole moment, whereupon Eqs. (6) and (7) are changed to

$$Q_F(t) = 4\pi cU\sum_i\int_{-\infty}^{t} dt_1 F_i^{[0]}(t_1)e^{-\frac{1}{2}\beta(t-t_1)} \times\cos\omega(t-t_1), \tag{22a}$$

$$Q_\gamma(t) = -\frac{4\pi c^2}{\omega}\sum_m\int_0^t dt_1\gamma_m(t_1)\cdot\nabla\times u e^{-\frac{1}{2}\beta(t-t_1)} \times\sin\omega(t-t_1), \tag{22b}$$

and the expressions for P satisfying the operator equations of motion for P and Q are

$$P_D(t) = P_{D_0}\cos(\omega t+\theta), \tag{23a}$$

$$P_F(t) = -\frac{\omega}{c}U\sum_i\int_{-\infty}^{t} dt_1 F_i^{[0]}(t_1)e^{-\frac{1}{2}\beta(t-t_1)} \times\sin\omega(t-t_1), \tag{23b}$$

$$P_\gamma(t) = -\sum_m\int_0^t dt_1\gamma_m(t_1)\cdot\nabla\times u e^{-\frac{1}{2}\beta(t-t_1)} \times\left[-\frac{\beta}{2\omega}\cdot\sin\omega(t-t_1)+\cos\omega(t-t_1)\right]. \tag{23c}$$

Only the component of γ_m (magnetic) parallel to the magnetic field pattern $\nabla\times u$ is relevant.

Equations (22) and (23) cannot be transformed identically into Eqs. (6) and (7) by means of appropriate changes of variables. However, Senitzky's expressions—and the expressions in our paper—for time-average energies and powers also describe (except for terms proportional to β) those energies and powers for the case of magnetic-dipole coupling if the following changes are made:

$$\gamma(\text{electric dipole})\rightarrow\gamma(\text{magnetic dipole})$$
$$P\rightarrow(\omega/4\pi c^2)Q,\qquad Q\rightarrow-(4\pi c^2/\omega)P,\qquad u\rightarrow(c/\omega)\nabla\times u. \tag{24}$$

Differences between the β terms of the electric-dipole and magnetic-dipole expressions are due to the fact that the *electric* field couples to the loss mechanism in both cases.

3. MODIFICATIONS TO ACCOUNT FOR TRANSIENT DRIVING FIELD AND AN "OFF-RESONANT" MOLECULAR DISTRIBUTION

We are primarily concerned with the equations of motion for field and molecular operators after the pumping source of power has been turned off and the stimulated emission of radiation has begun. Consequently, we must modify the work of the preceding section in order to provide for a transient rather than a driven coherent field. One resonant cavity mode still interacts with the molecules, each with only two energy levels, and the distribution in their transition frequencies may be either Lorentzian or Gaussian but the mean transition frequency ω_M is allowed to differ from ω, the resonant frequency of the cavity mode. In order that $\langle I_m\rangle$ be a slowly-varying function of frequency and the analysis be accurate, we require $\omega\simeq\omega_M$ in the sense that $(\gamma_m E_1/\hbar)^2\gg(\omega-\omega_M)^2$, E_1 being $\sqrt{2}$ times the root mean square (rms) value of the electric field. The necessity of this condition becomes evident after a study of the molecules driven by a classical field (see Sec. 7). Our equations will be written for electric-dipole coupling, but the ensuing expressions for time-average energies and powers will also be valid for magnetic-dipole coupling provided the changes of Eq. (24) are made.

The driving terms proportional to D in Eqs. (1) and (6a) are now removed. Otherwise, Eqs. (3)–(8) remain valid. In order to simplify the ensuing formulas and make some of them correspond to analogous expressions for Brownian motion,[11] we introduce now the following definitions for the remainder of this paper:

[10] I. R. Senitzky, Phys. Rev. **123**, 1525 (1961).

[11] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

$$p = -(4\pi c^2/\omega^2)P, \tag{25a}$$

$$q = Q, \tag{25b}$$

$$4\pi cUF^{[0]}(t) \equiv A(t). \tag{25c}$$

p corresponds to position and q to velocity of a particle performing one-dimensional Brownian motion under the action of a stochastic force $A(t)$. With these new variables, Eq. (9), modified for the presence of a coherent transient in place of a driving field, reads

$$\langle p^{[0]}(t)\rangle = \langle p^{(0)}e^{-\beta t/2} + p_A\rangle = \langle p^{(0)}(t)e^{-\frac{1}{2}\beta t}\rangle, \tag{26}$$

$p^{(0)}$ to be specified in a moment. Equation (10) remains valid when written in terms of p of Eq. (25a) and we begin the modifications with Eq. (11).

The zero-order decaying field in the absence of the molecules obeys the following differential equation[11]:

$$\ddot{p}^{[0]}(t) + \beta\dot{p}^{[0]} + \omega^2 p^{[0]} = A(t), \tag{27a}$$

along with

$$\dot{q} = p. \tag{27b}$$

The solution is

$$p^{[0]}(t) = p^{(0)}(t)e^{-\frac{1}{2}\beta t} + \omega_1^{-1}\int_0^t A(t_1)e^{-\frac{1}{2}\beta(t-t_1)} \times \sin\omega(t - t_1)\,dt_1, \tag{28a}$$

$$q^{[0]}(t) = q^{(0)}(t)e^{-\frac{1}{2}\beta t} + \omega_1^{-1}\int_0^t A(t_1)e^{-\frac{1}{2}\beta(t-t_1)} \times [-(\tfrac{1}{2}\beta)\sin\omega_1(t - t_1) + \omega_1\cos\omega_1(t - t_1)]\,dt_1, \tag{28b}$$

where

$$p^{(0)}(t) = [(\beta p_0 + 2q_0)/2\omega_1]\sin\omega_1 t + p_0\cos\omega_1 t, \tag{29a}$$

$$q^{(0)}(t) = -[(2\omega^2 p_0 + \beta q_0)/2\omega_1]\sin\omega_1 t + q_0\cos\omega_1 t, \tag{29b}$$

$$\omega_1^2 = \omega^2 - \tfrac{1}{4}\beta^2. \tag{29c}$$

We retain the small β term in Eq. (28b) in order to describe exactly the time-average expectation values of zero-order energies and powers. We neglect the slight shift of resonant frequency from ω to ω_1, though the validity of the analysis depends upon ω_M being sufficiently close to ω_1 according to the foregoing remark. The rule is: if an integral expression contains exponential factor(s) representing relaxation, then ω in the trigonometric terms is actually ω_1.

Evaluation of $p^{[1]}$, $p^{[2]}$, $q^{[1]}$, $q^{[2]}$, $\gamma_m^{[1]}$, and $\mathfrak{K}_m^{[1]}$ proceeds in the manner described in Sec. 2 except that $p^{[0]}$ is given by Eqs. (28a) and (29a). The expectation values of $p^{[2]}$ and $q^{[2]}$, as well as of $[p^{[0]}, p^{[2]}]_+$ and $[q^{[0]}, q^{[2]}]_+$ in the field energy, depend upon an ensemble average over the molecules of a molecular function proportional to matrix I_m of Eq. (10b); the average over a Gaussian distribution in transition frequency is performed with respect to the molecular density function of Eq. (12), as, for example,

$$\sum_m \langle I_m\rangle \sin\omega_m(t - t_1) = \int d\omega_m \rho(\omega_m)\langle I_m\rangle \sin\omega_m(t - t_1) = N\langle I_m\rangle_M e^{-(t-t_1)^2/4\alpha^2}\sin\omega_M(t - t_1), \tag{30a}$$

in which

$$\langle I_m\rangle_M = |a_{m2}|^2 - |a_{m1}|^2 \quad \text{at} \quad \omega_M. \tag{30b}$$

Henceforth we shall write $\langle I_m\rangle_M$ as simply $\langle I\rangle$, the time-average excess molecular population at centerfrequency ω_M.

For a Lorentzian distribution,

$$\rho_L(\omega_m) = (\alpha/2\pi)\frac{N}{(\omega_M - \omega_m)^2 + \frac{1}{4}\alpha^2}, \qquad \int \rho_L(\omega_m)\,d\omega_m = N, \tag{31}$$

with respect to which the expectation of Eq. (30a) would be

$$\sum_m \langle I_m\rangle \sin\omega_m(t - t_1) = N\langle I_m\rangle_M e^{-\frac{1}{2}\alpha(t-t_1)}\sin\omega_M(t - t_1). \tag{32}$$

To convert from Gaussian to Lorentzian distribution in the ensuing formulas, we need only change the exponential factor from that in Eq. (30b) to that in Eq. (32).

The various orders of field operators, averaged over the molecular ensemble (but not averaged over the loss mechanism), evaluate as

$$\langle p^{[1]}(t)\rangle_m = 0, \tag{33a}$$

$$\langle p^{[2]}(t)\rangle_m = (4\pi u^2\omega/\hbar)\tilde{\gamma}^2 N\langle I\rangle \times \left[\left(\frac{\beta p_0 + 2q_0}{2\omega}\right)F_3(t) - p_0 F_4(t)\right] - \frac{8\pi u^2\tilde{\gamma}^2}{\hbar} \times N\langle I\rangle \int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 A(t_3) \times e^{-\frac{1}{2}\beta(t-t_1)}e^{-\frac{1}{2}\beta(t_2-t_3)}e^{-(t_1-t_2)^2/4\alpha^2} \times \sin\omega_M(t_1 - t_2)\sin\omega(t - t_1)\sin\omega(t_2 - t_3), \tag{33b}$$

$$\langle q^{[1]}(t)\rangle_m = 0, \tag{34a}$$

$$\langle q^{[2]}(t)\rangle_m = \frac{4\pi(u\omega\tilde{\gamma})^2}{\hbar} N\langle I\rangle$$
$$\times \left[\left(\frac{\beta p_0 + 2q_0}{2\omega}\right)F_4 + p_0 F_3\right] - \frac{8\pi(u\omega\tilde{\gamma})^2}{\hbar\omega}$$
$$\times N\langle I\rangle \int_0^t dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 A(t_3)$$
$$\times e^{-\frac{1}{2}\beta(t-t_1)} e^{-\frac{1}{2}\beta(t_2-t_3)} e^{-(t_1-t_2)^2/4\alpha^2}$$
$$\times \sin\omega_M(t_1 - t_2)\cos\omega(t - t_1)\sin\omega(t_2 - t_3), \quad (34b)$$

in which

$$F_3(t) = e^{-\frac{1}{2}\beta t}\int_0^t d\eta e^{\frac{1}{2}\beta\eta - \eta^2/4\alpha^2}(t - \eta)$$
$$\times \sin\omega_M\eta\cos\omega(t - \eta), \quad (35a)$$

$$F_4(t) = e^{-\frac{1}{2}\beta t}\int_0^t d\eta e^{\frac{1}{2}\beta\eta - \eta^2/4\alpha^2}(t - \eta)$$
$$\times \sin\omega_M\eta\sin\omega(t - \eta). \quad (35b)$$

The fluctuation or A portions of Eqs. (33b) and (34b) are best kept in integral form for the evaluation of the time-average thermally induced portions of the field energy.

The time average of the zero-order electric-energy expectation, denoted $\langle \mathcal{H}_e^{[0]}\rangle_{ta}$, including thermally induced energy, is

$$\frac{\omega^4}{8\pi c^2}\tfrac{1}{2}\langle[p^{[0]}(t), p^{[0]}(t)]_+\rangle_{ta}$$
$$= \frac{\omega^4}{8\pi c^2}\frac{1}{2}\left\langle\left(\frac{\beta p_0 + 2q_0}{2\omega}\right)^2 + p_0^2\right\rangle e^{-\beta t}$$
$$+ \tfrac{1}{4}\hbar\omega[1 + 2(e^{\hbar\omega/kT} - 1)^{-1}](1 - e^{-\beta t}), \quad (36)$$

with the help of Eq. (28a). The zero-order magnetic energy is identical.

The first-order time-average electric (or magnetic) energy due to the molecules is spontaneous,

$$\frac{\omega^4}{8\pi c^2}\langle p^{[1]2}(t)\rangle_{ta} = \tfrac{1}{2}\pi(\omega u\tilde{\gamma})^2 N e^{-\beta t}\int_0^t dt_1\int_0^t dt_2$$
$$\times e^{\frac{1}{2}\beta(t_1+t_2)} e^{-(t_1-t_2)^2/4\alpha^2}\cos(\omega - \omega_M)(t_1 - t_2). \quad (37)$$

The $\langle p^{[0]}p^{[1]}\rangle$ portion of the electric energy proportional to $\langle P^2\rangle$ in Eq. (13) is zero if the molecules have random initial phases; the remaining second-order portion evaluates as follows. The second-order electric energy consists of two portions, one induced by the coherent decaying field and denoted by subscript "0" and the other induced by the thermal field, denoted by subscript "A". The time average of the first portion is, according to Eqs. (28a) and (33b),

$$\frac{\omega^4}{8\pi c^2}\langle[p_0^{[0]}(t), p_0^{[2]}(t)]_+\rangle_{ta}$$
$$= \frac{\omega^5 u^2\tilde{\gamma}^2}{4c^2\hbar} N\langle I\rangle e^{-\beta t}\left\langle\left(\frac{\beta p_0 + 2q_0}{2\omega}\right)^2 + q_0^2\right\rangle$$
$$\times \int_0^t dt_1\int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1-t_2)}$$
$$\times e^{-(t_1-t_2)^2/4\alpha^2}\cos(\omega - \omega_M)(t_1 - t_2). \quad (38)$$

The time-average value of the second or thermal portion is, according to Eqs. (28a) and (33b),

$$\frac{\omega^4}{8\pi c^2}\langle[p_A^{[0]}(t), p_A^{[2]}(t)]_+\rangle_{ta}$$
$$= -\tfrac{1}{2}\pi(\omega u\tilde{\gamma})^2 N\langle I\rangle e^{-\beta t}$$
$$\times \int_0^t dt_1\int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1-t_2)} e^{-(t_1-t_2)^2/4\alpha^2}$$
$$\times \cos(\omega - \omega_M)(t_1 - t_2)\{e^{\frac{1}{2}\beta t_2} - 1$$
$$+ (e^{\beta t_2} - e^{\frac{1}{2}\beta t})_{t\le 2t_2}\}. \quad (39)$$

Terms smaller than those in Eqs. (39) by the factor β/ω have been dropped. The time-average second-order magnetic energy is also given by Eqs. (38) and (39). The very last term in Eq. (39) is only present if $t \le 2t_2$ because of a pecularity in averaging over the states of the loss mechanism, explained as follows.

To evaluate the anticommutator in Eq. (39), as given by a 4-fold integral obtained by using Eqs. (28a) and (33b), we must average over the loss-mechanism states a double integral of the form

$$\int_0^t dt_0\int_0^{t_2} dt_3\langle[A(t_0), A(t_3)]_+\rangle e^{\frac{1}{2}\beta(t_0+t_3)}$$
$$\times \sin\omega(t - t_0)\sin\omega(t_2 - t_3), \quad (40a)$$

in which[8]

$$\langle[A(t_0), A(t_3)]_+\rangle$$
$$= A_0\sum_{ik} e^{-E_i/kT}|A_{ik}|^2\, 2\cos\omega_{ik}(t_0 - t_3). \quad (40b)$$

A_0 is the normalizing factor for the diagonal density matrix of the loss-mechanism states and $|A_{ik}|$ is the matrix element connecting states i and k. Upon converting to the variables

$$\zeta = t_0 + t_3, \qquad \eta = t_0 - t_3, \quad (41)$$

that portion of Eq. (40a) evaluated over region III of Fig. 1 involves the double integral

$$\int_t^{t+t_2} d\zeta\int_{\zeta-2t_2}^{2t-\zeta} d\eta e^{\frac{1}{2}\beta\zeta}\cos\omega_{ik}\eta$$
$$\times \sin\omega[t - \tfrac{1}{2}(\zeta + \eta)]\sin\omega[t_2 - \tfrac{1}{2}(\zeta - \eta)]. \quad (42)$$

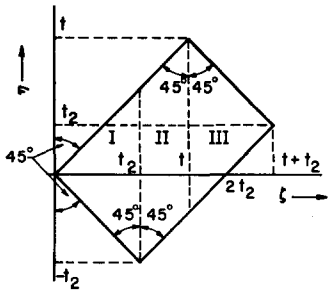

FIG. 1. Region of integration in the ζ–η plane.

After integrating this over η, it becomes apparent that one of the terms to be averaged over the loss-mechanism states before integration over the ζ range is proportional to

$$e^{\frac{1}{2}\beta t}\sin(\omega_{ik}-\omega)(\zeta-2t_2)/(\omega_{ik}-\omega),$$

so that the ζ integral must be split into two portions depending on whether or not $\zeta \leq 2t_2$. As a result, the last term in Eq. (39) is only present if $t \leq 2t_2$.

Equations (37)–(39) are written in forms which, when differentiated with respect to time, become equal (except for small terms proportional to β) to half of the following three portions of the time-average power flowing from the molecules, respectively,

$$-\frac{\omega^2 u}{2c}\cdot\begin{cases}\sum_m \langle[p^{[1]}(t),\dot{\gamma}_m^{[0]}(t)]_+\rangle_{ta}, & \text{spontaneous,}\\ \sum_m \langle[p_0^{[0]},\dot{\gamma}_{m0}^{[1]}]_+\rangle_{ta}, & \text{coherent induced,}\\ \sum_m \langle[p_A^{[0]},\dot{\gamma}_{mA}^{[1]}]_+\rangle_{ta}, & \text{thermal induced.}\end{cases} \quad (43)$$

The rate of increase of time-average magnetic field energy, $(\omega^2/8\pi c^2)\langle q^2\rangle_{ta}$ accounts for the remaining half of the molecular power flow.

Equations (37)–(39) may be integrated once with with respect to the variable $\zeta = t_1 + t_2$ of a transformation like Eq. (41).

4. TIME AVERAGES OF THE ANTICOMMUTATORS

Since the operation of taking a time average is not usually interchangeable with differentiation with respect to time we now list the various time averages of derivatives of expectation terms in the field energy. The expectations of various commutators are taken with respect to specified p_0 and q_0 at the start of the time interval of interest. p_0 and q_0 may be regarded as operators which measure the initial time-average electric (or magnetic) energy and its rate of change. We know by the formal correspondence of Eq. (27) with the equations for Brownian motion that

$$(d/dt\langle p_0^{[0]2}\rangle)_{ta} \equiv d/dt\langle p_0^{[0]2}\rangle_{ta}, \quad (44)$$

and similarly for $q_0^{[0]2}$, including all β terms. For the derivatives of the zero-order thermal energy,

$$(d/dt\langle p_A^{[0]2}\rangle)_{ta} = -\beta\langle p_A^{[0]2}\rangle_{ta} = d/dt\langle p_A^{[0]2}\rangle_{ta}, \quad (45)$$

and similarly for $q_A^{[0]2}$ [refer to Eq. (36)].

The derivatives of spontaneous energy terms have time averages determined by

$$(d/dt\langle p^{[1]2}\rangle)_{ta} = -\beta\langle p^{[1]2}\rangle_{ta}, \quad (46a)$$

$$(d/dt\langle q^{[1]2}\rangle)_{ta} = 2(d/dt)\langle q^{[1]2}\rangle_{ta} + \beta \text{ terms}. \quad (46b)$$

For the second-order induced powers we find

$$(d/dt\langle[p_0^{[0]}(t),p_0^{[2]}(t)]_+\rangle)_{ta} = -\beta\langle[p_0^{[0]},p_0^{[2]}]_+\rangle_{ta}, \quad (47a)$$

$$(d/dt\langle[q_0^{[0]},q_0^{[2]}]_+\rangle)_{ta} \simeq 2(d/dt)\langle[q_0^{[0]},q_0^{[2]}]_+\rangle_{ta}, \quad (47b)$$

in which the approximation becomes an equality if $\omega_1 = \omega_M$. Also,

$$(d/dt\langle[p_A^{[0]},p_A^{[2]}]_+\rangle)_{ta} = \text{terms} \propto \beta, \quad (48a)$$

$$(d/dt\langle[q_A^{[0]},q_A^{[2]}]_+\rangle)_{ta} = 2(d/dt)\langle[q_A^{[0]},q_A^{[2]}]_+\rangle_{ta} + \text{terms} \propto \beta. \quad (48b)$$

Since $\langle\omega^2p^2\rangle_{ta} \equiv \langle q^2\rangle_{ta}$ because the zero-order spontaneous, coherent-induced, and thermal-induced portions are separately equal, we see from Equations (44)–(48) that the time average of the derivative of *total* field energy

$$(\omega^2/8\pi c^2)(d/dt\langle\omega^2p^2+q^2\rangle)_{ta}$$

is identical, except for the β terms, to the derivative of the time average of total field energy. The identity holds for the separate portions of the total energy.

5. EQUATIONS OF MOTION FOR THE TIME-AVERAGE FIELD ENERGY AND MOLECULAR EXCESS POPULATION

We now obtain a second-order differential equation for the time-average electric field energy $W_e \equiv \langle\mathcal{H}_e\rangle_{ta}$ given by the sum of Eqs. (36)–(39). These expressions give the development of $W_e(t)$ from some arbitrary reference time $t = 0$. A first derivative yields

$$\dot{W}_e(t) = \dot{W}_e(0)e^{-\beta t} + \text{power flow from the molecules at time } t. \quad (49)$$

Notice the coherent-induced portion of the molecular power flow, Eq. (38), is proportional to

$$\langle I(0)\rangle\frac{\omega^4}{8\pi c^2}\frac{1}{2}\left\langle\left(\frac{\beta p_0+2q_0}{2\omega}\right)^2+q_0^2\right\rangle = \langle I(0)\rangle W_e(0),$$

in which $\langle I(0)\rangle$ is, by definition, a time-average ensemble average. The thermal-induced portion,

given by the derivative of Eq. (39), is proportional to t^2 and so it is negligible during an extremely short time interval for which Eq. (49) is written. Upon differentiating Eq. (49) once more and changing $\langle I(0)\rangle W_e(0)$ to $\langle I(t)\rangle W_e(t)$, we obtain the equation of motion for time-average field energy at the cavity resonant frequency during *any* short-time interval,

$$\ddot{W}_e(t) + \beta \dot{W}_e = \pi(\omega u \tilde{\gamma})^2 N \times [(4/\hbar\omega)\langle I(t)\rangle W_e(t) + 1]. \quad (50)$$

This remains valid even though $\langle I\rangle$ is driven by a pumping field at a steady-state frequency.

We now claim that Eq. (50), properly interpreted, serves to describe the behavior of $W_e(t)$ during the *whole interval* of stimulated emission. Recall first of all that Eq. (49) was obtained according to the time development of the field amplitude and energy operators dependent upon the *explicit* forms of the zero-order operators only insofar as $\gamma_m^{[0]}$ and $\mathfrak{IC}_m^{[0]}$ were off-diagonal and on-diagonal, respectively. During a short-time interval, they acquire diagonal [Eq. (10a)] and off-diagonal[7] components, respectively. Now a similarity transformation exists, as discussed in Appendix A, which we can apply to $\gamma_m(t)$ and $\mathfrak{IC}_m(t)$ at the end of the brief interval to restore them to their original forms. Of course the same transformation must be applied to all the other operators, including the system density matrix of direct-product form, but the transformation does not affect the field operators and, although it does transform ρ(molecules), it does not affect any molecular expectation values such as $\langle I_m\rangle$ = trace $(\rho_m I_m)$, with $\rho_{m1} = |a_{m1}|^2$ and $\rho_{m2} = |a_{m2}|^2$ of Eq. (10b). So we claim that if this so-called intermittent similarity transformation is applied to the molecular operators at the end of each short-time interval, the analysis which led us to Eq. (50) can be repeated for the next short-time interval, *ad infinitum*. It is important to realize that the details of the field portion of the density matrix, ρ(field), are also lost in this process because ρ(field) must also be changed at the end of each interval in order that the expectation value of each field operator match the zero-order expectation value in terms of the new density matrix at the beginning of the new interval.

Equation (50) states the correct thermal equilibrium relationship between $\langle I\rangle$ and W_e. In equilibrium, $\dot{W}_e = 0$ and $\rho_{m2} = e^{-\mu}\rho_{m1}$, $\mu = \hbar\omega/kT$, so that

$$\langle I\rangle_{\text{equil}} = (-\rho_{M1} + \rho_{M2})_{\text{equil}} = -[(e^{\mu} - 1)/(e^{\mu} + 1)], \quad (51a)$$

while

$$(W_e)_{\text{equil}} = \tfrac{1}{4}\hbar\omega + \tfrac{1}{2}\hbar\omega/(e^{\mu} - 1) = \tfrac{1}{4}\hbar\omega[(e^{\mu} + 1)/(e^{\mu} - 1)], \quad (51b)$$

so that the right side of Eq. (50) is indeed zero. We believe that the correct equilibrium solution of Eq. (50) is an important verification of the entire analysis leading to it; Eq. (50) is based upon a statement of the rate of change of field *power* during an arbitrary short-time interval, and its thermal equilibrium solution merely gives a relationship between the equilibrium values of $\langle I\rangle$ and W_e; but this relationship *is* the correct one.

Along with Eq. (50) we can obtain a second-order differential equation for the time-average molecular excess population, as follows. We observed in Sec. 3 that the derivative of the time-average electric energy, excluding initial energy, is precisely half of the time-average power flowing from the molecules during an arbitrary short-time interval:

$$(d/dt)W_e \text{ (spont + second-order induced)} = -\sum_m \langle \dot{\mathfrak{IC}}_m\rangle_{\text{ta}}. \quad (52)$$

The left side is obtained from Eqs. (37)–(39) with respect to the initial values of $\langle I(0)\rangle$ and $W_e(0)$. Upon differentiating Eq. (52) once, and changing $\langle I(0)\rangle$ and $W_e(0)$ to $\langle I(t)\rangle$ and $W_e(t)$, for an arbitrary short-time interval we obtain

$$-\frac{d}{dt}\sum \langle \dot{\mathfrak{IC}}_m(t)\rangle_{\text{ta}} = 2\pi(\omega u\tilde{\gamma})^2 N \times \left[\frac{4}{\hbar\omega}\langle I(t)\rangle W_e(t) + 1\right], \quad (53)$$

which implies conservation of field plus molecular energy except for the β decay in Eq. (50).

$\langle I(t)\rangle$ appears in both Eqs. (50) and (53); therefore we shall convert the left side of Eq. (53) into an expression involving $(d^2/dt^2)\langle I(t)\rangle$. First, we say that the left side should be identical when interpreted in the Schrödinger framework, viz.,

$$\sum_m \langle \dot{\mathfrak{IC}}_m\rangle_{\text{ta}} = N(\rho_{M1}\dot{\mathfrak{IC}}_{M1} + \rho_{M2}\dot{\mathfrak{IC}}_{M2}) \equiv N(\dot{\rho}_{M1}E_{M1} + \dot{\rho}_{M2}E_{M2})_{\text{Schrodinger}}, \quad (54)$$

an expression which serves to define the Schrödinger time-average populations $\rho_{M1}(t)$ and $\rho_{M2}(t)$ in energy levels 1 and 2 of the average molecules with transition frequency ω_M. If we define a convenient parameter n to measure the fractional excess molecular population,

$$n(t) = (\rho_{M2}(t) - 1/2)_{\text{Schrodinger}}$$

(a time-average quantity). (55)

Equations (50) and (53) can be written as a pair of nonlinear differential equations for the (time-average) W_e and n:

$$\ddot{W}_e(t) + \beta\dot{W}_e = \pi(\omega u\bar{\gamma})^2 N[(8/\omega\hbar)n(t)W_e(t) + 1] = -\tfrac{1}{2}N\hbar\omega_M\dot{n}. \quad (56)$$

Equation (56) is the most important result of this paper, valid when the electric-field pattern u parallel to the electric dipole moments of the molecules (or the magnetic-field pattern $\nabla \times u$ if coupling to the magnetic dipole moments is relevant) is uniform over all the molecules. The interpretation of Eq. (56), which furnishes a discription of envelope modulation in a laser beam emitted from a cavity many wavelengths long, is presented in Sec. 9.

The reader will observe that the parameter $(\omega - \omega_M)$, the difference between the mode resonant frequency and the molecular centerfrequency, does not appear in Eq. (56). The factor $\cos(\omega-\omega_M)(t_1-t_2)$ present in the integrals representing spontaneous and induced emission [Eqs. (37)–(39)] disappeared in the process of differentiating them twice. However the assumption of $(\gamma E_1/\hbar)^2 \gg (\omega - \omega_M)^2$ is implied according to the remark in the first paragraph of Sec. 3. If a pumping field is present, Eq. (56) describes the behavior of the time-average electric field energy at the natural frequencies $\pm\omega + \frac{1}{2}i\beta$ in terms of an excess molecular population which may be determined by the fields at both those frequencies and the pump frequency. These remarks are elaborated upon in Sec. 7.

6. DISPERSION OF ELECTRIC ENERGY WITH TIME

The basic Eq. (56) for the time-average expectation values of electric energy W_e and fractional excess molecular population n describe an ensemble of systems all with the same initial $W_e(0)$ and $n(0)$. How well the equation for W_e at the resonant frequency ω describes a *given* system depends upon the dispersion $D(t)$ of time-average electric field energy, defined as

$$D(t) = \langle[\mathcal{H}_e - \langle\mathcal{H}_e\rangle_{ta}]^2\rangle_{ta} = \langle\mathcal{H}_e^2\rangle_{ta} - (\langle\mathcal{H}_e\rangle_{ta})^2 = \langle\mathcal{H}_e^2\rangle_{ta} - W_e^2. \quad (57)$$

The smaller $D(t)$ is with respect to $W_e(t)$ the more accurately Eq. (56) describes a given system. It is more convenient to describe the dispersion in terms of p rather than $\mathcal{H}_e$ so we define a parameter d as

$$d(t) = (8\pi c^2/\omega^4)^2 D = \langle p^4\rangle_{ta} - (\langle p^2\rangle_{ta})^2. \quad (58)$$

We shall now obtain a differential equation for d, valid over an arbitrary time interval, by the same reasoning which led to Eq. (56). This differential equation, when compared to that for $\langle p^2\rangle_{ta} = (8\pi c^2/\omega^4)W_e$, will indicate that the relative dispersion $D(t)/W_e^2$ tends to remain less than the thermal equilibrium value during the pumping and stimulated emission intervals.

Expansion of operators like $p^4 = (p^{[0]}+p^{[1]}+p^{[2]})^4$ and retention of terms no higher than second-order (proportional to t^2 during a short-time interval) yields

$$d(t) = \{d_0(t) + 4\langle p_0^{[0]2}\rangle\langle p_A^{[0]2}\rangle + d_{th}(t) + \langle[p_0^{[0]2}, [p_0^{[0]}, p_0^{[2]}]_+]_+\rangle - 2\langle p_0^{[0]2}\rangle\langle[p_0^{[0]}, p_0^{[2]}]_+\rangle + 4\langle p_0^{[0]2}\rangle\langle p^{[1]2}\rangle\}_{ta}, \quad (59)$$

in which the initial-value and thermal dispersions are

$$d_0(t) = (\langle p_0^{[0]4}\rangle - \langle p_0^{[0]2}\rangle^2)_{ta} = e^{-2\beta t}d(0), \quad (60a)$$

$$d_{th}(t) = (\langle p_A^{[0]4}\rangle - \langle p_A^{[0]2}\rangle^2)_{ta} = 2\langle p_{th}^2\rangle^2(1 - e^{-\beta t})^2, \quad (60b)$$

$\langle p_{th}^2\rangle$ is for the thermal equilibrium field with a Boltzmann distribution of photons. d_{th} has been evaluated by the reasoning presented in Ref. 11, note I, for a particle in one-dimensional Brownian motion. With Eq. (60) we can write the first three terms of Eq. (59) as

$$[d_0(t) + 4\langle p_0^{[0]2}\rangle\langle p_A^{[0]2}\rangle + d_{th}(t)]_{ta} = e^{-2\beta t}d(0) + 4\langle p(0)^2\rangle_{ta}\langle p_{th}^2\rangle(e^{-\beta t} - e^{-2\beta t}) + 2\langle p_{th}^2\rangle^2(1 - e^{-\beta t})^2, \quad (61)$$

in which $d(0)$ and $\langle p(0)^2\rangle_{ta}$ stand for time-average values of these expectations at $t = 0^+$. If there are no molecules and the field is in thermal equilibrium, $d(0) = 2\langle p_{th}^2\rangle^2$, $\langle p(0)^2\rangle_{ta} = \langle p_{th}^2\rangle$, and Eq. (61) is just $2\langle p_{th}^2\rangle^2$. The next term in Eq. (59) evaluates as

$$\{\langle[p_0^{[0]2}, [p_0^{[0]}, p_0^{[2]}]_+]_+\rangle\}_{ta} = (8\pi\omega u^2\bar{\gamma}^2/\hbar)N\langle I\rangle e^{-2\beta t}\{d(0) + \langle p(0)^2\rangle^2\}_{ta} \times \int_0^t dt_1 \int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1-t_2)}e^{-(t_1-t_2)^2/4\alpha^2} \times \cos(\omega - \omega_M)(t_1 - t_2). \quad (62)$$

This expression has been evaluated by expanding the double anticommutator with regard for the proper sequence of operators, employing Eq. (28a) and the complete integral expression for $p_0^{[2]}$ in

place of Eq. (33), and then retaining only the slowly varying trigonometric terms. The next term in Eq. (59) evaluates as

$$\{-2\langle p_0^{[0]2}\rangle\langle[p_0^{[0]}, p_0^{[2]}]_+\rangle\}_{ta} = -(8\pi\omega u^2\tilde{\gamma}^2/\hbar)N\langle I\rangle e^{-2\beta t}\{\langle p(0)^2\rangle^2\}_{ta} \times \int_0^t dt_1 \int_0^{t_1} dt_2 e^{\frac{1}{2}\beta(t_1-t_2)} e^{-(t_1-t_2)^2/4\alpha^2} \times \cos(\omega - \omega_M)(t_1 - t_2). \quad (63)$$

The $\langle p(0)^2\rangle^2$ terms in Eqs. (62) and (63) conveniently cancel. The last term in Eq. (59) evaluates as

$$4\{\langle p_0^{[0]2}\rangle\langle p^{[1]2}\rangle\}_{ta} = (4\pi cu/\omega)^2 N\tilde{\gamma}^2\langle p(0)^2\rangle_{ta} e^{-2\beta t} \times \int_0^t dt_1 \int_0^t dt_2 e^{\frac{1}{2}\beta(t_1+t_2)} e^{-(t_1-t_2)^2/4\alpha^2} \times \cos(\omega - \omega_M)(t_1 - t_2). \quad (64)$$

We now substitute Eqs. (61)–(64) into Eq. (59), in terms of $d(0)$ and $\langle p(0)^2\rangle_{ta}$ as well as $\langle I(0)\rangle$. Upon differentiating twice, using the relation

$$\dot{d}(0) = -2\beta d(0) + 4\beta\langle p(0)^2\rangle_{ta}\langle p_{th}^2\rangle, \quad (65)$$

and finally changing the argument 0 to an arbitrary time t, we obtain the following differential equation for $d(t)$:

$$\begin{aligned}\ddot{d}(t) + 2\beta\dot{d}(t) &= \Omega^2\langle I(t)\rangle d \\ &\quad + 4\beta^2\langle p_{th}^2\rangle(\langle p_{th}^2\rangle - \langle p(t)^2\rangle_{ta}) \\ &\quad + (4\pi c^2/\omega^4)\hbar\omega\Omega^2\langle p(t)^2\rangle_{ta}, \\ \Omega^2 &= 8\pi\omega u^2\tilde{\gamma}^2 N/\hbar.\end{aligned} \quad (66)$$

Ω has the dimensions of a frequency.

Equation (66) predicts an equilibrium dispersion which is the thermal one, as may be verified by setting $\langle p(t)^2\rangle_{ta} = \langle p_{th}^2\rangle = (8\pi c^2/\omega^4)(W_e)_{equil}$ and $d_{equil} = 2\langle p_{th}^2\rangle^2$, whereupon the right side of Eq. (66) is zero just as the right side of Eq. (50) is zero.

We can describe the dispersion qualitatively with respect to W_e if $\langle p(t)^2\rangle_{ta} \gg \langle p_{th}^2\rangle$ and $d(t) \approx (\langle p(t)^2\rangle_{ta})^2$, such that only the first term on the right side of Eq. (66) is important. Then Eqs. (50) and (66) read, the latter in terms of energy dispersion D,

$$\ddot{W}_e + \beta\dot{W}_e = \tfrac{1}{2}\Omega^2\langle I(t)\rangle W_e(t), \quad (67a)$$

$$\ddot{D} + 2\beta\dot{D} = \Omega^2\langle I(t)\rangle D(t). \quad (67b)$$

Equations (67) describe the time-average electric energy and dispersion of the "transient" fields of complex frequencies $\pm\omega + \frac{1}{2}i\beta$, which are independent of those quantities established by the pump because the field operators of the transient frequencies commute with those of the classical pump frequency. However $\langle I(t)\rangle$ will be determined in general by both the transient and pumping fields. Neglecting the dispersion of the pump fields, the similarity of Eq. (67a) and (67b) descriptive of conditions at frequencies $\pm\omega + \frac{1}{2}i\beta$ enables us to say that, although $D(t)$ increases during the pumping period when $\langle I(t)\rangle$ is driven positive, it increases by about the same factor as $W_e(t)$ increases so that, at the end of the pumping period, the relative dispersion $D(t)/W_e(t)^2$ is actually *less* than the thermal equilibrium value of 2. After the pump is shut off both W_e and D decay toward their thermal equilibrium values, perhaps with some envelope modulation to be discussed in Sec. 9, and the ratio D/W_e^2 grows slowly back to the value 2.

Therefore we can make the important statement that the relative dispersion, neglecting the pumping field dispersion, remains on the order of or less than 2 during the pumping and stimulated-emission intervals, and hence the *ensemble behavior of time-average electric energy $W_e(t)$ is qualitatively descriptive of an individual system.* This statement is true when the pump operates directly between the two mean-energy levels of the molecules, but it may not be true in a three-level system for reasons discussed in the next section.

7. EFFECTS DURING THE PUMPING INTERVAL

So far we have developed the equations of motion for time-average field energy and excess molecular population in the system isolated from external excitation. We have described in Sec. 5 how the equations of motion during an arbitrary short-time interval can be interpreted as differential equations over an arbitrarily long interval by imposing an intermittent similarity transformation on ρ(molecules) to restore the molecular operator matrices at the end of each brief interval to their forms at the beginning of the interval. ρ(field) must also be changed for continuity of the time-average field operators. Fortunately we need not keep track of the individual elements of the density matrices in order to follow the time-average expectations of interest.

Now we let a pumping field be applied to the system which includes only two-level molecules. In most practical masers and lasers, the pump field excites molecules into a third level through a transition frequency higher than the mean radiative transition frequency ω_M. However, our equations for a system of two-level molecules, field, and dissipation mechanism is not directly applicable to a system in which a third level is used for pumping

purposes because of assumptions made early in the analysis about the forms of $\gamma_m^{[0]}$ and $\mathfrak{K}_m^{[0]}$. These 2×2 matrices are off- and on-diagonal, respectively, and solutions of the Heisenberg equations of motion yield first-order $\gamma_m^{[1]}$ and $\mathfrak{K}_m^{[1]}$ which are strictly on- and off-diagonal, respectively. Such is not the case in a system including three-level molecules, for which $\gamma_m^{[0]}$ has matrix elements between all pairs of levels. $\gamma_m^{[1]}$ is then a full matrix, which implies the expectation values of first- and second-order field and energy operators, as well as of the molecular power, must be recalculated and provision made for the presence of a classical pumping field. The equation of motion for an element of $\gamma_m^{[1]}$ connecting two energy levels now depends upon the elements of $\gamma_m^{[0]}$ and $\gamma_m^{[1]}$ connecting other pairs of levels. In view of the complexity of the solutions for $\gamma_m^{[1]}$ and $\mathfrak{K}_m^{[1]}$, we shall discuss the system of three-level molecules no further in this paper.

Now assume that a high-level pumping field $E_p \cos \omega_p t$ is applied to the molecules of $\omega_M = \omega_p$, during the interval when the "transient" field at frequencies $\pm\omega + \frac{1}{2}i\beta$ is negligible. The ensuing remarks will apply equally well for magnetic dipole coupling in a solid-state medium. The semiclassical and "neoclassical"[12,13] analyses predict a behavior of $n_m(t) = \langle \rho_{m2}(t) \rangle - \frac{1}{2} = \langle |a_{m2}|^2 \rangle - \frac{1}{2}$ for a molecule of transition frequency ω_m given by

$$n_m(t) = n_m(0)\left\{\frac{(\omega_m - \omega_p)^2 + (\gamma_m E_p/\hbar)^2 \cos\{[(\omega_m - \omega_p)^2 + (\gamma_m E_p/\hbar)^2]^{\frac{1}{2}}t\}}{(\omega_m - \omega_p)^2 + (\gamma_m E_p/\hbar)^2}\right\}. \tag{68}$$

In order that our assumption be valid, namely, that $\langle I_m \rangle = 2n_m$ is a slowly-varying function of ω_m, we must have $(\gamma_m E_p/\hbar)^2 \gg$ (molecular linewidth)^2, and this implies a simple harmonic motion of n with the frequency of nearly $\gamma_m E_p/\hbar$. This is precisely the behavior predicted by Eq. (56), in which $W_e u^2 = E_p^2/16\pi$. On the other hand, if the assumption is not valid, or if $\omega_p \neq \omega_M$, and $(\gamma_m E_p/\hbar)^2$ is not $\gg (\omega - \omega_M)^2$, Eqs. (56), based on second-order perturbation theory, are not descriptive of the phenomena which are better described by the semiclassical or "neoclassical" treatments.

In fact the application of $E_p \cos \omega_p t$ to the molecules (not a sinusoidal driving field to the cavity mode, which is a distinctly different drive) will tend to feed energy to the molecules, which, in turn will deliver energy to the "transient" field at frequencies $\pm\omega + \frac{1}{2}i\beta$. To investigate further let $\omega_p \approx \omega$ (actually ω_1, the real part of the cavity resonant frequency) and $\omega \simeq \omega_M$ such that $(\gamma_m E_1/\hbar)^2 \gg (\omega - \omega_M)^2$, E_1 being the peak value of the field at the complex resonant frequency. Then Eq. (56) describes the buildup of this "transient" field, as driven by $n(t)$. Equation (56) describes the transient-field behavior independently of any explicit dependence on the pump field because the transient-field operators commute with those of the pump field. Before the transient field becomes large, n_m is determined classically by Eq. (68); it may be convenient to let all the molecules have the same transition frequency ω_M if their spread is narrow. The moment the pump drives $n(t)$ positive, the transient $W_e(t)$ starts to grow exponentially, about as

$$e^{\alpha t}, \qquad \alpha = (8\pi\omega u^2 \tilde{\gamma}^2 N n(t)/\hbar)^{\frac{1}{2}}, \tag{69}$$

and the resonant field may extract energy from the molecules and force $n(t)$ to go negative again. Then W_e will diminish until the pump forces $n(t)$ positive again, whereupon the whole cycle repeats with a somewhat different period and amplitude. Figure 2 shows qualitatively how W_e might build up during an interval on the order of tens of microseconds. Let us estimate the exponential rise after $n(t)$ goes positive with the following numbers for a 1-cm $\times$ 1-cm $\times$ 1-m laser rod, mks units,

$$\begin{aligned} (u^2)_{\text{mean}} \rightarrow |(c/\omega)\nabla \times u|^2_{\text{mean}} \\ \approx 2(\text{volume})^{-1} = 2 \times 10^4, \\ \tilde{\gamma}^2_{\text{mag}} = 10^{-58}, \qquad \omega = 10^{15}, \\ N(\text{total}) = 10^{19}, \qquad \hbar \simeq 10^{-34}, \end{aligned} \tag{70a}$$

from which

$$\alpha = 7 \times 10^7 [n(t)]^{\frac{1}{2}}. \tag{70b}$$

As W_e rises, in Fig. 2, the resonant field furnishes

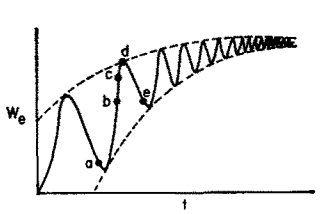

FIG. 2. Envelope fluctuations in the time-average electric energy W_e during the pumping period. At point a, the excess molecular population $n(t)$ is barely positive; at point b, n is maximum positive; point c, n is barely negative; point d, n is minimum negative; point e, n is barely positive again.

[12] E. T. Jaynes and F. W. Cummings, Proc. IEEE **51**, 89 (1963).
[13] F. W. Cummings, Am. J. Phys. **30**, 898 (1962).

an additional (classical) drive of the molecules and the phase relation between "transient" and pumping fields becomes important. The equilibrium level is determined by the condition that the time-average pump powers extracted equal the power dissipated into the cavity. If energy in the rod breaks through the end face, decay constant β in Eq. (56) is changed to a value typically on the order of 10^5. Actually the whole system composed of high-level transient and pump fields is best described by the "neoclassical" analysis.

The envelope modulation shown in Fig. 2 which is commonly observed during the stimulated emission period is discussed in Sec. 9.

8. JUSTIFICATION OF THE DIRECT-PRODUCT FORM FOR THE DENSITY MATRIX

Faced with the task of determining a matrix representation for the density operator of field, molecules, and loss mechanism in equilibrium before the pumping field is applied, we can resort to maximum entropy inference.[3] Maximum entropy inference yields that density matrix representing maximum system entropy consistent with constraints and/or known results of measurements, thereby providing the most unbiased description of the system. The density matrix, by definition, applies to an ensemble of identically prepared systems which are described in complete detail by one "global" density matrix. However, for the expectation values of interest, a "reduced" density matrix representing projections of the "global" matrix onto the field, molecular, and dissipation spaces is relevant. The process of removing correlations among various members of the ensemble consists in replacing ρ(global) by ρ(field) $\times$ ρ(molecules) $\times$ ρ(loss mechanism), whereupon the expectation of any operator acting on only one of the three subsystems is calculated as well by this reduced direct-product matrix as by the global one. First we obtain the direct-product form, then we relate its matrix elements to those of the (implicitly known) global one.

Assume for the moment only one molecule is present in the system at equilibrium. Our only measurement is of the temperature of the dissipation mechanism as a heat "sink". We know only that energy is conserved in the entire system, not how much energy is stored between the molecules and the field as compared to that stored between the field and the loss mechanism. Hence only the expectation value of total energy, $\langle \mathfrak{IC}_f + \mathfrak{IC}_m + \mathfrak{IC}_d \rangle$ should be constrained among the members of the ensemble. The entropy S is

$$S = -\langle \ln \rho \rangle = -\text{trace } (\rho \ln \rho). \tag{71}$$

We seek to maximize S subject to the constraint on total energy. It is convenient to use the Lagrange multiplier technique and maximize instead the quantity

$$S - \lambda \langle \mathfrak{IC}_f + \mathfrak{IC}_m + \mathfrak{IC}_d \rangle,$$

where λ will be determined by the measured temperature of the loss mechanism. The density matrix which maximizes S is

$$\rho_{max} = e^{\lambda_0} \exp [-\lambda(\mathfrak{IC}_f + \mathfrak{IC}_m + \mathfrak{IC}_d)], \qquad \lambda = 1/kT, \tag{72}$$

where

$$e^{-\lambda_0} = \text{trace } (\exp [-\lambda \mathfrak{IC}_f]) \times \text{trace } (\exp [-\lambda \mathfrak{IC}_m]) \text{ trace } (\exp [-\lambda \mathfrak{IC}_d]).$$

λ is determined by the temperature of the dissipation mechanism. The maximum entropy is

$$S_{max} = -\lambda_0 - \lambda(\partial/\partial\lambda) \text{ trace } (\rho) = -\lambda_0 + \lambda \langle \mathfrak{IC}_f + \mathfrak{IC}_m + \mathfrak{IC}_d \rangle. \tag{73}$$

An increment in S_{max}, δS_{max}, is caused by $\delta\lambda_0$ and $\delta\lambda$; in order that ρ remain properly normalized to trace $(\rho) = 1$, λ_0 is related to λ in such a way that $\delta S_{max} = 0$, thus proving the extremal property of S_{max}. The maximal property may be readily proved.

With many molecules present, we only replace $\mathfrak{IC}_m$ by $\sum \mathfrak{IC}_m$ in the absence of any detailed information about the relative energies stored by the various molecules.

Therefore the equilibrium density matrix is of the direct product form

$$\rho_{max} = \rho_f \times \rho_{mol} \times \rho_{diss}, \tag{74}$$

where the various factors are identified in Eq. (72). During the pumping period an additional term ρ(pump) which commutes with all other operators might in some instances be added to ρ(molecules) without destroying the direct-product form. The intermittent similarity transformation (Appendix A) changes ρ(field) to a new matrix ρ_f', which need not be explicitly known, restores ρ(molecules) to a function of molecular parameters only, and does not affect ρ(loss mechanism).

We now verify the statement that an operator, say $\mathcal{O}_f$, acting on only the field variables has an expectation value which may be calculated with the direct-product matrix provided the matrix elements

of ρ_f are properly related to the (implicitly) known ones of $\rho(\text{global}) \equiv \rho_g$.

$$\langle \mathcal{O}_f \rangle = \sum_{\substack{fmd \\ f'm'd'}} \langle fmd \,|\rho_g|\, f'm'd' \rangle \langle f'm'd' \,|\mathcal{O}_f|\, fmd \rangle, \quad (75)$$

in which $|fmd\rangle$ denotes a product of a trio of eigenfunctions forming a complete set for field, molecules, and dissipation mechanism. Since the eigenfunctions are orthogonal among themselves this is

$$\langle \mathcal{O}_f \rangle = \sum_{f,f'} \sum_{md} \langle fmd \,|\rho_g|\, f'md \rangle \langle f' \,|\mathcal{O}_f|\, f \rangle. \quad (76)$$

Using the direct-product form this is

$$\langle \mathcal{O}_f \rangle = \sum \langle fmd \,|\rho_f \times \rho_{\text{mol}} \times \rho_d|\, f'md \rangle \times \langle f' \,|\mathcal{O}_f|\, f \rangle. \quad (77)$$

We can break this up with the aid of relations like

$$1 = \sum_{f''m''d''} |f''m''d''\rangle\langle f''m''d''|$$

into

$$\langle \mathcal{O}_f \rangle = \sum_{ff'md} \langle f \,|\rho_f|\, f' \rangle \langle m \,|\rho_{\text{mol}}|\, m \rangle \times \langle d \,|\rho_d|\, d \rangle \langle f' \,|\mathcal{O}_f|\, f \rangle. \quad (78)$$

Since trace (ρ_{mol}) = trace (ρ_d) = 1, we see that Eq. (76) agrees with Eq. (78) if the elements of ρ_f are related to those of ρ_g as

$$\langle f \,|\rho_f|\, f' \rangle = \sum_{md} \langle fmd \,|\rho_g|\, f'md \rangle. \quad (79a)$$

Similarly, we may verify that the expectation value of any operator of the molecular or dissipation space may also be calculated with the direct-product density matrix provided

$$\langle m \,|\rho_{\text{mol}}|\, m' \rangle = \sum_{fd} \langle fmd \,|\rho_g|\, fm'd \rangle, \quad (79b)$$

$$\langle d \,|\rho_d|\, d' \rangle = \sum_{fm} \langle fmd \,|\rho_g|\, fmd' \rangle. \quad (79c)$$

9. ENVELOPE MODULATION DURING THE EMISSION PERIOD[14,15]

We shall now assume that the molecular distribution in a solid-state maser or laser has been inverted by the pump, $n > 0$, and the pump has been turned off, whereupon the system of molecules, high-energy field, and dissipation mechanism is free to interact according to Eq. (56), written for *magnetic*-dipole coupling as

$$\ddot{W}_m + \beta \dot{W}_m - B\,|(c/\omega)\nabla \times u|^2\, n(t) W_m(t) = 0, \quad (80a)$$

$$\ddot{n}(r,t) + (2B/\hbar\omega_M N)\,|(c/\omega)\nabla \times u|^2\, n W_m = 0, \quad B = 8\pi\omega^2\bar{\gamma}^2 N/\hbar\omega. \quad (80b)$$

$\nabla \times u$ is the magnetic field pattern, Eq. (3), and W_m is the time-average magnetic field energy in the entire resonant mode. β is the decay constant during the stimulated emission period. The molecular centerfrequency ω_M is exactly, or very nearly, ω according to previous remarks and the T_1 and T_2 time constants remain essentially infinity. Under these conditions one may readily verify—when β is very small—by expanding $W = 2W_m = W^{(0)}e^{-\beta t} + W^{(1)}(t) + \cdots$, and $n = n^{(0)}(t) + n^{(1)}(t) + \cdots$ that $W^{(1)}$ contains a first harmonic and $W^{(2)}$ contains four harmonics of an envelope modulation frequency Ω_{en},

$$\Omega_{en} = (BW^{(0)}/\hbar\omega N)^{\frac{1}{2}}\,|(c/\omega)\nabla \times u|\, e^{-\frac{1}{2}\beta t}. \quad (81)$$

The effect of lattice dissipation may be accounted for phenomenologically by appending the term $\dot{n}/T_1$ to the left side of Eq. (80b), whereupon Eq. (80) would describe qualitatively envelope modulation which decays with a time constant different from that of $1/\beta$. We shall now apply Eq. (80) to the situation in which the active material is an integral number of wavelengths long at the frequency ω, a laser rod for example.

Let us interprete Eq. (80) with the notation $W = 2W_m$, total time-average field energy, and

$$v(r) = |(c/\omega)\nabla \times u|.$$

Since various molecules experience different field strengths measured by $v(r)$, we must rewrite Eq. (80) as

$$\ddot{W} + \beta \dot{W} - B\left[\int_V v(r)^2 n(r,t)\, dV\right] W(t) = 0. \quad (82a)$$

N in coefficient B is the number of active molecules per unit volume. Equation (80b) reads

$$\ddot{n}(r,t) + (B/\hbar\omega N) v(r)^2 n(r,t) W(t) = 0. \quad (82b)$$

Suppose the normalized resonant cavity mode is a TM mode with the spacial dependence specified by

$$v(r) = (2/L)^{\frac{1}{2}} v_t(x,y) \cos kz, \quad \int_A v_t^2\, dA = 1. \quad (83)$$

The rod is L units long in the z direction and of cross sectional area A. From this point on we shall assume for simplicity that v_t is uniformly $A^{-\frac{1}{2}}$ over the cross section, whereupon n varies spatially only with z.

[14] R. J. Collins, D. F. Nelson, A. L. Schawlow, W. Bond, G. C. B. Garrett, and W. Kaiser, Phys. Rev. Letters 5, 303 (1960).

[15] P. P. Sorokin and M. J. Stevenson, Phys. Rev. Letters 5, 557 (1960).

Expand

$$n(z, t) = n^{(0)}(z, t) + n^{(1)}(z, t), \qquad W(t) = W^{(0)} + W^{(1)}(t), \tag{84}$$

and neglect the $\beta\dot{W}$ term. The solution of the zero-order portion of Eq. (82b),

$$\ddot{n}^{(0)}(z, t) + (2B/\hbar\omega NAL) \times \cos^2 kz W^{(0)} n^{(0)}(z, t) = 0, \tag{85}$$

is

$$n^{(0)}(z, t) = n^{(0)}(z) \cos(\Omega_0 t \cos kz), \qquad \Omega_0 = \{[BW^{(0)}/\hbar\omega N](2/AL)\}^{\frac{1}{2}}. \tag{86}$$

Ω_0 will turn out to be the reference frequency for the envelope modulation and can be expressed in terms of the field-molecular coupling by Eqs. (80b) and (24),

$$\Omega_0 = \frac{\omega}{c}\left(\frac{2}{AL}\right)^{\frac{1}{2}} \frac{\tilde{\gamma}_{\text{mag}} Q^{(0)}}{\hbar}, \tag{87}$$

the energy of coupling being $(\omega Q^{(0)} \gamma_{\text{mag}} v/c)$ according to Eq. (1) after the transformations of Eq. (24).

We now substitute $n^{(0)}$ into Eq. (82a), with $\beta = 0$, and solve for $W^{(1)}$:

$$\dot{W}^{(1)} - B\frac{2}{AL}\left[\int_V \cos^2 kz n^{(0)}(z) \times \cos(\Omega_0 t \cos kz)\, dV\right] W^{(0)} = 0, \tag{88}$$

If we neglect variations in initial excess population $n^{(0)}(z)$ along the rod and introduce the well known identity

$$\cos(\Omega_0 t \cos kz) = J_0(\Omega_0 t) - 2J_2(\Omega_0 t) \cos 2kz + \cdots, \tag{89}$$

in which J_n is a Bessel function, we notice only the J_0 and J_2 terms contribute to the integral. We obtain easily

$$\dot{W}^{(1)} - Bn^{(0)}(z) W^{(0)}[J_0(\Omega_0 t) - J_2(\Omega_0 t)] = 0. \tag{90}$$

The bracketed quantity can be integrated twice according to well-known relations, with the result

$$W^{(1)}(t) = -[2Bn^{(0)}(z) W^{(0)}/\Omega_0^2][J_0(\Omega_0 t) - 1]. \tag{91}$$

Equation (91) describes the envelope modulation as a nearly sinusoidal fluctuation decaying in magnitude about as $(\Omega_0 t)^{-\frac{1}{2}}$ for $\Omega_0 t \gtrsim 5$.

If now $W^{(1)}$ is substituted into the first-order portion of Eq. (82b),

$$\ddot{n}^{(1)}(z, t) + (2B/\hbar\omega NAL) \times \cos^2 kz(W^{(0)} n^{(1)} + n^{(0)} W^{(1)}) = 0, \tag{92}$$

The solution may be verified to be

$$n^{(1)}(z, t) = -\frac{2B}{\hbar\omega NAL\Omega_0} \cos kz \int_0^t W^{(1)}(\tau) n^{(0)}(z, \tau) \times \sin \Omega_0'(t - \tau)\, d\tau, \qquad \Omega_0' = \Omega_0 \cos kz. \tag{93}$$

Evaluation of $n^{(1)}$ and $W^{(2)}(t)$ in terms of

$$[n^{(0)}(z, t) W^{(1)}(t) + n^{(1)}(z, t) W^{(0)}]$$

yields still more fluctuations in the field energy.

APPENDIX A. THE INTERMITTENT SIMILARITY TRANSFORMATION

Assume that the electric-dipole matrix (or magnetic-dipole matrix if rf magnetic field coupling is involved, with no dc magnetic field present) is off-diagonal at a certain time $t = 0$.

$$\gamma_m(0) = \tilde{\gamma}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \tilde{\gamma} \text{ real.} \tag{A1}$$

During a short time (many periods of the maser or laser frequency but long compared to the time during which a time average is taken), γ_m develops according to Eq. (10a) into

$$\gamma_m(t) = \tilde{\gamma}\begin{bmatrix} -\gamma_1 + \mathcal{O}(t^2) & e^{-i\omega_m t} + \mathcal{O}(t^2) \\ e^{i\omega_m t} + \mathcal{O}(t^2) & \gamma_1 + \mathcal{O}(t^2) \end{bmatrix}, \tag{A2}$$

where

$$\gamma_1 = (8\pi c/\hbar) u\tilde{\gamma} \int_0^t dt_1 P^{[0]}(t_1) \sin \omega_m(t - t_1). \tag{A3}$$

$\mathcal{O}(t^2)$ represents terms proportional to t^2.

Consider the matrix T,

$$T = \begin{bmatrix} 1 & -\eta \\ \eta^* & 1 \end{bmatrix}, \tag{A4}$$

$$\eta = -\frac{4\pi c}{i\hbar} u\tilde{\gamma} \int_0^t dt_1 e^{-i\omega_m t_1} P^{[0]}(t_1)\, dt_1. \tag{A5}$$

T is unitary with error of $\mathcal{O}(t^2)$. If we transform $\gamma_m(t)$ we obtain

$$T\gamma_m(t)T^\dagger = \begin{bmatrix} 0 & \tilde{\gamma}e^{-i\omega_m t} + \mathcal{O}(t^2) \\ \tilde{\gamma}e^{i\omega_m t} + \mathcal{O}(t^2) & 0 \end{bmatrix}. \tag{A6}$$

This is of the form (A1); the $e^{\pm i\omega_m t}$ terms may be transformed away by using $T' = \text{diag}\,(e^{i\omega_m \frac{1}{2}t}, e^{-i\omega_m \frac{1}{2}t})$.

T also rediagonalizes $\mathfrak{K}_m$ which starts out

$$\mathfrak{K}_m(0) = \begin{bmatrix} E_{m1}' & 0 \\ 0 & E_{m2}' \end{bmatrix}, \tag{A7}$$

and ends up[7]

$$\mathfrak{K}_m(t) = \begin{bmatrix} E'_{m1} + \mathcal{O}(t^2) & H_1 + \mathcal{O}(t^2) \\ H_1^* + \mathcal{O}(t^2) & E'_{m2} + \mathcal{O}(t^2) \end{bmatrix}, \quad \text{(A8)}$$

in which

$$H_1 = \frac{4\pi c}{i\hbar} u\tilde{\gamma}(E'_{m1} - E'_{m2}) \int_0^t dt_1 e^{-i\omega_m t_1} P^{[0]}(t_1). \quad \text{(A9)}$$

We find, within $\mathcal{O}(t)$,

$$T\mathfrak{K}_m(t)T^\dagger = \begin{bmatrix} E'_{m1} & (E'_{m1} - E'_{m2})\eta + H_1 \\ (E'_{m1} - E'_{m2})\eta^* + H_1^* & E'_{m2} \end{bmatrix}. \quad \text{(A10)}$$

The off-diagonal elements are zero by (A5) and (A9), so the matrix of (A7) is restored.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Functional Differential Calculus of Operators

F. ROHRLICH*
Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa
(Received 13 May 1963)

The functional derivative with respect to operators of operator functionals is defined for operators which satisfy certain commutation relations of interest in quantum field theory. From this definition, a functional differential calculus is developed for functionals of tensor as well as spinor fields. It is noted that an implicit definition of the functional derivative can always be given while an explicit one seems to exist only for those operator fields which need not be restricted by supplementary operator conditions, and for which not more than one derivative occurs in the commutation relations.

INTRODUCTION

ATTEMPTS to define functional derivatives with respect to operators have in the past been beset with difficulties of various sorts. At the same time, these attempts were characterized by the great generality of the operators involved.

In the present paper a satisfactory functional differential calculus is developed by restricting the operators with respect to which the differentiations are carried out. This restriction is effectively limited to those operators which are of interest in quantum field theory. Functional differentiation was first used extensively in the theory by Schwinger,[1] who differentiated with respect to unquantized fields. The present formalism permits differentiation also with respect to quantized fields.

The finite formulation of quantum field theory[2] involves functional derivatives only with respect to the "in" and "out" field operators. The latter are characterized by well-known commutation relations. It is therefore natural to base the functional differential calculus on an essentially *algebraic* foundation of commutators.

In Sec. I, functionals of a real scalar field are studied. The generalizations to complex tensor fields are indicated in Sec. II. A very important extension of this differential calculus to functionals of Dirac spinor fields is carried through in Sec. III. The results are compared with the analytic definition of the functional derivative in Sec. IV, where it becomes evident why the algebraic definition is not beset by the same difficulties. Section V sums up the salient points of this study.

I. REAL SCALAR OPERATOR FIELDS

Consider an operator field $A(x)$ which is real,

$$A^*(x) = A(x), \quad \text{(I.1)}$$

and which satisfies the commutation relations

$$[A(x), A(y)]_- = -i\Delta(x - y). \quad \text{(I.2)}$$

While it would be sufficient to assume the independent variable x to refer to a real one-dimensional space, it is just as easy and much more convenient for the purpose of future applications to relativistic

* Present address: Department of Physics, Syracuse University, Syracuse, New York.

[1] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **37**, 452, 455 (1951).

[2] R. E. Pugh, Doctoral dissertation, University of Iowa (1963); also, Ann. Phys. **23**, 335 (1963).

and ends up[7]

$$\mathfrak{K}_m(t) = \begin{bmatrix} E'_{m1} + \mathcal{O}(t^2) & H_1 + \mathcal{O}(t^2) \\ H_1^* + \mathcal{O}(t^2) & E'_{m2} + \mathcal{O}(t^2) \end{bmatrix}, \quad \text{(A8)}$$

in which

$$H_1 = \frac{4\pi c}{i\hbar} u\tilde{\gamma}(E'_{m1} - E'_{m2}) \int_0^t dt_1 e^{-i\omega_m t_1} P^{[0]}(t_1). \quad \text{(A9)}$$

We find, within $\mathcal{O}(t)$,

$$T\mathfrak{K}_m(t)T^\dagger = \begin{bmatrix} E'_{m1} & (E'_{m1} - E'_{m2})\eta + H_1 \\ (E'_{m1} - E'_{m2})\eta^* + H_1^* & E'_{m2} \end{bmatrix}. \quad \text{(A10)}$$

The off-diagonal elements are zero by (A5) and (A9), so the matrix of (A7) is restored.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Functional Differential Calculus of Operators

F. ROHRLICH*
Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa
(Received 13 May 1963)

The functional derivative with respect to operators of operator functionals is defined for operators which satisfy certain commutation relations of interest in quantum field theory. From this definition, a functional differential calculus is developed for functionals of tensor as well as spinor fields. It is noted that an implicit definition of the functional derivative can always be given while an explicit one seems to exist only for those operator fields which need not be restricted by supplementary operator conditions, and for which not more than one derivative occurs in the commutation relations.

INTRODUCTION

ATTEMPTS to define functional derivatives with respect to operators have in the past been beset with difficulties of various sorts. At the same time, these attempts were characterized by the great generality of the operators involved.

In the present paper a satisfactory functional differential calculus is developed by restricting the operators with respect to which the differentiations are carried out. This restriction is effectively limited to those operators which are of interest in quantum field theory. Functional differentiation was first used extensively in the theory by Schwinger,[1] who differentiated with respect to unquantized fields. The present formalism permits differentiation also with respect to quantized fields.

The finite formulation of quantum field theory[2] involves functional derivatives only with respect to the "in" and "out" field operators. The latter are characterized by well-known commutation relations. It is therefore natural to base the functional differential calculus on an essentially *algebraic* foundation of commutators.

In Sec. I, functionals of a real scalar field are studied. The generalizations to complex tensor fields are indicated in Sec. II. A very important extension of this differential calculus to functionals of Dirac spinor fields is carried through in Sec. III. The results are compared with the analytic definition of the functional derivative in Sec. IV, where it becomes evident why the algebraic definition is not beset by the same difficulties. Section V sums up the salient points of this study.

I. REAL SCALAR OPERATOR FIELDS

Consider an operator field $A(x)$ which is real,

$$A^*(x) = A(x), \quad \text{(I.1)}$$

and which satisfies the commutation relations

$$[A(x), A(y)]_- = -i\Delta(x - y). \quad \text{(I.2)}$$

While it would be sufficient to assume the independent variable x to refer to a real one-dimensional space, it is just as easy and much more convenient for the purpose of future applications to relativistic

* Present address: Department of Physics, Syracuse University, Syracuse, New York.

[1] J. Schwinger, Proc. Natl. Acad. Sci. U. S. **37**, 452, 455 (1951).

[2] R. E. Pugh, Doctoral dissertation, University of Iowa (1963); also, Ann. Phys. **23**, 335 (1963).

quantum field theory to assume that x is a four-vector in four-dimensional Minkowski space. In that case, the Δ function in (I.2) is defined in the usual well-known manner.[3]

Let $F(A)$ be a functional[4] of A. Then we define the functional derivative of F with respect to $A(x)$ by the implicit relation[5]

$$[A(x), F(A)]_- = -i \int \Delta(x - y) \frac{\delta F(A)}{\delta A(y)} dy. \tag{I.3}$$

In particular, we require

$$\delta A(y)/\delta A(x) = \delta(y - x). \tag{I.4}$$

For example, if

$$F(A) = \int f(x_1, x_2, \cdots x_n) A(x_1) A(x_2) \cdots \times A(x_n)\, dx_1\, dx_2 \cdots dx_n, \tag{I.5}$$

with $f(x_1, x_2 \cdots x_n)$ a given symmetric function of its variables,

$$[A(x), F(A)]_- = \int f(x_1, \cdots x_n) \times [A(x), A(x_1) \cdots (Ax_n)]_- \, dx_1 \cdots dx_n.$$

Successive use of

$$[A(x), B(x_1)B(x_2)]_- = [A(x), B(x_1)]_- B(x_2) + B(x_1)[A(x), B(x_1)]_- \tag{I.6}$$

and (I.2) then yields an expression which has the form of the right-hand side of (I.3).

The dimensions of $\delta F/\delta A$ follow to be those of F/A divided by the dimensions of the volume involved in the integration (I.3). However, the implicit definition is not unique because one can always add terms of the type $t(y)$ which contribute a vanishing integral

$$\int \Delta(x - y) t(y)\, dy = 0.$$

It is therefore preferable to solve (I.3) for the derivative.

To this end a new commutator can be defined which has the following properties (I.7) and (I.8). First, the usual commutation relations hold,

$$\begin{aligned} [A, \lambda]^{\mathrm{I}}_- &= 0, \\ [A, B_1 + B_2]^{\mathrm{I}}_- &= [A, B_1]^{\mathrm{I}}_- + [A, B_2]^{\mathrm{I}}_-, \\ [A_1 + A_2, B]^{\mathrm{I}}_- &= [A_1, B]^{\mathrm{I}}_- + [A_2, B]^{\mathrm{I}}_-, \\ [A, B_1B_2]^{\mathrm{I}}_- &= [A, B_1]^{\mathrm{I}}_- B_2 + B_1[A, B_2]^{\mathrm{I}}_-, \\ [A_1A_2, B]^{\mathrm{I}}_- &= A_1[A_2, B]^{\mathrm{I}}_- + [A, B]^{\mathrm{I}}_- A_2, \end{aligned} \tag{I.7}$$

where A, A_1, B, B_1, B_2 are products of the operator field (I.1), while λ is not an operator. Secondly, we require that, in addition to (I.7), this commutator satisfies, with the help of (I.2), the inhomogeneous equation

$$K(x)[A(x), A(y)]^{\mathrm{I}}_- = -\delta(x - y). \tag{I.8}$$

With the notation of Ref. 3 for $\Delta(x)$, such commutators are given by

$$[A(x), A(y)]^{\mathrm{I}}_- = -i\Delta^{\mathrm{I}}(x - y),$$

and Δ^{I} is $\Delta_R(x)$, $\Delta_A(x)$, or any suitable linear combination of these, such as $\Delta_P(x)$,

$$\Delta_R(x) = \theta(x)\Delta(x), \quad \Delta_A(x) = -\theta(-x)\Delta(x) \tag{I.9}$$
$$\Delta_P(x) = \tfrac{1}{2}\epsilon(x)\Delta(x).$$

Thus, $[A(x), A(y)]^{\mathrm{I}}_-$ becomes a linear functional of $[A(x), A(y)]_-$. All inhomogeneous Δ functions Δ^{I} satisfy

$$K\Delta^{\mathrm{I}}(x) = (\Box - m^2)\Delta^{\mathrm{I}}(x) = -\delta(x), \tag{I.10}$$

so that $K(x)$ in (I.8) is to be identified with the Klein–Gordon operator in this case.

The implicit equation (I.3) now becomes

$$[A(x), F(A)]^{\mathrm{I}}_- = -i \int \Delta^{\mathrm{I}}(x - y) \frac{\delta F(A)}{\delta A(y)} dy \tag{I.11}$$

and can be solved in virtue of (I.8).

$$i\delta F(A)/\delta A(x) = K(x)[A(x), F(A)]^{\mathrm{I}}_-. \tag{I.12}$$

This is the explicit definition of the functional derivative with respect to an operator. While it requires a very special class of operators, this class is just the one which contains the operator field of importance in quantum field theory.

The differentiation operation thus defined results in a number of relations which follow immediately from (I.7):

$$\delta(\lambda F)/\delta A(x) = \lambda \delta F/\delta A(x) \quad (\lambda \text{ not an operator}), \tag{I.13a}$$

$$\frac{\delta(F + G)}{\delta A(x)} = \frac{\delta F}{\delta A(x)} + \frac{\delta G}{\delta A(x)}, \tag{I.13b}$$

$$\frac{\delta(FG)}{\delta A(x)} = \frac{\delta F}{\delta A(x)} G = F \frac{\delta G}{\delta A(x)},$$

---

[3] J. M. Jauch and F. Rohrlich, *Theory of Photons and Electrons* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1959).

[4] We use the notation $F(A)$ rather than $F[A]$, since square brackets will be used below with different meaning.

[5] When x refers to Minkowski space,

$$dx = d^4x = dx^0\, dx^1\, dx^2\, dx^3.$$

or

$$\left[\frac{\delta}{\delta A(x)}, F\right] = \frac{\delta F}{\delta A(x)}. \tag{I.13c}$$

As an example, the functional (I.5) gives

$$\frac{\delta F}{\delta A(x)} = n \int f(x, x_1, \cdots x_{n-1}) \times A(x_1) \cdots A(x_{n-1})\, dx_1 \cdots dx_{n-1}.$$

Higher derivatives are obtained by successive application of the defining equation. Thus, (I.3) implies

$$[A(y), [A(x), F(A)]_-]_- = -\int \Delta(y - y')\, dy' \frac{\delta}{\delta A(y')} \times \left[\int \Delta(x - x')\, dx' \frac{\delta F}{\delta A(x')}\right] = -\int \Delta(y - y')\, dy' \times \int \Delta(x - x')\, dx' \frac{\delta^2 F}{\delta A(y')\, \delta A(x')}.$$

From this equation follows immediately

$$\delta^2 F/\delta A(y)\, \delta A(x) = \delta^2 F/\delta A(x)\, \delta A(y) \tag{I.14}$$

for the integrand by means of the Jacobi identity,

$$[A(y), [A(x), F]_-]_- + [A(x), [F, A(y)]_-]_- + [F, [A(y), A(x)]_-]_- = 0.$$

Higher functional derivatives can thus be derived in an obvious way.

A theorem of special interest is the equation

$$e^{\lambda\delta/\delta A(x)} F(A) e^{-\lambda\delta/\delta A(x)} = F(A(y) + \lambda\delta(x - y)), \tag{I.15}$$

where the right side means that each $A(y)$ occurring in F is to be replaced by the indicated expression. The proof follows from the identity

$$e^{\lambda A} B e^{-\lambda A} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} [A, B]_n,$$

where

$$[A, B]_n = [A, [A, B]_{n-1}]_-, \qquad [A, B]_0 = B,$$

which is valid for any two operators A and B. One first obtains the series

$$\sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \left[\frac{\delta}{\delta A(x)}, F(A)\right]_n,$$

which is recognized as the "Taylor series" of the right side of (I.15).

Equations (I.14) and (I.4) can be written, using (I.13c),

$$[\delta/\delta A(x), \delta/\delta A(y)]_- = 0, \tag{I.16}$$

$$[\delta/\delta A(x), A(y)]_- = \delta(x - y). \tag{I.17}$$

II. GENERALIZATION TO OTHER TENSOR OPERATOR FIELDS

The *complex* scalar operator field satisfies

$$[A(x), A(y)]_- = 0, \qquad [A(x), A^*(y)]_- = [A^*(x), A(y)]_- = -i\Delta(x - y). \tag{II.1}$$

Correspondingly, the implicit definition (I.3) becomes

$$[A(x), F(A, A^*)]_- = -i \int \Delta(x - y) \frac{\delta F(A, A^*)}{\delta A^*(y)}\, dy, \tag{II.2}$$

and

$$[A^*(x), F(A, A^*)]_- = -i \int \Delta(x - y) \frac{\delta F(A, A^*)}{\delta A(y)}\, dy. \tag{II.3}$$

Its solution requires the inhomogeneous commutators which are defined in (I.7). Examples of $\Delta^{\mathrm{I}}(x)$ were given in (I.9). Thus, one has the same Eqs. (II.2) and (II.3) also for the inhomogeneous commutator when Δ is replaced by Δ^{I}. The inhomogeneous Eq. (I.10) then allows the solution

$$i\delta F(A, A^*)/\delta A^*(x) = K(x)[A(x), F(A, A^*)]^{\mathrm{I}}_-, \tag{II.4}$$

$$i\delta F(A, A^*)/\delta A(x) = K(x)[A^*(x), F(A, A^*)]^{\mathrm{I}}_-. \tag{II.5}$$

This provides the explicit form of the functional derivatives with respect to the complex scalar operator field. The same differentiation law (I.13) holds here as in the real case. Equations (I.16) and (I.17) are, however, generalized to yield

$$\left[\frac{\delta}{\delta A(x)}, \frac{\delta}{\delta A(y)}\right]_- = 0, \qquad \left[\frac{\delta}{\delta A^*(x)}, \frac{\delta}{\delta A(y)}\right]_- = 0, \tag{II.6}$$

$$\left[\frac{\delta}{\delta A(x)}, A(y)\right]_- = \delta(x - y), \qquad \left[\frac{\delta}{\delta A^*(x)}, A(y)\right]_- = 0. \tag{II.7}$$

This simple generalization from real to complex scalar fields can be followed by a further generalization of the results of the previous Section to *vector fields*.

First, consider the real massless field (electromagnetic field). It satisfies

$$A_\mu^* = A_\mu, \tag{II.8}$$

and

$$[A_\mu(x), A_\nu(y)]_- = -ig_{\mu\nu}D(x-y). \tag{II.9}$$

The implicit definition is therefore

$$[A_\mu(x), F]_- = -i\int g_{\mu\lambda}D(x-y)\frac{\delta F}{\delta A_\lambda(y)}\,dy$$
$$= -i\int D(x-y)\frac{\delta F}{\delta A^\mu(y)}\,dy. \tag{II.10}$$

Inhomogeneous commutators can be defined as in the real scalar case,

$$\Box[A_\mu(x), A_\nu(y)]_-^{\mathrm{I}} = -ig_{\mu\nu}\,\Box D^{\mathrm{I}}(x-y)$$
$$= +ig_{\mu\nu}\delta(x-y), \tag{II.11}$$

yielding

$$ig_{\mu\lambda}\frac{\delta F}{\delta A_\lambda(x)} = i\frac{\delta F}{\delta A^\mu(x)} = \Box[A_\mu(x), F]_-^{\mathrm{I}}. \tag{II.12}$$

The commutators of the functional derivative operator with itself and with A_μ are

$$\left[\frac{\delta}{\delta A^\mu(x)}, \frac{\delta}{\delta A^\nu(y)}\right]_- = 0,$$
$$\left[\frac{\delta}{\delta A^\mu(x)}, A_\nu(y)\right] = g_{\mu\nu}\delta(x-y), \tag{II.13}$$

and the translation theorem can be generalized analogously.

We now turn to the vector fields with mass $m \neq 0$. Here an essentially new situation arises. As we shall see presently, there is no simple way of solving the implicit equation into an explicit expression for the derivative.

Consider the complex vector field $\phi^\mu(x)$. It is assumed to satisfy the commutation relations

$$[\phi_\mu(x), \phi_\nu(y)]_- = 0,$$
$$[\phi_\mu^*(x), \phi_\nu(y)]_- = -i\Delta_{\mu\nu}(x-y), \tag{II.14}$$

where

$$\Delta_{\mu\nu}(x) = (g_{\mu\nu} - \partial_\mu\partial_\nu/m^2)\Delta(x). \tag{II.15}$$

Thus, we have

$$[\phi_\mu^*(x), F(\phi, \phi^*)]_-$$
$$= -i\int \Delta_{\mu\lambda}(x-y)\frac{\delta F}{\delta\phi_\lambda(y)}\,dy \tag{II.16}$$

as the implicit definition. However, *there does not seem to exist* an inhomogeneous commutator

$$[\phi_\mu^*(x), \phi_\nu(y)]_-^{\mathrm{I}} = -i\Delta_{\mu\nu}^{\mathrm{I}}(x-y)$$

satisfying an equation analogous to (I.10). This is related to the fact that the four components ϕ_μ are not independent and that, in fact, in vector meson theory, the canonical conjugate to ϕ_0 vanishes.

If one restricts μ in (II.16) to $k = 1, 2, 3$, inhomogeneous commutators $\Delta_{k\nu}^{\mathrm{I}}$ do exist, since e.g.,

$$[\phi_k^*(x), \phi_\nu(y)]_-^R$$
$$= -i\theta(x-y)(g_{k\nu} - \partial_k\partial_\nu/m^2)\Delta(x-y)$$
$$= -i(g_{k\nu} - \partial_k\partial_\nu/m^2)\Delta_R(x-y).$$

Correspondingly, (II.16) yields

$$K[\phi_k^*(x), F]_-^{\mathrm{I}}$$
$$= -i(g_{k\nu} - \partial_k\partial_\nu/m^2)\delta F/\delta\phi_\nu(x). \tag{II.17}$$

But this equation is neither covariant nor does it provide an explicit expression for the functional derivative.

The difficulty is apparently related to the number of degrees of freedom which is smaller than the number of components of the tensor operator, as is manifest by the need for a condition

$$\partial_\mu\phi^\mu = 0, \tag{II.18}$$

in addition to the field equations satisfied by ϕ_μ. Thus all higher-rank tensor operator fields will involve this difficulty.

For the real vector field with $m \neq 0$, the difficulty can be avoided in the same way as one makes this theory renormalizable, i.e., by removing the $\partial_\mu\partial_\nu$ term in (II.15) from the commutation relations (Stückelberg's method).

Adopting the implicit Eq. (II.16), the various differentiation rules derivable from the properties of the commutator will also be valid only in the corresponding implicit form.

III. DIRAC SPINOR OPERATOR FIELDS

The operator fields $\psi(x)$ and $\bar\psi(x)$ are characterized by[3]

$$[\psi(x), \psi(y)]_+ = 0, \qquad [\bar\psi(x), \bar\psi(y)]_+ = 0, \tag{III.1}$$

and

$$[\psi(x), \bar\psi(y)]_+ = iS(x-y), \tag{III.2}$$

with

$$S(x) = (\gamma\cdot\partial - m)\Delta(x). \tag{III.3}$$

The fact that we are now dealing with anticommutators rather than commutators has impor-

tant bearing on the signs of the following equations. In particular, it is essential to know whether $F(\psi, \bar{\psi})$ transforms like the product of an even or an odd number of spinors. This can be seen as follows. The commutator of an operator with a product can again be expressed in terms of commutators [see (I.7)], but the anticommutator of an operator with a product involves one commutator and one anticommutator,

$$[A, B_1B_2]_+ = B_1[A, B_2]_+ + [A, B_1]_-B_2 = [A, B_1]_+B_2 - B_1[A, B_2]_-.$$

A consistent procedure thus requires a distinction between products of even and odd numbers of spinors. Note that the commutator of an operator with a product *can* be written in terms of anticommutators,

$$[A, B_1B_2]_- = [A, B_1]_+B_2 - B_1[A, B_2]_+.$$

Thus,

$$[\psi_\alpha(x), \bar{\psi}_\beta(x')\bar{\psi}_\gamma(x'')]_- = iS_{\alpha\beta}(x - x')\bar{\psi}_\gamma(x'') - \bar{\psi}_\beta(x')iS_{\alpha\gamma}(x - x'').$$

Therefore, we can put

$$[\psi_\alpha(x), \bar{\psi}_\beta(x')\bar{\psi}_\gamma(x'')]_- = i\int S_{\alpha\lambda}(x - y)\frac{\delta(\bar{\psi}_\beta(x')\bar{\psi}_\gamma(x''))}{\delta\bar{\psi}_\lambda(y)}\,dy.$$

More generally,

$$\frac{\delta(\bar{\psi}_{\alpha_1}(x_1)\bar{\psi}_{\alpha_2}(x_2)\cdots\bar{\psi}_{\alpha_n}(x_n))}{\delta\bar{\psi}_\lambda(y)} = \sum_{k=1}^{n}(-1)^{k+1}\bar{\psi}_{\alpha_1}(x_1)\cdots\bar{\psi}_{\alpha_{k-1}}(x_{k-1}) \times \delta_{\alpha_k\lambda}\delta(x_k - y)\bar{\psi}_{\alpha_{k+1}}(x_{k+1})\cdots\bar{\psi}_{\alpha_n}(x_n). \quad \text{(III.4)}$$

If some of the $\bar{\psi}$ in the product are ψ, they will give no contribution to (III.4), except that they must be included in the counting to assure the correct sign factor. One is thus led to the generalization

$$[\psi_\alpha(x), F(\psi, \bar{\psi})]_{-*} = i\int S_{\alpha\lambda}(x - y)\frac{\delta F}{\delta\bar{\psi}_\lambda(y)}\,dy. \quad \text{(III.5)}$$

The "sign indicator" s is $+1$ or -1 when F transforms like the product of an even or odd number of spinors. The distinction between positive and negative s in (III.5) should not be surprising. If the underlying invariance group is the Lorentz group, it means that when F transforms according to an irreducible representation $\mathfrak{D}^{(jj')}$, the indicator s is $+1$ when $j + j'$ is an integer and -1 when $j + j'$ is a half-odd integer,

$$s = (-1)^{2(j+j')}. \quad \text{(III.6)}$$

Thus, $F(\psi, \bar{\psi})$, for which $s = +1$ (-1), belongs to a tensor (spinor) representation and must have the functional derivative defined by a commutator (anticommutator).

The solution of (III.5) is again possible due to the existence of *inhomogeneous* anticommutators, satisfying

$$[\psi, O_1 + O_2]^{\mathrm{I}}_+ = [\psi, O_1]^{\mathrm{I}}_+ + [\psi, O_2]^{\mathrm{I}}_+, \quad \text{(III.7a)}$$

$$[\psi, OE]^{\mathrm{I}}_+ = [\psi, O]^{\mathrm{I}}_+E - O[\psi, E]^{\mathrm{I}}_-, \quad [\psi, EO]^{\mathrm{I}}_+ = [\psi, E]^{\mathrm{I}}_-O + E[\psi, O]^{\mathrm{I}}_+. \quad \text{(III.7b)}$$

We also note

$$[\psi, O_1O_2]^{\mathrm{I}}_- = [\psi, O_1]^{\mathrm{I}}_+O_2 - O_1[\psi, O_2]^{\mathrm{I}}_+, \quad [\psi, E_1E_2]^{\mathrm{I}}_- = [\psi, E_1]^{\mathrm{I}}_-E_2 + E_1[\psi, E_2]^{\mathrm{I}}_-. \quad \text{(III.7c)}$$

In these relations E and O are functionals with $s = +1$ and -1, respectively, i.e., are of tensor or spinor type. Furthermore, in analogy to (I.8), we require

$$(\gamma\cdot\partial + m)[\psi(x), \bar{\psi}(y)]^{\mathrm{I}}_+ = -\delta(x - y). \quad \text{(III.8)}$$

Examples of inhomogeneous anticommutators

$$[\psi(x), \bar{\psi}(y)]^{\mathrm{I}}_+ = iS^{\mathrm{I}}(x - y) \quad \text{(III.9)}$$

include

$$S_R(x) = \theta(x)S(x), \quad S_A(x) = -\theta(-x)S(x), \quad S_P(x) = \tfrac{1}{2}\epsilon(x)S(x). \quad \text{(III.10)}$$

The defining equation (III.5) now becomes

$$[\psi_\alpha(x), F(\psi, \bar{\psi})]^{\mathrm{I}}_{-*} = i\int S^{\mathrm{I}}_{\alpha\lambda}(x - y)\frac{\delta F}{\delta\bar{\psi}_\lambda(y)}\,dy. \quad \text{(III.11)}$$

It permits an immediate solution. Using (III.8),

$$(\gamma\cdot\partial + m)[\psi_\alpha(x), F(\psi, \psi)]^{\mathrm{I}}_{-*} = -i\frac{\delta F(\psi, \bar{\psi})}{\delta\bar{\psi}_\alpha(x)}. \quad \text{(III.12)}$$

This is the explicit definition analogous to (I.12).

The properties of the inhomogeneous commutators and anticommutators (III.7) permit the following conclusions:

$$\delta(\lambda F)/\delta\bar{\psi} = \lambda(\delta F/\delta\bar{\psi}), \quad \text{(III.13)}$$

$$\frac{\delta(F + G)}{\delta\psi} = \frac{\delta F}{\delta\bar{\psi}} + \frac{\delta G}{\delta\bar{\psi}}, \quad \text{(III.14)}$$

$$\begin{aligned}
\frac{\delta(E_1E_2)}{\delta\bar\psi} &= \frac{\delta E_1}{\delta\bar\psi}E_2 + E_1\frac{\delta E_2}{\delta\bar\psi},\\
\frac{\delta(O_1O_2)}{\delta\bar\psi} &= \frac{\delta O_1}{\delta\bar\psi}O_2 - O_1\frac{\delta O_2}{\delta\bar\psi},\\
\frac{\delta(OE)}{\delta\bar\psi} &= \frac{\delta O}{\delta\bar\psi}E - O\frac{\delta E}{\delta\bar\psi},\\
\frac{\delta(EO)}{\delta\bar\psi} &= \frac{\delta E}{\delta\bar\psi}O + E\frac{\delta O}{\delta\bar\psi}.
\end{aligned} \tag{III.15}$$

This last set of equations can be written more elegantly as

$$\left[\frac{\delta}{\delta\bar\psi}, E\right]_- = \frac{\delta E}{\delta\bar\psi}, \quad \left[\frac{\delta}{\delta\psi}, O\right]_+ = \frac{\delta O}{\delta\bar\psi}, \tag{III.15′}$$

and as

$$G_s\left[F, \frac{\overleftarrow{\delta}}{\delta\bar\psi}\right]_- = sG_s\frac{\delta F}{\delta\bar\psi}, \tag{III.15″}$$

where G_s is a functional with indicator s.

The functional derivative with respect to the operator field $\psi(x)$ can be obtained in complete analogy. The implicit definition is [compare (III.5)]

$$[F(\psi, \bar\psi), \bar\psi_\alpha(x)]_{-s} = i\int \frac{\delta F}{\delta\psi_\lambda(y)} S_{\lambda\alpha}(y - x)\, dy, \tag{III.16}$$

with

$$\begin{aligned}
&\frac{\delta(\psi_{\alpha_n}(x_n)\cdots\psi_{\alpha_1}(x_1))}{\delta\psi_\lambda(y)}\\
&= \sum_{k=1}^{n}(-1)^{k+1}\psi_{\alpha_n}(x_n)\cdots\psi_{\alpha_{k+1}}(x_{k+1})\\
&\times \delta_{\alpha_k\lambda}\delta(x_k - y)\psi_{k-1}(x_{k-1})\cdots\psi_{\alpha_1}(x_1).
\end{aligned} \tag{III.17}$$

The inhomogeneous anticommutators of interest here satisfy

$$[O_1 + O_2, \bar\psi]^{\mathrm{I}}_+ = [O_1, \bar\psi]^{\mathrm{I}}_+ + [O_2, \bar\psi]^{\mathrm{I}}_+, \tag{III.18a}$$

$$\begin{aligned}
[OE, \bar\psi]^{\mathrm{I}}_+ &= O[E, \bar\psi]^{\mathrm{I}}_- + [O, \bar\psi]^{\mathrm{I}}_+E,\\
[EO, \bar\psi]^{\mathrm{I}}_+ &= E[O, \bar\psi]^{\mathrm{I}}_+ - [E, \bar\psi]^{\mathrm{I}}_-O.
\end{aligned} \tag{III.18b}$$

We also note

$$\begin{aligned}
[O_1O_2, \bar\psi]^{\mathrm{I}}_- &= O_1[O_2, \bar\psi]^{\mathrm{I}}_+ - [O_1, \bar\psi]^{\mathrm{I}}_+O_2,\\
[E_1E_2, \bar\psi]^{\mathrm{I}}_- &= E_1[E_2, \bar\psi]^{\mathrm{I}}_- + [E_1, \bar\psi]^{\mathrm{I}}_-E_2.
\end{aligned} \tag{III.18c}$$

Again, an inhomogeneous equation is satisfied,

$$[\psi(y), \bar\psi(x)]^{\mathrm{I}}_+(-\gamma\cdot\overleftarrow{\partial}_x + m) = -\delta(y - x). \tag{III.19}$$

The same inhomogeneous anticommutators (III.9) and (III.10), as before, are examples also for the present case.

The implicit equation (III.16) becomes

$$[F(\psi, \bar\psi), \bar\psi_\alpha(x)]^{\mathrm{I}}_{-s} = i\int \frac{\delta F}{\delta\psi_\lambda(y)} S^{\mathrm{I}}_{\lambda\alpha}(y - x)\, dy, \tag{III.20}$$

and results in

$$\begin{aligned}
&-i\frac{\delta F(\psi, \bar\psi)}{\delta\psi_\alpha(x)}\\
&\quad = [F(\psi, \bar\psi), \bar\psi_\alpha(x)]^{\mathrm{I}}_{-s}(-\gamma\cdot\overleftarrow{\partial}_x + m).
\end{aligned} \tag{III.21}$$

The following basic differentiation rules obtain as a consequence of (III.18),

$$\delta(\lambda F)/\delta\psi = \lambda(\delta F/\delta\psi), \tag{III.22}$$

$$\frac{\delta(F + G)}{\delta\psi} = \frac{\delta F}{\delta\psi} + \frac{\delta G}{\delta\psi}, \tag{III.23}$$

$$\begin{aligned}
\frac{\delta(E_1E_2)}{\delta\psi} &= E_1\frac{\delta E_2}{\delta\psi} + \frac{\delta E_1}{\delta\psi}E_2,\\
\frac{\delta(O_1O_2)}{\delta\psi} &= O_1\frac{\delta O_2}{\delta\psi} - \frac{\delta O_1}{\delta\psi}O_2,\\
\frac{\delta(OE)}{\delta\psi} &= O\frac{\delta E}{\delta\psi} + \frac{\delta O}{\delta\psi}E,\\
\frac{\delta(EO)}{\delta\psi} &= E\frac{\delta O}{\delta\psi} - \frac{\delta E}{\delta\psi}O.
\end{aligned} \tag{III.24}$$

More elegantly, this last set of equations is

$$\left[E, \frac{\overleftarrow{\delta}}{\delta\psi}\right]_- = \frac{\delta E}{\delta\psi}, \quad \left[O, \frac{\overleftarrow{\delta}}{\delta\psi}\right]_+ = \frac{\delta O}{\delta\psi}, \tag{III.24′}$$

or

$$\left[\frac{\delta}{\delta\psi}, F\right]_- G_s = s\frac{\delta F}{\delta\psi}G_s, \tag{III.24″}$$

where G_s is a functional of spinor operators with indicator s. The difference in the signs of (III.24) and (III.15) should be noted.

As a check on the explicit definitions of the functional derivatives (III.12) and (III.21) one sees that, if the functional F is replaced by a function as in (III.4) and in (III.17), these equations are recovered.

It is clear that the derivative with respect to ψ or $\bar\psi$ of a functional of tensor type (spinor type) is of spinor type (tensor type). One can therefore make use of the identities which are similar to the Jacobi identity,

$$[A, [B, C]_-]_+ + [B, [A, C]_-]_+ = [[A, B]_+, C]_-, \tag{III.25}$$

and

$$[A, [B, C]_+]_- + [B, [C, A]_+]_- + [C, [A, B]_+]_- = 0. \quad \text{(III.26)}$$

From these one concludes that the following anticommutation relations must hold:

$$\left[\frac{\delta}{\delta\psi(x)}, \frac{\delta}{\delta\psi(y)}\right]_+ = 0, \qquad \left[\frac{\delta}{\delta\bar{\psi}(x)}, \frac{\delta}{\delta\bar{\psi}(y)}\right]_+ = 0. \quad \text{(III.27)}$$

To these equations should be added special cases of (III.15) and (III.24):

$$\left[\frac{\delta}{\delta\bar{\psi}_\alpha(x)}, \bar{\psi}_\beta(y)\right]_+ = \delta_{\alpha\beta}\delta(x-y), \qquad \left[\frac{\delta}{\delta\bar{\psi}(x)}, \psi(y)\right]_+ = 0, \quad \text{(III.28)}$$

$$\left[\bar{\psi}(x), \frac{\overleftarrow{\delta}}{\delta\psi(y)}\right]_+ = 0, \qquad \left[\psi_\alpha(x), \frac{\overleftarrow{\delta}}{\delta\psi_\beta(y)}\right]_+ = \delta_{\alpha\beta}\delta(x-y). \quad \text{(III.29)}$$

On the other hand, the following relations involve *commutators*:

$$\left[\frac{\delta}{\delta\bar{\psi}(x)}, \frac{\delta}{\delta\psi(y)}\right]_- = 0, \quad \text{(III.30)}$$

$$\left[\bar{\psi}_\alpha(x), \frac{\overleftarrow{\delta}}{\delta\bar{\psi}_\beta(y)}\right]_- = s\delta_{\alpha\beta}\delta(x-y), \qquad \left[\psi(x), \frac{\overleftarrow{\delta}}{\delta\bar{\psi}(y)}\right]_- = 0, \quad \text{(III.31)}$$

$$\left[\frac{\delta}{\delta\psi(x)}, \bar{\psi}(y)\right]_- = 0, \qquad \left[\frac{\delta}{\delta\psi_\alpha(x)}, \psi_\beta(y)\right]_- = s\delta_{\alpha\beta}\delta(x-y). \quad \text{(III.32)}$$

In the last two equations, s is the indicator of the functional to the left and right, respectively.

It is not surprising that the symmetry of the anticommutation relations (III.1) and (III.2) results in the need to introduce functional derivatives which can act to the left in addition to those which act to the right.[6]

Equations (III.30)–(III.32) are very useful. In particular, they permit the derivation of the "translation theorem" (I.15) also for the spinor case. Using the same identity as in the derivation of (I.15) one finds with (III.32)

$$e^{\lambda\delta/\delta\psi_\alpha(x)}\psi_\mu(y)e^{-\lambda\delta/\delta\psi_\alpha(x)} = \psi_\mu(y) + s\lambda\delta_{\alpha\mu}\delta(x-y), \quad \text{(III.33)}$$

where s refers again to the functional (to the right!) on which this expression operates. Its generalization, obtained by means of (III.24), is

$$e^{\lambda\delta/\delta\psi_\alpha(x)}F(\psi, \bar{\psi})e^{-\lambda\delta/\delta\psi_\alpha(x)} = F(\psi + s\lambda 1, \bar{\psi}). \quad \text{(III.34)}$$

It depends on the indicator of the functional on which it operates. The meaning of the unit operator 1 is evident from (III.33).

Analogous to this result is the "translation theorem" for $\bar{\psi}$,

$$\exp(\lambda[\overleftarrow{\delta}/\delta\bar{\psi}_\alpha(x)])F(\psi, \bar{\psi})\exp(-\lambda[\overleftarrow{\delta}/\delta\bar{\psi}_\alpha(x)]) = F(\psi, \bar{\psi} - s\lambda 1), \quad \text{(III.35)}$$

where s is the indicator of the functional to the *left* of this expression.

IV. ANALYTIC DEFINITION OF THE FUNCTIONAL DERIVATIVE

From (I.15), for small λ,

$$e^{\lambda\delta/\delta A(x)}F(A)e^{-\lambda\delta/\delta A(x)} = F(A) + \lambda\delta F/\delta A(x) + O(\lambda^2),$$

follows

$$\frac{\delta F(A)}{\delta A(x)} = \lim_{\lambda\to 0}\frac{F(A(y) + \lambda\delta(x-y)) - F(A)}{\lambda}. \quad \text{(IV.1)}$$

A completely analogous equation is obtanied for the tensor operator fields discussed in Sec. II. This equation is, in fact, often used as the defining equation of the functional derivative.

The usual justification for the definition (IV.1) is given in terms of the functional derivative with respect to a c number: Let $a(x)$ be a c-number function. Then one defines[6]

$$\frac{\delta F(A)}{\delta A(x)} = \frac{\delta F(A + a)}{\delta a(x)}\bigg|_{a=0}. \quad \text{(IV.2)}$$

The deduction of (IV.1) from our algebraic definition of the functional derivative justifies this procedure for the tensor operators in question.

The situation is different in the case of spinor operators. The definition (IV.2) cannot be applied consistently to *anti*commuting operators unless one admits the existence of c numbers $a(x)$ which also

[6] This possibility which also exists with ordinary derivatives (as in (III.19) must be clearly distinguished from the "right derivative" and "left derivative" defined by N. N. Bogoliubov and D. V. Shirkov in *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York 1959), p. 547. This ambiguity is eliminated in our treatment.

anticommute. Our algebraic definition is not beset with this difficulty. In fact, one deduces from (III.34),

$$\frac{\delta F(\psi, \bar{\psi})}{\delta \psi_\alpha(x)} = \lim_{\epsilon \to 0} \frac{F(\psi_\mu(y) + \epsilon \delta_{\mu\alpha} \delta(y-x), \bar{\psi}) - F(\psi, \bar{\psi})}{\epsilon}, \quad \text{(IV.3)}$$

with $\epsilon = s\lambda$. Similarly, from (III.35),

$$\frac{\delta F(\psi, \bar{\psi})}{\delta \bar{\psi}_\alpha(x)} = \lim_{\epsilon \to 0} \frac{F(\psi, \bar{\psi}_\mu(y) + \epsilon \delta_{\mu\alpha} \delta(y-x)) - F(\psi, \bar{\psi})}{\epsilon}, \quad \text{(IV.4)}$$

with $\epsilon = -s\lambda$. Both Eqs., (IV.3) and (IV.4) are independent of s.

The operator functional

$$\delta_\psi F(\psi, \bar{\psi}) = F(\psi + \epsilon 1, \bar{\psi}) - F(\psi, \bar{\psi})$$

is also well defined. But no relationship between $\delta_\psi F$ and $\delta F/\delta\psi$ exists other than (IV.3). No quantity $\delta\psi$ has been defined, and there is no question whether δF is given by $\delta\psi(\delta F/\delta\psi)$ or by $(\delta F/\delta\psi)\delta\psi$. This difficulty arises only when one starts with a *variational* expression δF involving $\delta\psi$. The above definitions are free of this difficulty.

V. CONCLUSION

The essential features of the above presentation can be stated as follows. Let $u_\alpha(x)$ be an operator field which transforms according to the irreducible Lorentz group representation $\mathfrak{D}^{(ii')}$. Let $\bar{u}_\alpha(x)$ be its adjoint, such that the commutators

$$[u_\alpha(x), \bar{u}_\beta(y)]_{(-1)^{2(i+i')+1}} = -i\Delta_{\alpha\beta}(x-y). \quad \text{(V.1)}$$

Then the functional derivative of a functional F of u and $\bar{u}$ is given implicitly by

$$[u_\alpha(x), F]_{-s} = -i \int \Delta_{\alpha\lambda}(x-y) \frac{\delta F}{\delta \bar{u}_\lambda(y)} dy, \quad \text{(V.2)}$$

where $s = +1$ (-1) when F transforms like a tensor (spinor) representation, i.e., $2(j + j')$ is even (odd).

Under certain conditions, (V.2) can be solved and an explicit expression can be obtained for the functional derivative. These conditions are satisfied for fields for which $\Delta_{\alpha\beta}$ contains no more than one derivative of the Δ function.

When the explicit form exists, the translation theorem can be derived and the analytic form of the functional derivative can be obtained from it. The problem of the so-called "anticommuting c numbers" does not arise.

ACKNOWLEDGMENT

Support for this work by the National Science Foundation and the Air Force Office of Scientific Research is gratefully acknowledged.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Feynman Integrals and the Schrödinger Equation*

EDWARD NELSON
Department of Mathematics, Princeton University, Princeton, New Jersey
(Received 19 June 1963)

Feynman integrals, in the context of the Schrödinger equation with a scalar potential, are defined by means of an analytic continuation in the mass parameter from the corresponding Wiener integrals. The method yields a new interpretation of highly singular attractive potentials in quantum mechanics. For the example of the attractive $1/r^2$ potential, this interpretation agrees with classical mechanics in the correspondence limit.

INTRODUCTION

It has been recognized since Feynman introduced his integrals[1] that there is a close formal analogy with Wiener integrals. In this paper we make the connection precise by using an analytic continuation to connect the two integrals. This idea is not new,[2] but by performing the analytic continuation in the mass parameter, rather than in the time or $\hbar$, we are able to treat attractive potentials as well as repulsive ones. In fact, the method yields new results concerning highly singular attractive potentials V which are not amenable to treatment by standard methods [such that there is no uniquely determined self-adjoint operator corresponding in a natural way to the expression $-(1/2m)\hbar^2\Delta + V$].

The proper mathematical context of the theory involves abstract Markoff processes. Here we shall develop the theory only for the special case of Brownian motion on Euclidean space without striving for maximum generality on the nature of the potential V. The relevant mathematical properties of Brownian motion are discussed in Appendix A, which is entirely expository. Similar properties of Markoff processes in general are well-known to the specialist in probability theory, but are not very accessible to the nonspecialist.

Appendix B is a largely expository account of a perturbation theory[3] due to Kato and Trotter. In Sec. 1 we show that this theory gives immediately the existence of the Feynman integrals for the class of potentials discussed by Kato.

1. FEYNMAN INTEGRALS FOR THE SCHRÖDINGER EQUATION WITH REGULAR POTENTIALS

Consider the initial-value problem for the Schrödinger equation

$$\partial u/\partial t = i[(1/2m)\Delta - V]u(t), \qquad u(0) = \psi. \tag{1}$$

Here Δ is the Laplace operator $\partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_l^2$ on l-dimensional Euclidean space $\mathbf{R}^l$, V is a real measurable function on $\mathbf{R}^l$, ψ and each $u(t)$ are elements of $\mathfrak{L}^2(\mathbf{R}^l)$, and m is a constant. [We have set $\hbar = 1$, so that if $l = 3$ then (1) is the Schrödinger equation for a single particle of mass m with the Hamiltonian $-(\hbar^2/2m)\Delta + V$.]

First suppose that $V = 0$. Then (1) may be solved explicitly by use of the Fourier transform. We recall briefly how this is done in order to fix some notation. Let $\mathfrak{F}$ denote the Fourier transformation, $\mathfrak{F}^{-1}$ its inverse. We define as usual

$$\Delta\psi = \mathfrak{F}^{-1}(-|\lambda|^2)\mathfrak{F}\psi \tag{2}$$

on the domain $\mathfrak{D}(\Delta)$ of all square-integrable ψ, such that $-|\lambda|^2\ \mathfrak{F}\psi$ is also square-integrable. (Here λ denotes the variable in momentum space and $|\lambda|^2 = \lambda_1^2 + \cdots + \lambda_l^2$.) Then Δ is self-adjoint, and if

$$K_m^t = \exp it(1/2m)\Delta, \tag{3}$$

then $u(t) = K_m^t\psi$ is the solution of (1) for $V = 0$. Explicitly,

$$K_m^t\psi(x) = (2\pi it/m)^{-\frac{1}{2}l} \times \int \exp\left[i\tfrac{1}{2}m(|x - y|^2/t)\right]\psi(y)\, dy, \tag{4}$$

where $z^{-\frac{1}{2}l}$ has that determination which is positive for $z = 2\pi it/m$ positive, the complex plane being cut along the negative half-axis, and where $dy = dy_1 \cdots dy_l$. The integral in (4) exists for all ψ in $\mathfrak{L}^1 \cap \mathfrak{L}^2$—for a general ψ in $\mathfrak{L}^2$ we interpret

* This work was supported by a fellowship from the Alfred P. Sloan Foundation.

[1] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).

[2] See I. M. Gelfand and A. M. Yaglom, Uspekhi Mat. Nauk **11**, 77 (1956) (English transl.: J. Math. Phys. **1**, 48 (1960); R. H. Cameron, J. Math. & Phys. **39**, 126 (1960); D. G. Babbitt, J. Math. Phys. **4**, 36 (1963); and J. Feldman, Trans. Am. Math. Soc. **108**, 251 (1963). A preliminary version of the present paper appears in Colloques internationaux du C. N. R. S., Colloque sur les équations aux dérivées partielles (Paris, 1962).

[3] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951); H. F. Trotter, Proc. Am. Math. Soc. **10**, 345 (1959), and Pacific J. Math. **8**, 887 (1958). An error in a proof in the last reference has been corrected by T. Kato, Proc. Japan Acad. **35**, 467 (1959).

$K_m^t\psi$ as the limit in $\mathfrak{L}^2$ as $K \to \infty$ of the integral restricted to $|y| \leq K$.

Secondly, suppose that $1/2m$ is replaced by 0 in (1). The operator V of multiplication by the function V, on the domain $\mathfrak{D}(V)$ of all ψ in $\mathfrak{L}^2$ such that $V\psi$ is also in $\mathfrak{L}^2$, is self-adjoint, and if

$$M_V^t = \exp(-itV), \tag{5}$$

then $u(t) = M_V^t\psi$ is the solution of (1) with $1/2m$ replaced by 0.

Kato[3] has found conditions under which the operator $(1/2m)\Delta - V$ is self-adjoint, and under these conditions if we let

$$U_{m,V}^t = \exp it[(1/2m)\Delta - V], \tag{6}$$

then a theorem of Trotter[3] asserts that for all ψ in $\mathfrak{L}^2$,

$$U_{m,V}^t\psi = \lim_{n\to\infty} (K_m^{t/n}M_V^{t/n})^n\psi. \tag{7}$$

This is discussed in detail in Appendix B.

Using (4) and (5) we find that

$$(K_m^{t/n}M_V^{t/n})^n\psi(x) = (2\pi it/nm)^{-\frac{1}{2}ln} \times \int \cdots \int e^{iS(x_0,\cdots,x_n;t)}\psi(x_n)\,dx_1 \cdots dx_n, \tag{8}$$

where we have set $x_0 = x$ and

$$S(x_0, \cdots, x_n; t) = \sum_{j=1}^{n} \left[\frac{m}{2}\frac{|x_j - x_{j-1}|^2}{(t/n)^2} - V(x_j)\right]\frac{t}{n}. \tag{9}$$

Let ω be a trajectory; that is, a function from $[0, t]$ to $\mathbf{R}^l$, with $\omega(0) = x$ and set $x_j = \omega(jt/n)$ for $j = 0, \cdots, n$. Then (9) is, formally, the Riemann sum for the action

$$S(\omega; t) = \int_0^t \left[\frac{m}{2}\dot{\omega}^2 - V(\omega(s))\right] ds \tag{10}$$

of a particle of mass m with the trajectory ω in the presence of the potential V. Taking the limit formally in (8) yields the Feynman integral

$$\text{constant} \int_{\Omega_x} e^{iS(\omega;t)}\psi(\omega(t))\,\mathfrak{D}\omega, \tag{11}$$

where Ω_x is the set of all trajectories with $\omega(0) = x$.

As is well-known, there are difficulties in interpreting (11): the constant is infinite (of infinite order), the action (10) exists only if ω and V are sufficiently smooth, and $\mathfrak{D}\omega = \prod_{0\leq s\leq t} dx_s$ has no meaning. However, (8) is well-defined, and according to (7) converges in $\mathfrak{L}^2$ to the solution of the Schrödinger equation. Thus the results of Kato and Trotter give a precise meaning to the Feynman integral when the potential V is sufficiently regular. In the following sections we give a different interpretation of the Feynman integral, relating it to a genuine integration over trajectories, which is valid for a much wider class of potentials.

2. THE SCHRÖDINGER EQUATION WITH IMAGINARY MASS

The results of this section which are used in the sequel are summarized in Theorem 1 near the end of the section.

If m is purely imaginary with $\operatorname{Im} m > 0$, let us put $D = i/2m$, so that $D > 0$. Then (1) becomes the heat equation with zero-order term iV,

$$\partial u/\partial t = D\Delta u - iVu; \qquad u(0) = \psi. \tag{12}$$

Kac[4] has shown how to solve such equations by means of Wiener integrals. We shall make the following assumption on the function V:

> There is a closed set F in $\mathbf{R}^l$ of capacity 0 such that V is continuous and real on the complement of F. (13)

For example, if $l > 1$ and V is a function which is continuous except at one point, say the origin (with an arbitrary singularity at the origin and arbitrary growth at infinity), then V satisfies (13). Let x be any point in the complement of F and Pr_x the corresponding Wiener measure, with diffusion constant $D = i/2m > 0$, on the space Ω of all trajectories. Then for almost every ω in Ω, $V(\omega(t))$ is a continuous function of t for $0 \leq t < \infty$ (this is discussed in detail in Appendix A), and so we may form the integral

$$U_{m,V}^t\psi(x) = \int_\Omega \exp\left[-i\int_0^t V(\omega(s))\,ds\right] \times \psi(\omega(t))Pr_x(d\omega) \tag{14}$$

for any ψ in $\mathfrak{L}^2$ and any $t \geq 0$. In fact, the integrand is defined for almost every ω in Ω, and it is bounded in absolute value by $|\psi(\omega(t))|$. But this is integrable, for by definition[5] $\int_\Omega |\psi(\omega(t))|\,Pr_x(d\omega) = K_m^t|\psi|(x)$, and this is finite by (4).

Since $V(\omega(s))$ is continuous for almost every ω,

$$\int_0^t V(\omega(s))\,ds = \lim_{n\to\infty} \sum_{j=1}^{n} V\left(\omega\left(j\frac{t}{n}\right)\right)\frac{t}{n}$$

[4] M. Kac, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley and Los Angeles, 1951), p. 189.

[5] The operator which in Appendix A is denoted P^t is here denoted $K_m{}^t$, as in (4), with $D = i/2m > 0$.

for almost every ω. By the Lebesgue dominated convergence theorem,[6] this implies that

$$U^t_{m,V}\psi(x) = \lim_{n\to\infty} \int_\Omega \exp\left[-i \sum_{j=1}^{n} V\left(\omega\left(j\frac{t}{n}\right)\right)\frac{t}{n}\right] \times \psi(\omega(t))Pr_x(d\omega)$$

for all $x \notin F$, and by definition of the Wiener integral, this means that for all $x \notin F$

$$U^t_{m,V}\psi(x) = \lim_{n\to\infty} (K_m^{t/n}M_V^{t/n})^n\psi(x).$$

A set F of capacity 0 has Lebesgue measure 0, and so if $\psi_n = (K_m^{t/n}M_V^{t/n})^n\psi$, then ψ_n converges almost everywhere in $\mathbf{R}^l$ to $U^t_{m,V}\psi$. By (4), $|K_m^t\varphi| \leq K_m^t|\varphi|$ for $i/2m > 0$ and by (5), M_V^t is multiplication by a function of modulus 1. Consequently, $|\psi_n| \leq K_m^t|\psi|$, and so $|\psi_n - \psi_{n'}|^2 \leq 4(K_m^t|\psi|)^2$, which is in $\mathfrak{L}^1$. Therefore, by the Lebesgue dominated convergence theorem, ψ_n is a Cauchy sequence in $\mathfrak{L}^2$, and so

$$U^t_{m,V}\psi = \lim_{n\to\infty} (K_m^{t/n}M_V^{t/n})^n\psi, \tag{15}$$

the limit now being taken in $\mathfrak{L}^2$.

Now let ψ be of class $\mathcal{C}^2$ with compact support in the complement of F, and let $\mathfrak{D}_0$ be the set of all such ψ. Then, by (14),

$$U^t_{m,V}\psi(x) = \int_\Omega [1 - itV(x)]\psi(\omega(t))Pr_x(d\omega) + o(t).$$

The term $o(t)$ is uniform in x, for x in a neighborhood of the support of ψ, by Lemma 3 of Appendix A. Therefore,

$$U^t_{m,V}\psi(x) = K_m^t\psi(x) - itVK_m^t\psi(x) + o(t),$$

and so

$$\lim_{t\to 0} \frac{1}{t}(U^t_{m,V}\psi - \psi) = i\left(\frac{1}{2m}\Delta - V\right)\psi, \tag{16}$$

for all ψ in $\mathfrak{D}_0$. The limit in (16) is uniform in x for x in a neighborhood N of the support of ψ. On the complement of N, $(1/t)(U^t_{m,V}\psi - \psi) = (1/t)U^t_{m,V}\psi$ converges to $0 = i[(1/2m)\Delta - V]\psi$ in $L^2(R^l - N)$, by (14), so that (16) holds with the limit taken in $L^2(R^l)$. In particular,

$$\lim_{t\to 0} U^t_{m,V}\psi = \psi \tag{17}$$

for all ψ in $\mathfrak{D}_0$, and since $\mathfrak{D}_0$ is dense in $\mathfrak{L}^2$ and $||U^t_{m,V}|| \leq 1$, (17) holds for all ψ in $\mathfrak{L}^2$.

An easy consequence of this fact and (15) is that $U^t_{m,V}$ is a strongly continuous semigroup on $\mathfrak{L}^2$; that is,

$$U^t_{m,V}U^s_{m,V} = U^{t+s}_{m,V}, \qquad 0 \leq t, s < \infty, \tag{18}$$

and for all ψ in $\mathfrak{L}^2$, $t \to U^t_{m,V}\psi$ is continuous from $[0, \infty)$ to $\mathfrak{L}^2$. Let $A_{m,V}$ be the infinitesimal generator of the semigroup $U^t_{m,V}$. That is,

$$A_{m,V}\psi = \lim_{t\to 0} \frac{1}{t}(U^t_{m,V}\psi - \psi) \tag{19}$$

on the domain $\mathfrak{D}$ of all ψ in $\mathfrak{L}^2$ for which the limit exists in $\mathfrak{L}^2$. By (16), $\mathfrak{D}_0 \subset \mathfrak{D}$ and $A_{m,V}\psi = i[(1/2m)\Delta - V]\psi$ for all ψ in $\mathfrak{D}_0$.

We shall need more precise information on the domain $\mathfrak{D}$. To obtain this, let

$$V_j(x) = \begin{cases} j, & j \leq V(x), \\ V(x), & -j < V(x) < j, \\ -j, & V(x) \leq -j. \end{cases} \tag{20}$$

Thus V_j is a bounded continuous function on the complement of F, and $\lim_{j\to\infty} V_j(x) = V(x)$ for $x \notin F$. By the same kind of argument used to establish (15),

$$\lim_{j\to\infty} \int_0^t V_j(\omega(s))\,ds = \int_0^t V(\omega(s))\,ds$$

for almost every ω, and

$$\lim_{j\to\infty} U^t_{m,V_j}\psi = U^t_{m,V}\psi \tag{21}$$

for all ψ in $\mathfrak{L}^2$. Let

$$R_{m,V}\psi = \int_0^\infty e^{-t}U^t_{m,V}\psi\,dt \tag{22}$$

for all ψ in $\mathfrak{L}^2$, and define R_{m,V_j} similarly. Then clearly

$$\lim_{j\to\infty} R_{m,V_j}\psi = R_{m,V}\psi \tag{23}$$

for all ψ in $\mathfrak{L}^2$. However, each V_j is bounded and it is easily seen that the infinitesimal generator A_{m,V_j} of U^t_{m,V_j} is $i[(1/2m)\Delta - V_j] = D\Delta - iV_j$, with the same domain as Δ.

Define

$$||\text{grad } \varphi||^2 = \sum_{k=1}^{l} \int \left|\frac{\partial\varphi}{\partial x_k}(x)\right|^2 dx, \tag{24}$$

where $\partial\varphi/\partial x_k$ is an $\mathfrak{L}^2$ derivative, which may be defined by $\partial\varphi/\partial x_k = \mathfrak{F}^{-1}i\lambda_k\mathfrak{F}\varphi$ on the domain of all square-integrable φ such that $\lambda_k\mathfrak{F}\varphi$ is also square-integrable. If $\varphi_j \to \varphi$ in $\mathfrak{L}^2$, and $||\text{grad } \varphi_j||$ is uniformly bounded, then it is easy to see that $||\text{grad } \varphi|| < \infty$. For all φ in the domain of A_{m,V_j},

[6] See P. R. Halmos, *Measure Theory* (D. Van Nostrand, Inc., New York, 1950).

$$\text{Re}\,((1 - A_{m,V_i})\varphi, \varphi) = D\,||\text{grad}\,\varphi||^2 + ||\varphi||^2.$$

In particular, for $\varphi = R_{m,V_i}\psi$, since[7] $R_{m,V_i} = (1 - A_{m,V_i})^{-1}$, we have $\text{Re}\,((1 - A_{m,V_i})\varphi, \varphi) = \text{Re}\,(\psi, R_{m,V_i}\psi) \leq ||\psi||^2$, and consequently

$$||\text{grad}\,R_{m,V_i}\psi||^2 \leq \frac{1}{D}\,||\psi||^2. \tag{25}$$

By (23) and (25),

$$||\text{grad}\,R_{m,V}\psi||^2 \leq \frac{1}{D}\,||\psi||^2. \tag{26}$$

Since the range of $R_{m,V}$ is equal to the domain $\mathfrak{D}$ of $A_{m,V}$, we conclude that every element φ in $\mathfrak{D}$ has the property that $||\text{grad}\,\varphi|| < \infty$.

We have proved the following theorem.

Theorem 1. Let m be purely imaginary with $\text{Im}\,m > 0$ *and let V satisfy* (13). *Then for all ψ in $\mathfrak{L}^2$ and $t \geq 0$, the limit* (15) *exists (where K_m^t and M_V^t are defined by* (4) *and* (5), *respectively). The operators $U_{m,V}^t$ so defined have the semigroup property* (18) *and for all ψ in $\mathfrak{L}^2$, $t \to U_{m,V}^t\psi$ is continuous from $[0, \infty)$ to $\mathfrak{L}^2$. If $A_{m,V}$ is defined by* (19), *then for all φ in the domain of $A_{m,V}$, $||\text{grad}\,\varphi|| < \infty$, where $||\text{grad}\,\varphi||$ is defined by* (24).

We remark without proof that, if $u(t, x) = U_{m,V}^t\psi(x)$ then u satisfies (12) in the sense of distributions for x in the complement of F.

3. THE SCHRÖDINGER EQUATION WITH COMPLEX MASS

Theorem 2. Theorem 1 remains true for all complex m with $\text{Im}\,m > 0$.

To prove this, observe that from the definition (3) of K_m^t and the fact that Δ is a negative operator, K_m^t is a holomorphic operator-valued function[7] of m in the upper half-plane $\text{Im}\,m > 0$, for each $t \geq 0$. Consequently if ψ is in $\mathfrak{L}^2$, $(K_m^{t/n}M_V^{t/n})^n\psi$ is a holomorphic $\mathfrak{L}^2$-valued function of m for $\text{Im}\,m > 0$, and the sequence is uniformly bounded in norm by $||\psi||$. By Theorem 1, the sequence converges for m on the imaginary upper half-axis. But a uniformly bounded sequence of holomorphic functions which converges on a determining set, converges on the entire domain, so that the limit (15) exists for all ψ in $\mathfrak{L}^2$, $t \geq 0$, and $\text{Im}\,m > 0$. We take (15) as the definition of $U_{m,V}^t$ for $\text{Im}\,m > 0$—it is not defined by a functional integral when m is not purely imaginary, as Cameron[2] showed. The same argument shows that $t \to U_{m,V}^t\psi$ is continuous from $[0, \infty)$ to $\mathfrak{L}^2$, whenever $\text{Im}\,m > 0$ and ψ is in $\mathfrak{L}^2$. The semigroup property (18) continues to hold by analytic continuation. If $A_{m,V}$ is defined by (19), then it can be shown (as before by considering the bounded functions V_i and obtaining an estimate independent of j) that each φ in the domain of $A_{m,V}$, for $\text{Im}\,m > 0$, satisfies $||\text{grad}\,\varphi|| < \infty$. As we shall not need this last fact, we omit the details.

4. THE SCHRÖDINGER EQUATION WITH REAL MASS

Theorem 3. Let V satisfy (13). *There is a set N of real numbers of Lebesgue measure* 0 *such that for all $m \notin N$, ψ in $\mathfrak{L}^2$ and $t \geq 0$*

$$U_{m,V}^t\psi = \lim_{m'} U_{m',V}^t\psi \tag{27}$$

exists when m' approaches m nontangentially from the upper half-plane (that is, $\text{Im}\,m' > 0$, $m' \to m$, *and* $|\text{Re}\,(m' - m)|/\text{Im}\,m'$ *is bounded). The operators $U_{m,V}^t$ so defined have the semigroup property* (18) *and for all ψ in $\mathfrak{L}^2$, $t \to U_{m,V}^t$ is continuous from $0[, \infty)$ to $\mathfrak{L}^2$.*

The restriction to almost every real value of the mass parameter is an unsatisfactory feature of the theory. It would be interesting to know whether $U_{m',V}^t$ always has continuous boundary values for $1/2m$ real. The example of Sec. 5 shows that, for m real, $(K_m^{t/n}M_V^{t/n})^n\psi$ does not always have a limit. As a function of m it may have a highly oscillatory part which in the limit as $n \to \infty$ makes no contribution to the analytic extension to the upper half-plane. The proof of Theorem 3 is based on the Fatou–Privaloff theorem[8] which asserts that a bounded holomorphic function has a nontangential limit almost everywhere on the real axis. The remainder of this section is devoted to the measure-theoretic details of the proof.

Let ψ and φ be in $\mathfrak{L}^2$, so that $(U_{m',V}^t\psi, \varphi)$ is a bounded holomorphic[9] complex-valued function of m' for $\text{Im}\,m' > 0$. By the Fatou–Privaloff theorem, for almost every real m,

$$(U_{m,V}^t\psi, \varphi) = \lim_{m'} (U_{m',V}^t\psi, \varphi) \tag{28}$$

exists. (Throughout this section a limit on m' will mean a nontangential limit.) That is, $U_{m',V}^t\psi$ has a nontangential limit almost everywhere in the weak topology of $\mathfrak{L}^2$. The limit $U_{m,V}^t\psi$ is a weakly

[7] See E. Hille and R. S. Phillips, *Functional Analysis and Semi-Groups* (American Mathematical Society Colloquium Publications, Providence, Rhode Island, 1957), Vol. 31.

[8] See L. P. Priwalow, *Randeigenschaften Analytischer Funktionen* (VEB Deutscher Verlag der Wissenschaften, Berlin, 1958).

[9] Our inner products are linear in the first vector and conjugate linear in the second.

measurable function of m, and since $\mathfrak{L}^2$ is separable it is a strongly measurable[7] function of m. If $m' = x + iy$, $y > 0$, then by the Poisson formula

$$U^t_{m',V}\psi = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{y^2 + (x-m)^2}\, U^t_{m,V}\psi \, dm. \tag{29}$$

Since $U^t_{m,V}\psi$ is strongly measurable, (29) holds as a strong (Bochner) integral, and using the Schwarz inequality we find that

$$\begin{aligned} ||U^t_{m',V}\psi||^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{y}{y^2+(x-m_1)^2}\frac{1}{\pi} \\ &\quad \times \frac{y}{y^2+(x-m_2)^2}\,(U^t_{m_1,V}\psi,\, U^t_{m_2,V}\psi)\, dm_1\, dm_2 \\ &\leq \left[\int_{-\infty}^{\infty} \frac{1}{\pi}\frac{y}{y^2+(x-m)^2}\,||U^t_{m,V}\psi||\, dm\right]^2. \end{aligned}$$

Therefore,

$$||U^t_{m,V}\psi|| \geq \liminf_{m'} ||U^t_{m',V}\psi|| \tag{30}$$

for almost every m. In a Hilbert space, (28) and (30) together imply strong convergence, so that for every ψ in $\mathfrak{L}^2$ and $t \geq 0$, (27) holds for almost every m.

The exceptional set may depend on ψ and t. However, (27) must hold almost everywhere simultaneously for a countable dense set of ψ's, and since the $U^t_{m',V}$ are linear operators with uniformly bounded norms ($||U^t_{m',V}|| \leq 1$), (27) holds almost everywhere simultaneously for all ψ in $\mathfrak{L}^2$. That is, for each $t \geq 0$, there is a set N_t of measure 0 such that for all $m \notin N_t$ and all ψ in $\mathfrak{L}^2$, (27) holds. By Fubini's theorem,[6] there is a set N of measure 0 such that for all $m \notin N$ and all ψ in $\mathfrak{L}^2$, (27) holds for almost every $t \geq 0$. Now $U^t_{m',V}U^s_{m',V} = U^{t+s}_{m',V}$ for Im $m' > 0$, and if $\lim_{m'} U^t_{m',V}$ and $\lim_{m'} U^s_{m',V}$ exist in the strong operator topology, so does $\lim_{m'} U^t_{m',V}U^s_{m',V}$. But if we have a subset of $(0, \infty)$ whose complement is of measure 0, then every number in $(0, \infty)$ is of the form $t + s$ for some t and s in the subset. Consequently (27) holds for all ψ in $\mathfrak{L}^2$ and all $t \geq 0$ whenever $m \notin N$, and the semigroup property (18) holds whenever $m \notin N$.

By their construction, the $U^t_{m,V}$ are strongly measurable functions of t for $m \notin N$, and this together with (18) implies[7] that $U^t_{m,V}\psi$ is continuous in t for $t > 0$. To finish the proof of the theorem, we need only show that $U^t_{m,V}\psi \to \psi$ as $t \to 0$ for all ψ in $\mathfrak{L}^2$. By the Poisson formula,

$$\begin{aligned} &\text{Re}\,(U^t_{m',V}\psi,\, \psi) \\ &\quad = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{y^2+(x-m)^2}\,\text{Re}\,(U^t_{m,V}\psi,\, \psi)\, dm, \end{aligned}$$

where $m' = x + iy$, $y > 0$. By the Lebesgue dominated convergence theorem,

$$\begin{aligned} (\psi,\, \psi) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{y}{y^2+(x-m)^2} \\ &\quad \times \liminf_{t\to 0} \text{Re}\,(U^t_{m,V}\psi,\, \psi)\, dm, \end{aligned}$$

and so $\liminf_{t\to 0} \text{Re}\,(U^t_{m,V}\psi,\, \psi) = (\psi,\, \psi)$ for almost every m. Since $||U^t_{m,V}\psi - \psi||^2 = 2(\psi,\, \psi) - 2\,\text{Re}\,(U^t_{m,V}\psi,\, \psi)$, $\limsup_{t\to 0} ||U_{m,V}\psi - \psi|| = 0$ for almost every m. This is true simultaneously for a countable dense set of ψ's and consequently simultaneously for all ψ in $\mathfrak{L}^2$. Therefore, by enlarging the set N if necessary, for all $m \notin N$, $t \to U^t_{m,V}\psi$ is continuous from $[0, \infty)$ to $\mathfrak{L}^2$, for all ψ in $\mathfrak{L}^2$. This concludes the proof of Theorem 3.

It is not difficult to show that if $u(x, t) = U^t_{m,V}\psi(x)$, then u satisfies (1) in the sense of distributions for $m \notin N$ and $x \notin F$.

Notice that if V is a Kato potential as in Sec. 1 then $\lim_{n\to\infty} (K^{t/n}_m M^{t/n}_V)^n\psi$ exists for all m with Im $m \geq 0$, $m \neq 0$ and so the Feynman integral as constructed in Secs. 2–4 agrees with the Feynman integral of Sec. 1 whenever the latter exists.

We remark that a completely analogous theory for $t \leq 0$ may be developed by taking m in the lower half-plane. Since $(K^t_m M^t_V)^* = M^{-t}_V K^{-t}_m$, we find that

$$(U^t_{m,V})^* = U^{-t}_{m,V}. \tag{31}$$

5. THE $1/r^2$ POTENTIAL

The $1/r^2$ potential has been discussed by a number of authors.[10] The conclusions of Case are essentially that for certain values of the parameters the Hamiltonian is semibounded and the usual extension (the Friedrichs extension) is a self-adjoint operator, while for the other values of the parameters this is no longer true; one is in the limit circle case[11] and there is a one-parameter family of self-adjoint extensions, no one of which is singled out in any natural way. The solution of the Schrödinger equation by means of Feynman integrals, as described in the preceding sections, leads to very different results which bear out the interpretation of Landau and Lifschitz[10] of the possibility of a collision with the center of attraction.

We consider now a three-dimensional Euclidean space, and write $\mathfrak{L}^2(\mathbf{R}^3)$ as the direct sum of the subspaces $\mathfrak{H}_l$ of wavefunctions with azimuthal

[10] K. M. Case, Phys. Rev. **80**, 797 (1950); L. D. Landau and E. M. Lifshitz, *Quantum Mechanics—Non-Relativistic Theory*, translated by J. B. Sykes and J. S. Bell, (Addison–Wesley Publishing Company, Inc., Reading, Massachusetts, 1958).

[11] H. Weyl, Math. Ann. **68**, 220 (1910).

quantum number l. Let us define u by $\psi_0 = ur$, where r is the distance to the origin and ψ_0 is the radial part of ψ in $\mathcal{H}_l$. Then Schrödinger's equation with a centrally symmetric potential V becomes (see, e.g., Landau and Lifschitz[10])

$$\frac{\partial u}{\partial t} = \frac{i}{2m}\frac{\partial^2}{\partial r^2}u - \frac{i}{2m}\frac{l(l+1)}{r^2}u - iVu,$$

where for each t, $u(t)$ lies in $\mathfrak{L}^2(0, \infty)$. We shall use the same notation $U^t_{m,V}$, etc. for operators on $\mathcal{H}_l$ and the corresponding operators on $\mathfrak{L}^2(0, \infty)$. Consider the potential $V = -1/r^2$. Let us find the kernel of the integral operator $R_{m,V}$ defined by (22); that is, we wish to find a function u such that

$$(d^2/dr^2)u + \{2mi + [2m - l(l+1)]/r^2\}u = -2mi\delta(r - r_0), \qquad (32)$$

where r_0 is some fixed point in $(0, \infty)$. Then (32) for $r \neq r_0$ is a form of Bessel's equation[12] and has the solutions

$$r^{\frac{1}{2}}\mathcal{C}_\nu((2mi)^{\frac{1}{2}}r),$$

where $\nu^2 = \frac{1}{4} + l(l+1) - 2m$ and $\mathcal{C}_\nu$ denotes any linear combination of the Hankel functions H^1_ν and H^2_ν of order ν; for example, the Bessel function J_ν. The determination of $(2mi)^{\frac{1}{2}}$ is unimportant. We choose that determination which is positive imaginary when m is positive imaginary, the m plane being cut along the negative imaginary half-axis. Then to the right of r_0 there is no problem—u must be square-integrable near ∞, and so

$$u(r) = \text{constant}\, r^{\frac{1}{2}}H^1_\nu((2mi)^{\frac{1}{2}}r), \qquad r > r_0. \qquad (33)$$

The determination of ν is crucial. We use the fact, established in Theorem 1, that every element φ in the domain of $A_{m,V}$, for m purely imaginary with $\operatorname{Im} m > 0$, satisfies $||\operatorname{grad} \varphi|| < \infty$. This implies that du/dr is square-integrable near 0. Now $r^{\frac{1}{2}}J_\nu((2mi)^{\frac{1}{2}}r)$ has an expansion beginning

$$cr^{\frac{1}{2}+\nu}, \qquad (34)$$

where c is a nonzero constant. When m is purely imaginary with $\operatorname{Im} m > 0$, $\nu^2 = \frac{1}{4} + l(l+1) - 2m$ lies in the fourth quandrant, and so ν lies in the second or fourth quadrant. To ensure that du/dr be square-integrable near 0, we must choose ν to lie in the fourth quandrant, by (34). By analytic continuation, $\operatorname{Im} \nu < 0$ when $\operatorname{Im} m > 0$, and when m is real, ν lies on the boundary of the fourth quadrant. That is,

$$u(r) = \text{constant}\, r^{\frac{1}{2}}J_\nu((2mi)^{\frac{1}{2}}r), \qquad r < r_0, \qquad (35)$$

where $\nu^2 = \frac{1}{4} + l(l+1) - 2m$, $\nu \geq 0$ if $\nu^2 \geq 0$, and ν is negative imaginary if $\nu^2 < 0$. Thus we have determined the Green's function for our problem, by (33) and (35) (except for the values of the constants, which we will not need).

Now let ζ be of class $\mathcal{C}^2$ on $[0, \infty)$, $\zeta = 1$ near 0 and $\zeta = 0$ near ∞. Then

$$f(r) = r^{\frac{1}{2}}J_\nu((2mi)^{\frac{1}{2}}r)\zeta(r)$$

is in the domain of the infinitesimal generator $A_{m,V}$ of $U^t_{m,V}$. Now $iA_{m,V}$ is an extension of $-(1/2m)\Delta + V$ on the domain of all $\mathcal{C}^2$ functions with compact support in $\mathbf{R}^3 - \{0\}$, by (16), but

$$2m(iA_{m,V}f, f) = \int_0^\infty \left(\frac{d^2}{dr^2} + \frac{2m - l(l+1)}{r^2}\right)f\cdot\bar{f}\, dr$$

$$= \lim_{\epsilon\to 0}\int_\epsilon^\infty \left(\frac{2m - l(l+1)}{r^2}f\bar{f} - \frac{df}{dr}\frac{d\bar{f}}{dr}\right)dr$$

$$- \lim_{\epsilon\to 0}\frac{df}{dr}(\epsilon)\bar{f}(\epsilon).$$

If $\nu^2 < 0$, this is equal to $a + |c|^2\nu$, by (34), where a is real. Consequently, if $\nu^2 < 0$, $iA_{m,V}$ is not symmetric, let alone self-adjoint, even though it is an extension of a symmetric operator, and so $U^t_{m,V}$ is not unitary. In fact,

$$(d/dt)\,||U^t_{m,V}f||^2\,|_{t=0} = -i\nu\,|c|^2 < 0,$$

so that $||U^t_{m,V}f|| < ||f||$ for $t > 0$.

It is easily verified that if $\nu^2 > 0$ our $U^t_{m,V}$ is the usual unitary group of operators generated by the Friedrichs extension of $-(1/2m)\Delta + V$. Now $-(1/2m)\Delta + V$ is unitarily equivalent to any strictly positive constant multiple of itself, since for $k > 0$, $\psi(x) \to k^{\frac{3}{2}}\psi(kx)$ is unitary on $\mathfrak{L}^2(\mathbf{R}^3)$, and Δ corresponds to $k^{-2}\Delta$ and $V = -1/r^2$ corresponds to $k^{-2}V$. Thus when $\nu^2 > 0$, $iA_{m,V}$ is self-adjoint with spectrum $[0, \infty)$ and no bound states. When $\nu^2 = 0$, $U^t_{m,V}$ is isometric, being a limit of unitary operators. The same argument works for negative t, and so by (31) $U^t_{m,V}$ is still unitary when $\nu^2 = 0$. It may be verified that the operator $iA_{m,V}$ is still bounded below by 0 when $\nu = 0$, but since du/dr is not square-integrable near 0, by (34), there are functions ψ in the domain of $iA_{m,V}$ such that the kinetic energy $(1/2m)\,||\operatorname{grad}\psi||^2$ is ∞ and the potential energy $(V\psi, \psi)$ is $-\infty$. For $\nu^2 < 0$ we have seen that $iA_{m,V}$ is not self-adjoint.

Let us insert dimensional parameters, so that our Hamiltonian becomes $-(1/2m)\hbar^2\Delta + (\beta/r^2)$,

[12] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1944).

where β has the dimensions ML^4T^{-2}. Then $\nu^2 = \hbar^2 l(l+1) + \frac{1}{4}\hbar^2 - 2m\beta$. Let $J^2 = \hbar^2 l(l+1)$, so that J is the angular momentum, and let ψ be a state vector with angular momentum J ($\psi \in \mathfrak{K}_l$, $||\psi|| = 1$). Then our conclusions on the motion of a Schrödinger particle of mass m in the presence of the potential $-\beta/r^2$ may be summarized as follows:

Case I. $2m\beta < J^2 + \frac{1}{4}\hbar^2$. The motion is described by a unitary group of operators and the energy is bounded below.

Case II. $2m\beta = J^2 + \frac{1}{4}\hbar^2$. The motion is described by a unitary group of operators but the potential energy is not bounded below in terms of the total energy.

Case III. $2m\beta > J^2 + \frac{1}{4}\hbar^2$. The motion is described by a nonunitary semigroup of operators $U^t_{m,\nu}$, and in general $1 - ||U^t_{m,\nu}\psi||^2 > 0$. We interpret $1 - ||U^t_{m,\nu}\psi||^2$ as the probability that the particle has collided with the center of attraction by time t. There is no lower bound on the energy.

It is interesting to compare these results with the known results for the motion of a classical particle of mass m in the presence of the potential $-\beta/r^2$; i.e., in the presence of the central force $2\beta/r^3$. According to Whittaker[13] the orbits in polar coordinates are

$$\frac{1}{r} = A\cos(k\theta + \epsilon); \qquad k^2 = 1 - 2m\beta/J^2, \quad 2m\beta < J^2, \tag{36}$$

$$\frac{1}{r} = A\theta + \epsilon, \qquad 2m\beta = J^2, \tag{37}$$

$$\frac{1}{r} = A\cosh(k\theta + \epsilon); \qquad k^2 = 2m\beta/J^2 - 1, \quad 2m\beta > J^2. \tag{38}$$

To obtain a complete list we should adjoin the limiting case of (38):

$$\frac{1}{r} = A\sinh(k\theta + \epsilon); \qquad k^2 = 2m\beta/J^2 - 1, \quad 2m\beta > J^2, \tag{39}$$

and θ = constant for $J = 0$. Here A and ϵ are constants of integration which may be arbitrary real numbers except that $A \neq 0$ for (36), (38), and (39), and J is the classical angular momentum, $J = mr^2\dot{\theta}$, so that the time is given by

$$t = \frac{m}{J}\int r^2\, d\theta + \text{constant}.$$

[13] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, (Cambridge University Press, Cambridge, England, 1927), 3rd ed., p. 82.

The motion of a classical particle of mass m in the presence of the potential $-\beta/r^2$ may be summarized as follows:

Case I. $2m/\beta < J^2$. The motion is described by a group of transformations and the potential energy is bounded below.

Case II. $2m/\beta = J^2$. The motion is described by a group of transformations but the potential energy is not bounded below [unless $A = 0$ in (37)].

Case III. $2m\beta > J^2$. The particle collides with the center of attraction at a finite time. There is no lower bound on the potential energy.

This example suggests that, for highly attractive singular potentials, the solution of the Schrödinger equation obtained by Feynman integrals as developed here is physically relevant even though the operators may not be unitary. If the energy is not bounded below, the potential may produce collisions as well as scattering.

ACKNOWLEDGMENTS

It is a pleasure to thank Arthur S. Wightman and Hale F. Trotter for a number of helpful suggestions and valuable discussions.

APPENDIX A

This appendix is devoted to an exposition of the definition of Wiener integrals, a proof of the continuity of trajectories, and the connection with potential theory. It is hoped that this will be useful to the nonspecialist in probability theory who is interested in the mathematical basis of functional integration. The material is by its nature technical.

Wiener Integrals

We use the notation

$$p^t(x, dy) = (4\pi Dt)^{-l/2}\exp(-|x-y|^2/4Dt)\,dy, \tag{A1}$$

where x and y are in l-dimensional Euclidean space $\mathbf{R}^l$, $dy = dy_1 \cdots dy_l$, $|x|^2 = x_1^2 + \cdots + x_l^2$, and D is a strictly positive constant. This is interpreted as the probability that a particle, performing Brownian motion with diffusion constant D, which starts at x at time 0, will be in dy at time t. We let

$$P^t f(x) = \int p^t(x, dy) f(y).$$

We wish to construct a probability measure Pr_x on the space Ω of all trajectories. It is convenient to introduce the one-point compactification[14] $\dot{\mathbf{R}}^l$ of $\mathbf{R}^l$ by a point ∞, and let Ω be the Cartesian product

[14] John L. Kelley, *General Topology*, (D. Van Nostrand, Inc., New York, 1955).

space

$$\Omega = \prod_{0 \le t < \infty} \dot{\mathbf{R}}^l.$$

Thus an element ω of Ω is an entirely arbitrary function from time $[0, \infty)$ to space $\dot{\mathbf{R}}^l$. Let us give Ω the product topology, so that by the Tychonoff theorem[14] Ω is a compact Hausdorff space.

Now we can define the Wiener integral of a functional φ; i.e., of a function φ defined on Ω. Consider first a function φ of the simple form

$$\varphi(\omega) = F(\omega(t_1), \cdots, \omega(t_n)), \quad t_1 < \cdots < t_n. \tag{A2}$$

Let x be a point in $\mathbf{R}^l$ and let

$$Pr_x(\varphi) = \int \cdots \int p^{t_1}(x, dx_1) p^{t_2 - t_1}(x_1, dx_2) \cdots \times p^{t_n - t_{n-1}}(x_{n-1}, dx_n) F(x_1, \cdots, x_n), \tag{A3}$$

provided the integrals exist. For example, if $E_1, \cdots, E_n$ are Borel sets in $\mathbf{R}^l$ and $F(x_1, \cdots, x_n) = 1$ if $x_j \in E_j$ for $j = 1, \cdots, n$, $F(x_1, \cdots, x_n) = 0$ otherwise, then (A3) is

$$\int_{E_1} \cdots \int_{E_n} p^{t_1}(x, dx_1) \cdots p^{t_n - t_{n-1}}(x_{n-1}, dx_n),$$

and is interpreted as the probability that a particle starting at x at time 0 is in E_j at time t_j, for all $j = 1, \cdots, n$. Let $\mathcal{C}(\Omega)$ be the set of all continuous functions on Ω, $\mathcal{C}_{fin}(\Omega)$ the set of all those of the form (A2) where F is continuous. Any function in $\mathcal{C}(\Omega)$ may be uniformly approximated by functions in $\mathcal{C}_{fin}(\Omega)$, by the Stone–Weierstrass theorem.[14] The mapping $\varphi \to Pr_x(\varphi)$ is a linear functional on $\mathcal{C}_{fin}(\Omega)$, which is such that $Pr_x(1) = 1$ and $Pr_x(\varphi) \ge 0$ whenever $\varphi \ge 0$. Consequently, $|Pr_x(\varphi)| \le \sup_\omega |\varphi(\omega)|$, and so the mapping $\varphi \to Pr_x(\varphi)$ has a unique extension to a positive linear functional defined for all φ in $\mathcal{C}(\Omega)$. The Riesz representation theorem[6] asserts that there is a regular[6] probability measure, also denoted Pr_x, such that

$$Pr_x(\varphi) = \int_\Omega \varphi(\omega) Pr_x(d\omega)$$

for all φ in $\mathcal{C}(\Omega)$. This is Wiener measure. A consequence of regularity which we shall use is this: if Γ_S is an arbitrary (possibly uncountable) collection of open subsets of Ω, such that the union of finitely many of the Γ_S is always again in the collection, and if Γ denotes the union of all of the Γ_S, then

$$Pr_x(\Gamma) = \sup_S Pr_x(\Gamma_S). \tag{A4}$$

The positivity of the kernel $p^t(x, dy)$ is essential in the above proof of the existence of Wiener measure. There is no corresponding measure for complex values of D, despite the contrary assertion of Gelfand and Yaglom, as Cameron shows.[2]

Continuity of Trajectories

The elements of Ω were allowed to be arbitrary trajectories, even taking the value ∞. We turn now to a proof of the well-known, but nontrivial, result that the measure Pr_x is concentrated on the continuous trajectories taking values in $\mathbf{R}^l$. This is based on the fact that if $\epsilon > 0$ and

$$\rho(\epsilon, \delta) = \sup_{t \le \delta} \int_{|y-x| > \epsilon} p^t(x, dy)$$

(this is clearly independent of x), then by (A1),

$$\rho(\epsilon, \delta) = o(\delta). \tag{A5}$$

Lemma 1. Let $\epsilon > 0$, $\delta > 0$, $x \in \mathbf{R}^l$, and let $0 \le t_1 < \cdots < t_n$ with $t_n - t_1 \le \delta$. Let

$$A = \{\omega: |\omega(t_1) - \omega(t_j)| > \epsilon \text{ for some } j = 1, \cdots, n\}.$$

Then $Pr_x(A) \le 2\rho(\frac{1}{2}\epsilon, \delta)$.

The crucial aspect of this estimate is that it is independent of the number n of times considered. To prove the lemma, let

$$B = \{\omega: |\omega(t_1) - \omega(t_n)| > \tfrac{1}{2}\epsilon\},$$

$$C_j = \{\omega: |\omega(t_j) - \omega(t_n)| > \tfrac{1}{2}\epsilon\},$$

$$D_j = \{\omega: |\omega(t_1) - \omega(t_j)| > \epsilon, \text{ and } |\omega(t_1) - \omega(t_k)| \le \epsilon \text{ for } k = 1, \cdots, j-1\}.$$

Then $A \subset B \cup \bigcup_{j=1}^n (C_j \cap D_j)$, for if ω is in A, there is a least j such that $|\omega(t_1) - \omega(t_j)| > \epsilon$, so that ω is in D_j. If now ω is not in B, then ω must be in C_j, since $|\omega(t_1) - \omega(t_n)| \le \frac{1}{2}\epsilon$ and $|\omega(t_1) - \omega(t_j)| > \epsilon$, so that $|\omega(t_j) - \omega(t_n)| > \frac{1}{2}\epsilon$. Therefore,

$$Pr_x(A) \le Pr_x(B) + \sum_{j=1}^n Pr_x(C_j \cap D_j). \tag{A6}$$

Let $D(\omega(t_1), \cdots, \omega(t_j)) = 1$ if $\omega \in D_j$, 0 otherwise, and let $C(\omega(t_j), \omega(t_n)) = 1$ if $\omega \in C_j$, 0 otherwise. Then by (A3),

$$Pr_x(C_j \cap D_j) = \int \cdots \int p^{t_1}(x, dx_1) \cdots \times p^{t_j - t_{j-1}}(x_{j-1}, dx_j) p^{t_n - t_j}(x_j, dx_n)\, D(x_1, \cdots, x_j) \times C(x_j, x_n) \le \rho(\epsilon, \delta) \int \cdots \int p^{t_1}(x, dx_1) \cdots \times p^{t_j - t_{j-1}}(x_{j-1}, dx_j)\, D(x_1, \cdots, x_j) = \rho(\epsilon, \delta) Pr_x(D_j).$$

Since the D_j are disjoint,

$$\sum_{j=1}^{n} Pr_x(C_j \cap D_j) \leq \sum_{j=1}^{n} \rho(\epsilon, \delta) Pr_x(D_j) \leq \rho(\epsilon, \delta),$$

and since $Pr_x(B) \leq \rho(\epsilon, \delta)$, the lemma follows by (A6).

The intuitive idea back of this proof is that the particle has no memory. Given that the particle is at x_j at time t_j, the past (that ω lies in D_j) has no influence on the future (whether ω lies in C_j). This lack of memory, called the Markoff property, is built into Wiener measure via the definition (A3).

Lemma 2. With the notation of Lemma 1, let $E = \{\omega: |\omega(t_j) - \omega(t_k)| > 2\epsilon$ *for some* j *and* k, $1 \leq j, k \leq n\}$. *Then* $Pr_x(E) \leq 2\rho(\frac{1}{2}\epsilon, \delta)$.

This follows at once from Lemma 1, since $E \subset A$, for if $|\omega(t_j) - \omega(t_k)| > 2\epsilon$, then $|\omega(t_1) - \omega(t_j)| > \epsilon$ or $|\omega(t_1) - \omega(t_k)| > \epsilon$.

The rest of the proof of the continuity of trajectories is merely technical, for Lemma 2 says that the probability of a 2ϵ variation in a given interval of length δ is $o(\delta)$, and hence the probability of a 2ϵ variation in some interval of length δ is $o(1)$, and since this is so for all $\epsilon > 0$, the trajectories must be continuous with probability 1.

Lemma 3. Let $0 \leq a < b$ *with* $b - a \leq \delta$. *Let* $E(a, b, \epsilon) = \{\omega: |\omega(t) - \omega(s)| > 2\epsilon$ *for some* t *and* s *in* $[a, b]\}$. *Then* $Pr_x(E(a, b, \epsilon)) \leq 2\rho(\frac{1}{2}\epsilon, \delta)$.

To prove this, let S denote a finite subset of $[a, b]$ and let $E(a, b, \epsilon, S) = \{\omega: |\omega(t) - \omega(s)| > 2\epsilon$ for some t and s in $S\}$. Then the $E(a, b, \epsilon, S)$ are a collection of open subsets of Ω whose union is $E(a, b, \epsilon)$, so that by (A4), $Pr_x(E(a, b, \epsilon)) = \sup_S Pr_x(E(A, b, \epsilon, S))$. By Lemma 2, for each S, $Pr_x(E(a, b, \epsilon, S)) \leq 2\rho(\frac{1}{2}\epsilon, \delta)$, concluding the proof.

Lemma 4. Let k *be a positive integer,* $\epsilon > 0$, $\delta > 0$, *and assume that* $1/\delta$ *is an integer. Let* $F(k, \epsilon, \delta) = \{\omega: |\omega(t) - \omega(s)| > 4\epsilon$ *for some* t *and* s *in* $[0, k]$ *with* $|t - s| < \delta\}$. *Then* $Pr_x(F(k, \epsilon, \delta)) \leq 2k\rho(\epsilon, \delta)/\delta$.

The interval $[0, k]$ is the union of the k/δ subintervals $[0, \delta]$, $[\delta, 2\delta]$, $\cdots$, $[k - \delta, k]$. If ω is in $F(k, \epsilon, \delta)$ then $|\omega(t) - \omega(s)| > 4\epsilon$ for some t and s lying in the same or adjacent subintervals, and so, in the notation of Lemma 3, ω is in $E(a, b, \epsilon)$ for $[a, b]$ one of the subintervals. Hence Lemma 4 follows from Lemma 3.

Theorem 4. (Wiener[15]*) Let* Φ *be the set of all* ω *which are continuous and such that* $\omega(t)$ *lies in* $\mathbf{R}^l$ *for* $0 \leq t < \infty$, *and let* x *be in* $\mathbf{R}^l$. *Then* $Pr_x(\Phi) = 1$.

[15] N. Wiener, J. Math. and Phys. **2**, 131 (1923).

Observe that

$$\Phi = \bigcap_{k=1}^{\infty} \bigcap_{\epsilon>0} \bigcup_{\delta>0} F'(k, \epsilon, \delta), \tag{A7}$$

where $F'(k, \epsilon, \delta)$ is the complement of $F(k, \epsilon, \delta)$ of Lemma 4. This merely expresses the fact that ω is continuous from $[0, \infty)$ to $\mathbf{R}^l$ if and only if it is uniformly continuous on each interval $[0, k]$. The ϵ and δ in (A7) may be restricted to be reciprocals of integers. Since each $F'(k, \epsilon, \delta)$ is a closed set, Φ is a Borel set, and $Pr_x(\Phi)$ is meaningful. To show that $Pr_x(\Phi) = 1$, we need only show for each k and ϵ that $\lim_{\delta \to 0} Pr_x(F(k, \epsilon, \delta)) = 0$. But this is true by Lemma 4 and (A5).

Potential Theory

Next we develop briefly the connection between Brownian motion and potential theory.[16]

Lemma 5. Let f *be a positive measurable function on* $\mathbf{R}^l$ *such that for all* $t \geq 0$, $P^t f \leq f$. *Let* $t_1 < \cdots < t_n$ *and let* $\lambda > 0$. *Then for all* x *in* $\mathbf{R}^l$,

$$Pr_x(\{\omega: \max_{j=1,\cdots,n} f(\omega(t_j)) \geq \lambda\}) \leq f(x)/\lambda. \tag{A8}$$

Again, the point is that the estimate is independent of n. We may think of $f(\omega(t))$ as the fortune of a gambler playing a game which is unfavorable to him since $P^t f \leq f$. From this point of view (which is Doob's[16]), (A8) is quite intuitive.

To prove Lemma 5, let Φ be the set occurring in (A8), and for all ω in Φ, let $m(\omega)$ be the smallest value of j such that $f(\omega(t_j)) \geq \lambda$. Then

$$\lambda Pr_x(\Phi) \leq \int_{\Phi} f(\omega(t_{m(\omega)})) Pr_x(d\omega). \tag{A9}$$

Let χ be the operator of multiplication by the function which is 1 where $f(y) \geq \lambda$ and 0 elsewhere. Then by the definition of Wiener integrals, the right-hand side of (A9) is the value at x of

$$P^{t_1}\chi f + P^{t_1}(1 - \chi)P^{t_2 - t_1}\chi f + \cdots$$
$$+ P^{t_1}(1 - \chi) \cdots P^{t_{n-1} - t_{n-2}}(1 - \chi)P^{t_n - t_{n-1}}\chi f. \tag{A10}$$

But (A10) telescopes and is $\leq f$. To see this, suppose that $n = 3$. Then, since $P^{t_3 - t_2}\chi f \leq P^{t_3 - t_2} f \leq f$,

$$P^{t_1}\chi f + P^{t_1}(1 - \chi)P^{t_2 - t_1}\chi f$$
$$+ P^{t_1}(1 - \chi)P^{t_2 - t_1}(1 - \chi)P^{t_3 - t_2}\chi f$$
$$\leq P^{t_1}\chi f + P^{t_1}(1 - \chi)P^{t_2 - t_1} f \leq P^{t_1} f \leq f.$$

Therefore, by (A9), $\lambda Pr_x(\Phi) \leq f(x)$, concluding the proof.

[16] J. L. Doob, Trans. Am. Math. Soc. **77**, 86 (1954).

Lemma 6. Let f be a lower semicontinuous function from $\mathbf{R}^l$ to $[0, \infty]$ (the value ∞ is allowed) such that $P^t f \leq f$ for all $t \geq 0$. Then for all $\lambda > 0$ and x in $\mathbf{R}^l$,

$$Pr_x(\{\omega: \sup_{0 \leq t < \infty} f(\omega(t)) > \lambda\}) \leq f(x)/\lambda.$$

Lower semicontinuity means that $\{y: f(y) > \lambda\}$ is open. Lemma 6 follows immediately from (A4) and Lemma 5.

Theorem 5. (Doob[16]) Let F be a closed subset of $\mathbf{R}^l$ of capacity 0 and let x be a point which is not in F. Then

$$Pr_x(\{\omega: \omega(t) \notin F \quad \text{for} \quad 0 \leq t < \infty\}) = 1.$$

There is no loss of generality in assuming that F is compact, since any closed set in $\mathbf{R}^l$ is a countable union of compact subsets. If $l \geq 3$, a theorem of Cartan[17] asserts that, given a compact set F in $\mathbf{R}^l$ of capacity 0, there is a positive superharmonic function f, equal to ∞ on F and finite on the complement of F, and continuous from $\mathbf{R}^l$ to $[0, \infty]$. Since f is superharmonic, $f(x)$ is greater than the average of f on any sphere with x as center, and so by (A1), $P^t f \leq f$ for all $t \geq 0$. By Lemma 6 the probability that $\omega(t) \in F$ for some t is $\leq f(x)/\lambda$ for all $\lambda > 0$, and so is 0, concluding the proof in the case $l \geq 3$. It is not difficult to modify the proof for the case[16] $l = 2$. If $l = 1$ there are no nonempty sets of capacity 0.

We remark that the assumption that F is closed is not necessary. The existence of the function f which Cartan's theory gives is in practice as easy to verify as the usual definition of capacity 0 (which we have not stated). For example, if $l \geq 3$ and $F = \{0\}$ consists of the origin alone (as in Sec. 5), then $f(x) = 1/r^{l-2}$, where $r = |x|$, has the required properties.

APPENDIX B

This appendix develops the perturbation theory[3] which is due to Kato and to Trotter. Although the principal ideas are those of these authors, some of the results are new and there is some simplification of the proofs. We begin by recasting the basic theorem on the generation of contraction semigroups, the Hille–Yosida theorem,[18] in a form which is more convenient for our purposes.

Dissipative Operators

By a *contraction semigroup* on a Banach space $\mathfrak{X}$ we shall mean a family of bounded everywhere-defined linear operators P^t, $0 \leq t < \infty$, mapping $\mathfrak{X}$ into itself such that

$$P^0 = 1,$$

$$P^t P^s = P^{t+s}, \qquad 0 \leq t, s < \infty,$$

$$||P^t|| \leq 1, \qquad 0 \leq t < \infty,$$

$$\lim_{t \to 0} P^t \psi = \psi, \qquad \psi \in \mathfrak{X}.$$

The *infinitesimal generator* A of a contraction semigroup P^t is defined by

$$A\psi = \lim_{t \to 0} \frac{1}{t} (P^t \psi - \psi) \tag{B1}$$

on the domain $\mathfrak{D}(A)$ of all ψ in $\mathfrak{X}$ for which the limit exists. The well-known theorem of Yosida[18] asserts that A is the infinitesimal generator of a contraction semigroup if and only if A is densely defined and there is a real λ_0 such that for all $\lambda > \lambda_0$, λ is in the resolvent set[7] of A and $||(\lambda - A)^{-1}|| \leq 1/\lambda$. Note that since A has a nonempty resolvent set, A is a closed operator. Also, $(\lambda - A)^{-1}\psi = \int_0^\infty e^{-\lambda t} P^t \psi \, dt$ for all ψ in $\mathfrak{X}$ whenever $\mathrm{Re}\ \lambda > 0$, from which it follows that $||(\lambda - A)^{-1}|| \leq 1/\mathrm{Re}\ \lambda$ whenever $\mathrm{Re}\ \lambda > 0$.

We shall call an operator A on a Banach space $\mathfrak{X}$ *dissipative* in case for all ψ in $\mathfrak{D}(A)$, if φ is in the space $\mathfrak{X}^*$ of continuous linear functionals on $\mathfrak{X}$, $||\varphi|| = 1$ and $(\psi, \varphi) = ||\psi||$ then $\mathrm{Re}\ (A\psi, \varphi) \leq 0$. Geometrically such a φ represents a supporting hyperplane at the point ψ on the sphere of radius $||\psi||$ in $\mathfrak{X}$ (or more precisely $\mathrm{Re}\ \varphi$ does on $\mathfrak{X}$ regarded as a real Banach space) and the condition that A be dissipative is that $\psi + tA\psi$ lie on the same side of this hyperplane as the origin, for $t \geq 0$. If $\mathfrak{X}$ is a Hilbert space, we must have $\varphi = \psi/||\psi||$ for $\psi \neq 0$, and so an operator A on a Hilbert space is dissipative if and only if $\mathrm{Re}\ (A\psi, \psi) \leq 0$ for all ψ on $\mathfrak{D}(A)$. Thus our terminology agrees with that of Phillips[19] for Hilbert spaces.

If A is a dissipative operator on a Banach space $\mathfrak{X}$, and $\psi \in \mathfrak{X}$, there always exists a φ in $\mathfrak{X}^*$ such that $||\varphi|| = 1$ and $(\psi, \varphi) = ||\psi||$, by the Hahn–Banach theorem.[7] (The linear functional φ is obtained by extending the linear functional $a\psi \to a\,||\psi||$ defined on the one-dimensional subspace of $\mathfrak{X}$ containing ψ.) Since $||\varphi|| = 1$, $||(\lambda - A)\psi|| \geq \mathrm{Re}\ ((\lambda - A)\psi, \varphi) = \lambda\,||\psi|| - \mathrm{Re}\ (A\psi, \varphi) \geq \lambda\,||\psi||$ for all real λ. That is, if A is dissipative,

[17] H. Cartan, Bull. Soc. Math. France **73**, 74 (1945).

[18] We shall use the form of the theorem due to K. Yosida, J. Math. Soc. Japan **1**, 15 (1948).

[19] R. S. Phillips, Trans. Am. Math. Soc. **90**, 193 (1959). This notion of a dissipative operator is very closely related to the dissipative operators in a Banach space of G. Lumer and R. S. Phillips, Pacific J. Math. **11**, 679 (1961).

$$||(\lambda - A)\psi|| \geq \lambda\, ||\psi|| \qquad \text{(B2)}$$

whenever $\psi \in \mathfrak{D}(A)$ and λ is real.

Theorem 6. A densely defined linear operator A on a Banach space $\mathfrak{X}$ is the infinitesimal generator of a contraction semigroup if and only if it is dissipative and there is a real λ_0 such that λ is in the resolvent set of A whenever $\lambda > \lambda_0$.

To prove this, let A be the infinitesimal generator of a contraction semigroup P^t, and let ψ be in $\mathfrak{D}(A)$. Then λ is in the resolvent set of A whenever $\lambda > 0$. Let ψ be in $\mathfrak{D}(A)$ and suppose that φ is in $\mathfrak{X}^*$, $||\varphi|| = 1$, and $(\psi, \varphi) = ||\psi||$. Then $|(P^t\psi, \varphi)| \leq ||\psi|| = (\psi, \varphi)$ and so $t^{-1}\,\mathrm{Re}\,(P^t\psi - \psi, \varphi) \leq 0$. By (B1), A is dissipative.

Conversely, suppose that A is densely defined, dissipative, and that λ is in the resolvent set of A whenever $\lambda > \lambda_0$. By Yosida's theorem, we need only show that $||(\lambda - A)^{-1}|| \leq 1/\lambda$ for $\lambda > \lambda_0$, but this is true by (B2).

The advantage of this version of Yosida's theorem is that the set of dissipative operators clearly forms a convex cone, whereas this is not true of the set of all operators satisfying (B2).

Perturbation of Dissipative Operators

The following theorem is a slight extension of a result of Trotter.[3]

Theorem 7. Let A be the infinitesimal generator of a contraction semigroup on the Banach space $\mathfrak{X}$ and let B be a dissipative operator with $\mathfrak{D}(B) \supset \mathfrak{D}(A)$. If there exist constants a and b with $a < \frac{1}{2}$ such that for all ψ in $\mathfrak{D}(A)$

$$||B\psi|| \leq a\, ||A\psi|| + b\, ||\psi||, \qquad \text{(B3)}$$

then $A + B$ is the infinitesimal generator of a contraction semigroup.

The operator A is dissipative by Theorem 6, and B is dissipative by assumption, so that $A + B$ is dissipative. By Theorem 6 we need only show that there is a λ_0 such that λ is in the resolvent set of $A + B$ whenever $\lambda > \lambda_0$. Let $\lambda_0 = b/(1 - 2a)$, so that if $\lambda > \lambda_0$ then $2a + b/\lambda < 1$. Since $\mathfrak{D}(B) \supset \mathfrak{D}(A)$, $B(\lambda - A)^{-1}$ is everywhere-defined and $||B(\lambda - A)^{-1}|| \leq a\, ||A(\lambda - A)^{-1}|| + b\, ||(\lambda - A)^{-1}||$ by (B3). Now $||(\lambda - A)^{-1}|| \leq 1/\lambda$ and $A(\lambda - A) = \lambda(\lambda - A)^{-1} - 1$, so that $||A(\lambda - A)^{-1}|| \leq 2$. Therefore, $||B(\lambda - A)^{-1}|| \leq 2a + b/\lambda < 1$ for $\lambda > \lambda_0$. Consequently, the range of $1 - B(\lambda - A)^{-1}$ is $\mathfrak{X}$ for $\lambda > \lambda_0$. But, using $\mathfrak{R}$ to denote the range of an operator, $\mathfrak{R}(\lambda - A - B) = \mathfrak{R}((\lambda - A - B)(\lambda - A)^{-1}) = \mathfrak{R}(1 - B(\lambda - A)^{-1}) = \mathfrak{X}$ for $\lambda > \lambda_0$. Since $A + B$ is dissipative, (B2) holds for it, and so $(\lambda - A - B)^{-1}$ exists and is a bounded everywhere-defined operator, concluding the proof.

Kato Potentials

It is well-known and easy to see that, on a Hilbert space, A is self-adjoint if and only if iA and $-iA$ are infinitesimal generators of contraction semigroups [which must be $U(t)$ and $U(-t)$ respectively, for $t \geq 0$, where $U(t)$ for $-\infty < t < \infty$ is a strongly continuous one-parameter group of unitary operators]. In this case, $||A(\lambda - A)|| \leq 1$ by the spectral theorem, so that we need only require $a < 1$ in Theorem 7. Therefore, Theorem 7 has the following corollary, due to Kato.[3]

Corollary. Let A be a self-adjoint operator on the Hilbert space $\mathfrak{K}$, and let B be a symmetric operator with $\mathfrak{D}(B) \supset \mathfrak{D}(A)$. If there exist constants a and b with $a < 1$ such that (B3) *holds for all ψ in $\mathfrak{D}(A)$, then $A + B$ is self-adjoint.*

Following Kato, we apply this to show the self-adjointness of the Hamiltonian $-\Delta + V$ for a certain class of potentials V. The Laplace operator Δ and the multiplication operator V are understood to be defined as in Sec. 1. We call V a *Kato potential* in case $\mathfrak{D}(V) \supset \mathfrak{D}(\Delta)$ and, for all $a > 0$, there is a b such that (B3) holds for all ψ in $\mathfrak{D}(\Delta)$. Thus $-\Delta + V$ is self-adjoint if V is a Kato potential. The advantage of requiring (B3) to hold for all $a > 0$ (for some b depending on a) is that this makes the set of Kato potentials a vector space. Furthermore, if A is any self-adjoint operator with $\mathfrak{D}(A) \subset \mathfrak{D}(\Delta)$, there is a constant c such that $||\Delta\psi|| \leq c\, ||A\psi|| + c\, ||\psi||$ for all ψ in $\mathfrak{D}(A)$ (by the closed-graph theorem[7] in principle, and by inspection in all cases which arise in practice), so that $A + V$ is self-adjoint if V is a Kato potential.

The Soboleff inequalities show that V is a Kato potential in case V is in $\mathfrak{L}^p(\mathbf{R}^l)$ whenever $p > \frac{1}{2}l$ and $p \geq 2$. This may be proved by using Fourier analysis as follows. If ψ in $\mathfrak{L}^2(\mathbf{R}^l)$ is in $\mathfrak{D}(\Delta)$, then $(r^2 + 1)\mathfrak{F}\psi \in \mathfrak{L}^2(\mathbf{R}^l)$. But $(r^2 + 1)^{-1} \in \mathfrak{L}^p(\mathbf{R}^l)$ if $p > \frac{1}{2}l$, so that by the Hölder inequality, if $p > \frac{1}{2}l$ and $p \geq 2$, then $\mathfrak{F}\psi$ is in $\mathfrak{L}^q(\mathbf{R}^l)$, where $1/q = 1/p + \frac{1}{2}$. By the Hausdorff–Young theorem,[20] this implies that ψ is in $\mathfrak{L}^s(\mathbf{R}^l)$, where $1/s = 1 - 1/q = \frac{1}{2} - 1/p$. By the Hölder inequality, if V is in $\mathfrak{L}^p(\mathbf{R}^l)$ for $p > \frac{1}{2}l$ and $p \geq 2$, $V\psi$ is in $\mathfrak{L}^2(\mathbf{R}^l)$. Therefore (B3) holds for some values of a and b.

[20] A. Zygmund, *Trigonometric Series* (Cambridge University Press, Cambridge, England, 1959), 2nd ed., Vol. II, p. 254.

A dimensional analysis [consider the effect of the transformation $\psi(x) \to \psi(kx)$ for $k > 0$] shows that a may be made arbitrarily small by taking b sufficiently large.

For applications to problems involving several particles, notice that if, after an inhomogeneous linear change of variables, V depends on only j of the coordinates, say $V(x_1, \cdots, x_l) = W(x_1, \cdots, x_j)$, and W is in $\mathfrak{L}^p(R^j)$ for $p > \frac{1}{2}j$ and $p \geq 2$, then V is a Kato potential.

The following theorem summarizes this discussion of Kato's results.

Theorem 8. Let V be a real function on $\mathbf{R}^l$ which is a finite sum of functions in $\mathfrak{L}^p(\mathbf{R}^l)$ for values of p satisfying $p > \frac{1}{2}l$ and $p \geq 2$, and functions which after an inhomogeneous linear change of variables are of the form $W(x_1, \cdots, x_j)$ where W is in $\mathfrak{L}^p(\mathbf{R}^j)$ for values of p satisfying $p > \frac{1}{2}j$ and $p \geq 2$. Then V is a Kato potential.

If V is a Kato potential and A is a self-adjoint operator on $\mathfrak{L}^2(\mathbf{R}^l)$ with $\mathfrak{D}(A) \subset \mathfrak{D}(\Delta)$, then $A + V$ is self-adjoint.

An example is the Hamiltonian of a finite number of Schrödinger particles with Coulomb interactions.

Trotter's Product Formula

Theorem 9. Let A and B be linear operators on the Banach space $\mathfrak{X}$ such that A, B, and $A + B$ are the infinitesimal generators of the contraction semigroups P^t, Q^t, and R^t, respectively. Then for all ψ in $\mathfrak{X}$,

$$R^t\psi = \lim_{n\to\infty} (P^{t/n}Q^{t/n})^n\psi. \tag{B4}$$

Notice that Theorems 7 and 8 give conditions under which the hypotheses of this theorem are satisfied. Theorem 9 is a special case of Trotter's result, since Trotter proves that (B4) holds if it is merely assumed that the closure of $A + B$ is the infinitesimal generator of R^t. The following proof of the special case is considerably shorter than the proof of the stronger result, which uses the full force of Trotter's 1958 paper.[3]

It is sufficient to prove (B4) for a dense set of ψ's, since all of the operators in question have norm ≤ 1. Let ψ be in $\mathfrak{D}(A+B)$. Since $(P^hQ^h-1)\psi = (P^h - 1)\psi + P^h(Q^h - 1)\psi$, we have $(P^hQ^h - 1)\psi = h(A + B)\psi + o(h)$, where $o(h)$ denotes vectors in $\mathfrak{X}$ such that $\lim_{h\to 0} ||o(h)||/h = 0$. Also, $(R^h-1)\psi = h(A + B)\psi + o(h)$. Therefore, $(P^hQ^h - R^h)\psi = o(h)$ for all ψ in $\mathfrak{D}(A + B)$, since $(P^hQ^h - R^h) = (P^hQ^h - 1) - (R^h - 1)$.

Since $A+B$ is a closed operator, $\mathfrak{D}(A+B)$ is a Banach space in the norm $||\psi||_{A+B} = ||(A + B)\psi|| + ||\psi||$. For each ψ in $\mathfrak{D}(A + B)$, $h^{-1}(P^hQ^h - R^h)\varphi$ is a bounded set in $\mathfrak{X}$. By the principle of uniform boundedness,[7] there is a constant c such that $||h^{-1}(P^hQ^h - R^h)\psi|| \leq c\,||\psi||_{A+B}$ for all $h > 0$ and ψ in $\mathfrak{D}(A+B)$. That is, the operators $h^{-1}(P^hQ^h-R^h)$, $h > 0$, are a uniformly bounded set of operators from $\mathfrak{D}(A + B)$ to $\mathfrak{X}$ which tend strongly to 0 as $h \to 0$. Consequently, $||h^{-1}(P^hQ^h - R^h)\psi|| \to 0$ uniformly on any compact set in $\mathfrak{D}(A + B)$. Let $t \geq 0$. Then, since $A + B$ is the infinitesimal generator of R^t, if ψ is any element of $\mathfrak{D}(A + B)$ the set of all $R^s\psi$, for $0 \leq s \leq t$, is a compact set in $\mathfrak{D}(A + B)$. Therefore,

$$||(P^hQ^h - R^h)R^s\psi|| = o(h), \tag{B5}$$

uniformly for $0 \leq s \leq t$. Set $h = t/n$. We need to show that $||((P^hQ^h)^n - R^{hn})\psi|| \to 0$ as $n \to \infty$. But

$$\begin{aligned}(P^hQ^h)^n - R^{hn} = {} & (P^hQ^h - R^h)R^{h(n-1)} \\ & + P^hQ^h(P^hQ^h - R^h)R^{h(n-2)} + \cdots \\ & + (P^hQ^h)^{(n-1)}(P^hQ^h - R^h),\end{aligned}$$

and since $||P^hQ^h|| \leq 1$,

$$\begin{aligned}||((P^hQ^h)^n - R^{hn})\psi|| \leq {} & ||(P^hQ^h - R^h)R^{h(n-1)}\psi|| \\ & + ||(P^hQ^h - R^h)R^{h(n-2)}\psi|| + \cdots \\ & + ||(P^hQ^h - R^h)\psi||.\end{aligned}$$

This is the sum of n terms which are uniformly $o(t/n)$, by (B5), and the proof is complete.

An Application of Sommerfeld's Complex-Order Wavefunctions to an Antenna Problem*

C. YEH
Electrical Engineering Department, University of Southern California, Los Angeles, California
(Received 27 February 1963; and in revised form 1 November 1963)

Using the orthogonality relations of Sommerfeld's complex-order wavefunctions, the exact solution for the problem of electromagnetic radiation from a circularly symmetric slot on the conducting surface of a dielectric-coated cone is obtained. The results are valid for the near-zone region as well as for the far-zone region, and they are applicable for arbitrary-angle cones. It is noted that the technique used to solve this problem may be applied to similar types of problems involving conical structure, such as the diffraction of waves by a dielectric-coated, spherically tipped cone.

I. INTRODUCTION

THE problems of scattering of waves by a perfectly conducting conical obstacle or radiation from such a structure have been considered by many authors.[1-5] The exact mathematical solution to the problem of the diffraction of waves by a finite, perfectly conducting cone has recently been obtained by Northover.[6] However, the corresponding solution for the diffraction by or radiation from a dielectric-coated, semi-infinite conical structure has not been found. It is the purpose of this paper to present the exact solution of the radiation from this dielectric-coated structure. It is shown that certain mathematical difficulties can be overcome by the use of Sommerfeld's complex-order wavefunctions[7] and their orthogonality properties.

II. FORMULATION OF THE PROBLEM

To analyze this problem, the spherical coordinates (r, θ, ϕ) are used. The geometry of this conical structure is shown in Fig. 1. The vertex of the cone is taken to coincide with the origin of the spherical polar coordinates. To eliminate the singularity at the vertex, a small perfectly conducting spherical boss of radius a, with its center at the origin, is situated at the tip of the cone. The outer boundary of the dielectric-coated cone is assumed to coincide with $\theta = \theta_1$; the inner boundary is assumed to coincide with $\theta = \theta_0$. The dielectric coating has a permittivity of ϵ_1, a permeability μ, and a conductivity of zero. It is assumed that this radiating structure is embedded in a homogeneous, perfect dielectric medium (ϵ_0, μ; $\sigma_0 = 0$), and that the applied electric-field intensity across the slot, which is located on the perfectly conducting conical surface $\theta = \theta_0$, is circularly symmetric about the axis of the cone and linearly polarized in the radial direction.

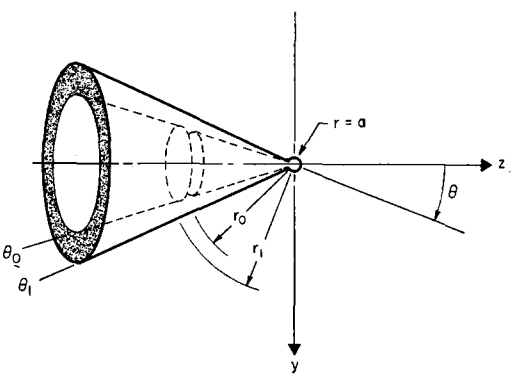

FIG. 1 The dielectric coated spherically tipped cone.

Due to the symmetrical characteristics of this problem, all components of the electromagnetic field are independent of the azimuthal angle ϕ. For a TM wave, the nonvanishing components are E_r, E_θ, and H_ϕ. The wave equation in spherical coordinates takes the form

$$\frac{\partial^2}{\partial r^2}(rH_\phi) + \frac{1}{r}\frac{\partial}{\partial \theta}\left[\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(H_\phi \sin\theta)\right] + k^2 r H_\phi = 0, \quad (1)$$

where $k^2 = \omega^2\mu\epsilon$ and the steady-state time depend-

* This work was supported by the Air Force Cambridge Research Laboratories.

[1] H. S. Carslaw, Math Ann. **75**, 133 (1914).
[2] W. W. Hansen and L. I. Schiff, "Theoretical Study of Electromagnetic Waves Scattered from Shaped Metal Surfaces," Quarterly Progress Rept. No. 4, Stanford University Microwave Laboratory, September, 1948.
[3] L. B. Felsen, J. Appl. Phys. **26**, 138 (1955).
[4] K. M. Siegel, R. F. Goodrich, and V. H. Weston, Appl. Sci. Res. **B8**, 8 (1959).
[5] J. B. Keller, "Back Scattering from a Finite Cone," New York University, Courant Institute of Mathematical Sciences, Division of Electromagnetic Research, Res. Rept. No. EM-127, (1959).
[6] F. H. Northover, Quart. J. Mech. Appl. Math. **15**, 1 (1962).
[7] A. Sommerfeld, *Partial Differential Equations* (Academic Press Inc., New York, 1949).

ence $e^{-i\omega t}$ has been assumed. Setting

$$H_\phi = i\omega\epsilon(\partial u/\partial\theta) \tag{2}$$

in Eq. (1) gives

$$(\nabla^2 + k^2)u(r, \theta) = 0. \tag{3}$$

A possible solution of Eq. (3) is then $u(r, \theta) = R(r)\Theta(\theta)$, where R and Θ satisfy the differential equations

$$d/dr\,[r^2(dR/dr)] + (k^2r^2 - c)R = 0, \tag{4}$$

$$d/d\theta\,[\sin\theta(d\Theta/d\theta)] + c\sin\theta\Theta = 0, \tag{5}$$

in which c is the separation constant. If one chooses $c = \nu(\nu + 1)$ where ν may be a complex number, the solutions of Eq. (5) are the Legendre functions

$$\begin{Bmatrix} P_\nu(\cos\theta) \\ Q_\nu(\cos\theta) \end{Bmatrix}.$$

The corresponding solutions of Eq. (4) are the spherical Hankel functions

$$\begin{Bmatrix} h_\nu^{(1)}(kr) \\ h_\nu^{(2)}(kr) \end{Bmatrix}.$$

The proper choice of these functions to represent the electromagnetic fields depends upon the boundary conditions. All field components must be finite in all regions (i.e., the region within the dielectric sheath and the region outside the sheath). In addition, all field components for the radiated wave must satisfy Sommerfeld's radiation condition at infinity. Consequently, the appropriate solution for the region inside the dielectric sheath is

$$u_s(r, \theta) = \sum_\nu h_\nu^{(1)}(k_1r) \times [A_\nu P_\nu(\cos\theta) + B_\nu Q_\nu(\cos\theta)], \tag{6}$$

and that for the radiated wave,

$$u_r(r, \theta) = \sum_{\nu'} G_{\nu'} h_{\nu'}^{(1)}(k_0r) P_{\nu'}(\cos\theta), \tag{7}$$

where $k_1^2 = \omega^2\mu\epsilon_1$ and $k_0^2 = \omega^2\mu\epsilon_0$. A_ν, B_ν, and $G_{\nu'}$ are arbitrary constants to be determined by the boundary conditions. The summation is over all values of ν which are determined by the boundary condition on the spherical boss at the tip of the cone; i.e., the tangential electric field must vanish on the perfectly conducting spherical boss:

$$\partial/\partial r\,[rh_\nu^{(1)}(k_1r)]|_{r=a} = 0, \tag{8}$$

$$\partial/\partial r\,[rh_{\nu'}^{(1)}(k_0r)]|_{r=a} = 0. \tag{9}$$

The roots of ν from Eqs. (8) and (9) will be designated, respectively, by ν_n and ν_m'. It should be noted that $h_\nu^{(1)}(k_1r)$ or $h_{\nu'}^{(1)}(k_0r)$ are orthogonal over the range k_1a to ∞ or k_0a to ∞, respectively. (The proof is given in Appendix A.) It is because of this orthogonality property of these radial functions that they are so useful for the conical problems. The orthogonality characteristic of the Hankel functions with complex order was first investigated by Sommerfeld.[7] Hence these functions are also called Sommerfeld's complex-order wavefunctions.[8]

III. THE MATHEMATICAL SOLUTION

The boundary conditions require the continuity of the tangential electric and magnetic fields at the boundary surface $\theta = \theta_1$. On the conducting surface $\theta = \theta_0$, the tangential electric field must be zero everywhere except across the gap where it is equal to the applied field. Let the applied field be defined by

$$E_r^{\text{app}} = E_0 d(r) e^{-i\omega t}, \tag{10}$$

where $d(r)$ is defined as follows:

$$d(r) = \begin{cases} 1 & \text{for } r_0 < r < r_1, \\ 0 & \text{for } a < r < r_0, \text{ and } r > r_1. \end{cases} \tag{11}$$

$|r_1 - r_0|$ is the gap width (see Fig. 1). Expanding the applied field in terms of Sommerfeld's complex-order wavefunctions, gives

$$E_r^{\text{app}} = \frac{1}{r}\sum_{\nu_n} L_{\nu_n}\nu_n(\nu_n + 1)h_{\nu_n}^{(1)}(k_1r)P_{\nu_n}(\cos\theta_0) \times e^{-i\omega t}, \tag{12}$$

where

$$L_{\nu_n} = \frac{1}{\nu_n(\nu_n + 1)P_{\nu_n}(\cos\theta_0)N_{\nu_n}(k_1a)} \times \int_{r_0}^{r_1} E_0 r h_{\nu_n}^{(1)}(k_1r)\,d(k_1r), \tag{13}$$

in which the normalizing factor is

$$N_{\nu_n}(k_1a) = \int_{k_1a}^{\infty} [h_{\nu_n}^{(1)}(k_1r)]^2\,d(k_1r). \tag{14}$$

Expression (12) represents E_r^{app} for all values of r between a and ∞.

Upon matching the tangential magnetic and electric fields at $\theta = \theta_1$, one obtains

$$i\omega\epsilon_0 \sum_{\nu_m'} G_{\nu_m'} h_{\nu_m'}^{(1)}(k_0r)\left[\frac{d}{d\theta_1}P_{\nu_m'}(\cos\theta_1)\right] = i\omega\epsilon_1 \sum_{\nu_n} h_{\nu_n}^{(1)}(k_1r)\left[A_{\nu_n}\frac{d}{d\theta_1}P_{\nu_n}(\cos\theta_1) + B_{\nu_n}\frac{d}{d\theta_1}Q_{\nu_n}(\cos\theta_1)\right], \tag{15}$$

[8] C. H. Papas, J. Math. and Phys. **33**, 269 (1954).

$$\sum_{\nu_{m'}} G_{\nu_m,\nu'_m}(\nu'_m + 1)h^{(1)}_{\nu'_m}(k_0 r)P_{\nu'_m}(\cos\theta_1)$$
$$= \sum_{\nu_n} \nu_n(\nu_n + 1)h^{(1)}_{\nu_n}(k_1 r)$$
$$\times [A_{\nu_n}P_{\nu_n}(\cos\theta_1) + B_{\nu_n}Q_{\nu_n}(\cos\theta_1)]. \quad (16)$$

It is noted that, in contrast with the spherical and circular cylindrical boundary-value problems, the boundary conditions cannot be satisfied by equating each term of the series expansion. For the present case, the above equations must be satisfied for all values of r from $r = a$ to $r = \infty$. Consequently, the orthogonality properties of the radial function must be utilized to overcome the difficulty. Substituting the expansion

$$h^{(1)}_{\nu_n}(k_1 r) = \sum_{\nu_{m'}} \alpha_{\nu_n,\nu_m} h^{(1)}_{\nu_m}(k_0 r) \quad (17)$$

into Eqs. (15) and (16), and applying the orthogonality relations of the radial function, leads to the following expressions:

$$\frac{\epsilon_0}{\epsilon_1} G_{\nu_m,\nu'_m} = \sum_{\nu_n} (A_{\nu_n}a'_{\nu_n} + B_{\nu_n}b'_{\nu_n})\alpha_{\nu_n,\nu_m}, \quad (18)$$

$$G_{\nu_m,\nu'_m}(\nu'_m + 1)g_{\nu_m,\nu'_m} = \sum_{\nu_n} (A_{\nu_n}a_{\nu_n} + B_{\nu_n}b_{\nu_n})$$
$$\times \nu_n(\nu_n + 1)\alpha_{\nu_n,\nu_m}, \quad (\nu'_m = \nu'_0, \nu'_1, \nu'_2, \cdots), \quad (19)$$

where the abbreviations

$$a_{\nu_n} = P_{\nu_n}(\cos\theta_1), \quad a'_{\nu_n} = (d/d\theta_1)P_{\nu_n}(\cos\theta_1),$$
$$b_{\nu_n} = Q_{\nu_n}(\cos\theta_1), \quad b'_{\nu_n} = (d/d\theta_1)Q_{\nu_n}(\cos\theta_1),$$
$$g_{\nu_m,\nu'_m} = P_{\nu_m,\nu'_m}(\cos\theta_1), \quad g'_{\nu_m,\nu'_m} = (d/d\theta_1)P_{\nu_m,\nu'_m}(\cos\theta_1) \quad (20)$$

have been used. α_{ν_n,ν_m} is given in Appendix B. Expressing B_{ν_n} in terms of A_{ν_n} gives (in matrix notation)

$$B_{\nu_n} = R^{-1}_{\nu_n,\nu_m,\nu'_m} D_{\nu_m,\nu'_m}, \quad (21)$$

where $R^{-1}_{\nu_n,\nu_m,\nu'_m}$ is the inverse of the matrix

$$[R_{\nu_m,\nu'_m,\nu_n}] = \left[\left(\frac{\epsilon_1}{\epsilon_0} b'_{\nu_n}g_{\nu_m,\nu'_m}(\nu'_m + 1)\right.\right.$$
$$\left.\left. - b_{\nu_n}\nu_n(\nu_n + 1)g'_{\nu_m,\nu'_m}\right)\alpha_{\nu_n,\nu_m,\nu'_m}\right],$$

and D_{ν_m,ν'_m} is a column matrix

$$\left[\sum_{\nu_e} A_{\nu_e}\left(a_{\nu_e}\nu_e(\nu_e + 1)g'_{\nu_m,\nu'_m}\right.\right.$$
$$\left.\left. - \frac{\epsilon_1}{\epsilon_0} a'_{\nu_e}g_{\nu_m,\nu'_m}\nu'_m(\nu'_m + 1)\right)\alpha_{\nu_e,\nu_m,\nu'_m}\right].$$

Equation (21) can also be written in the form

$$B_{\nu_n} = \sum_{\nu_e} h_{\nu_n,\nu_e}A_{\nu_e}, \quad (22)$$

where h_{ν_n,ν_e} are obtained using Eq. (21).

At the surface of the conducting cone, $\theta = \theta_0$, E_r in the dielectric sheath and Eq. (12) must be identically equal for all values of r between a and ∞. Therefore,

$$A_{\nu_n}P_{\nu_n}(\cos\theta_0) + B_{\nu_n}Q_{\nu_n}(\cos\theta_0)$$
$$= L_{\nu_n}P_{\nu_n}(\cos\theta_0), \quad (23)$$

where L_{ν_n} is given by Eq. (13). Making the identification

$$d_{\nu_n} = P_{\nu_n}(\cos\theta_0), \quad f_{\nu_n} = Q_{\nu_n}(\cos\theta_0), \quad (24)$$

and substituting Eq. (22) into Eq. (23), one finds

$$A_{\nu_n}d_{\nu_n} + f_{\nu_n}\sum_{\nu_e} h_{\nu_n,\nu_e}A_{\nu_e} = L_{\nu_n}d_{\nu_n}. \quad (25)$$

Solving for A_{ν_n} gives

$$A_{\nu_n} = [Q^{-1}_{\nu_n,\nu_m}][L_{\nu_m}d_{\nu_m}], \quad (26)$$

where $Q^{-1}_{\nu_n,\nu_m}$ is the inverse of the matrix

$$[Q_{\nu_m,\nu_n}] = [(d_{\nu_n}\delta_{\nu_n,\nu_m} + h_{\nu_n,\nu_m}f_{\nu_n})], \quad (27)$$

and δ_{ν_n,ν_m} is the Kronecker delta which is equal to zero when $\nu_n \neq \nu_m$, and is equal to unity when $\nu_n = \nu_m$. $[L_{\nu_m}d_{\nu_m}]$ is a column matrix. With the knowledge of A_{ν_n} and B_{ν_n}, the coefficient G_{ν_m,ν'_m} can easily be computed using either Eq. (18) or Eq. (19).

The electromagnetic fields in the dielectric shell and in the free space are now completely determined. At large distances from the radiating source, the asymptotic expressions for $h^{(1)}_{\nu'}(k_0 r)$ which is

$$(e^{ik_0 r}/k_0 r)e^{-i(\nu'+1)\frac{1}{2}\pi},$$

leads to

$$H'_\phi = (i\omega\epsilon_0 e^{ik_0 r}/k_0 r)$$
$$\times \sum_{\nu'} G_{\nu'}\left[\frac{\partial}{\partial\theta}P_{\nu'}(\cos\theta)\right]e^{-i(\nu'+1)\frac{1}{2}\pi}. \quad (28)$$

IV. NUMERICAL COMPUTATIONS

The influence of the presence of a dielectric sheath upon the electromagnetic field radiated from a spherically tipped cone can now be computed. However, it is noted that the computation is by no means trivial since the required Sommerfeld's complex-order wavefunctions have not been tabulated; only certain limiting values are known at present. Hence, the task of tabulating these complex-order wavefunctions was first carried out for $kr \leq 10$ and $|\nu| \leq 10$. This tabulation is given elsewhere.[9]

In a recent article by Keller, Rubinow, and

[9] C. Yeh, "Tabulation of Complex Order Spherical Hankel Functions," USCEC Rept., Electrical Engineering Dept., University of Southern California, (June, 1963).

Goldstein,[10] the complex zero $\nu_n(z)$, $n = 1, 2 \cdots$ of $H_\nu^{(1)}(z)$, $dH_\nu^{(1)}(z)/dz$ and $dH_\nu^{(1)}(z)/dz + iZH_\nu^{(1)}(z)$ were investigated. They obtained approximate expressions for ν_n for small and large values of z by using appropriate approximate representations for Bessel functions. Also given were the numerical solutions of ν_n $n = 1, 2, 3, 4, 5$ for $H_\nu^{(1)}(z)$ and $dH_\nu^{(1)}(z)/dz$, with $0.01 \leq z \leq 7$. Similar procedures as those given by Keller, Rubinow, and Goldstein (KRG) will be used to find the roots of Eqs. (8) and (9). However it will be pointed out that the approximate expressions of ν_n for $H_\nu^{(1)}(z)$ and $dH_\nu^{(1)}(z)/dz$ given here for $z \ll 1$ are better approximations than those given by KRG. Starting with the expression

$$H_\nu^{(1)}(z) = (J_{-\nu}(z) - J_\nu(z)e^{-i\nu\pi})/i \sin \nu\pi \tag{29}$$

with

$$J_{\pm\nu}(z) = \sum_{m=0}^{\infty} \frac{(-1)^m(\frac{1}{2}z)^{\pm\nu+2m}}{m!\,\Gamma(\pm\nu + 1 + m)}, \tag{30}$$

and making the appropriate approximation for $0 < z \ll 1$, one has

$$-\sum_{m=1}^{\infty} \frac{\zeta(2m+1)}{(2m+1)} \nu^{2m+1} + (\tfrac{1}{2}i\pi - \gamma - \log \tfrac{1}{2}z)\nu - in\pi + O(z^2) = 0, \tag{31}$$

where γ is the Euler's constant and ζ is the Riemann zeta function. The condition $H_\nu^{(1)}(z) = 0$ has been used to obtain Eq. (31). It should be noted that Eq. (31) is identical to Eq. (9) in KRG's paper. However instead of expanding ν as a power series in $[\log \frac{1}{2}z]^{-1}$ as given by KRG, successive-approximation method is used here to solve for ν. We have

$$\nu_n = \frac{1}{[\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}z]} \left\{ in\pi + \tfrac{1}{3}[\zeta(3)] \times \left[\frac{in\pi}{(\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}z)}\right]^3 + O(z^2)\right\} \quad z \ll 1. \tag{32}$$

This expression is different from KRG's Eq. (10). To get an idea of the improvement, a numerical example is carried out. For $z = 0.01$, KRG's Eq. (10) yields Re $\nu_1 = 0.205$, Im $\nu_1 = 0.613$, while Eq. (32) gives Re $\nu_1 = 0.180$, Im $\nu_1 = 0.593$. The numerical solution is Re $\nu_1 = 0.184$, Im $\nu_1 = 0.592$.

In the same way, when $z \ll 1$, we have

$$\nu_n = \frac{1}{(\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}z)} \left\{ i\pi(n - \tfrac{1}{2}) + \tfrac{1}{3}[\zeta(3)]\left[\frac{i\pi(n-\frac{1}{2})}{(\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}z)}\right]^3 + O(z^2)\right\} \tag{33}$$

for the roots of $dH_\nu^{(1)}(z)/dz = 0$. This Eq. (33) is also different from KRG's Eq. (16).

Returning now to the problem of finding the roots of Eq. (8) or (9), we note that Eq. (8) or (9) can be simplified to give

$$H_\mu^{(1)}(v) + 2v(d/dv)H_\mu^{(1)}(v) = 0, \tag{34}$$

where $\mu = \nu + \frac{1}{2}$ and v may be k_0a or k_1a. When $v \ll 1$, the following expression can be obtained using Eqs. (29) and (30):

$$\log \frac{\Gamma(\mu + 1)}{\Gamma(-\mu + 1)} = \log\left(\frac{1 + 2\mu}{1 - 2\mu}\right) + 2n\pi i + 2\mu \log \tfrac{1}{2}v - i\mu\pi + O(v^2). \tag{35}$$

Substituting the following series:

$$\log \frac{\Gamma(\mu + 1)}{\Gamma(-\mu + 1)} = -2\gamma\mu - 2\sum_{m=1}^{\infty} \frac{\zeta(2m+1)}{2m+1} \mu^{2m+1}, \tag{36}$$

$$\log\left(\frac{1 + 2\mu}{1 - 2\mu}\right) = 2\sum_{m=1}^{\infty} \frac{\mu^{2m+1}}{2m+1} 2^{2m+1} \tag{37}$$

into Eq. (35) and simplifying, gives

$$\mu(2 \log \tfrac{1}{2}v - i\pi) = -2n\pi i - 2\gamma\mu - 2\sum_{m=1}^{\infty} \frac{2^{2m+1} + \zeta(2m+1)}{2m+1} \mu^{2m+1} + O(u^2). \tag{38}$$

When $v \ll 1$, using the method of successive approximation, one obtains

$$\mu_n = \frac{1}{\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}v} \left\{ in\pi + \tfrac{1}{3}[8 + \zeta(3)] \times \left[\frac{1}{\frac{1}{2}i\pi - \gamma - \log \frac{1}{2}v}\right]^3 + O(v^2)\right\}. \tag{39}$$

It is noted that for very small values of v, Eq. (39) approaches Eq. (32). Hence the roots for Eq. (34) approaches the roots for $H_\mu^{(1)}(v) = 0$ when v is very small.

Numerical solutions of Eq. (34) were carried out for moderate values of $v(0.1 < v < 1.0)$ using the IBM 7090 computer. It is found that the roots μ_n given by Eq. (39) are within 15% of those given by Eq. (34) provided that $v \leq 0.1$ and $n \leq 5$, and the roots computed from the simplified Eq. (35) are within 3% of those obtained from Eq. (34) when $v \leq 0.1$. As a matter of fact, the roots from the simplified Eq. (35) are in agreement within 5% of those obtained from Eq. (34) even when $v \leq 0.3$. It should be noted that the roots are symmetric about the origin since $H_{-\mu}^{(1)}(v) = e^{-i\mu\pi}H_\mu^{(1)}(v)$.[10]

[10] J. B. Keller, S. I. Rubinow, and M. Goldstein, J. Math. Phys. 4, 829 (1963).

For large values of v and $|\nu_n|$, the approximate expression for the roots of Eq. (34) can be derived with the help of Debye's expansion for the Bessel functions.[10,11] Since we are concerned with the problem of a cone with a small spherical boss at the tip, the roots of Eq. (34) for large value of v will not be considered.

Since the tables for complex-order Legendre functions are also not available, numerical values are computed from the series expansions for the Legendre functions.[12] To compute $\alpha_{\nu_n,\nu_m'}$ from Eq. (B1) (see Appendix B), one must evaluate the integral

$$I = \int_{k_0 a}^{\infty} h_{\nu_n}^{(1)}(k_1 r) h_{\nu_m'}^{(1)}(k_0 r)\, d(k_0 r). \tag{40}$$

This integral may be separated into two parts as follows:

$$I = I_1 + I_2, \tag{41}$$

where

$$I_1 = \int_{k_0 a}^{\xi} h_{\nu_n}^{(1)}(k_1 r) h_{\nu_m'}^{(1)}(k_0 r)\, d(k_0 r), \tag{42}$$

$$I_2 = \int_{\xi}^{\infty} h_{\nu_n}^{(1)}(k_1 r) h_{\nu_m'}^{(1)}(k_0 r)\, d(k_0 r). \tag{43}$$

The integral I_2 may be evaluated approximately

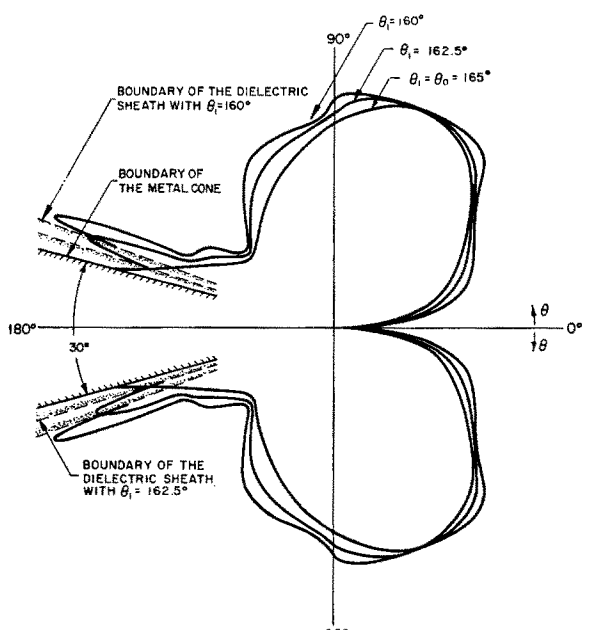

FIG. 2. Radiation patterns for dielectric coated cone excited by axially symmetric circumferential slot. Thickness of the dielectric sheath is indicated by θ_1.

[11] W. Franz, Z. Naturforsch. 9A, 705 (1954).

[12] Bateman manuscript project, *Higher Transcendental Functions*, edited by A. Erdély (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1.

using the asymptotic expression for the spherical Hankel function,

$$h_{\nu}^{(1)}(kr) \approx (1/kr) e^{ikr} e^{-i(\nu+1)\frac{1}{2}\pi}, \tag{44}$$

provided that $kr \gg |\nu|$. ξ is chosen such that $\xi \gg |\nu_n|$ and $|\nu_m'|$. Substituting Eq. (44) into Eq. (43) and carrying out the integration gives

$$I_2 \approx \frac{k_0}{k_1} e^{-i(\nu_n+\nu_m'+2)\frac{1}{2}\pi} \left\{ \frac{1}{\xi} \cos(\eta\xi) - \eta[\tfrac{1}{2}\pi - \mathrm{Si}(\eta\xi)] + i\left[\frac{1}{\xi} \sin(\eta\xi) - \eta\, \mathrm{Ci}(\eta\xi) \right] \right\}, \tag{45}$$

with

$$\eta = 1 + k_1/k_0.$$

$\mathrm{Si}(x)$ and $\mathrm{Ci}(x)$ are, respectively, Sine integral and Cosine integral,[13] i.e.,

$$\mathrm{Si}(x) = \tfrac{1}{2}\pi - \int_x^{\infty} \frac{\sin u}{u}\, du,$$

$$\mathrm{Ci}(x) = \int_{\infty}^{x} \frac{\cos u}{u}\, du. \tag{46}$$

The integral I_1 cannot be integrated analytically and must therefore be evaluated numerically. Simpson's rule was used. It is found from the computations that the coefficients $\alpha_{\nu_n,\nu_m'}$ converge quite rapidly for small values of $\chi = k_1/k_0 - 1$; only the first few terms of the infinite series [Eq. (17)] are needed as long as χ is less than 3.

To qualitatively illustrate how the solutions behave, the radiation patterns of $|H_\phi^r|$ given by Eq. (28) are computed. It is assumed that the conical structure with $\theta_0 = 165°$ (i.e., a cone angle of 30°) is excited by a delta slot located at $k_0 r_1$. Two cases are considered: one with $k_0 r_1 = \pi$, and the other with $k_0 r_1 = 3\pi$. The assumption of a delta-slot source simplifies considerably the expression for L_{ν_n} in Eq. (13); we have

$$L_{\nu_n} = \frac{E_0 r_1 k_1 h_{\nu_n}^{(1)}(k_1 r_1)}{\nu_n(\nu_n+1) P_{\nu_n}(\cos\theta_0) N_{\nu_n}(k_1 a)}.$$

Three cases of the thickness of the dielectric coating with $\epsilon_1/\epsilon_0 = 2.0$ are considered: $\theta_1 = 165°$, $\theta_1 = 162.5°$, and $\theta_1 = 160°$. The spherical boss at the tip of the cone has a radius of $k_0 a = 0.1$. The roots ν_n and ν_m' are computed, respectively, from Eq. (8) with $k_1 a = 0.1414$, and from Eq. (9) with $k_0 a = 0.1$. A total of 20 roots each were found for ν_n and ν_m'; i.e., $n = \pm 1, \pm 2, \cdots \pm 10$ and $m = \pm 1, \pm 2, \cdots \pm 10$. Results for the computed radiation patterns are shown in Figs. 2 and 3. Numerical investigation

[13] E. Jahnke and F. Emde, *Tables of Functions*, (Dover Publications, Inc., New York, 1945).

shows that the infinite series representing these expansion coefficients for the radiated wave converge quite rapidly for small values of k_1a and k_1r_1. Upon comparison of the radiation pattern for the uncoated, spherically tipped cone (i.e., the $\theta_1 = 165°$ case in Fig. 2) with the pattern given by Bailin and Silver[14] for the uncoated cone, one notes that if the spherical boss is small (i.e., $ka \ll 1$), and if $k_0r_1 > 1$, the presence of the spherical tip alters only slightly the radiation patterns for the cone. Only the pattern in the forward direction is somewhat enlarged; this is due to the induced currents on the spherical boss. The presence of the dielectric sheath tends to spread out the radiation pattern, and to diffract more field to the rear of the cone; the radiated field in the forward direction was not altered significantly. It can also be observed that the dielectric coating seems to enhance the ripples in the curves.

It is remarked here that the formal exact solutions given in Sec. III are valid for the near zone (i.e., near the conical structure) as well as for the far zone.

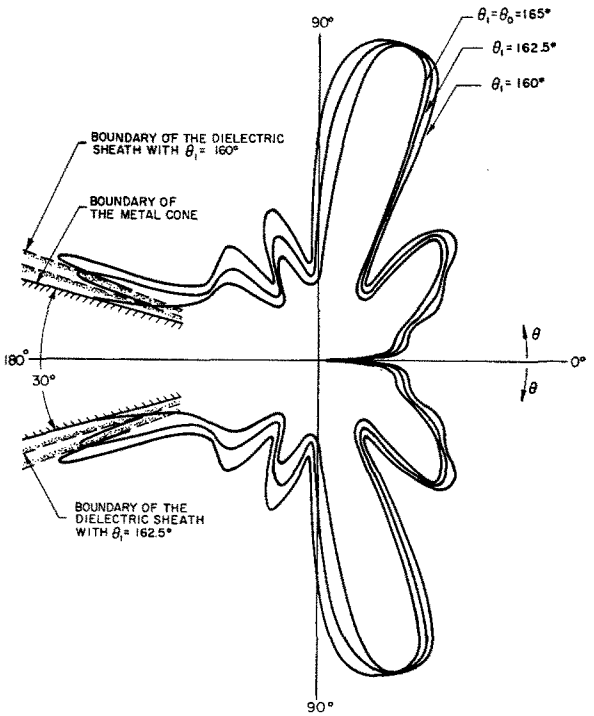

FIG. 3. Radiation patterns for dielectric coated cone excited by axially symmetric circumferential slot. Thickness of the dielectric sheath is indicated by θ_1.

APPENDIX A. ORTHOGONALITY CHARACTERISTICS OF $h_\nu^{(1)}(kr)$

To show that the radial functions $h_\nu^{(1)}(kr)$ are orthogonal over the range $kr = ka$ to $kr = \infty$,[7] one notes that for any ν, say, ν_n,

$$(kr)(d^2/d(kr)^2)(krh_{\nu_n}^{(1)}(kr)) + ((kr)^2 - \nu_n(\nu_n + 1))h_{\nu_n}^{(1)}(kr) = 0, \quad \text{(A1)}$$

and for any other value of ν, say, ν_m,

$$(kr)(d^2/d(kr)^2)(krh_{\nu_m}^{(1)}(kr)) + ((kr)^2 - \nu_m(\nu_m + 1))h_{\nu_m}^{(1)}(kr) = 0. \quad \text{(A2)}$$

Multiplying the first equation by $h_{\nu_m}^{(1)}(kr)$ and the second by $h_{\nu_n}^{(1)}(kr)$ and integrating the difference from $kr = ka$ to $kr = \infty$, one gets

$$[\nu_m(\nu_m + 1) - \nu_n(\nu_n + 1)] \int_{ka}^{\infty} h_{\nu_n}^{(1)}(kr)h_{\nu_m}^{(1)}(kr)\, d(kr) = \int_{ka}^{\infty} \left[krh_{\nu_n}^{(1)}(kr) \frac{d^2}{d(kr)^2}(krh_{\nu_m}^{(1)}(kr)) - krh_{\nu_m}^{(1)}(kr) \frac{d^2}{d(kr)^2}(krh_{\nu_n}^{(1)}(kr)) \right] d(kr). \quad \text{(A3)}$$

Integrating the above by parts gives

$$[\nu_m(\nu_m + 1) - \nu_n(\nu_n + 1)] \int_{ka}^{\infty} h_{\nu_n}^{(1)}(kr)h_{\nu_m}^{(1)}(kr)\, d(kr) = krh_{\nu_n}^{(1)}(kr) \frac{d}{d(kr)}(krh_{\nu_m}^{(1)}(kr)) - krh_{\nu_m}^{(1)}(kr) \frac{d}{d(kr)}(krh_{\nu_n}^{(1)}(kr))\Big|_{ka}^{\infty}. \quad \text{(A4)}$$

The terms on the right-hand side of the equal sign are zero by virtue of the boundary condition (8) or (9) and the asymptotic behavior of $h_{\nu_n}^{(1)}(kr)$. Hence,

$$\int_{ka}^{\infty} h_{\nu_n}^{(1)}(kr)h_{\nu_m}^{(1)}(kr)\, d(kr) = \delta_{\nu_n,\nu_m}N_{\nu_n}(ka), \quad \text{(A5)}$$

where δ_{ν_n,ν_m} is the Kronecker delta and $N_{\nu_n}(ka)$ is a normalization factor which can be obtained from Eq. (A4) by an application of de l'Hospital's rule for the limit $\nu_n \to \nu_m$. Substituting the relations

$$\Delta\nu_m = \nu_n - \nu_m,$$

and

$$krh_{\nu_n}^{(1)}(kr) = krh_{\nu_n}^{(1)}(kr) + (\partial/\partial\nu_m)(krh_{\nu_m}^{(1)}(kr))\Delta\nu_m$$

into Eq. (A4) and using the boundary condition

$$[\partial/\partial(ka)][kah_{\nu_m}^{(1)}(ka)] = 0,$$

one obtains the normalization factor

$$N_{\nu_n}(ka) = \int_{ka}^{\infty} [h_{\nu_n}^{(1)}(kr)]^2\, d(kr) = \frac{1}{2\nu_n + 1}\left\{kah_{\nu_n}^{(1)}(ka) \frac{\partial^2}{\partial(ka)\,\partial(\nu_n)}[kah_{\nu_n}^{(1)}(ka)]\right\}. \quad \text{(A6)}$$

[14] L. L. Bailin and S. Silver, IRE Trans. Antennas Propagation 4, 5 (1956).

APPENDIX B. FORMULA FOR $\alpha_{\nu_n\nu_m'}$

Multiplying both sides of (17) by $h_{\nu_m'}^{(1)}(k_0 r)$, integrating with respect to $k_0 r$ from $k_0 a$ to ∞, and using the orthogonality relation for the radial function (A5), one obtains

$$\alpha_{\nu_n,\nu_m'} = \frac{1}{M_{\nu_m'}} \int_{k_0 a}^{\infty} h_{\nu_n}^{(1)}(k_1 r) h_{\nu_m'}^{(1)}(k_0 r)\, d(k_0 r), \qquad \text{(B1)}$$

where

$$M_{\nu_m'} = \int_{k_0 a}^{\infty} [h_{\nu_m'}(k_0 r)]^2\, d(k_0 r)$$

$$= \frac{1}{2\nu_m' + 1} \left\{ k_0 a h_{\nu_m'}^{(1)}(k_0 a) \frac{\partial^2}{\partial(k_0 a)\, \partial(\nu_m')} [k_0 a h_{\nu_m'}^{(1)}(k_0 a)] \right\}. \qquad \text{(B2)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Regge Poles and Potentials with Cores*†

JOHN W. DURSO AND PETER SIGNELL
Physics Department, The Pennsylvania State University, University Park, Pennsylvania
(Received 8 October 1963)

The analytic behavior of the S matrix as a function of real momentum and complex angular momentum is examined for the case of nonrelativistic scattering by potentials with hard-core interiors. The presence of a hard core in a potential introduces a type of symmetry which, combined with unitarity, places restrictions upon the movement of the Regge poles and upon the allowable representations of the scattering amplitude. In particular, it is found that the Watson–Sommerfeld transform representation of the scattering amplitude is not valid. An alternate representation of the S function, more practical for such cases, is postulated and examined numerically. Numerical results are presented for the case when the hard core is replaced by a finite barrier added to a repulsive Yukawa potential interior. It is found that additional trajectories appear. The mechanism for producing the threshold behavior of the phase shifts is compared to that of the case where a hard core is not present.

INTRODUCTION

A GREAT deal of study has recently been devoted to the analytic behavior of the scattering amplitude as a function of complex angular momentum in potential scattering. Regge, Bottino, and Longoni[1] (hereafter referred to as RBL) have proved several theorems concerning the analyticity of the wavefunctions for complex values of l in the nonrelativistic Schrödinger equation when the potential may be expressed as a superposition of Yukawa potentials. They have also shown that the scattering amplitude for a given value of the momentum may be expressed by an integral along the line Re $\lambda = 0$, where $\lambda = l + \frac{1}{2}$, and the sum of residues of poles in the half-plane Re $\lambda > 0$. Most of these proofs depend upon the assumption that the potential is a superposition of Yukawa potentials. However, phenomenological nucleon–nucleon potentials are not of this type; rather, they include a hard core in order to fit the nucleon–nucleon phase shifts over a wide energy range. For this reason, it is of interest to examine the effect of a hard core upon the S matrix (or S function) in the complex angular momentum plane. In this paper we examine the Regge pole trajectories for such potentials; for real, positive values of the momentum variable k. The Cauchy theorem representation is found not to be practical; we postulate a new representation of the S function, which we check by numerical calculation.

* This work was assisted by support from the U. S. Atomic Energy Commission.
† The computations reported here were carried out at The Pennsylvania State University Computation Center.
[1] T. Regge, A. Bottino, and A. M. Longoni, Nuovo Cimento **23**, 954 (1962).

I. HARD-CORE SYMMETRY

Consider a potential

$$V(x) = \infty \qquad (0 \le x \le a), \qquad a > 0$$

$$\le \int_{m>0}^{\infty} \sigma(\mu) \frac{\exp(-\mu x)}{x}\, d\mu \qquad (a < x),$$

where $\sigma(\mu)$ satisfies all the criteria enumerated in RBL.

The solution of the radial Schrödinger equation,

APPENDIX B. FORMULA FOR $\alpha_{\nu_n\nu_m'}$

Multiplying both sides of (17) by $h_{\nu_m'}^{(1)}(k_0 r)$, integrating with respect to $k_0 r$ from $k_0 a$ to ∞, and using the orthogonality relation for the radial function (A5), one obtains

$$\alpha_{\nu_n,\nu_m'} = \frac{1}{M_{\nu_m'}} \int_{k_0 a}^{\infty} h_{\nu_n}^{(1)}(k_1 r) h_{\nu_m'}^{(1)}(k_0 r)\, d(k_0 r), \qquad \text{(B1)}$$

where

$$M_{\nu_m'} = \int_{k_0 a}^{\infty} [h_{\nu_m'}(k_0 r)]^2\, d(k_0 r)$$

$$= \frac{1}{2\nu_m' + 1} \left\{ k_0 a h_{\nu_m'}^{(1)}(k_0 a) \frac{\partial^2}{\partial(k_0 a)\, \partial(\nu_m')} [k_0 a h_{\nu_m'}^{(1)}(k_0 a)] \right\}. \qquad \text{(B2)}$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Regge Poles and Potentials with Cores*†

JOHN W. DURSO AND PETER SIGNELL
Physics Department, The Pennsylvania State University, University Park, Pennsylvania
(Received 8 October 1963)

The analytic behavior of the S matrix as a function of real momentum and complex angular momentum is examined for the case of nonrelativistic scattering by potentials with hard-core interiors. The presence of a hard core in a potential introduces a type of symmetry which, combined with unitarity, places restrictions upon the movement of the Regge poles and upon the allowable representations of the scattering amplitude. In particular, it is found that the Watson–Sommerfeld transform representation of the scattering amplitude is not valid. An alternate representation of the S function, more practical for such cases, is postulated and examined numerically. Numerical results are presented for the case when the hard core is replaced by a finite barrier added to a repulsive Yukawa potential interior. It is found that additional trajectories appear. The mechanism for producing the threshold behavior of the phase shifts is compared to that of the case where a hard core is not present.

INTRODUCTION

A GREAT deal of study has recently been devoted to the analytic behavior of the scattering amplitude as a function of complex angular momentum in potential scattering. Regge, Bottino, and Longoni[1] (hereafter referred to as RBL) have proved several theorems concerning the analyticity of the wavefunctions for complex values of l in the nonrelativistic Schrödinger equation when the potential may be expressed as a superposition of Yukawa potentials. They have also shown that the scattering amplitude for a given value of the momentum may be expressed by an integral along the line Re $\lambda = 0$, where $\lambda = l + \frac{1}{2}$, and the sum of residues of poles in the half-plane Re $\lambda > 0$. Most of these proofs depend upon the assumption that the potential is a superposition of Yukawa potentials. However, phenomenological nucleon–nucleon potentials are not of this type; rather, they include a hard core in order to fit the nucleon–nucleon phase shifts over a wide energy range. For this reason, it is of interest to examine the effect of a hard core upon the S matrix (or S function) in the complex angular momentum plane. In this paper we examine the Regge pole trajectories for such potentials; for real, positive values of the momentum variable k. The Cauchy theorem representation is found not to be practical; we postulate a new representation of the S function, which we check by numerical calculation.

* This work was assisted by support from the U. S. Atomic Energy Commission.
† The computations reported here were carried out at The Pennsylvania State University Computation Center.
[1] T. Regge, A. Bottino, and A. M. Longoni, Nuovo Cimento **23**, 954 (1962).

I. HARD-CORE SYMMETRY

Consider a potential

$$V(x) = \infty \qquad (0 \le x \le a), \qquad a > 0$$

$$\le \int_{m>0}^{\infty} \sigma(\mu) \frac{\exp(-\mu x)}{x}\, d\mu \qquad (a < x),$$

where $\sigma(\mu)$ satisfies all the criteria enumerated in RBL.

The solution of the radial Schrödinger equation,

$$d^2\psi/dx^2 + (k^2 - (\lambda^2 - \tfrac{1}{4})/x^2 - V(x))\psi = 0,$$

for $x > a$ is given by a superposition of the two Jost solutions as defined in RBL:

$$\psi(\lambda, k, x) = f(\lambda, k, x) - \exp[-i\pi(\lambda - \tfrac{1}{2})]S(\lambda, k)f(\lambda, -k, x), \quad \text{(I.1)}$$

where $S(\lambda, k)$ is the usual S function, and $f(\lambda, \pm k, x) \to \exp(\mp ikx)$ as $x \to \infty$. The Jost solutions also have the following properties:

$$f(\lambda, k, x) = f(-\lambda, k, x), \qquad f(\lambda, k, x) = f^*(\lambda^*, -k^*, x). \quad \text{(I.2)}$$

Up to this point, no assumptions have been made about the wavefunction for $x \leq a$. Now define the hard core as the limit of a finite square barrier as the height goes to infinity. Consider the case where the barrier height is V_0 and the potential outside the barrier is zero. (The proof for the general case is essentially the same but is lengthier and does not bring out any additional concepts.)

The wavefunction for the barrier potential is

$$\psi(x) = AF_\lambda(i\gamma x) \quad (x \leq a) = K_\lambda^{(1)}(kx) + S(\lambda, k)K_\lambda^{(2)}(kx) \quad (x > a), \quad \text{(I.3)}$$

where $\gamma = +(V_0 - k^2)^{\frac{1}{2}}$, and

$$F_\lambda(z) = (\tfrac{1}{2}\pi z)^{\frac{1}{2}}J_\lambda(z), \qquad K_\lambda^{(1,2)}(z) = (\tfrac{1}{2}\pi z)^{\frac{1}{2}}H_\lambda^{(1,2)}(z).$$

$J_\lambda(z)$ and $H_\lambda^{(1,2)}(z)$ are the ordinary Bessel and Hankel functions.

Matching wavefunctions and derivatives at $x = a$ and solving for A, one obtains

$$A = 2ik/[kF_\lambda(i\gamma a)K_\lambda^{(1)\prime}(ka) - i\gamma F_\lambda'(i\gamma a)K_\lambda^{(1)}(ka)]. \quad \text{(I.4)}$$

Using recurrence relations for the derivatives and noting that

$$F(i\gamma a) \to \tfrac{1}{2}i \exp(\gamma a) \exp\{i[\tfrac{1}{2}(\lambda - \tfrac{1}{2})]\pi\} \text{ as } \gamma \to +\infty$$

(i.e., $V_0 \to \infty$), one obtains

$$A \sim \frac{4 \exp(-\gamma a) \exp\{-i[\tfrac{1}{2}(\lambda - \tfrac{1}{2})]\pi\}}{(\gamma/k)K_\lambda^{(1)}(ka) - K_{\lambda+1}^{(1)}(ka)}. \quad \text{(I.5)}$$

Hence for $0 < x \leq a$, in the limit as the barrier height approaches infinity,

$$\psi(x) \sim \frac{2i \exp[-\gamma(a - x)]}{(\gamma/k)K_\lambda^{(1)}(ka) - K_{\lambda+1}^{(1)}(ka)}. \quad \text{(I.6)}$$

Thus the wavefunction approaches zero everywhere inside the barrier as the core is made hard. The only difficulty might occur when $x = a$ and $K_\lambda^{(1)}(ka) = 0$. In that case one need only use the continuity of $\psi(x)$ to show that $\psi(a) \to 0$ in the limit as V_0 approaches infinity.

Hence, even for Re $\lambda < -\frac{1}{2}$, in which case the wavefunction is singular at the origin, the wavefunction at any finite point inside the barrier approaches zero as V_0 approaches infinity. Specifically, the wavefunction at a given point inside the barrier becomes infinitesimal compared to the coefficient of the exp $(-ikx)$ term in the asymptotic form of the wavefunction when $V_0 \gg k^2, 1/x$.

Setting (I.1) equal to zero at the core radius,

$$S(\lambda, k) = [f(\lambda, k, a)/f(\lambda, -k, a)] \times \exp[i\pi(\lambda - \tfrac{1}{2})]. \quad \text{(I.7)}$$

Since the $f(\lambda, \pm k, x)$ are entire functions of λ and k for $x > 0$, $S(\lambda, k)$ is a meromorphic function of λ and k outside cuts on the positive and negative imaginary k axes extending from $\pm i\frac{1}{2}m$ to $\pm i\infty$ due to $f(\lambda, \pm k, a)$.

Using (I.2),

$$S(-\lambda, k) = S(\lambda, k) \exp(-i2\pi\lambda), \quad \text{(I.8)}$$

and

$$S^*(\lambda, k) = [f(\lambda^*, -k^*, a)/f(\lambda^*, k^*, a)] \times \exp[-i\pi(\lambda^* - \tfrac{1}{2})] = [S(\lambda^*, k^*)]^{-1}, \quad \text{(I.9)}$$

which is the statement that S is "unitary."

The result (I.8) can easily be demonstrated without reference to the Jost solutions. Consider the asymptotic forms of the wavefunctions for $\pm\lambda$ as $x \to \infty$:

$$\psi(\lambda, k, x) \to \exp(-ikx) - \exp[-i\pi(\lambda - \tfrac{1}{2})]S(\lambda, k) \exp(ikx),$$

and

$$\psi(-\lambda, k, x) \to \exp(-ikx) - \exp[-i\pi(-\lambda - \tfrac{1}{2})]S(-\lambda, k) \exp(ikx). \quad \text{(I.10)}$$

Since the Schrödinger equation depends only upon λ^2, and the boundary condition at $x = a$ is independent of λ, the two wavefunctions must be identical, whereby (I.8) follows immediately.

II. RESTRICTIONS DUE TO THE HARD CORE

Equipped now with the unitarity relation and the symmetry condition (I.9), we may proceed to an interesting consequence concerning the Regge

trajectories; that is, concerning the location of the Regge poles for real, positive values of k.

Let k be real and greater than zero and let $\lambda = i\eta$, η real. Then the unitarity condition states

$$S^*(i\eta, k) = [S(-i\eta, k)]^{-1}. \quad \text{(II.1)}$$

The symmetry condition states

$$S(-i\eta, k) = S(i\eta, k) \exp(2\pi\eta), \quad \text{(II.2)}$$

so

$$S^*(i\eta, k) = [S(i\eta, k) \exp(2\pi\eta)]^{-1}, \quad \text{(II.3)}$$

and

$$S^*(i\eta, k)S(i\eta, k) = |S(i\eta, k)|^2 = \exp(-2\pi\eta). \quad \text{(II.4)}$$

This result states that on the imaginary axis the absolute value of the S function is *independent of the external potential* so long as the potential for $x > a$ satisfies the criterion mentioned in Sec. I.

Thus we can say that the Regge trajectories for k real and not equal to zero cannot cross the imaginary λ axis unless the residue of the pole approaches zero on the imaginary axis in such a way as to satisfy (II.4). If the latter does not happen (it does not with Yukawa-type potentials), then the trajectories are either unbounded or approach a finite limit point, but in either case remain only in one half-plane. By the "symmetry" condition (I.9), there are an even number of trajectories, occurring in pairs, and each trajectory probably remains in the half-plane in which it started at zero energy. More important, (II.4) immediately precludes the use of the Watson–Sommerfeld transform, since the integral along the imaginary axis is now divergent.

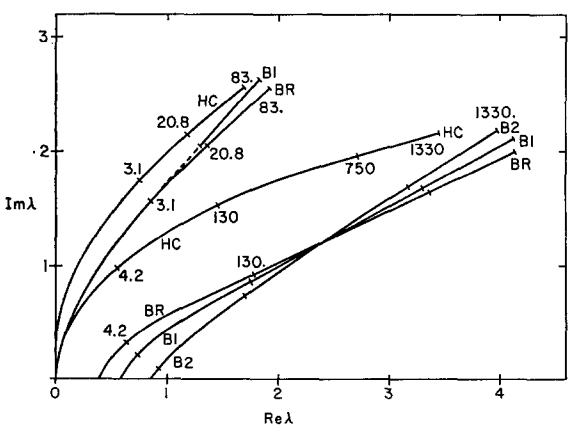

FIG. 1. The first two Regge trajectories in the first quadrant for several potentials. HC indicates the trajectories for a pure hard core, BR the trajectories for the unmodified Bryan potential, B1 the Bryan exterior potential with a 1-GeV core plus repulsive Yukawa interior, and B2 the same but with a 100-MeV core. Numbers next to the curves indicate energies in MeV.

In the preceding analysis one must exclude the values $a = 0$ and $k = 0$. The value $a = 0$ is excluded because $f(\lambda, k, a)$ is not analytic at that point. If, as before, the hard core is the limit of a finite barrier with V_0 the height of the barrier, then in order to prove (I.7), γa must be able to be made arbitrarily large. If a is made to approach zero, a discontinuity develops across the line Re $\lambda = 0$. The value $k = 0$ has been excluded since in that case $S(\lambda, k)$ becomes unity in the right half-plane and $\exp(-i2\pi\lambda)$ in the left half-plane. Thus a discontinuity develops in this case also.

To illustrate the above remarks, consider the case of a pure hard core ($V(x) = 0$, $x > a$). The wavefunction may be written

$$\psi_{hc}(\lambda, k, x) = B(\tfrac{1}{2}\pi x)^{\frac{1}{2}}[H_\lambda^{(2)}(kx) + S_{hc}(\lambda, k)H_\lambda^{(1)}(kx)] \quad (x > a), \quad \text{(II.5)}$$

so that the condition $\psi = 0$ at $x = a$ determines S:

$$S_{hc}(\lambda, k) = -H_\lambda^{(2)}(ka)/H_\lambda^{(1)}(ka). \quad \text{(II.6)}$$

We see that the S function depends only upon λ and ka. One may show simply that in the limit as k or a goes to zero, for this case,

$$\lim_{\epsilon\to 0}\lim_{ka\to 0}[S_{hc}(i\eta + \epsilon, k) - S_{hc}(i\eta - \epsilon, k)] = 1 - \exp(-2\pi\eta). \quad \text{(II.7)}$$

The Regge trajectories in the first quadrant and hence in the third quadrant of the complex λ plane for a pure hard core have been shown to be infinite in number by Kotin and Magnus.[2,3] We have used a numerical computer program to determine the trajectories for both the pure-hard core case and for the singlet even Bryan potential, which is a typical physical nucleon–nucleon potential possessing a hard core. For the pure hard-core case, the radius of the core was set equal to the radius of the Bryan core. The first two trajectories in the first quandrant for both cases are shown by the curves HC and BR in Fig. 1. At high energies the Bryan trajectories approach the pure hard-core trajectories, as one would expect, since the potential is finite everywhere beyond the core radius.

III. REPRESENTATION OF $S(\lambda, k)$

The "unitarity" and "symmetry" of $S(\lambda, k)$ suggest that a convenient way to express the S function for the pure hard-core case is

[2] W. Magnus and L. Kotin, Numer. Math. 2, 228 (1960).

[3] Compare A. O. Barut and F. Calogero (University of California, 1963 preprint), where it is assumed that the hard-core potential has no Regge trajectories.

$$S_{hc}(\lambda, k) = \left\{\prod_{i=1}^{\infty}\left[\frac{\lambda^2 - \alpha_i^{*2}(k)}{\lambda^2 - \alpha_i^2(k)}\right]\right\}\xi_{hc}(\lambda, k), \qquad \text{(III.1)}$$

where the α's are the pole locations; ξ_{hc} is an entire function of λ defined by (III.1). The product runs only over poles in the first quadrant and the dependence upon a, the core radius, is suppressed since a is assumed to be fixed throughout this discussion. Examining (III.1), one can see that ξ_{hc} also satisfies the unitarity and symmetry conditions (I.8) and (I.9).

Now assume that the S function for a hard core plus an exterior potential may be written in the same *form* as above:

$$S(\lambda, k) = \left\{\prod_{i=1}^{\infty}\left[\frac{\lambda^2 - \beta_i^{*2}(k)}{\lambda^2 - \beta_i^2(k)}\right]\right\}\xi(\lambda, k). \qquad \text{(III.2)}$$

Since $S(\lambda, k) = S_{hc}(\lambda, k)$ as $k \to \infty$, and we have observed numerically that $\beta_i(k) \to \alpha_i(k)$ in this limit, $\xi(\lambda, k)$ must approach $\xi_{hc}(\lambda, k)$ in this limit. Furthermore $S(\lambda, k) \to S_{hc}(\lambda, k) \to 1$ as $\lambda \to +\infty$ means that $\xi(\lambda, k) \to \xi_{hc}(\lambda, k)$ in the latter limit as well.

We now suggest equality of the two entire functions, $\xi(\lambda, k) = \xi_{hc}(\lambda, k)$. Since $\xi_{hc}(\lambda, k)$ is independent of the external potential, $S(\lambda, k)$ is now determined only by the hard-core function $\xi_{hc}(\lambda, k)$ and by pole locations. Although we have no analytic proof as yet, the numerical results seem to strongly confirm this assumption. Using a numerical program to calculate the locations of the Regge poles on several trajectories, the validity of the assumption was tested in the following way.

Abbreviating finite numbers of pole contributions by

$$P_{hc}^{(n)}(\lambda, k) = \prod_{i=1}^{n}\left[\frac{\lambda^2 - \alpha_i^{*2}(k)}{\lambda^2 - \alpha_i^2(k)}\right],$$
$$P^{(n)}(\lambda, k) = \prod_{i=1}^{n}\left[\frac{\lambda^2 - \beta_i^{*2}(k)}{\lambda^2 - \beta_i^2(k)}\right], \qquad \text{(III.3)}$$

where n is the number of poles used, ordered by proximity to the real axis,

$$S^{(n)}(\lambda, k) = [P^{(n)}(\lambda, k)/P_{hc}^{(n)}(\lambda, k)]S_{hc}(\lambda, k). \qquad \text{(III.4)}$$

This just replaces the first n poles of the pure hard-core case by those computed using the physical potential. The higher poles are close to the hard-core positions, so they may be left as hard core without undue error. The phase shifts are easily calculable in this approximation.

Phases were calculated for the Bryan potential for three different lab energies, 10, 50, and 210 MeV. In each case, first one, then two poles were used, and in several cases, three and four. In order to test the effect of poles further away from the real axis, and the effect of ξ_{hc}, phase shifts were computed from the pole functions $P^{(n)}(\lambda, k)$ alone. Exact solutions were obtained by calculating the phase shifts directly from the potential. The results are presented in Table I. Numerical difficulties made it impossible to calculate the positions of higher poles, but one can see that the fit to the exact result in all three cases is quite good. Note that in each case the addition of one more pole consistently improved the approximation. Furthermore, the corrections became smaller and smaller as more poles were added, as can readily be seen in the 10- and 50-MeV cases. That the pole terms alone are insufficient is also obvious from Table I.

TABLE I. Approximations to phase shifts using pole terms multiplied by the entire function $\xi_{hc}(\lambda, x)$, and phase shifts using pole contributions only.

Poles	Phase Shifts			
	S	P	D	F
E lab = 10 MeV				
1	40.2°	−7.25°	−4.7°	
2	49.7°	−2.8°	−3.45°	
3	52.45°	−0.03°	−1.9°	
4	53.6°	0.77°	−0.98°	
Exact	55.2°	2.2°	0.18°	
$P^{(4)}$ without ξ_{hc}	67.7°	−49.6°	85.7°	
E lab = 50 MeV				
1	28.6°	1.8°	−4.95°	−3.5°
2	37.2°	8.4°	−1.21°	−1.61°
3	39.9°	10.7°	0.49°	−0.52°
Exact	40.88°	11.7°	1.49°	0.43°
$P^{(3)}$ without ξ_{hc}	54.3°	−39.8°	−3.44°	67.8°
E lab = 210 MeV				
1	3.98°	23.3°	2.42°	−2.16°
2	8.0°	28.1°	6.1°	0.69°
Exact	9.47°	29.6°	7.6°	2.07°
$P^{(2)}$ without ξ_{hc}	14.7°	59.7°	84.5°	51.7°

IV. BEHAVIOR OF TRAJECTORIES FOR "SOFT" CORES

The numerical calculations of Sec. III have indicated that the Regge trajectories for potentials with hard cores are unbounded. Bethe and Kinoshita[4] have shown that when the potential is a superposition of Yukawa potentials, the negative half-integers should be high-energy limit points of Regge trajectories. This has been demonstrated numerically for the case of attractive Yukawa potentials.[5] Bethe and Kinoshita have also shown

[4] H. A. Bethe and T. Kinoshita, Phys. Rev. **128**, 1418 (1962).

[5] A. Ahmadzadeh, P. G. Burke, and C. Tate, UCRL-10216 (1962).

that the high-energy pole near $\lambda = -\frac{1}{2}$ is due to the $1/r$ behavior of the potential near $r = 0$.

With this in mind, we examined numerically the behavior of the Regge trajectories when the core of the Bryan potential used in Secs. II and III was replaced by finite barriers of first 1 GeV and then 100 MeV (inside 0.39-pion Compton wavelengths), added to a repulsive Yukawa potential there of strength equal to the one-pion-exchange potential. For each barrier height the first two trajectories in the first quadrant were computed, and for the case of the 1-GeV barrier, the first two in the third quadrant. Since in both cases the potential behaved as $1/r$ near the origin, one would expect the high-energy pole near $\lambda = -\frac{1}{2}$ to be present. The computed trajectories are shown in Figs. 1 and 2.

The trajectory indicated by the small loop in Fig. 2 was found to be the one which approached $\lambda = -\frac{1}{2}$ at high energy. The behavior of this trajectory is interesting in that as zero energy is approached the trajectory approaches $\lambda = -\frac{1}{2}$ tangent to the real axis. Even at its furthest point from the real axis, the imaginary part of the pole location was only -0.14. Presumably, approximately the same kind of trajectory[5] would occur near $\lambda = -\frac{3}{2}$. It would seem that increasing the height of the barrier (making it more closely approach the hard core) would decrease the size of the loop until it disappeared when the core became hard. At moderate energies the small loop trajectories contribute little to the S function.

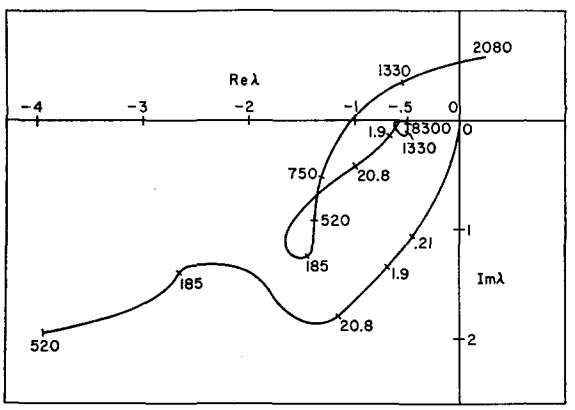

FIG. 2. The first two Regge trajectories in the third quadrant for case B1 of Fig. 1, and the trajectory which approaches $\lambda = -\frac{1}{2}$ at high energy. Note the approximate low-energy λ symmetry with respect to the corresponding trajectories in Fig. 1.

V. DISCUSSION AND SUMMARY

The results of Keller *et al.*[6] and of the present numerical calculations suggest that the mechanism for producing the low-energy behavior of the S function in the hard-core case and in the "no core" case is essentially the same; in both cases the accumulation of an infinite number of poles of vanishing residue at $\lambda = 0$ (for $k^2 \to 0$) produces the threshold behavior.[5,7,8] However, the presence of a hard core in the potential produces restrictions upon the movement of Regge poles which limits a given trajectory to the left or right half-plane. The presence of a hard core also invalidates the use of the Watson–Sommerfeld transform to express the scattering amplitude. Also, it appears that S functions from strong-force-type potentials containing hard cores can be represented by a few Regge pole trajectories suitably combined with calculable Hankel functions and pole positions.

[6] J. B. Keller, S. I. Rubinow, and M. Goldstein, J. Math. Phys. 4, 829 (1963).
[7] V. N. Gribov and I. Ya. Pomeranchuk, Phys. Rev. Letters 9, 238 (1962).
[8] B. R. Desai and R. G. Newton, Phys. Rev. 130, 2109 (1963).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Generalized Adiabatic Invariance*

L. M. GARRIDO
Facultad de Ciencias, Universidad de Barcelona, Barcelona, Spain
(Received 16 August 1963)

In this paper we find the quantities that are adiabatic invariants of any desired order for a general slowly time-dependent Hamiltonian. In a preceding paper, we chose a quantity that was initially an adiabatic invariant to first order, and sought the conditions to be imposed upon the Hamiltonian so that the quantum mechanical adiabatic theorem would be valid to mth order. [We found that this occurs when the first $(m-1)$ time derivatives of the Hamiltonian at the initial and final time instants are equal to zero.] Here we look for a quantity that is an adiabatic invariant to mth order for any Hamiltonian that changes slowly in time, and that does not fulfill any special condition (its first time derivatives are not zero initially and finally).

I. INTRODUCTION

IN many cases it is possible to obtain an asymptotic solution of the equations of motion by perturbation theory. For many problems of quantum mechanics, it is too much to require the convergence in the mathematical sense of the formal series of perturbation method. In all practical problems, only the first several terms are calculated and the whole series may ultimately be divergent. Thus we are led to regard them as asymptotic rather than power series. The range of applicability of perturbation methods is much extended by this new interpretation.

Kato[1] shows also that the perturbation method gives asymptotic series which are correct so far as the coefficients can be calculated by means of operators within the Hilbert space. It is important to note that this was established independently of the convergence or divergence of the formal series. In fact it is rather usual that the series has only a finite number of significant terms.

The significance of the adiabatic theory can only be appreciated by noting that the adiabatic expansion in the appropriate expansion parameter is asymptotic rather than convergent. Thus, the higher-order adiabatic theory may give very great accuracy when the first term is already quite good, but it usually makes matters worse when the first term is mediocre.[2]

Therefore, in many cases it is possible to obtain an asymptotic solution of the equations of motion by perturbation theory. We are going to show in this paper how we can construct an asymptotic integral of equations of motion of a quantum mechanical system which is constant to any desired order.

The adiabatic theorem is divided into two parts. In the first place it states the existence of a virtual change in the system which may be called adiabatic transformation. Secondly, it asserts that the dynamical transformation defined by the Schrödinger equation goes over to the adiabatic transformation in the limit when the time dependence of the Hamiltonian is infinitely slow. Therefore, first we have to find the unitary operator representing the adiabatic transformation.

The construction of this adiabatic transformation constitutes the main part of the present note. Our proof is rather formal and not faultless from the point of view of mathematical rigor.

Kruskal[3] has studied the gyration of a charged particle in a magnetic field when the magnetic field at the position of the particle changes only a little during one gyration period. The guiding center approximation is then formulated by considering the ratio $\epsilon = e/m$ to be numerically small, where m is the mass of the particle and e its charge. An asymptotic analysis is called for in Kruskal's work. Once having completed the expansion of the equations of motion, he claims that there is an adiabatic invariant which is constant to all orders, an invariant that is given by the magnetic moment to lowest order. Of course this result is only asymptotic, i.e., the constancy to all orders does not mean

* This research was supported in part by the United States Air Force under Grant No. AF-EOAR, 62,89 and monitored by the European Office, Offite of Aerospace Research, and in part by the "Fundación March, Ayuda a Ciencias Físicas, 1961".

[1] T. Kato, Progr. Theoret. Phys. (Kyoto) 5, 95 (1950); 5, 207 (1950).

[2] H. Grad, Nucl. Fusion Suppl. 1, 61(1962).

[3] M. Kruskal, U. S. Atomic Energy Commission, Rept. NYO 7903 (1958).

exact constancy, but merely that the deviation from constancy goes to zero faster than any power of ϵ.

Kruskal's argument is valid quite generally for any classical system describable by a Hamiltonian, of which the particle moving in a given electromagnetic field constitutes a particular example.[4,5]

Berkowitz and Gardner[6] have shown that, in effect, the expression presented by Kruskal in Ref. 1 is indeed an asymptotic representation of the exact solution of the initial-value problem, valid as $\epsilon \to 0$. These authors have given, therefore, mathematical rigor to Kruskal's deductions.

In this paper we fix our attention on the quantum mechanical domain. We introduce the idea of generalized adiabatic invariants, which are the operators that are adiabatic invariants to any desired order in the parameter that measures the slow time variation of any explicitly time-dependent Hamiltonian. Let us observe that the generalized adiabatic invariants to mth order are not equivalent to the adiabatic invariants to mth order that were studied in a preceding paper[7] in which we gave the conditions that make the quantum mechanical adiabatic theorem valid to mth order. In Ref. 7 we chose a quantity that initially was an adiabatic invariant to first order, and sought the conditions to be imposed upon the Hamiltonian so that the quantum mechanical adiabatic theorem would be valid to mth order. [We found out that this occurs when the first $(m - 1)$ time derivatives of the Hamiltonian at the initial and final time instants are equal to zero.] Here we look for a quantity that is an adiabatic invariant to mth order for any Hamiltonian that changes slowly in time and that does not fulfill any special condition (its first time derivatives are not zero initially and finally). The new results of this paper are essentially contained in Secs. 3, 4, and 6.

In Sec. 5 we show how the generalized adiabatic invariant to mth order becomes the adiabatic invariant to mth order when the appropriate conditions are imposed upon the Hamiltonian, i.e., when we demand that the first $m - 1$ time derivatives of the Hamiltonian be zero initially and finally.

The research presented in this paper is carried on for a quantum mechanical system. However, following the operational techniques developed elsewhere[8-11] for classical mechanics, quite similar theorems can be proved for the classical domain; some clarifications may be needed, however.

In contrast with the conventional use of the operator calculus in quantum mechanics, it is generally not appropriate to apply that method to classical mechanics unless considerable care is taken. The Hilbert space upon which a quantum mechanical Hamiltonian operator is defined is simple, and is assumed not to change its properties in a time flight, or according to other changes of parameters implicitly included in the system, such as constants of motion, initial conditions and so on. The Hilbert space for classical mechanics[12,13] is, however, not so simple. It is impossible to know its Hermitian character when the energy surfaces of $(2n - 1)$ dimensions, defined by the equation $H = E$ in $2n$-dimensional phase space, are not closed. But we should mention that the formal developments of classical and quantum mechanics in Hilbert space can be rigorously used in classical mechanics if we restrict the classical system to a multiperiodic system.

Our paper starts now with a short exposition of the quantum mechanical adiabatic theorem, presented for the sake of completeness and in order to fix the notation. We generalize the method inmediately, and define the generalized adiabatic invariants to mth order. The paper ends by showing how the first part of Ref. 7 can be deduced from the present theorem by simply imposing the appropriate restrictions to the Hamiltonian.

2. QUANTUM MECHANICAL ADIABATIC THEOREM

Let us suppose that the Hamiltonian $H(t)$ of the system changes continuously from $H_0 = H(t_0)$ at the instant $t = t_0$, to $H_1 = H(t_1)$ corresponding to $t = t_1$.

We call $T = t_1 - t_0$ the time interval during which the evolution of the system takes place. We introduce the fictitious time τ that results when we measure the physical time t with the parameter T as unity. The Hamiltonian of the system at the instant $t = t_0 + \tau T$ is $H(\tau)$, an operator that is a given continuous function of τ,

[4] A. Lenard, Ann. Phys. (NY) **6**, 261 (1959); C. S. Gardner, Phys. Rev. **115**, 791 (1959).
[5] M. Kruskal, J. Math. Phys. **3**, 806 (1962).
[6] J. Berkowitz and C. S. Gardner, U. S. Atomic Energy Commission, Rept. NYO 7975 (1957).
[7] L. M. Garrido and F. J. Sancho, Physica **28**, 553 (1962).
[8] L. M. Garrido, Proc. Phys. Soc. (London) **76**, 33 (1960). L. M. Garrido and F. Gascón, Proc. Phys. Soc. (London) **81**, 1115 (1963).
[9] L. M. Garrido, J. Math. Anal. Appl. **3**, 295 (1961).
[10] L. M. Garrido, Progr. Theoret. Phys. (Kyoto) **26**, 577 (1961).
[11] L. M. Garrido and F. Gascón, Progr. Theoret. Phys. (Kyoto) **28**, 573 (1962).
[12] B. O. Koopman, Proc. Nat. Acad. Sci. U. S. **17**, 315 (1931).
[13] J. V. von Neumann, Ann. Math. **33**, 587 (1932).

and such that

$$H(0) = H_0, \qquad H(1) = H_1.$$

We are going to study the case when T is large and the evolution of the system takes place while the fictitious time changes from $\tau = 0$ to $\tau = 1$.

Let us call $U_T(\tau)$ the evolution operator where τ is the fictitious time that, together with T, was defined above.

$$i\hbar(d/d\tau)U_T(\tau) = TH(\tau)U_T(\tau), \tag{1}$$

where $H(\tau)$ is the slowly time-dependent Hamiltonian, given by the expression

$$H(\tau) = \sum_i E_i^{(1)}(\tau)P_i^{(1)}(\tau), \tag{2}$$

where $P_i^{(1)}(\tau)$ are the projection operators of the stationary states, states that we suppose discrete and nondegenerate.

Let us call $R^{(1)}(\tau)$ a unitary operator such that

$$P_i^{(1)}(\tau) = R^{(1)}(\tau)P_i^{(1)}(0)R^{(1)\dagger}(\tau). \tag{3}$$

It is completely defined by the initial condition $R^{(1)}(0) = I$ and the differential equation

$$i\hbar(d/d\tau)R^{(1)}(\tau) = K^{(1)}(\tau)R^{(1)}(\tau). \tag{4}$$

The operator $K^{(1)}(\tau)$ obeys the following commutation relations:

$$[K^{(1)}(\tau), P_i^{(1)}(\tau)] = i\hbar(d/d\tau)P_i^{(1)}(\tau), \qquad (j = 1, 2, 3, \cdots), \tag{5}$$

and is determined without ambiguity if we add the following supplementary condition:

$$P_i^{(1)}(\tau)K^{(1)}(\tau)P_i^{(1)}(\tau) = 0, \quad (j = 1, 2, 3, \cdots), \tag{6}$$

equations that yield the following expression for $K^{(1)}(\tau)$:

$$K^{(1)}(\tau) = i\hbar \sum_i [(d/d\tau)P_i^{(1)}(\tau)]P_i^{(1)}(\tau). \tag{7}$$

The unitary transformation $R^{(1)\dagger}(\tau)$, applied to the operators and vectors of Schrödinger's picture, produces a new picture: the picture of the rotating axis,

$$H^{(1)}(\tau) = R^{(1)\dagger}(\tau)H(\tau)R^{(1)}(\tau) = \sum_i E_i^{(1)}(\tau)P_i^{(1)}(0), \tag{8}$$

$$K_1^{(1)}(\tau) = R^{(1)\dagger}(\tau)K^{(1)}(\tau)R^{(1)}(\tau). \tag{9}$$

The evolution operator $U^{(1)}(\tau) \equiv R^{(1)}(\tau)U_T(\tau)$ in the new picture is defined by the initial condition $U^{(1)}(0) = I$ and the equation

$$i\hbar(d/d\tau)U^{(1)}(\tau) = [TH^{(1)}(\tau) - K_1^{(1)}(\tau)]U^{(1)}(\tau), \tag{10}$$

an equation that, in the demonstration of the adiabatic theorem in quantum mechanics, is treated by the method of perturbations considering $K_1^{(1)}(\tau)$ as the perturbation of $TH^{(1)}(\tau)$.

3. GENERALIZATION OF THE PROCEDURE

To generalize the preceding procedure we have to treat (10) as we have treated (1), i.e., as if it were the initial Schrödinger equation. We have to find the fictitious time-dependent projection operators corresponding to the Hamiltonian $H_1^{(1)}(\tau) \equiv H^{(1)}(\tau) - (1/T)K_1^{(1)}(\tau)$, operators that we will call $P_i^{(2)}(\tau)$. The subspace subtended by $P_i^{(2)}(\tau)$, according to the theorems of T. Kato, should have the same number of dimensions that the subspace projected by $P_i^{(1)}(\tau)$, i.e., the instantaneous eigenstates of $H_1^{(1)}(\tau)$ should not be degenerate. Indeed, Kato[1] says that the Hamiltonian operator $H_\lambda = H_0 + \lambda H_1$ that is a function of the real parameter λ, is a regular function of λ if the propagator $G_\lambda(z) = -[1/(H_\lambda - z)]$ is regular in the proximity of $\lambda = 0$, for some fixed z; then it is shown that the same is true for every z not belonging to the spectrum of H_0. In this paper we assume that $H_1^{(1)}(\tau)$, and all the Hamiltonians whose eigenstates and projection operators will be used, are regular in the real parameter $1/T$. When these conditions are satisfied, T. Kato shows that the multiplicity of the eigenvalues is independent of λ, or, in our case, of $1/T$. Correspondingly, all along the present paper we suppose that the successive series of instantaneous eigenvalues that will be defined are not degenerate. Perhaps this condition is much too restrictive and, as a consequence, we may leave out some cases for which the concept of generalized adiabatic invariance could still be valid. We should like to remark that it is not our intention to present here the most general case; we rather intend to introduce the generalized adiabatic invariants, taking as a model the steps developed in Sec. 1, which constitute the usual demonstration of the adiabatic theorem. Other conditions require different procedure, as can be seen in another paper of T. Kato.[14]

Following the theory of perturbations in quantum mechanics, we can write the projection operators $P_i^{(2)}(\tau)$ as a function of the projection operators of $H^{(1)}(\tau)$ and of the perturbation $(1/T)K_1^{(1)}(\tau)$. And, because the projection operators of $H^{(1)}(\tau)$ are independent of τ as can be seen from (8), we have

$$P_i^{(2)}(\tau) = P_i^{(1)}(0) + (1/T)F_i^{(1)}(\tau), \qquad (j = 1, 2, 3, \cdots), \tag{11}$$

where $F_i^{(1)}(\tau)$ contains only powers of $1/T$.

[14] T. Kato, J. Phys. Soc. Japan 5, 435 (1950).

We try now to establish a second rotating-axis picture for the operators $P_i^{(2)}(\tau)$ as follows:

$$P_i^{(2)}(\tau) = R^{(2)}(\tau)P_i^{(2)}(0)R^{(2)\dagger}(\tau), \quad (j = 1, 2, 3, \cdots), \tag{12}$$

where the unitary operator $R^{(2)}(\tau)$ is defined by the condition $R^{(2)}(0) = I$, and the differential equation

$$i\hbar(d/d\tau)R^{(2)}(\tau) = K^{(2)}(\tau)R^{(2)}(\tau), \tag{13}$$

in which the operator $K^{(2)}(\tau)$ obeys, similarly to $K^{(1)}(\tau)$, the commutation relations

$$[K^{(2)}(\tau), P_i^{(2)}(\tau)] = i\hbar(d/d\tau)P_i^{(2)}(\tau), \quad (j = 1, 2, 3, \cdots), \tag{14}$$

and complementary conditions

$$P_i^{(2)}(\tau)K^{(2)}(\tau)P_i^{(2)}(\tau) = 0, \quad (j = 1, 2, 3, \cdots). \tag{15}$$

Therefore, such an operator has the form

$$K^{(2)}(\tau) = i\hbar \sum_i [(d/d\tau)P_i^{(2)}(\tau)]P_i^{(2)}(\tau). \tag{16}$$

In the new picture, we define

$$H^{(2)}(\tau) \equiv R^{(2)\dagger}(\tau)H_1^{(1)}(\tau)R^{(2)}(\tau) = \sum_i E_i^{(2)}(\tau)P_i^{(2)}(0), \tag{17}$$

where $E_i^{(2)}(\tau)$ are the instantaneous eigenvalues of $H_1^{(1)}(\tau)$, and $P_i^{(2)}(0)$ are its initial projectors.

The new evolution operator

$$U^{(2)}(\tau) \equiv R^{(2)\dagger}(\tau)U^{(1)}(\tau)$$

obeys the equation

$$i\hbar(d/d\tau)U^{(2)}(\tau) = [TH^{(2)}(\tau) - K_2^{(2)}(\tau)]U^{(2)}(\tau), \tag{18}$$

when

$$K_2^{(2)}(\tau) \equiv R^{(2)\dagger}(\tau)K^{(2)}(\tau)R^{(2)}(\tau). \tag{19}$$

Evidently, the original evolution operator in relation to the new one is given by the expression

$$U_T(\tau) = R^{(1)}(\tau)R^{(2)}(\tau)U^{(2)}(\tau). \tag{20}$$

This process can be repeated indefinitely, and so we get the procedure to arrive at the generalized adiabatic invariance.

Let us observe that $K_2^{(2)}(\tau)$ is of the order of magnitude of $1/T$. Indeed, combining (16) with (11) we find

$$K^{(2)}(\tau) = i\hbar \sum_i [(d/d\tau)F_i^{(1)}(\tau)]P_i^{(2)}(\tau), \tag{21}$$

and $K_2^{(2)}(\tau)$ is of the same order of magnitude of $K^{(2)}(\tau)$. Therefore, in the Hamiltonian

$$H_2^{(2)}(\tau) \equiv H^{(2)}(\tau) - (1/T)K_2^{(2)}(\tau), \tag{22}$$

the term $-(1/T)K_2^{(2)}(\tau)$ is, at least, proportional to $1/T^2$. We said before that $F_i^{(1)}(\tau)$ contains only powers of $1/T$; if the first terms of the expansion of $F_i^{(1)}(\tau)$ in powers of $1/T$ do not depend on τ, expression (21) tells us that $K^{(2)}(\tau)$ is proportional to even higher powers of $1/T$ than the second.

The equation for $U^{(2)}(\tau)$ could be integrated without difficulty if we could neglect $K_2^{(2)}(\tau)$ in comparison with $TH^{(2)}(\tau)$. The solution of the Schrödinger equation that appears when we do so,

$$i\hbar(d/d\tau)\phi^{(2)}(\tau) = TH^{(2)}(\tau)\phi^{(2)}(\tau), \tag{23}$$

may be written, with the initial condition $\phi^{(2)}(0) = I$,

$$\phi^{(2)}(\tau) = \sum_i \exp(-iT\varphi_i^{(2)}(\tau)\hbar^{-1})P_i^{(2)}(0), \tag{24}$$

where

$$\varphi_i^{(2)}(\tau) = \int_0^\tau E_i^{(2)}(\tau')\,d\tau'. \tag{25}$$

If, as we will see immediately, $U^{(2)}(\tau)$ tends toward $\phi^{(2)}(\tau)$ for large T, we will have approximately

$$U_T(\tau) \simeq R^{(1)}(\tau)R^{(2)}(\tau)\phi^{(2)}(\tau). \tag{26}$$

We are now going to show that, indeed, (26) is a good approximation for $U_T(\tau)$ for large T. The complete solution of (18) can be written by means of the formula for time-dependent perturbations as follows:

$$U^{(2)}(\tau) = \phi^{(2)}(\tau)U_2^{(2)}[\tau], \tag{27}$$

where the equation satisfied by $U_2^{(2)}[\tau]$ is

$$U_2^{(2)}[\tau] = 1 + \frac{i}{\hbar}\int_0^\tau K_2^{(2)}[\tau']U_2^{(2)}[\tau']\,d\tau', \tag{28}$$

with

$$K_2^{(2)}[\tau] \equiv \phi^{(2)\dagger}(\tau)K_2^{(2)}(\tau)\phi^{(2)}(\tau).$$

We plan to show now that the kernel $K_2^{(2)}[\tau']$ is a sum of oscillating functions whose frequencies increase with T, and that therefore the integral in the second member of the Volterra equation (28) goes to zero when $T \to \infty$.

Any operator $\mathfrak{L}$ admits the following decomposition

$$\mathfrak{L} = \sum_{j,k} \mathfrak{L}_{jk}, \tag{29}$$

where we use the following notation:

$$\mathfrak{L}_{jk} = P_j^{(2)}(0)\mathfrak{L}P_k^{(2)}(0). \tag{30}$$

We shall use this decomposition for the kernel of

the integral equation (28),

$$K_2^{(2)}[\tau] = \sum_{j,k} (\exp\{(iT/\hbar)[\varphi_j^{(2)}(\tau) - \varphi_k(\tau)]\} P_j^{(2)}(0) K_2^{(2)}(\tau) P_k^{(2)}(0)) = \sum_{\substack{j,k \\ j \neq k}} (\exp\{(iT/\hbar)[\varphi_j^{(2)}(\tau) - \varphi_k^{(2)}(\tau)]\} K_2^{(2)}(\tau)_{jk}), \quad (31)$$

an expression in which we have introduced the condition $j \neq k$ because, from (15), we deduce

$$K_2^{(2)}(\tau)_{jj} = 0 \qquad (j = 1, 2, 3, \cdots). \quad (32)$$

The frequency of the oscillations can be obtained by calculating the derivative of the phase of the exponentials with respect to τ. And so we see that the frequency is never zero because the system is not degenerate, and that it is proportional to T.

Let us now consider the operator

$$F(\tau) = \int_0^\tau K_2^{(2)}[\sigma]\, d\sigma, \quad (33)$$

whose diagonal elements are all zero while their nondiagonal elements are

$$F_{j,k} = \int_0^\tau \exp\left[\frac{iT}{\hbar}(\varphi_j^{(2)}(\sigma) - \varphi_k^{(2)}(\sigma))\right] \times K_2^{(2)}(\sigma)_{jk}\, d\sigma \qquad (j \neq k), \quad (34)$$

where $K_2^{(2)}(\sigma)$ only contains negative powers of T and is, at least, proportional to $1/T$. Integrating by parts it is easy to see that $F_{jk}(\tau)$ goes as $1/T^2$. With this result we can deduce immediately, also integrating by parts, that

$$U_2^{(2)}[\tau] = 1 + O(1/T^2). \quad (35)$$

With this conclusion we arrive at the following expression for $U_T(\tau)$:

$$U_T(\tau) = R^{(1)}(\tau)R^{(2)}(\tau)\phi^{(2)}(\tau)[1 + O(1/T^2)] \qquad (T \to \infty). \quad (36)$$

4. GENERALIZED ADIABATIC INVARIANTS

If we perform only the first transformation $R^{(1)}(\tau)$, the result admits a simple physical interpretation: a system that initially is in the state selected by $P_j^{(1)}(0)$, will end up in the state selected by $P_j^{(1)}(1)$. This is the statement of the well-known theorem.

To state the generalized adiabatic theorem, we have to perform successively more than one transformations from a rotating axis picture to another rotating axis picture. Then, after the lth transformation, we have the following approximation for the evolution operator of the system:

$$U_T(\tau) = R^{(1)}(\tau)R^{(2)}(\tau) \cdots R^{(l)}(\tau)\phi^{(l)}(\tau) \times [1 + O(1/T^l)] \qquad (T \to \infty). \quad (37)$$

Let us remark now that this expression is an excellent approximation for the evolution operator of a system with any Hamiltonian that has a slow time dependence. No further requirement is imposed upon the Hamiltonian so far as its time dependence is concerned, and we can approximate its evolution operator to any desired power of $1/T$.

We may now present another consequence of the result (37), a consequence which is deduced by evaluating the following product:

$$\begin{aligned} U_T(\tau)P_j^{(l)}(0) &\simeq R^{(1)}(\tau)R^{(2)}(\tau) \cdots \\ &\quad \times R^{(l)}(\tau)P_j^{(l)}(0)R^{(l)\dagger}(\tau)R^{(l)}(\tau)\phi^{(l)}(\tau) \\ &= R^{(1)}(\tau)R^{(2)}(\tau) \cdots \\ &\quad \times R^{(l-1)}(\tau)P_j^{(l)}(\tau)R^{(l)}(\tau)\phi^{(l)}(\tau) \\ &= R^{(1)}(\tau)R^{(2)}(\tau) \cdots R^{(l-1)}(\tau)P_j^{(l)}(\tau)R^{(l-1)\dagger}(\tau) \cdots \\ &\quad \times R^{(2)\dagger}(\tau)R^{(1)\dagger}(\tau)U_T(\tau). \end{aligned} \quad (38)$$

Due to the above relation, we may now state the generalized adiabatic theorem in the following manner. A system whose state is initially a vector $|l\rangle$ whose projector is $P_j^{(l)}(0)$ will end up its evolution in a state that is a vector of the Hilbert subspace projected by the following projection operator:

$$S_j^{(l)}(1) = R^{(1)}(1)R^{(2)}(1) \cdots \times R^{(l-1)}(1)P_j^{(l-1)\dagger}(1)R^{(l-1)\dagger}(1) \cdots R^{(2)\dagger}(1)R^{(1)\dagger}(1), \quad (39)$$

with an approximation of the order of $1/T^l$.

Let us now find out the observables that are generalized adiabatic invariants. Suppose that we have performed l transformations. We get the following evolution operator in the resulting picture:

$$U^{(l)}(\tau) = \phi^{(l)}(\tau)U_l^{(l)}[\tau]. \quad (40)$$

With a procedure similar to that presented before, this will show that

$$U_l^{(l)}[\tau] = 1 + O(1/T^l). \quad (41)$$

Therefore an observable $\mathcal{L}^{(l)}$ which is a constant of the motion generated by the unitary operator $\phi^{(l)}(\tau)$ in this picture will become

$$\mathcal{L}^{(l)}(\tau) = U^{(l)\dagger}(\tau)\mathcal{L}^{(l)}U^{(l)}(\tau) = U_l^{(l)\dagger}[\tau]\mathcal{L}^{(l)}U_l^{(l)}[\tau] = \mathcal{L}^{(l)} + O(1/T^l), \quad (42)$$

where $\mathcal{L}^{(l)} = \mathcal{L}^{(l)}(0)$. Therefore the observables that commute with $H^{(l)}(\tau)$, i.e., those that commute with $H^{(l)}(0)$, are adiabatic invariants of lth order in the lth rotating-axis picture. Throughout this

paper we suppose that the projection operators $P_j^{(l)}(0)$ (for all j and any l) form a complete and orthonormal set.

Now, the evolution operator of the system is (37). Therefore the quantity

$$I^{(l)}(\tau) = R^{(1)}(\tau)R^{(2)}(\tau)\cdots \times R^{(l)}(\tau)\mathfrak{L}^{(l)}R^{(l)\dagger}(\tau)\cdots R^{(2)\dagger}(\tau)R^{(1)\dagger}(\tau) \quad (43)$$

is the generalized adiabatic invariant of lth order. It depends explicitly on time but its expected value is constant during the evolution of the system with an approximation of the order $1/T^l$. In general, therefore, the generalized adiabatic invariants depend explicitly on time; later we shall study the case when such an explicit dependence does not appear.

5. COMPARATION WITH THE ADIABATIC THEOREM OF lTH ORDER

In a preceding paper[7] we have shown that a system that initially is in a state belonging to the subspace projected by $P_j^{(1)}(0)$ will end up in the state belonging to the subspace projected by $P_j^{(1)}(1)$ with an error of the order $(1/T^l)$ when its $(l-1)$ first time derivatives of $H(\tau)$ are zero initially and finally. This is the statement of the adiabatic theorem of lth order. We now want to compare this result with the present generalized adiabatic invariance of lth order.

The comparison will be reduced to showing that we obtain the adiabatic theorem of lth order from the generalized adiabatic theorem of the same order when we add to the second theorem the extra conditions that the first $(l-1)$ time derivatives of the Hamiltonian $H(\tau)$ are zero initially and finally.

To achieve our aim, we have to show, at first, two properties of the operators that generate the successive changes of pictures, which are directly due to the fact that the first $(l-1)$ time derivatives of the Hamiltonian are zero at certain time instants.

The first one is concerned with the behavior of the unitary operators $R^{(1)}(\tau)$, $R^{(2)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$ and of their first time derivatives for a certain τ at which the first $(l-1)$ time derivatives of $H(\tau)$ are zero. Indeed in Ref. 7 we have shown that when the first $(l-1)$ time derivatives of $H(\tau)$ are zero, $K^{(1)}(\tau)$ and its first $(l-2)$ time derivatives are zero. From (9) we immediately deduce that $K_1^{(1)}(\tau)$ and its $(l-2)$ time derivatives are also zero for the same values of τ.

The unitary operator $R^{(1)}(\tau)$ satisfies the differential equation (4), from whose successive differentiation with respect to the parameter τ we find that the unitary operator $R^{(1)}(\tau)$ has its $(l-1)$ time derivatives equal to zero for the values of τ for which the first $(l-1)$ time derivatives of $H(\tau)$ are equal to zero. This result, together with the fact that the Hamiltonian $H_1^{(1)}(\tau)$ was defined by means of the relation

$$H_1^{(1)}(\tau) = H^{(1)}(\tau) - (1/T)K_1^{(1)}(\tau) = R^{(1)\dagger}(\tau)H(\tau)R^{(1)}(\tau) - (1/T)K_1^{(1)}(\tau), \quad (44)$$

allow us to show that, for the above-mentioned values of τ, the $(l-2)$ first time derivatives of $H_1^{(1)}(\tau)$ are zero while the $(l-1)$st time derivative of the same operator is

$$\frac{d^{l-1}}{d\tau^{l-1}}H_1^{(1)}(\tau) = -\frac{1}{T}\frac{d^{l-1}}{d\tau^{l-1}}K_1^{(1)}(\tau). \quad (45)$$

The results of Ref. 7 and the above properties of the operators $K_1^{(1)}(\tau)$, $R^{(1)}(\tau)$, and $H_1^{(1)}(\tau)$ can be extended further by the same procedure to all the series of similar operators that we have introduced in Sec. 3. And so we arrive at the first statement that we needed, i.e., to the fact that, for the values of τ for which the first $(l-1)$ time derivatives of $H(\tau)$ are zero, $K_1^{(1)}(\tau)$ and its first $(l-2)$ time derivatives, the first $(l-1)$ time derivatives of $R^{(1)}(\tau)$, and the first $(l-2)$ time derivatives of $H_1^{(1)}(\tau)$ are zero; $K_2^{(2)}(\tau)$ and its first $(l-3)$ time derivatives, the first $(l-2)$ time derivatives of $R^{(2)}(\tau)$, and the first $(l-3)$ time derivatives of $H_2^{(2)}(\tau)$ are zero; and so on.

Given this first statement, and remembering that $(1/T)K_1^{(1)}(\tau)$ is the perturbation that, by the methods of time-independent perturbations, yields $P_j^{(2)}(\tau)$ from $P_j^{(1)}(0)$, in agreement with (11) and that similarly $P_j^{(3)}(\tau)$ is deduced from $P_j^{(2)}(0)$ by a perturbation proportional to $K_2^{(2)}(\tau)$, we deduce that

$$P_j^{(1)}(0) = P_j^{(2)}(\tau) = P_j^{(3)}(\tau) = \cdots = P_j^{(l-1)}(\tau) = P_j^{(l)}(\tau), \quad (46)$$

for the values of τ for which the $(l-1)$ time derivatives of $H(\tau)$ are zero; the operators $K_1^{(1)}(\tau)$; $K_2^{(2)}(\tau)$; $\cdots$; $K_{l-1}^{(l-1)}(\tau)$ will be zero for the same values of τ, where $K_{l-1}^{(l-1)}(\tau)$ characterizes the change of rotating-axis picture.

The second property we must show regarding the unitary operators $R^{(2)}(\tau)$, $R^{(3)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$ is that, for the values of τ for which the first $(l-1)$ time derivatives of $H(\tau)$ are zero, the unitary operators $R^{(2)}(\tau)$, $R^{(3)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$ [$R^{(1)}(\tau)$ is not included] commute with any of the projectors $P_j^{(1)}(0)$. Indeed, to show this theorem it is sufficient to apply (46) to the relations like (3), (12), and

so on, successively. Therefore, from (12) and (46) we have

$$P_i^{(2)}(\tau) = P_i^{(1)}(0) = R^{(2)}(\tau)P_i^{(2)}(0)R^{(2)\dagger}(\tau), \quad (47)$$

which is equivalent to

$$[R^{(2)}(\tau), P_i^{(1)}(0)] = 0 \qquad (j = 1, 2, 3, \cdots), \quad (48)$$

and similarly for the other unitary operators $R^{(3)}(\tau)$, $R^{(4)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$. However, we cannot deduce that, for these values of τ, $R^{(1)}(\tau)$ also commutes with any $P_i^{(1)}(0)$. If the projection operators $P_i^{(1)}(0)$ for all j constitute a complete set, relation (48) implies that the unitary operators $R^{(2)}(\tau)$, which satisfy the same, are unity. Therefore, if the $(l-1)$ first time derivatives of $H(\tau)$ are zero at the initial and final instants, the unitary operators $R^{(2)}(\tau)$, $R^{(3)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$ will commute with all $P_i^{(1)}(0)$, and then we will deduce that the system that initially was at a state whose projector is $P_i^{(1)}(0)$ will end up at a state whose projector is $P_i^{(1)}(1)$, with an error smaller that $1/T^l$. This is so because at the final instant, in this case, the projector $S_i^{(l)}(1)$ of (39) becomes

$$\begin{aligned} &R^{(1)}(1)R^{(2)}(1)\cdots R^{(l-1)}(1)P_i^{(1)}(0) \\ &\qquad \times R^{(l-1)\dagger}(1)\cdots R^{(2)\dagger}(1)R^{(1)\dagger}(1) \\ &\quad = R^{(1)}(1)P_i^{(1)}(0)R^{(1)\dagger}(1) = P_i^{(1)}(1). \end{aligned} \quad (49)$$

This concludes our comparison of the generalized adiabatic invariance with the adiabatic invariance of order l. Let us study now the relations between the corresponding adiabatic invariants in both cases. Indeed, because in this case $K_{l-1}^{(l-1)}(\tau)$, $\cdots$, $K_2^{(2)}(\tau)$, $K_1^{(1)}(\tau)$ are zero for the values of τ when the first $(l-1)$ time derivatives of $H(\tau)$ are zero, we deduce, for those values of τ,

$$\begin{aligned} &H_1^{(1)}(\tau) = H^{(1)}(\tau), \qquad H_2^{(2)}(\tau) = H^{(2)}(\tau), \cdots \\ &\qquad\qquad H_{l-1}^{(l-1)}(\tau) = H^{(l-1)}(\tau), \\ &E_i^{(2)}(\tau) = E_i^{(1)}(\tau), \qquad E_i^{(3)}(\tau) = E_i^{(2)}(\tau), \cdots \\ &\qquad\qquad E_i^{(l)}(\tau) = E_i^{(l-1)}(\tau), \end{aligned} \quad (50)$$

and from relations like (8), (17), and so on, we get

$$\begin{aligned} H^{(2)}(\tau) &\equiv R^{(2)\dagger}(\tau)H_1^{(1)}(\tau)R^{(2)}(\tau) \\ &= R^{(2)\dagger}(\tau)H^{(1)}(\tau)R^{(2)}(\tau) \\ &= \sum_i E_i^{(1)}(\tau)R^{(2)\dagger}(\tau)P_i^{(1)}(0)R^{(2)}(\tau) = H^{(1)}(\tau), \end{aligned} \quad (51)$$

given (48) and similar relations. Continuing the same procedure, we deduce, for these particular values of τ,

$$H^{(1)}(\tau) = H^{(2)}(\tau) = \cdots = H^{(l-1)}(\tau) = H^{(l)}(\tau), \quad (52)$$

which are $(l-1)$ equations.

Therefore, the constants of the motion $\mathfrak{L}^{(l)}$ of $H^{(l)}(\tau)$, are also, in this case, the constants of the motion of $H^{(1)}(\tau)$, constants which, since $P_i^{(1)}(0)$ in (8) are a complete set of projectors, are also the constants of the motion of $H^{(1)}(0)$,

$$\mathfrak{L}^{(l)} = \mathfrak{L}^{(1)}. \quad (53)$$

But we have shown before that $R^{(2)}(\tau)$, $R^{(3)}(\tau)$, $\cdots$, $R^{(l)}(\tau)$, for these particular values of τ, are unity. Therefore, in this case, (43) becomes

$$I^{(1)}(\tau) = R^{(1)}(\tau)\mathfrak{L}^{(1)}R^{(1)\dagger}(\tau), \quad (54)$$

a quantity that now is the generalized adiabatic invariant of lth order.

There remains to identify

$$I^{(1)}(\tau) = R^{(1)}(\tau)\mathfrak{L}^{(1)}R^{(1)\dagger}(\tau),$$

where $\mathfrak{L}^{(1)}$ are the constants of the motion of $H^{(1)}(\tau)$ with the constants of the motion of $H(\tau)$. But from (8) we have that the commutator $[H^{(1)}(\tau), I^{(1)}(\tau)]=0$ implies

$$\begin{aligned} &[H(\tau), R^{(1)}(\tau)\mathfrak{L}^{(1)}R^{(1)\dagger}(\tau)] \\ &\qquad = [H(\tau), I^{(1)}(\tau)] = 0, \end{aligned} \quad (55)$$

and therefore for these values of τ which make zero the first $(l-1)$ time derivatives of $H(\tau)$, the instantaneous constants of the motion of $H(\tau)$ are adiabatic invariants of lth order. This is the meaning of the adiabatic theorem of lth order.

6. GENERAL CASE

The most general case is that situation in which we study the generalized adiabatic invariance of order l when the first $(l'-1)$ time derivatives of $H(\tau)$ are zero initially and finally. We can study two alternatives: either $l > l'$ or $l < l'$. The results for the third alternative $l = l'$ were studied in the preceding section.

Let's begin with the case $l > l'$. We deduce, as before, that for the values of τ for which the first $(l'-1)$ time derivatives of $H(\tau)$ are zero, $K_1^{(1)}(\tau)$ and its first $(l'-2)$ time derivatives, the first $(l'-1)$ time derivatives of $R^{(1)}(\tau)$, and the first $(l'-2)$ time derivatives of $H_1^{(1)}(\tau)$, are zero; $K_2^{(2)}(\tau)$ and its first $(l'-3)$ time derivatives, the first $(l'-2)$ time derivatives of $R^{(2)}(\tau)$, and the first $(l'-3)$ time derivatives of $H_2^{(2)}(\tau)$ are zero, and so on.

We will deduce that, as in (46),

$$P_i^{(1)}(0) = P_i^{(2)}(\tau) = \cdots = P_i^{(l')}(\tau), \quad (56)$$

and since the $P_i^{(1)}(0)$ form a complete set, we will deduce that

$$R^{(2)}(\tau) = R^{(3)}(\tau) = \cdots = R^{(l')}(\tau) = I \qquad (57)$$

for these particular values of τ. Nothing can be said about $R^{(1)}(\tau)$ and about $R^{(h)}(\tau)$, where $h > l'$. Therefore in this case the adiabatic theorem may be stated as follows; a system whose state is initially a vector $|l_i\rangle$ whose projector is $P_i^{(1)}(0)$ will end up its evolution in a state that is a vector of the Hilbert space whose projector is

$$S_i^{(l)}(1) = R^{(1)}(1)R^{(l'+1)}(1) \cdots \times R^{(l-1)}(1)P_i^{(l)}(1)R^{(l-1)\dagger}(1) \cdots \times R^{(l'+1)\dagger}(1)R^{(1)\dagger}(d), \qquad (58)$$

with an approximation of the order $1/T^l$. Similarly, we will obtain an expression for the adiabatic invariant in this case.

The last case to study corresponds to $l < l'$. Relation (57) is also valid now initially and finally, and here the generalized adiabatic theorem of order l is completely equivalent to the adiabatic theorem of l'th order.

7. CONCLUSION

We would like to remark that even better approximation for the generalized adiabatic invariance can be obtained in the case when the expansion in powers of $1/T$ of $F_i^{(1)}(\tau)$ will make that the first term of this expansion independent of the fictitious time τ, because then relations like (16) would give a $K_2^{(2)}(\tau)$ that will be smaller than the one used in our exposition. Besides this point, there remain others whose study may be of some profit. We mention, for instance, the combination of the generalized adiabatic invariance with the extended adiabatic invariance which, following the paper of S. Tamor,[15] will be published by us soon.[16] The reader may like to compare the present new concept of generalized adiabatic invariance with some relevant previous work on adiabatic invariants of any order.[4, 17, 18]

---

[15] S. Tamor, J. Nucl. Energy PC **1**, 199 (1960).
[16] L. M. Garrido (to be published).
[17] A. Messiah, *Mécanique Quantique* (Dunod Cie., Paris, 1960).
[18] L. M. Garrido, Coll Math. **13**, 219 (1961).

Convergence of Yukawa Theories with a Finite Number of Interacting Boson Modes*

WILLIAM M. FRANK
U. S. Naval Ordnance Laboratory, Silver Spring, Maryland
(Received 7 August 1963)

Convergence of the perturbation expansion of the S matrix is demonstrated for a field theory where a quantized fermion field with regularized propagator interacts with a finite number of quantized boson modes. Conventional graphical techniques and combinatorial analysis are used to establish this result, derived earlier by Edwards, but we believe not properly clarified. A zero radius of convergence is demonstrated for the theory where a Yukawa-type interaction takes place between one boson field interacting in only a finite number of modes with another boson field coupled to it bilinearly. The relation of this result to the convergence of infinite coupled mode theories is discussed.

I. INTRODUCTION

THE subject of the convergence of field-theoretic perturbation expansions has received relatively little substantial treatment beyond the speculative or qualitative stage. In the opinion of the writer, the strongest and most relevant result which as appeared to date is that of Edwards[1] who, in a brief note with the barest detail, claims the convergence of a Yukawa interaction theory with a momentum "cutoff". Edwards' result is obtained through functional integration techniques and the cutoff procedure consists of representing the classical boson potential as a superposition of a finite number of modes. In this paper a detailed proof of the same result is presented, based on more conventional graphical techniques. Our results apply to a class of infinite coupled boson mode theories, but will not apply to the crucial case where the boson field propagation is a function of only the distance between the points. We also calculate the S matrix for the case where the Yukawa interaction couples two boson fields, one linearly and one bilinearly, and we show that such a theory has a zero radius of convergence. This is in agreement with similar results obtained by other authors for the coupling of three boson fields,[2] and conclusively as well as concisely demonstrates the crucial role of the phase cancellations due to Fermi statistics for convergence of the perturbation expansions.

Edwards claims that his proof demonstrates the convergence of relativistic field theories in the presence of a cutoff. We believe this interpretation must be qualified, as the "cutoff" procedure he uses, i.e., coupling to only a finite number of boson modes, is not quite the same thing as a form-factor cutoff in an interaction which allows an infinite number of modes. Thus, the trigometric polynomials of maximum finite order; i.e., a finite number of modes, do not constitute a complete set for a finite interval (i.e., a cutoff range of integration). The Hilbert space for an infinite number of boson modes is nonseparable, while that for a finite number is separable. The mathematical difference between such theories is likely to be profound. The inequivalence of a finite number of modes to a momentum cutoff is particularly apparent when one considers fermions. A finite number of modes by the exclusion principle implies a finite maximum number of interacting particles in the field at once, which is not the case if the interaction is merely cutoff. A proof of the convergence of perturbation expansions for cutoff theories would be a major interest as such nonlocal theories would be candidates for a "correct" physical theory. Recent investigations indicate that unitarity[3] and perhaps some form of macroscopic causality[4] can be realized in such theories. The straightforward estimation procedures carried out for the case of a finite number of boson modes would require careful refinement for applicability to an infinite number of modes. We discuss this in Sec. V.

The convergence of the vacuum-to-vacuum S matrix and energy shift will be presented. This amplitude consists of vacuum polarization effects. The vacuum-to-vacuum S matrix S_v is found to

* Preliminary report presented at New York Physical Society Meeting, 1963.

[1] S. F. Edwards, Phil. Mag. **45**, 758 (1954); **46**, 569 (1955).

[2] C. A. Hurst, Proc. Cambridge Phil. Soc. **48**, 625 (1952). W. Thirring, Helv. Phys. Acta **26**, 33 (1953). A. Petermann, Arch. Sci. Phys. Nat. **6**, 5 (1953). R. Utiyama and T. Imamura, Progr. Theoret. Phys. (Kyoto) **9**, 431 (1953).

[3] E. C. G. Sudarshan, Phys. Rev. **123**, 2183 (1961); K. Yokoyama, Progr. Theoret. Phys. (Kyoto) **26**, 131 (1961); D. A. Kirzhnits, Zh. Eksperim. i Teor. Fiz. **41**, 551 (1961) [English transl.: Soviet Phys.—JETP **14**, 395 (1962)].

[4] B. Ferretti, Nuovo Cimento **27**, 1503 (1963).

be an analytic function of the coupling constant g, and under certain conditions it is entire. The estimation technique of Caianiello[5] using the Hadamard inequality always gives a finite radius of convergence for S_v. The vacuum self-energy (E_v) is found to have logarithmic singularities. Application of these methods to other than vacuum-to-vacuum matrix elements of the S matrix is straightforward and is briefly discussed in Sec. IV. An upper bound to S_v is found which can be evaluated by generating function techniques and the analytic properties follow from theorems in Fredholm theory and complex analysis. A proof is constructed in Sec. II for the simple model of one interacting boson mode. The technique consists of relating the S matrix for a quantized boson field to that for an unquantized boson field. In Sec. III we show how S_v for a finite number of quantized interacting modes can be related to that for an unquantized real boson field. The convergence of the unquantized-field case and the consequent result for the quantized case are derived in Sec. IV. The results are discussed in Sec. V.

II. THE ONE-MODE MODEL

In the interest of simplicity we shall consider neutral scalar or pseudoscalar mesons interacting with fermions. First, we assume that only one four-dimensionally square-integrable boson mode interacts with the fermions.

We proceed from the expression for the S-matrix perturbation expansion in terms of propagators of the fermion and boson fields. Loosely speaking, we derive this from the standard perturbation expansion of the S matrix in Hamiltonian form,

$$S = \sum_{n=0}^{\infty} \frac{(-ig)^n}{n!} \times \int dx_1 \cdots \int dx_n \, T[H_I(x_1) \cdots H_I(x_n)] \,, \tag{1}$$

where g is the coupling constant and the interaction density without g is given by

$$H_I(x) = :\bar{\psi}(x)\Gamma\psi(x):\phi(x), \tag{2}$$

where Γ is some matrix operator. We are not, however, committing ourselves to the existence of a Hamiltonian. All that is relevant about the fermion field $\psi(x)$ is the propagator $K_{\alpha\beta}(x - x')$, which we assume to be regularized sufficiently to make integrals of interest finite. Indeed, such a regularized propagator does not follow from the coupling in Eq. (2) if $\psi(x)$ is the true free-field interaction picture operator. Equations (1) and (2) are written as a heuristic basis for the expression of the S matrix in terms of propagators which is our essential starting point. We presently restrict the boson field $\phi(x)$ to have the simple form

$$\phi(x) = (a + a^*)u(x), \tag{3}$$

where a, a^* are the usual annihilation and creation operators for the boson mode $u(x)$ which we assume is L^{x1}. The boson propagator is $\Delta(x, x') = u(x)u(x')$. Equation (3) does not correspond to a positive–negative frequency decomposition of the field. In Sec. III a more general form of boson propagation will be treated. The vacuum-to-vacuum S matrix element obtained from Eqs. (1) and (2) is

$$S_v \equiv \sum_{n=0}^{\infty} S_v^{(n)} = \sum_{\substack{n=0 \\ \text{even}}}^{\infty} \frac{(-ig)^n}{n!} \sum_{\substack{\beta_1 \cdots \beta_n \\ \alpha_1 \cdots \alpha_n}} \Gamma_{\beta_1\alpha_1} \cdots \Gamma_{\beta_n\alpha_n}$$

$$\times \int \cdots \int dx_1 \cdots dx_n \langle T[:\bar{\psi}_{\beta_1}(x_1)\psi_{\alpha_1}(x_1): \cdots$$

$$\times :\bar{\psi}_{\beta_n}(x_n)\psi_{\alpha_n}(x_n):]\rangle_0 \langle T[\varphi(x_1) \cdots \varphi(x_n)]\rangle_0. \tag{4}$$

The fermion Green's functions in Eq. (4) can be expressed as a determinant of fermion propagators,[6]

$$K_{\beta\alpha}(x_1, \cdots x_n) \equiv \langle T[:\bar{\psi}_{\beta_n}(x_n)\psi_{\alpha_n}(x_n): \cdots :\bar{\psi}_{\beta_1}(x_1)\psi_{\alpha_1}(x_1):]\rangle_0$$

$$= \begin{vmatrix} 0 & K_{12} & K_{13} & \cdots & K_{1n} \\ K_{21} & 0 & K_{23} & \cdots & \\ K_{31} & K_{32} & 0 & \cdots & \\ \vdots & \vdots & & & \\ K_{n1} & & & & 0 \end{vmatrix}, \tag{5}$$

where $K_{ij} \equiv K(x_i - x_j)$. The matrix indices have been suppressed. The progapator $K(x - y)$ one assumes to be sufficiently regularized so that $\int dx \, |K(x)|^2 < \infty$. From Eq. (3),

$$\langle T[\phi(x_1) \cdots \phi(x_n)]\rangle_0 = (n-1)!! \, u(x_1) \cdots u(x_n), \tag{6}$$

where

$$(n-1)!! = n!/2^{\frac{1}{2}n}(\tfrac{1}{2}n)! = 1 \cdot 3 \cdot 5 \cdots (n-1). \tag{7}$$

Had $\phi(x)$ been a classical (i.e., unquantized) field, $S_v^{(n)}$ would have, in place of the factor

$$(n-1)!!u(x_1) \cdots u(x_n),$$

merely been $u(x_1) \cdots u(x_n)$. By means of the

[5] E. R. Caianiello, Nuovo Cimento **3**, 223 (1956).

[6] See, e.g., E. R. Caianiello, Nuovo Cimento **13**, 637 (1959).

relation

$$(n-1)!! = \pi^{-\frac{1}{2}} 2^{\frac{1}{2}n} \int_0^\infty dv\, e^{-v} v^{\frac{1}{2}(n-1)} \quad (n \text{ even}), \tag{8}$$

one can rewrite Eq. (4) as

$$S_e \equiv S(g) = \pi^{-\frac{1}{2}} \int_0^\infty dv\, e^{-v} \Bigg[\sum_{n=0}^\infty \frac{[-ig(2v)^{\frac{1}{2}}]^n}{n!} \times \sum_{\substack{\beta_1 \cdots \beta_n \\ \alpha_1 \cdots \alpha_n}} \Gamma_{\beta_1 \alpha_1} \cdots \Gamma_{\beta_n \alpha_n} \int \cdots \int dx_1 \cdots dx_n \times u(x_1) \cdots u(x_n) K_{\beta\alpha}(x_1, \cdots x_n) \Bigg], \tag{9}$$

provided the interchange of summation and integration is legitimate. From Eq. (9)

$$\begin{aligned} S(g) &= \pi^{-\frac{1}{2}} \int_0^\infty dv\, v^{-\frac{1}{2}} e^{-v} S_e^c[g(2v)^{\frac{1}{2}}] \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty dt\, e^{-\frac{1}{2}t^2} S_e^c(tg) \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{g} \int_0^\infty dt\, e^{-t^2/2g^2} S_e^c(t), \end{aligned} \tag{10}$$

where the last expression holds in a wedgelike region in the complex g plane. The function $S_e^c(g)$ of Eq. (9) is the even part of the vaccum-to-vacuum S matrix $S^c(g)$ for the interaction of a quantized fermion field with the c-number field $u(x)$.[7] The latter is, however, the modified Fredholm determinant[8] based on the (matrix) kernel

$$i\Gamma K(x - x')u(x'), \tag{11}$$

which, with proper regularization of $K(x - x')$, is L^2. The quantity $S^c(g)$ is known from the Fredholm theory for L^2 kernels[9] to be an entire function of g. It can be expressed (see Sec. IV) in the closed form

$$S^c(g) = \exp -\left[\sum_{m=2}^\infty \frac{i^m g^m}{m} \sigma_m \right], \tag{12}$$

where σ_m denotes the ring integral

$$\sigma_m \equiv \int \cdots \int dx_1 \cdots dx_m \,\mathrm{tr}\, [\Gamma K(x_1 - x_2) \times \Gamma K(x_2 - x_3) \cdots \Gamma K(x_m - x_1)] u(x_1) \cdots u(x_m). \tag{13}$$

Equation (12) is merely a manifestation of the formal operator relation[10]

$$\det (1 - A) = \exp [\mathrm{tr} \ln (1 - A)].$$

It remains to be seen how the relation between $S(g)$ and $S^c(g)$ as given by Eq. (10) reflects on the analytic properties in g of $S(g)$.

Were we to consider the $\psi(x)$ field of Eq. (2) to be a boson field (of spin zero, say) so that we had a three boson field coupling in place of a yukawa coupling, the relation of the type Eq. (10) would continue to hold between $S_B(g)$ (the vacuum-to-vacuum S matrix for the boson case) and $S_{B,e}^c(g)$ (the even part of the vacuum S matrix) for coupling to a c-number source (see Sec. IV). The vacuum S matrix for this case would be given by Eq. (5) where $K(x - y)$ corresponds to the propagator for boson ψ field (which is a one-dimensional matrix) and a permanent[6] replaces the determinant. For this case, Eq. (12) is simply replaced by

$$S_B^c(g) = \exp \left[\sum_{m=2}^\infty \frac{(-i)^m g^m \sigma_m}{m} \right] \tag{14}$$

with σ_m again defined by Eq. (13) (without the superfluous tr symbol). We shall see how significant is the difference in the signs of the exponential between Eqs. (12) and (14).

An upper bound to $S(g)$ of Eq. (4) is clearly

$$\bar{S}(g) = \pi^{-\frac{1}{2}} \int_0^\infty dv\, v^{-\frac{1}{2}} e^{-v} \Bigg\{ \sum_{n=0}^\infty \frac{[\gamma g(2v)^{\frac{1}{2}}]^n}{n!} \times \sum_{\substack{\beta_1 \cdots \beta_n \\ \alpha_1 \cdots \alpha_n}} |K_{\beta\alpha}(x_1, \cdots x_n)|\, |u(x_1) \cdots u(x_n)| \Bigg\}, \tag{15}$$

where γ is the absolute value of the maximum matrix element of the matrix Γ. It is somewhat more convenient to study, in place of $\bar{S}(g)$, the related quantity $\mathfrak{S}(g)$ which will have at least the same radius of convergence. It is defined by

$$\mathfrak{S}(g) \equiv \sum_{n=0}^\infty \frac{g^n}{n!} (n-1)!! \sum_{\substack{\beta_1 \cdots \beta_n \\ \alpha_1 \cdots \alpha_n}} \int \cdots \int dx_1 \cdots dx_n \times |u(x_1) \cdots u(x_n)|\, K_{\beta\alpha}(x_1, \cdots x_n), \tag{16}$$

where the absolute value sign has been removed from $K_{\beta\alpha}(x_1 \cdots x_n)$, and we have set $\gamma = 1$ for simplicity. From Eq. (8),

$$\mathfrak{S}(g) = \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^\infty dt\, e^{-\frac{1}{2}t^2} \mathfrak{S}_e^c(tg), \tag{17}$$

where $\mathfrak{S}_e^c(g)$ is the even part of the modified Fredholm determinant[8] based on the Hermitian (matrix) kernel

$$\mathcal{K}(x, y) = K(x - y)\, |u(x)u(y)|^{\frac{1}{2}}, \tag{18}$$

and is given by Eq. (16) with the $(n-1)!!$ factor removed from the right-hand side. The $S^c(g)$ is

[7] A. Salam and P. T. Matthews, Phys. Rev. **90**, 690 (1953).
[8] F. Smithies, in *Integral Equations* (Cambridge University Press, Cambridge, England, 1962), p. 97.
[9] Ref. 8, p. 96, Theorem 6.5.2.
[10] See, e.g., M. Baker, Ann. Phys. (N. Y.) **4**, 271 (1958).

also given by (see Sec. IV)

$$\mathcal{S}^c(g) = \exp -\left[\sum_{m=2}^{\infty} \frac{(-g)^m \sigma_m}{m}\right], \qquad (19)$$

where the σ_m would be defined accordingly. The kernel $K(x, y)$ of Eq. (18) is L^2 in $x \times y$ space if $K(x)$, $u(x)$ separately are in x space. In Sec. IV we show that $S^c(g)$ is analytic in g and that $S(g)$ will have nonvanishing radius of convergence if $K(x, y)$ is L^2 and will have an infinite radius of convergence if $K(x, y)$ also obeys some other conditions. The same will follow for $S(g)$.

The vacuum self-energy defined as i times the sum of all connected loop diagrams is just $i \ln S$, and always has a finite radius of convergence.

The convergence of $S_v(g)$ in this one-mode model can also be established by application of the Hadamard inequality to Eq. (4) as done by Caianiello.[5] For $K(x)$ regularized to be finite everywhere and an L^1 mode $u(x)$, one would majorize the fermion Green's function in Eq. (4) by the upper bound

$$|K(x)| < \kappa(n-1)^{\frac{1}{2}n},$$

where $\kappa = \max |K(x)|$. Then

$$|S| \leq \sum_{n=0}^{\infty} (4\kappa\xi g)^n \frac{(n-1)!!\,(n-1)^{\frac{1}{2}n}}{n!} < \infty, \qquad (20)$$

where $\xi = \int dx\, |u(x)|$. The majorizing series in Eq. (20) has a finite radius of convergence.

The Hadamard inequality thus implies convergence but under stronger conditions [i.e., finite bound on $K(x)$]. The L^2 nature of the boson mode replaces the additional assumption of finite space–time volume of Caianiello. Buccafurri and Caianiello[11] specifically devote themselves to demonstrating an infinite radius of convergence for the S matrix when one is restricted to a finite space–time region and only a finite number of boson and fermion modes interact. This result will be shown for a class of infinite mode interactions in Sec. IV. The condition of a finite number of interacting fermion modes, as we have pointed out, means only a finite total number of fermions and therefore does not correspond to a field theory as far as the fermions go. The convergence of theories which have only a finite number of fermions present obtains when even an infinite number of boson modes interact with a form factor in the interaction and no restriction on the time modes at all.[12] Caianiello's[5,11] limitation to a finite number of plane-wave modes interacting in a finite space–time volume is an unnecessary restriction concealing the logic of the situation.

In the case where ψ is chosen to be a boson field, Eq. (17) continues to relate the corresponding quantities and

$$\mathcal{S}_B^c(g) = \exp\left[\sum_{m=2}^{\infty} \frac{g^m \sigma_m}{m}\right], \qquad (21)$$

with the σ_m correspondingly defined (see Sec. IV). The quantity of Eq. (21) necessarily has a finite (but nonvanishing) radius of convergence and a lemma of Watson[13] tells us that $\mathcal{S}(g)$ defined by Eq. (17) must then have a zero radius of convergence. This is seen straightforwardly from the way in which Eq. (17) relates the power series coefficients of $\mathcal{S}_B^c(g)$ and $\mathcal{S}_B(g)$. The coefficients of $\mathcal{S}_B^c(g)$ because of the finite radius of convergence have geometrical growth, and the Gaussian transform of Eq. (17) changes this to an $(n!)^{\frac{1}{2}}$ type of growth giving divergence. This result emphasizes the significance of the statistics to convergence considerations.[14] As will be discussed elsewhere, the divergence of perturbation expansions for theories where three boson fields are coupled, has nothing to do with field theory and the higher manifold of infinitude of states that it entails. A cubic anharmonic perturbation in a simple one-dimensional harmonic oscillator has zero radius of convergence in the anharmonic coupling constant.

III. THE MANY-MODE MODEL

Convergence of $S(g)$ can now be established for a boson propagator $\Delta(x, x')$ which is a degenerate symmetric kernel; i.e., of the form

$$\Delta(x, x') = \sum_{i=1}^{N} u_i(x) u_i(x'). \qquad (22)$$

The $u(x)$ need not be real. Only symmetric kernels will be considered. These are not the same as complex Hermitian kernels. The true propagator resulting from the interaction of an infinite number of modes when transformed to the Euclidean metric is, but for a trivial phase factor, a symmetric kernel, and can be represented in form Eq. (22) if $N = \infty$. Convergence for unsymmetric kernels will not follow

[11] A. Buccafurri and E. R. Caianiello, Nuovo Cimento **8**, 171 (1958).

[12] The result follows using the same methods as in W. M. Frank, J. Math. Phys. **3**, 272 (1962).

[13] H. Jeffreys and B. Jeffreys, *Methods of Mathematical Physics* (Cambridge University Press, Cambridge, England, 1956), 3rd ed., Chap. 17.

[14] Misconceptions about this continue to appear in the literature. See, e.g., S. Frautschi, Progr. Theoret. Phys. (Kyoto) **22**, 882 (1952), p. 882, footnote*; A. Peres, J. Math. Phys. **4**, 332 (1963).

by the method to be presented, but this extension will not be considered here.[15]

A boson propagator of the form Eq. (23) derives from the field operator,

$$\varphi(x) = \sum_i (a_i + a_i^*)u_i(x), \quad [a_i^*, a_j] = \delta_{ij}. \tag{23}$$

Again the quantized boson S matrix can be derived from an unquantized-field S matrix by means of an integral transformation. To do this we consider the fermion field in interaction with the classical field

$$u_{v\theta}(x) = \sum_{i=1}^{N} (2v_i)^{\frac{1}{2}} \cos\theta_i u_i(x), \tag{24}$$

where v_i and θ_i are parameters introduced for integral transformation purposes. The vacuum-to-vacuum S matrix for fermions interacting with the classical field $u_{v\theta}(x)$ will be denoted by $S^c(g; v, \theta)$. The vacuum-to-vacuum S matrix for interaction of the fermion field with the quantized boson field $\varphi(x)$ of Eq. (23) is expressible as

$$S(g) = \frac{1}{(2\pi)^N} \int_0^\infty dv_1 \cdots \int_0^\infty dv_N \times \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_N \exp -\left(\sum_{i=1}^{N} v_i\right) S_e^c(g; v, \theta), \tag{25}$$

where $S_e^c(g; v, \theta)$ is the even part in g of $S^c(g; v, \theta)$ which is the vacuum S matrix for the interaction of the ψ field with the single boson mode $u_{v\theta}(x)$. Eq. (25) follows by expanding the nth-order boson Green's function

$$\Delta_n(x_1, \cdots x_n) = \langle T[\varphi(x_1) \cdots \varphi(x_n)]\rangle_0 = \sum_{\substack{\text{appropriate} \\ \text{permutations}}} \prod_{l=1}^{\frac{1}{2}n} \Delta(x_1^{(l)}, x_2^{(l)}), \tag{26}$$

where $x_1^{(1)}, x_2^{(1)}, x_1^{(2)}, x_2^{(2)}, \cdots x_1^{(l)}, x_2^{(l)}$ ranges over all permutations of $x_1, x_2 \cdots x_n$ counting only once permutations which transform into each other under interchange of any $x_1^{(k)}$ with $x_2^{(k)}$. Substitution for $\Delta(x_1^{(l)}, x_2^{(l)})$ from Eq. (22) into Eq. (26) and distributing the product of Eq. (26) leads to an expression for $\Delta_n(x_1, x_2, \cdots x_n)$ as a sum of "subterms." Each subterm corresponds to a partition of the vertex coordinates $x_1 \cdots x_n$ among the N modes, where the coordinates belonging to the jth mode ($j = 1, 2 \cdots N$) are those which are the arguments of $u_j(x)$ in the given subterm. Thus for $N = 2$, $n=4$, $\Delta_4(x_1, x_2, x_3, x_4)$ is a sum of twelve "subterms",

$$\Delta_4(x_1, x_2, x_3, x_4) = u_1(x_1)u_1(x_2)u_1(x_3)u_1(x_4) + u_1(x_1)u_1(x_2)u_2(x_3)u_2(x_4) + u_2(x_1)u_2(x_2)u_1(x_3)u_1(x_4) + u_2(x_1)u_2(x_2)u_2(x_3)u_2(x_4) + (x_2 \rightleftarrows x_3) + (x_2 \rightleftarrows x_4).$$

For a given subterm let there be s_1 vertices as arguments of u_1, s_2 vertices which are arguments of u_2, $\cdots$, s_n vertices belonging to the mode u_n. The s_j must be even and $\sum_j s_j = N$. A particular subterm is of the form

$$u_1(k_1^1, \cdots k_{s_1}^1)u_2(k_1^2, \cdots k_{s_2}^2) \cdots u_N(k_1^N, \cdots k_{s_N}^N), \tag{27}$$

where

$$u(p_1, \cdots p_m) \equiv \prod_{j=1}^{m} u(p_j), \qquad u(p) \equiv u(x_p). \tag{28}$$

The subterm in Eq. (27) appears $\prod_{j=1}^{N} (s_j - 1)!!$ times in the expression for $\Delta_n(x_1, \cdots x_n)$ which counts the number of pairings within the sets of s_j vertices, each pairing representing a particular propagation. Thus

$$\Delta_n(x_1, \cdots x_n) = \sum_{\substack{s_j \text{ even} \\ s_1+s_2+\cdots s_j=N}} \prod_{j=1}^{N} (s_j - 1)!! \times \sum_{\substack{\text{appropriate} \\ \text{permutations}}} u_1(k_1^1, \cdots k_{s_1}^1) \times u_2(k_1^2, \cdots k_{s_2}^2) \cdots u_N(k_1^N, \cdots k_{s_N}^N), \tag{29}$$

where the sequence $k_1^1, k_2^1 \cdots k_{s_1}^1, k_1^2 \cdots, k_1^N \cdots k_{s_N}^N$ ranges over all ordered permutations of $1, 2 \cdots n$, counting only once permutations which transform into each other under any succession of interchanges of any vertex coordinates belonging to the same mode.

We note now the equality of $\Delta_n(x_1, \cdots x_n)$ given in Eq. (29) with the quantity

$$\frac{1}{(2\pi)^N} \int_0^\infty dv_1 \cdots \int_0^\infty dv_N \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_N \times \left[\exp -\sum_{i=1}^{N} v_i\right] u_{v\theta}(x_1) \cdots u_{v\theta}(x_n), \tag{30}$$

which follows by using the relation

$$(m-1)!! = \frac{1}{2\pi} \int_0^\infty dv\, e^{-v} \left[\int_0^{2\pi} d\theta [(2v)^{\frac{1}{2}} \cos\theta]^m \right. \tag{31}$$

applied to the distributed expansion of $u_{v\theta}(x_1) \cdots u_{v\theta}(x_n)$. Equation (25) follows from this, Eq. (4), and the consideration that only even powers of g can appear in S.

In Eq. (25) we now substitute $v_i = t^2 \cos^2 \gamma_i$,

[15] Convergence in the many-mode case for general forms of the boson propagator can be proved by application of a general comparison theorem for convergence of field theories which will be presented elsewhere. This comparison theorem enables one to establish convergence in the case of an infinite number of interacting boson modes but where the boson propagator can be bounded in the Euclidean metric by a finite interacting mode propagator.

where

$$t^2 = \sum_{i=1}^{N} v_i,$$

and we find

$$S(g) = \frac{1}{\pi^N}\int_0^\infty dt\, t^{2N-1}e^{-t^2}\int_+ d\Omega_\gamma \prod_{i=1}^{N}\cos\gamma_i \times \int_0^{2\pi} d\theta_1 \cdots \int_0^{2\pi} d\theta_N \widetilde{S}_e^c(\sqrt{2}\, tg : \gamma, \theta), \quad (33)$$

where $\widetilde{S}_e^c(G;\ \gamma,\ \theta)$ is the S matrix corresponding to the interaction of the fermion field with the c-number source

$$\tilde{u}_{\gamma\theta} \equiv \sum_{i=1}^{N} \cos\gamma_i \cos\theta_i u_i(x), \quad (34)$$

i.e.,

$$\tilde{S}^c(G) \equiv \sum_{n=0}^{\infty}\frac{(-iG)^n}{n!}\sum_{\substack{\beta_1\cdots\beta_n\\ \alpha_1\cdots\alpha_n}}\Gamma_{\beta_1\alpha_1}\cdots\Gamma_{\beta_n\alpha_n} \times\int\cdots\int dx_1\cdots dx_n K_{\beta\alpha}(x_1,\cdots x_n) \times \tilde{u}_{\gamma\theta}(x_1)\cdots\tilde{u}_{\gamma\theta}(x_n). \quad (35)$$

$\int_+ d\Omega_\gamma$ denotes an integral over the first "quadrant" (i.e., $0 \le \gamma_i \le \frac{1}{2}\pi$, $i = 1, 2, \cdots N$) of rays in N-dimensional Euclidean space, where the direction cosines of the ray are $\cos\gamma_i$. Again one would consider $\mathfrak{s}(g)$ and $\mathfrak{s}^c(g)$ defined by

$$\tilde{\mathfrak{s}}(g) = \frac{1}{\pi^N}\int_0^\infty dt\, e^{-t^2}t^{2N-1}\int_+ d\Omega_\gamma \prod_{i=1}^{N}\cos\gamma_i \times \prod_{j=1}^{N}\int_0^{2\pi} d\theta_j \tilde{\mathfrak{s}}^c(\sqrt{2}\, gt : \gamma, \theta), \quad (36)$$

where

$$\tilde{\mathfrak{s}}^c(G : \gamma, \theta) \equiv \sum_{n=0}^{\infty}\frac{G^n}{n!}\sum_{\substack{\beta_1\cdots\beta_n\\ \alpha_1\cdots\alpha_n}}\int dx_1\cdots dx_n \times K_{\beta\alpha}(x_1,\cdots x_n)\,|\tilde{u}_{\gamma\theta}(x_1)\cdots\tilde{u}_{\gamma\theta}(x_n)|. \quad (37)$$

This is essentially the vacuum S matrix of a c-number boson source theory and its convergence which follows as does that of the $S^c(g)$ of Eqs. (17), (19) is uniform in γ_i and θ. The equation (36) connecting $\tilde{\mathfrak{s}}(g)$ with $\tilde{\mathfrak{s}}^c(g:\gamma,\theta)$ relates their analytic properties (in g) in essentially the same way that Eq. (17) does for the corresponding quantities $\mathfrak{s}(g)$ and $\mathfrak{s}^c(g)$. The same relation also connects the corresponding quantities in the case where ψ is a boson field.

In the next section, we show that the relations Eqs. (17) and (36) imply analyticity in g of the quantities $\mathfrak{s}(g)$ and $\tilde{\mathfrak{s}}(g)$, and therefore also the quantity $S(g)$. In the boson ψ field case, $S_B(g)$ has zero radius of convergence.

IV. ANALYTIC PROPERTIES IN g

We shall now demonstrate:

(a) The quantity $\mathfrak{s}^c(g)$ of Eq. (17) and the corresponding quantity $\tilde{\mathfrak{s}}^c(g;\ \gamma,\ \theta)$ of Eq. (37) are analytic as a function of g. The quantities $\mathfrak{s}(g)$ and $\tilde{\mathfrak{s}}(g)$, respectively related by Eqs. (17) and (36), have a nonvanishing radius of convergence when $\mathcal{K}(x)$ and $u(x)$ are L^2. The radius of convergence is infinite under further conditions on these quantities. The same results obtain for $S(g)$.

(b) When ψ is a boson field, relations like Eqs. (17) and (36) hold for $\mathfrak{s}_B(g)$ in the one- and many-mode cases, respectively, and their radius of convergence is zero as a result.

(c) The analytic properties of $S(g)$ imply essentially the same property for other matrix elements of the theory.

We start from the expression

$$\mathfrak{s}^c(g) = \sum_{n=0}^{\infty}\frac{g^n}{n!}\int dx_1\cdots dx_n\,|u(x_1)\cdots u(x_n)| \times \sum_{\substack{\beta_1\cdots\beta_n\\ \alpha_1\cdots\alpha_n}} K_{\beta\alpha}(x_1\cdots x_n), \quad (38)$$

referring to the definition in Eq. (5). We have indicated that the analyticity of $\mathfrak{s}(g)$ in a circle implies the analyticity of $\tilde{S}(g)$, which in turn implies the analyticity, and therefore convergence, of $S(g)$ in the same circle of the complex g plane.

As stated earlier, $\mathfrak{s}^c(g)$ is merely the modified Fredholm determinant[7] based on the L^2 kernel $\mathcal{K}(x, y)$ of Eq. (18), and is known from the Fredholm theory for L^2 kernels to be an entire function of g; i.e., analytic everywhere. We now prove Eqs. (12), (14), and (19) which are instrumental in determining the analytic properties of $S(g)$, $S_B(g)$.

The determinant[16] $K_{\beta\alpha}(x_1\cdots x_n)$ of Eq. (5) can be expressed in terms of the well-known expansion in permutations,

$$K_{\beta\alpha}(x_1,\cdots x_n) = \sum_{\substack{\text{permutations}\\ p}}(-)^p K_{1p_1}K_{2p_2}\cdots K_{np_n}, \quad (39)$$

where the permutation labeled p consists of the replacements $1 \to p_1$, $2 \to p_2$, $\cdots$, $n \to p_n$ and $(-)^p$ is the signature of this permutation. Any permutation of 1, 2, $\cdots$, n can be factored into a product of cycles where a cyle is a permutation of a subset of 1, 2, $\cdots$, n into itself which is no longer so factorable. We shall use the notation

[16] The method which follows appears in a number of other investigations: See, e.g., E. Montroll and J. Ward, Phys. Fluids 1, 55 (1958); M. Hamermesh, *Group Theory* (Addison-Wesley Publishing Company, Reading, Massachusetts, 1962), Sec. 7.2.

$(i_1, i_2 \cdots i_k)$ for the permutation cycle which takes i_1, into i_2, i_2 into i_3, $\cdots$, i_{k-1} into i_k, and i_k into i_1. The cycle $(i_1 i_2 \cdots i_k)$ will be said to be of length k. There corresponds to each permutation a set of integers giving the length of its component cycle factors. This set of integers constitutes a partition of the integer n in the sense that their sum is n. Such a set of integers will be termed a "partition," and the set of permutations having the same set of such integers will be termed a "partition class". Cycles of length 1 are excluded due to the zero diagonal elements in Eq. (5) which are the result of the normal ordering of the interaction in Eq. (2). The sum in Eq. (39) can be written as a double sum,

$$\sum_{\text{permutations}} = \sum_{\substack{\text{partition} \\ \text{classes } \pi}} \sum_{\substack{\text{permutations} \\ \text{in } \pi}} . \tag{40}$$

The signature is the same for all permutations in a partition class and is given by $(-1)^{n+\nu}$ where ν is the number of cycles in the cycle factorization of a permutation. Each cycle corresponds to a closed fermion loop, and a factor of $(-1)^{\nu}$ in the signature of the particular permutation corresponds to a factor -1 for each closed fermion loop in the corresponding diagram. These factors of -1 are contributed by the antisymmetrization in their fermion statistics, and, as will be seen, they are crucial for convergence. We use the abbreviations

$$u_i \equiv u(x_i), \tag{41}$$

and

$$K_m(1, 2, \cdots m) \equiv K(x_1 - x_2)K(x_2 - x_3) \cdots \times K(x_{m-1} - x_m)K(x_m - x_1), \tag{42}$$

which is the product of the fermion propagators for a closed loop with m vertices. Equation (39) can therefore be rewritten as

$$K = (-)^n \sum_{\substack{\text{partition} \\ \text{classes } \pi}} (-)^{\nu(\pi)} \sum_{\substack{\text{permutations} \\ \text{in } \pi}} K_{n_1}(i_1^1 \cdots i_{n_1}^1) \times K_{n_2}(i_1^2 \cdots i_{n_2}^2) \cdots K_{n_{\nu(\pi)}}(i_1^{\nu(\pi)} \cdots i_{n_{\nu(\pi)}}^{\nu(\pi)}), \tag{43}$$

where the partition class π corresponds to the partition $(n_1, n_2, \cdots n_{\nu(\pi)})$ and a typical permutation in π factors into the cycles $(i_1^1 \cdots i_{n_1}^1)(i_1^2 \cdots i_{n_2}^2) \cdots (i_1^{\nu(\pi)} \cdots i_{n_{\nu(\pi)}}^{\nu(\pi)})$. The coefficient of g^n in Eq. (38) is

$$\mathcal{S}_n^c = \frac{(-)^n}{n!} \sum_{\substack{\text{partition} \\ \text{classes } \pi}} (-)^{\nu(\pi)} \times \sum_{\substack{\text{permutations} \\ \text{in } \pi}} \int dx_1 \cdots dx_n \, |u_1 \cdots u_n| \times K_{n_1}(i_1^1 \cdots i_{n_1}^1)K_{n_2}(i_1^2 \cdots i_{n_2}^2) \cdots \times K_{n_{\nu(\pi)}}(i_1^{\nu(\pi)} \cdots i_{n_{\nu(\pi)}}^{\nu(\pi)}). \tag{44}$$

We note that the integral in Eq. (44) has the same value for all permutations belonging to the same partition class, since any relabeling of the integration variables does not change the value of the integrals. The value of each such integral in Eq. (44) for a given partition π after performing the spin sum is simply

$$\prod_{i=1}^{\nu(\pi)} \int dx_1 \cdots dx_{n_i} \operatorname{tr} K_{n_i}(12 \cdots n_i) \, |u_1 \cdots u_{n_i}|. \tag{45}$$

We introduce the ring integrals

$$\sigma_m \equiv \int dx_1 \cdots dx_m \operatorname{tr} K_m(12 \cdots m) \, |u_1 \cdots u_m|. \tag{46}$$

We find

$$\mathcal{S}_n^c = \frac{(-)^n}{n!} \sum_{\substack{\text{partition} \\ \text{classes } \pi}} (-)^{\nu(\pi)} N_\pi \sigma_{n_1} \sigma_{n_2} \cdots \sigma_{n_{\nu(\pi)}}, \tag{47}$$

where N_π is the number of permutations in the partition class π. This is expressible in terms of the integers in the partition corresponding to the partition class π. For a particular partition π let the integers in this partition be $n_1, n_2, \cdots n_{\nu(\pi)}$. Let this set consist of distinct integers $m_1, m_2, \cdots m_{\mu(\pi)}$ where there are r_1 integers $m_1, \cdots, r_{\mu(\pi)}$ integers $m_{\mu(\pi)}$. N_π is then given by

$$N_\pi = n!/(m_1)^{r_1}(m_2)^{r_2} \cdots \times (m_{\mu(\pi)})^{r_{\mu(\pi)}} r_1! \, r_2! \cdots r_{\mu(\pi)}!. \tag{48}$$

The quantity $n!/n_1!n_2! \cdots n_{\nu(\pi)}!$ gives the number of ways of distributing n integers into distinguishable unordered classes consisting of $n_1, n_2 \cdots n_{\nu(\pi)}$ elements, respectively. A factor of $r_1!r_2! \cdots r_{\mu(\pi)}!$ reduces this number to the number of ways of distributing the integers into indistinguishable classes. An additional factor of $(n_j - 1)!$ for each class j gives the number of cyclic orderings of the integers in this class, counting the number of permutation cycles that can be found from this set of integers. Thus

$$\mathcal{S}_n^c = (-)^n \sum_{\substack{\text{partition} \\ \text{classes } \pi}} \frac{(-)^{\nu(\pi)}(\sigma_{m_1})^{r_1}(\sigma_{m_2})^{r_2} \cdots (\sigma_{m_{\mu(\pi)}})^{r_{\mu(\pi)}}}{(m_1)^{r_1}(m_2)^{r_2} \cdots (m_{\mu(\pi)})^{r_{\mu(\pi)}} r_1! \, r_2! \cdots r_{\mu(\pi)}!}, \tag{49}$$

$$\mathcal{S}^c(g) = \sum_{n=0}^{\infty} (-)^n \sum_{\substack{\text{partition} \\ \text{classes } \pi}} \frac{(-)^{\nu(\pi)}(\sigma_{m_1})^{r_1}(\sigma_{m_2})^{r_2} \cdots (\sigma_{m_{\mu(\pi)}})^{r_{\mu(\pi)}}}{(m_1)^{r_1}(m_2)^{r_2} \cdots (m_{\mu(\pi)})^{r_{\mu(\pi)}} r_1! \, r_2! \cdots r_{\mu(\pi)}!}, \tag{50}$$

where

$$m_1r_1 + m_2r_2 + \cdots m_{\mu(\pi)}r_{\mu(\pi)} = n \qquad (m_i \neq 0, 1),$$
$$r_1 + r_2 + \cdots r_{\mu(\pi)} = \nu(\pi). \tag{51}$$

The sum of Eq. (50) can be evaluated in closed form. One may verify that it is equal to the infinite product

$$\mathcal{S}^c(g) = \prod_{m=2}^{\infty}\left\{1 - \frac{(-g)^m\sigma_m}{m} + \frac{1}{2!}\left[\frac{(-g)^m\sigma_m}{m}\right]^2 - \frac{1}{3!}\left[\frac{(-g)^m\sigma_m}{m}\right]^3 + \cdots\right\}$$
$$= \exp - \left[\sum_{m=2}^{\infty}\frac{(-g)^m\sigma_m}{m}\right]. \tag{52}$$

This is Eq. (19) and is the same as Eq. (12) with g replaced by ig, and the σ_m correspondingly defined as in Eq. (13).

Note that the negative sign in the exponential originates in the $(-)^{\nu(\pi)}$ factor in Eq. (50) expressing the Fermi statistics. Had we considered the $\psi(x)$ field to be a boson field, so that we had a three-boson field coupling in place of a Yukawa coupling the only change in $\mathcal{S}^c(g)$ would be the replacement of $(-)^{\nu(\pi)}$ in Eq. (50) by $+1$ and the exponential function in Eq. (52) would have a positive sign before the summation in the exponent. Thus

$$\mathcal{S}_B^c(g) = \exp\sum_{m=2}^{\infty}\frac{g^m\sigma_m}{m}, \tag{53}$$

where the σ_m would be defined correspondingly.

To appreciate the significance of this sign let us first consider a model where $\sigma_m = \beta^m$ (β some number). This would be the case for example if the ψ field had only one mode and its propagator were therefore separable. (In any case the σ_m have a leading asymptotic dependence of the form β^m.) Then

$$\sum_{m=2}^{\infty}\frac{(-g)^m\beta^m}{m} = \beta g - \ln(1 + g\beta), \tag{54}$$

and for a fermion ψ field, we find

$$\mathcal{S}^c(g) = (1 + g\beta)e^{-\beta g}, \tag{55}$$

while for a boson field

$$\mathcal{S}_B^c(g) = e^{-\beta g}/(1 - \beta g) = [\mathcal{S}^c(g)]^{-1}. \tag{56}$$

Thus the Yukawa interaction S matrix in this simple case has an infinite radius of convergence while the three-boson interaction has a finite radius.

We have mentioned that $\mathcal{S}^c(g)$ has an infinite radius of convergence. It also has an infinite number of zeros. From the Hermiticity of $\mathcal{K}(x, y)$, follows follows the fact that $\mathcal{S}^c(g)$ has at least one zero.[17] This tells us that $\mathcal{S}_B^c(g)$ has poles and therefore only a finite radius of convergence. If the L^2 kernel $\mathcal{K}(x)$ is finite everywhere (we have regularized it) and goes to zero continuously for large values of $|x|$, then $\mathcal{S}^c(g)$ has in fact an infinite number of real zeros with no finite accumulation point. If $\mathcal{S}^c(g)$ had only a finite number of zeros, say M (where we weight with the multiplicity), then the Hermitian kernel $\mathcal{K}(x, y)$ is expressible as[18]

$$\mathcal{K}(x, y) = \sum_{s=0}^{m}\frac{\bar{\psi}_s(y)\psi_s(x)}{g_s}, \tag{58}$$

where g_s, $\psi_s(x)$ are, respectively, the eigenvalue and eigensolution of the homogeneous integral equation

$$\psi_s(x) = g_s\int dy\mathcal{K}(x, y)\psi_s(y) \tag{59a}$$

or

$$\psi_s(x) = g_s\int dyK(x - y)|u(x)u(y)|^{\frac{1}{2}}\psi_s(y), \tag{59b}$$

and the $\psi_s(x)$, ($s = 1, 2, \cdots M$) are an orthornormal set. This is so because the g_s are identical with zeros of $\mathcal{S}^c(g)$.[19] Therefore,

$$K(x - y) = \sum_{s=0}^{M}\frac{\bar{\psi}_s(y)\psi_s(x)}{g_s\,|u(x)u(y)|^{\frac{1}{2}}}$$
$$= \sum_{s=0}^{M} g_s\iint dz_1\,dz_2\bar{K}(y - z_2)\bar{\psi}_s(z_2)$$
$$\times\,|u(z_2)|^{\frac{1}{2}}\,K(x - z_1)\psi_s(z_1)\,|u(z_1)|^{\frac{1}{2}}, \tag{60}$$

from Eqs. (59), (58), and (59b). Let ξ be such that $K(\xi) \neq 0$. One finds, setting $y = x + \xi$ in Eq. (60), that the left side $K(\xi) \neq 0$, while the right side goes to zero as $|x|$, $|y| \to \infty$. The last fact follows readily because

$$\lim_{|x|\to\infty}\int dzK(x - z)\,|u(z)|^{\frac{1}{2}}\psi_s(z) = 0, \tag{61}$$

which is an easily verified consequence of the fact that $|u(z)|^{\frac{1}{2}}\,\psi_s(z)$ is L^1 and $K(x)$ goes to zero continuously for large values of its argument.

We next consider the analytical properties implied by Eqs. (17) and (36), which connect the respective function pairs, $\mathcal{S}(g)$, $\mathcal{S}^c(g)$; $\tilde{\mathcal{S}}(g)$, $\tilde{\mathcal{S}}^c(g;\ \gamma,\ \theta)$; $\mathcal{S}_B(g)$,

[17] Ref. 8, Theorem 7.2.1.

[18] This relation must in general be understood to hold in the mean in L^2. See, e.g., Ref. 8, Theorem 7.4.1; F. G. Tricomi, in *Integral Equations* (Interscience Publishers, Inc., New York, 1957), Sec. 3.9.

[19] Ref. 8, Corollary to Theorem 6.7.1.

$S_B^c(g)$. [We denote the generic pair by $S(g)$, $S^c(g)$.] The convergence of the power series for $S(g)$ depends on the convergence of the term by term integrated series for $S^c(g)$,[20] the latter depending on the behavior of the coefficients in the power series expansion for $S^c(g)$. Considering the Gaussian transform in Eqs. (17) and (36), it is clear that a nonvanishing radius of convergence will follow for $S(g)$ only if

$$\frac{1}{\rho} \equiv \lim_{n\to\infty} \frac{\ln |S_n^c|^{-1}}{n \ln n} \geq \frac{1}{2}, \tag{62}$$

where S_n^c is the coefficient of g^n in the expansion of $S^c(g)$. From the finite radius of convergence of $S_B^c(g)$ it is immediately clear that $S_B(g)$ has a zero radius of convergence. Such a result also goes under the name of Watson's lemma[13] to the effect that the Gaussian transform of a function with a finite radius of convergence gives a power series which is necessarily asymptotic. For the case of the fermion ψ field, $S^c(g)$ is an entire function and the validity of Eq. (62) becomes necessary and sufficient for analyticity of $S(g)$. Appeal to certain theorems in the theory of integral functions will now establish the validity of Eq. (62).

The quantity ρ can be shown to be[21] the order of the integral function $S^c(g)$ which means that

$$|S^c(g)| = O(e^{|g|^{\rho+\epsilon}}), \tag{63}$$

for all positive and no negative value of ϵ. Equation (62) requires that $\rho \leq 2$. The order ρ of an entire function $f(z)$ is related to ρ_1, the exponent of convergence of its zeros; i.e., ρ_1 is the lower bound of α for which $\sum_s |z_s|^{<\infty}$ where z_s are the zeros of $f(z)$. As $S^c(g)$ is the modified Fredholm determinant deriving from the L^2 kernel $\mathcal{K}(x, y)$ as given in Eq. (57), its exponent of convergence $\rho_1 \leq 2$. This follows from[22]

$$\sum_s \frac{1}{g_s^2} = \int dx\, dy\, |\mathcal{K}(x, y)|^2 < \infty. \tag{64}$$

In general $\rho_1 \leq \rho$, and we must show that indeed $\rho \leq 2$. If ρ_1 is not an integer, then it can be shown[23] that $\rho_1 = \rho$, and the desired result obtains. In fact in this case $\rho_1 < 2$ so that $\rho < 2$ and $S(g)$ has an infinite radius of convergence.

For the case where ρ_1 is an integer we shall show that the "genus" r (to be defined) of $S^c(g)$ is equal to one. The genus satisfies the inequality[24]

$$\rho - 1 \leq r \leq \rho, \tag{65}$$

and $r = 1$ therefore implies that $\rho \leq 2$. From Eq. (52),

$$S^c(g) = \exp -\sum_{m=2}^{\infty} \frac{(-g)^m \sigma_m}{m}, \tag{52}$$

where

$$\sigma_m = \int dx_1 \cdots dx_m \operatorname{tr} [\mathcal{K}(x_1, x_2) \cdots \mathcal{K}(x_m, x_1)] \tag{66}$$

[what follows for $S^c(g)$ will also hold for $\tilde{S}^c(g)$]. Now

$$\sigma_m = \int dx \operatorname{tr} \mathcal{K}^m(x, x), \tag{67}$$

where $\mathcal{K}^m(x, y)$ is the m-times iterated kernel $\mathcal{K}(x, y)$. For $\mathcal{K}(x, y)$ continuous, the uniform convergence of the spectral expansion of $\mathcal{K}^m(x, y)$[25] (for $m \geq 2$) implies that

$$\sigma_m = \sum_s \frac{1}{g_s^m}. \tag{68}$$

Then

$$-\ln S^c(g) = \sum_{m=2}^{\infty} \frac{(-g)^m}{m} \sum_s \frac{1}{g_s^m}. \tag{69}$$

Since no $g_s = 0$ and only a finite number of g_s lie between -1 and 1, the above double sum is absolutely convergent for sufficiently small g and can be rearranged to

$$-\ln S^c(g) = \sum_s \left[\frac{g}{g_s} - \ln\left(1 + \frac{g}{g_s}\right)\right], \tag{70}$$

where in Eq. (70) we may not distribute the summation sign separately over the two terms in the bracket. Consequently,

$$S^c(g) = \prod_s \left[\left(1 + \frac{g}{g_s}\right) e^{-g/g_s}\right], \tag{71}$$

where the same restriction holds on distributing the product symbol in Eq. (71) over the two factors in the bracket. Eq. (71) is a canonical factorization of the entire function $S^c(g)$ into a product of "primary factors",[26] The exponential factor in each primary factor contains a first degree polynomial in the exponential and this defines the genus r to be 1.[27] In fact, since $S^c(g)$ is given as a canonical product in terms of its zeros, $\rho_1 = \rho$,[28] and therefore if $\rho_1 = 1$, so does ρ, which would then imply an infinite radius of convergence. There are conditions under which $\rho_1 = \rho = 1$, and consequently an infinite

[20] See, e.g., E. C. Titschmarsh, *The Theory of Functions* (Oxford University Press, London and New York, 1939), 2nd ed. See 1.79 for proof of a slightly different result which is easily extended to this case.
[21] Ref. 20, Sec. 8.3.
[22] Ref. 8, Corollary, p. 116.
[23] Ref. 20, Sec. 8.26.
[24] Ref. 20, Sec. 8.27.
[25] Ref. 8, Theorem 7.6.1.
[26] Ref. 20, Sec. 8.1.
[27] Ref. 20, Sec. 8.23.
[28] Ref. 20, Sec. 8.25.

radius of convergence obtains. This is the case when the fermion field has a finite number of modes and we thus get the result of Buccafurri and Caianiello.[22] Another example is when the kernel $\mathcal{K}(x, y)$ has only a finite number of negative eigenvalues[29] which is true, say in a model of scalar fermions if $K(x - y)$ is represented in a Euclidean metric. We shall not dwell on these here but for mentioning that the existence of $\int dx\mathcal{K}(x, x)$ is necessary in this case.

We have concerned ourselves till now with the convergence only of the vacuum-to-vacuum S matrix. Other matrix elements of the theory are related through functional derivatives to the vacuum S matrix of a theory containing c-number source functions. Thus the n-point boson Green's functions can be obtained from the vacuum S matrix corresponding (schematically) to the Yukawa interaction,

$$H_I = g[:\bar{\psi}(x)\Gamma\psi(x): + j(x)]\varphi(x), \tag{72}$$

where $j(x)$ is a c-number source. The quantity $\mathcal{S}_j^c(g)$ based on a c-number $\varphi(x)$ for H given by Eq. (72) is merely given by

$$\mathcal{S}_j^c(g) = \mathcal{S}^c(g) \exp g \int dxj(x)\varphi(x), \tag{73}$$

and entirely analogous considerations to those outlined in this paper will establish convergence of the perturbation expansions.

V. DISCUSSION

The principle of this proof is the relation of the quantized boson-field vacuum S matrix to the S matrix describing fermions interacting with a classical boson field. The latter quantity is merely a Fredholm determinant whose known attributes have been exploited. This relation is a manifestation of the functional equation[30] relating a functional of the quantized boson field to a functional of a c-number field

$$TS[\varphi_{\text{quant}}] = :\exp -\frac{i}{2}\iint dx\, dy \frac{\delta}{\delta\varphi(x)} \times \Delta(x - y)\frac{\delta}{\delta\varphi(y)} S[\varphi_{\text{class}}]: \quad . \tag{74}$$

One might hope to develop a refined technique for extending the convergence demonstration to an infinite number of interacting boson modes. The considerations are not entirely straightforward. The introduction of an infinite number of modes or rather what is more crucial, the introduction of boson propagators which are functions of only the invariant distance between points leads to an infinite amplitude for each closed loop arising from the δ function of overall energy–momentum conservation. From the physics point of view, one knows how to handle these infinities unambiguously. The mathematical limiting procedures, however, will encounter these infinities, and must be so performed as not to be invalidated by them. While presumably this can be done, the added complication calls for more refined techniques. One might, for example, consider the vacuum self-energy which involves only connected diagrams and therefore has only one overall δ function. This and other approaches are presently under consideration.

In conclusion, this paper has presented an explicit demonstration of the convergence of Yukawa interaction theories where only a finite number of boson modes can interact with a relativistic quantized fermion field. This result is presented very sketchily in an earlier paper of Edwards, and to the writer's knowledge and understanding is the strongest result so far appearing in the literature on the subject of convergence. The role of statistics in convergence is clarified by explicit demonstration of the divergence for a three-boson coupling under the same conditions. It has also been pointed out that the restriction to a finite number of interacting boson modes is not equivalent to a Feynman cutoff and constitutes a significantly greater limitation.

Notes added in proof. Considerably stronger results have recently been derived by the author. They will be reported elsewhere. The reciprocal relation Eq. (56) between the Yukawa and three-boson vacuum S matrices is derived as a general relation by R. Feynman in Phys. Rev. **76**, 749 (1949), Sec. 5.

ACKNOWLEDGMENT

The author would like to acknowledge helpful discussions with Dr. Z. Fried and Dr. H. Reiss.

[29] F. G. Tricomi, Ref. 18, Sec. 3.12, Mercer's Theorem.

[30] See, e.g., N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience Publishers, Inc., New York, 1959), Sec. 39.2; B. Zumino, "Notes on the Quantum Theory of Fields"(New York University Lecture Notes, Lecture 8, 1958) (unpublished).

The Connection between Conservation Laws and Laws of Motion in Affine Spaces*

PETER HAVAS
Department of Physics, Lehigh University, Bethlehem, Pennsylvania
(Received 5 September 1963)

It was pointed out recently that, for any theory describing matter as a collection of mass points in a metric space and subject to a covariant conservation law for a symmetric tensor density $\mathfrak{P}^{\mu\nu}$, the geodesic law of motion as well as the form of $\mathfrak{P}^{\mu\nu}$ follow from the conservation law *alone*, independent of any equations obeyed by the metric. This result is shown to be valid in any affine space, independent of any equations obeyed by the affine connection; conversely, the geodesic law implies a conservation law for a singular symmetric tensor density. Similarly, the existence in any affine space of a covariant conservation law for a vector density $\mathfrak{I}^{\nu}$ describing a collection of point charges is shown to imply the constancy of charge, and the form of $\mathfrak{I}^{\nu}$; conversely, the constancy of charge implies a conservation law for a singular vector density. Some applications of these results are presented. An Appendix contains a discussion of the laws of motion for particles with an intrinsic dipole moment.

INTRODUCTION

ONE of the most notable achievements of the general theory of relativity is its interrelation of the gravitational fields and of the motion of their sources. The basic equations of the theory are differential equations for the gravitational field (described by the metric tensor $g_{\mu\nu}$), as determined by the distribution of matter (described by an energy–momentum tensor $P^{\mu\nu}$). It was at first assumed that, as in other field theories, the motion of these sources is independent of the field equations, and it was postulated that a test particle was moving along a geodesic of the background metric.[1] However, it was soon realized by Weyl[2] and by Einstein and Grommer[3] that the field equations impose limitations on the motion of particles, and several authors were able to show that Einstein's field equations imply that a test particle has to move along a geodesic.[2,4] Later, several methods were devised to obtain approximate equations of motion of bodies with comparable masses[5]; while the resulting equations differ according to the method of approximation used, they have in common that the particles move along geodesics[6,7] (though no longer of a fixed background metric).

Einstein's field equations were obtained originally by starting from the requirement of an invariant conservation law for the symmetric energy–momentum tensor of matter $P^{\mu\nu}$; we can equivalently state this conservation law for the corresponding tensor density $\mathfrak{P}^{\mu\nu}$ as

$$\mathfrak{P}^{\mu\nu}{}_{;\nu} = 0. \tag{1}$$

Here $;_\nu$ denotes covariant differentiation with respect to x^ν; summation over repeated co- and contravariant indices is understood. Einstein then obtained his field equations by equating $P^{\mu\nu}$ to an expression formed from the $g_{\mu\nu}$ which satisfies differential identities entailing Eq. (1).

It was pointed out recently[7] that for matter represented by a collection of mass points (an assumption common to most methods of calculation[5]), the geodesic law as well as the form of $\mathfrak{P}^{\mu\nu}$ follow from the conservation law (1) *alone*, without any consideration of the equations obeyed by the metric. However, the *existence* of a metric was made use of in the proof presented.

It is the main purpose of this paper to show that the assumption of the existence of a metric is not necessary for the proof, but only that of an affine connection. It is also shown that the method developed for the proof can be applied to a conserva-

* This research was supported by the Aerospace Research Laboratories of the Office of Aerospace Research, United States Air Force.

[1] A. Einstein, Ann. Physik **49**, 769 (1916).

[2] H. Weyl, *Raum. Zeit. Materie* (Springer–Verlag, Berlin, 1921), 4th ed., Sec. 36; in more detail in the 5th ed. (1923).

[3] A. Einstein and J. Grommer, Sitzber. Preuss. Akad. Wiss., Physik. Math. Kl., 2 (1927); A. Einstein, *ibid.*, 235.

[4] M. von Laue, *Die Relativitätstheorie* (Friedrich Vieweg und Sohn, Braunschweig, Germany, 1921), Vol. 2, Sec. 15; A. S. Eddington, *The Mathematical Theory of Relativity* (Cambridge University Press, Cambridge, England, 1923), Sec. 56; C. Lanczos, Z. Physik **59**, 514 (1930); M. Mathisson, *ibid.* **67**, 270 (1931); H. P. Robertson, Proc. Edinburgh Math. Soc. **5**, 63 (1936); L. Infeld and A. Schild, Rev. Mod. Phys. **21**, 408 (1949).

[5] For a brief review see J. N. Goldberg, in *Gravitation*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962), p. 102.

[6] L. Infeld and J. Plebański, Bull. Acad. Polon. Sci. III, 4, 757 (1956).

[7] P. Havas and J. N. Goldberg, Phys. Rev. **128**, 398 (1962).

tion law for a vector density (equation of continuity). In the Appendix we discuss the problem of the laws of motion in affine spaces of particles with an intrinsic dipole moment.

II. TENSOR CONSERVATION LAW AND LAW OF GEODESIC

We consider an affine n-space[8] with coordinates x^ρ ($\rho = 0, 1, \cdots n-1$) and affine connection $\Gamma^\rho_{\mu\nu}$. One of these coordinates, say x^0, will be interpreted as the time. We shall be concerned with the motion of N singularities. The coordinates of the ith particle are denoted by z_i^ρ, and are functions of some scalars λ_i parametrizing the world lines. Their time coordinates z_i^0 are assumed to increase monotonically with the λ_i's from $-\infty$ to ∞. We do not consider any fields other than those described by the affine connections. Matter is to be described by a symmetric tensor density $\mathfrak{P}^{\mu\nu}$, subject to the conservation law (1), and of the form

$$\mathfrak{P}^{\mu\nu} = \sum_i \int_{-\infty}^{\infty} \mathfrak{p}_i{}^{\mu\nu}(\lambda_i)\delta^n[x^\rho - z_i{}^\rho(\lambda_i)]\, d\lambda_i, \qquad (2)$$

where the exact form of the $\mathfrak{p}_i{}^{\mu\nu}$ remains to be determined. δ^n is an n-fold product of Dirac δ functions. The representation of singular quantities by integrals of the type (2) is well known from special relativity.[9] We shall first restrict our considerations to simple poles, i.e., integrals not involving derivatives of δ functions; the problems arising in the case of dipoles are discussed in the Appendix. As usual an expression involving δ functions is a shorthand notation for an expression which acquires meaning only on integration over the coordinates.

In general relativity, $\mathfrak{P}^{\mu\nu}$ represents the density of energy and momentum and the stresses, and Eq. (1) expresses the conservation of energy and momentum. We shall not need such an interpretation for the derivation of our general law of motion, however.

In a metric space we can associate a tensor density of weight 1 with a tensor by $\mathfrak{P}^{\mu\nu} = |g|^{\frac{1}{2}}P^{\mu\nu}$ (where g is the determinant of the metric tensor $g_{\mu\nu}$), and conversely. Since $|g|^{\frac{1}{2}}{}_{;\nu} = 0$, the conservation law (1) for $\mathfrak{P}^{\mu\nu}$ implies a similar law for $P^{\mu\nu}$, and conversely. Although a similar association of a tensor and a tensor density is possible in any affine space by means of an arbitrary symmetric tensor $g_{\mu\nu}$ with nonvanishing determinant, we no longer have necessarily $|g|^{\frac{1}{2}}{}_{;\nu} = 0$, and thus the requirement of a conservation law for $\mathfrak{P}^{\mu\nu}$ is not necessarily equivalent to such a requirement for $P^{\mu\nu}$, and conversely. Thus we have to distinguish between these two cases in the following.

To obtain the law of motion we shall use a method based on one due to Mathisson,[10] which for a four-dimensional metric space has been described previously.[7,11-13] We multiply Eq. (1) by a function ξ_μ and integrate over all x^ρ to obtain

$$\sum_i \iint \{\mathfrak{p}_i{}^{\mu\nu}(\lambda_i)\delta^n[x^\rho - z_i^\rho(\lambda_i)]\}_{;\nu}\, \xi_\mu\, d\lambda_i\, d^nx = 0, \qquad (3)$$

where ξ_μ is assumed to be completely arbitrary except for vanishing at the limits of the λ_i integrations together with all its derivatives.

To allow an invariant volume integration, the integrand in (3) must be a scalar density,[8] and thus ξ_μ must be a vector. Alternatively, if we had required a conservation law (1) for a tensor $P^{\mu\nu}$ rather than a tensor density, ξ_μ would be a vector density. The subsequent calculations do not depend explicitly on the transformation properties of ξ_μ; the difference would only be in the transformation properties of the expressions entering Eq. (8).

We can transfer the derivative to ξ_μ by an integration by parts, and then carry out the x integration. Then ξ_μ and its derivatives (including the Γ's implied by the covariant differentiation), being evaluated at the positions $z_i{}^\rho(\lambda_i)$, become functions of the λ_i's. Thus we get

$$\sum_i \int (-\mathfrak{p}_i{}^{\mu\nu}\xi_{\mu;\nu_i})\, d\lambda_i = 0, \qquad (4)$$

where $;\nu_i$ indicates covariant differentiation with respect to $z_i{}^\nu$.

An essential part of Mathisson's method is the decomposition of tensors such as $\mathfrak{p}_i{}^{\mu\nu}$ in components parallel and perpendicular to the n-velocity $v_i{}^\rho$ defined by

$$v_i{}^\rho \equiv dz_i^\rho/d\lambda_i. \qquad (5)$$

Such a decomposition is desirable, since from Eq. (5),

$$\xi_{\mu;\nu_i}v_i{}^\nu = D\xi_\mu/d\lambda_i, \qquad (6)$$

where $D/d\lambda_i$ denotes covariant differentiation with respect to λ_i, and the derivative (6) can be removed from ξ_μ in Eq. (4) by an integration by parts.

[8] For a brief discussion of affine spaces see, e.g., E. Schrödinger, *Space-Time Structure* (Cambridge University Press, Cambridge, England, 1950), or Ref. 2.

[9] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938).

[10] M. Mathisson, Acta Phys. Polon. **6**, 163 (1937).

[11] M. Mathisson, Proc. Cambridge Phil. Soc. **36**, 331 (1940).

[12] W. Tulczyjew, Acta Phys. Polon. **18**, 393 (1959).

[13] P. Havas, in *Recent Developments in General Relativity* (Pergamon Press, Inc.-PWN, New York-Warsaw, 1962), p. 259.

Then the integrand in Eq. (4) separates into parts involving ξ_μ or $\xi_{\mu;\nu_i}$ as a factor; because of the arbitrariness of these functions, their coefficients must vanish separately for all i.

This procedure can be carried out without difficulty in a metric space, where a unique covariant vector $v_{i\mu} = g_{\mu\nu}v_i^{\ \nu}$ is associated with $v_i^{\ \nu}$ and a contravariant vector $n_i^{\ \nu}$ is orthogonal to $v_i^{\ \nu}$ if $n_i^{\ \nu}v_{i\nu}$ vanishes. No such unique vector is available in the absence of a metric. However, we can introduce covariant vectors $w_{i\nu}(\lambda_i)$ such that

$$v_i^{\ \nu}w_{i\nu} \neq 0 \tag{7}$$

at each point λ_i, since by our assumption on the z_i^0 none of the $v_i^{\ 0}$ can vanish anywhere; this will be sufficient to carry through Mathisson's procedure, and the results will be independent of any further specification of $w_{i\nu}$.[14]

Now we put

$$\begin{aligned}\mathfrak{p}_i^{\ \mu\nu} &= M_i(\lambda_i)v_i^{\ \mu}v_i^{\ \nu} + n_i^{\ \mu}(\lambda_i)v_i^{\ \nu} \\ &\quad + n_i^{\ \nu}(\lambda_i)v_i^{\ \mu} + {}^*\mathfrak{p}_i^{\ \mu\nu}(\lambda_i), \\ {}^*\mathfrak{p}_i^{\ \mu\nu} &= {}^*\mathfrak{p}_i^{\ \nu\mu}, \qquad {}^*\mathfrak{p}_i^{\ \mu\nu}w_{i\nu} = 0, \quad n_i^{\ \nu}w_{i\nu} = 0.\end{aligned} \tag{8}$$

Substituting these expressions into Eq. (4), using (6), and carrying out an integration by parts, we obtain

$$\sum_i \int \left[-({}^*\mathfrak{p}_i^{\ \mu\nu} + n_i^{\ \nu}v_i^{\ \mu})\xi_{\mu;\nu_i} + \frac{D}{d\lambda_i}(M_i v_i^{\ \mu} + n_i^{\ \mu})\xi_\mu \right] d\lambda_i = 0. \tag{9}$$

The vanishing of the coefficient of $\xi_{\mu;\nu_i}$ requires

$${}^*\mathfrak{p}_i^{\ \mu\nu} + n_i^{\ \nu}v_i^{\ \mu} = 0, \tag{10}$$

from which we get by contraction with $w_{i\mu}$, using (7) and (8), that ${}^*\mathfrak{p}_i^{\ \mu\nu}$ and $n_i^{\ \nu}$ must vanish separately, and thus

$$\mathfrak{p}_i^{\ \mu\nu} = M_i v_i^{\ \mu}v_i^{\ \nu}. \tag{11}$$

Then the vanishing of the coefficients of ξ_μ requires

$$D(M_i v_i^{\ \mu})/d\lambda_i = 0. \tag{12}$$

Here M_i is a scalar if Eq. (1) holds for a tensor density of weight 1, and a scalar density of weight -1 if we start from a similar equation for a tensor instead. From our initial assumption on z_i^0, v_i^0 cannot vanish any where on the ith world line; therefore Eq. (12) implies that M_i also can not vanish anywhere on this line except if it vanishes everywhere [which from Eq. (11) corresponds to the trivial case of the absence of the ith singularity].

We now investigate whether Eq. (12) is indeed the equation of a geodesic. Introducing a new parameter τ_i we can write it as

$$D\left[M_i \frac{dz_i^{\ \mu}}{d\tau_i}\frac{d\tau_i}{d\lambda_i}\right] \Big/ d\lambda_i = 0.$$

This can be reduced to the standard form of a geodesic provided we can choose τ_i such that

$$D\left[M_i \frac{d\tau_i}{d\lambda_i}\right] \Big/ d\lambda_i = 0, \tag{13}$$

or

$$M_i \frac{d\tau_i}{d\lambda_i} = m_i, \tag{14}$$

where m_i is a constant scalar if Eq. (1) holds for a tensor density, or a scalar density of weight -1 with vanishing covariant derivative if Eq. (1) was assumed for a tensor. If our affine space admits such an m_i, the required τ_i can be determined from Eq. (14) by integration, M_i not being zero anywhere. Then Eq. (12) reduces to the standard form

$$\frac{D}{d\tau_i}\frac{dz_i^{\ \mu}}{d\tau_i} \equiv \frac{d^2z_i^{\ \mu}}{d\tau_i^{\ 2}} + \Gamma^\mu_{\rho\sigma}\frac{dz_i^{\ \rho}}{d\tau_i}\frac{dz_i^{\ \sigma}}{d\tau_i} = 0, \tag{15}$$

and from Eqs. (2), (11), and (14) we get

$$\mathfrak{P}^{\mu\nu} = \sum_i \int_{-\infty}^{\infty} m_i \frac{dz_i^{\ \mu}}{d\tau_i}\frac{dz_i^{\ \nu}}{d\tau_i}\,\delta^n[x^\rho - z_i^{\ \rho}(\tau_i)]\,d\tau_i. \tag{16}$$

While any space whatever admits the existence of a constant scalar, not all of them admit the existence of a scalar density of weight -1 with vanishing covariant derivative. Thus Eq. (13) can always be integrated along the entire ith world line, but this is not necessarily the case if it is a scalar density of weight -1. However, this is still possible provided the affine connection of the space under consideration is *symmetric*. By Fermi's theorem as generalized by Eisenhart,[15] we can introduce a coordinate system in such a space such that the Γ's vanish along the entire world line; then the covariant derivative in Eq. (13) reduces to an ordinary one, just as in the case of a scalar M_i, and thus Eq. (14) can be established as in that case.

Thus a conservation law (1) for a symmetric tensor density always implies the geodesic law for simple poles. On the other hand, such a conservation law for a tensor leads to the geodesic law only for spaces with a symmetric affine connection, and

[14] We could actually require that $v_i^{\ \nu}w_{i\nu} = 1$, and specify $w_{i\nu}$ even further by requiring say that in the "rest system," in which $v_i^{\ \nu}$ has the components $v_i^{\ 0}$, 0, $\cdots$ 0, $w_{i\nu}$ should have components $1/v_i^{\ 0}$, 0, $\cdots$ 0, but this is not needed.

[15] L. P. Eisenhart, *Non-Riemannian Geometry* (American Mathematical Society, New York, 1927), Sec. 25.

otherwise only implies a law of motion (12) which cannot be brought into a form involving no properties of the singularities other than their coordinates.

The geodesic (15) was obtained starting from the tensor density (2) required to satisfy the conservation law (1). Conversely, given the geodesic (15), we can construct tensor densities (2) with (11) or (16) which satisfy this conservation law, as can be easily verified by direct calculation.

III. APPLICATIONS

In the special case of a metric space considered earlier,[7] the Γ's can be expressed in terms of the metric tensor $g_{\mu\nu}$. Just as was the case there for $g_{\mu\nu}$, the very method of our general derivation implies that the affine connection $\Gamma^{\mu}_{\rho\sigma}$ entering the law of motion (15) remains undetermined. Furthermore, we have tacitly assumed that $\Gamma^{\mu}_{\rho\sigma}$ is finite. However, it was not necessary to assume that it is independent of m_i, and thus our result is not restricted to test particles. The problem of constructing a theory determining the dependence of the Γ's on the m_i's is separate from the problem of the form of the law of motion. Such a theory may, but does not necessarily have to, take the form of field equations for $\Gamma^{\mu}_{\rho\sigma}$.[16]

Examples in which the Γ's are determined by a different condition are the Whitehead-type theories of gravitation.[17] These are action-at-a-distance theories in the flat space–time of special relativity, where the interactions are determined by a Fokker-type variational principle. Although the physical metric is the Minkowski metric $\eta_{\mu\nu}$, the variational principle leads to equations of motion which are formally those of a geodesic in a Riemannian space with a metric tensor $g_{\mu\nu}$ given as an explicit function of the coordinates of the particles. From our results it follows immediately that there exists a conservation law (1), with the covariant derivatives to be taken using the affine connections following from $g_{\mu\nu}$.

An even simpler example is that of Newtonian mechanics without gravitation. The four-dimensional space–time of this theory is not metric, but only affine. However, it is flat, i.e., we can introduce coordinates such that $\Gamma^{\mu}_{\rho\sigma}$ vanishes everywhere. Assuming a four-dimensional conservation law (1), we obtain (with an appropriate parametrization),

$$m_i = \text{constant}, \qquad d^2 z_i^{\mu}/d\tau_i^2 = 0, \tag{17}$$

where $z_i^{\mu} = (t_i, z_i^1, z_i^2, z_i^3)$, and the space coordinates are Cartesian. Integration of these equations and subsequent elimination of τ_i leads to the result that the ordinary three-velocity must be constant. Thus the physical content of Eqs. (1) with (16) and of (17) is simply the constancy of mass and three-velocity (or three-momentum). The same Eqs. (17) with the same physical content hold in the metric flat space–time of special relativity. This is as expected, as the behavior of free particles is identical in both theories; the differences appear only in the presence of forces.[18] In both theories we can obtain a four-dimensional conservation law (1) using (17) as a starting point.

We can also apply our results to Newtonian mechanics *including* gravitation, which can be described by a curved space–time with an affine connection obeying field equations very similar to those of general relativity, as will be shown elsewhere.[19,20] Similarly they can be applied to any unified field theory containing a conservation law (1), such as Einstein's.[21]

IV. VECTOR CONSERVATION LAW

The method developed above can also be applied to a conservation law of the form

$$\mathfrak{I}^{\nu}{}_{;\nu} = 0 \tag{18}$$

for a current density $\mathfrak{I}^{\nu}$ (the equation of continuity). We take $\mathfrak{I}^{\nu}$ to be of the form analogous to (2),

$$\mathfrak{I}^{\nu} = \sum_i \int_{-\infty}^{\infty} j_i{}^{\nu}(\lambda_i)\, \delta^n[x^{\rho} - z_i{}^{\rho}(\lambda_i)]\, d\lambda_i, \tag{19}$$

substitute this into Eq. (18), multiply by an arbitrary scalar ξ and integrate over all x^{ρ}. Performing an integration by parts we obtain instead of (4)

$$\sum_i \int (-j_i{}^{\nu}\xi_{;\nu i})\, d\lambda_i = 0. \tag{20}$$

Now we break up $j_i{}^{\nu}$, putting

$$j_i{}^{\nu} = e_i(\lambda_i) v_i{}^{\nu} + n_i{}^{\nu}(\lambda_i), \qquad n_i{}^{\nu} w_{i\nu} = 0. \tag{21}$$

Substituting this into Eq. (20), using (6), and integrating by parts we get

[16] For a discussion of a very general class of such theories, see D. G. B. Edelen, *The Structure of Field Space* (University of California Press, Berkeley and Los Angeles, 1962).

[17] For a review see A. Schild, in *Proceedings of the International School of Physics "Enrico Fermi," Course XX*, (Academic Press Inc., New York, 1962), p. 69.

[18] For the derivation of the equations of motion of special relativity from the conservation laws including force fields, see P. Havas, Ref. 13, and references given there. A development of the Newtonian equations along similar lines (with the modifications due to the absence of a metric discussed here) can be carried out without difficulty. For a particular case it is discussed in Ref. 20.

[19] For empty space this was shown by K. Friedrichs, Ann. Math. **98**, 566 (1927); equivalent results were obtained by A. Trautman, Compt. rend. **257**, 617 (1963).

[20] P. Havas (to be published).

[21] V. Hlavatý, *Geometry of Einstein's Unified Field Theory* (P. Noordhoff Ltd., Groningen, The Netherlands, 1957), Chap. V.

$$\sum_i \int \left(n_i{}^{\nu} \xi_{;\nu i} - \frac{De_i}{d\lambda_i} \xi \right) d\lambda_i = 0. \tag{22}$$

Therefore $n_i{}^{\nu}$ and $De_i/d\lambda_i$ must vanish separately. Thus, independent of the parametrization used, we must have

$$j_i{}^{\nu} = e_i v_i{}^{\nu}, \qquad e_i = \text{constant}, \tag{23}$$

where the e_i are scalars, with no restriction on the motion, in complete analogy to the usual results of electrodynamics in metric spaces. If $\mathfrak{J}^{\nu}$ had been taken to be a vector rather than a vector density of weight 1, we would have had to take ξ to be a scalar density; the calculation would have been identical, the e_i's now being scalar densities of weight -1 with vanishing covariant derivative rather than constant scalars. The difference between the two cases is similar to that encountered in Sec. III, and remarks analogous to those made after Eq. (16) apply.

Just as the geodesic law is sufficient to assure a conservation law (1) for a singular tensor, the existence of quantities e_i constant along arbitrary world lines is sufficient to assure the existence of a conservation law (18) with a singular vector density (19) with (23). The proof is immediate.

The result (23) is immediately applicable to the space–time of Newtonian mechanics with or without gravitation, as well as to the various field theories discussed before. Substituting (23) into (19) we obtain an expression for the singular current density $\mathfrak{J}^{\nu}$ which, being parameter-independent, has the form familiar from electrodynamics in all four-dimensional theories.

ACKNOWLEDGMENT

The author is indebted to Dr. W. Tulczyjew for his valuable criticism of the manuscript.

APPENDIX

The auxiliary covariant vector $w_{i\nu}$ used in our derivations is arbitrary except for the restriction (7). Nevertheless, the result (15) for simple poles is free from arbitrariness. However, this is no longer the case for particles with an intrinsic dipole moment. In general relativity the equations of motion of such particles in a fixed background metric were first obtained by Mathisson.[10] A simplified derivation was given by Tulczyjew,[12] whose calculations we can follow with minor modifications in the case of an affine rather than a metric space. We take instead of (2)

$$\mathfrak{P}^{\mu\nu} = \sum_i \int_{-\infty}^{\infty} \{ \mathfrak{p}_i{}^{\mu\nu}(\lambda_i)\, \delta^n - [\mathfrak{p}_i{}^{\rho\mu\nu}(\lambda_i)\, \delta^n]_{;\rho} \}\, d\lambda_i. \tag{24}$$

Then $\mathfrak{p}_i{}^{\mu\nu}$ and $\mathfrak{p}_i^{\rho\mu\nu}$ are decomposed in components parallel to the $v_i{}^{\sigma}$ and perpendicular to the $w_{i\sigma}$. The resultant equations are

$$\frac{DS_i{}^{\mu\nu}}{d\lambda_i} + v_i{}^{\mu} \frac{w_{i\rho}}{w_{i\sigma} v_i{}^{\sigma}} \frac{DS_i{}^{\nu\rho}}{d\lambda_i} + v_i{}^{\nu} \frac{w_{i\rho}}{w_{i\sigma} v_i{}^{\sigma}} \frac{DS_i{}^{\rho\mu}}{d\lambda_i} = 0, \qquad S_i{}^{\mu\nu} = -S_i{}^{\nu\mu}, \tag{25}$$

$$\frac{Dp_i{}^{\mu}}{d\lambda_1} + \tfrac{1}{2} S_i{}^{\kappa\lambda} R^{\mu}{}_{\nu\kappa\lambda} v_i{}^{\nu} = 0,$$

$$p_i{}^{\mu} \equiv M_i v_i{}^{\mu} + \frac{w_{i\rho}}{w_{i\sigma} v_i{}^{\sigma}} \frac{DS_i{}^{\mu\rho}}{d\lambda_i}, \tag{26}$$

where $S_i{}^{\mu\nu}$ is a quantity characterizing the dipole, and $R^{\mu}{}_{\nu\kappa\lambda}$ is the Riemann–Christoffel curvature tensor; again, if Eq. (1) holds for a tensor density of weight 1, $S_i{}^{\mu\nu}$ is a tensor; if it holds for a tensor, $S_i{}^{\mu\nu}$ is a tensor density of weight -1. Equations (25) and (26) differ from those of Mathisson and Tulczyjew only by the appearance of $w_{i\rho}(w_{i\sigma} v_i^{\sigma})^{-1}$ in place of $v_{i\rho}$. As in the case of the simple poles considered above, we are not restricted to test particles, but our equations hold even if the Γ's depend on the variables characterizing the particles.

We could again reparametrize by Eqs. (13) and (14) if desired. We could also make Eqs. (25) and (26) more definite by imposing a condition such as that of footnote 14 on the w_i's, but even then they can not in general be simplified as is possible in the metric case.

There the last term of (26) vanishes on contraction with $v_{i\mu}$ because of the symmetry properties of the totally covariant $R_{\mu\nu\kappa\lambda}$; as this quantity can not be constructed in the absence of a metric tensor, such a simplification is no longer possible except under special conditions. Similarly, the condition $S_i{}^{\mu\nu} p_{i\nu} = 0$ suggested by Tulczyjew to obtain an invariant definition of the center of mass cannot be imposed, because $p_{i\nu}$ cannot be constructed. Although a condition $S_i{}^{\mu\nu} w_{i\nu} = 0$ analogous to Mathisson's condition $S_i{}^{\mu\nu} v_{i\nu} = 0$ (which is necessary to allow us to interpret $S_i{}^{\mu\nu}$ as an intrinsic angular momentum) could be imposed, unlike the metric case this does not allow us either to deduce the constancy of the magnitude of $S_i{}^{\mu\nu}$ from Eq. (25) or to simplify the structure of Eq. (26). Thus, unlike the case of simple poles subject to a conservation law (1), the laws of motion of dipoles in affine spaces are not in general identical with those in metric spaces, and no completely satisfactory physical interpretation of the results appears possible.

However, under special conditions such as exist, e.g., in the case of Newtonian mechanics considered in Sec. III, we can still obtain unambiguous and physically satisfactory equations. In this case, as discussed in Ref. 20, we can define unique vectors $w_{i\nu}$ with vanishing covariant derivatives both in the absence and in the presence of gravitation. Then the imposition of the condition $S_i^{\mu\nu}w_{i\nu} = 0$ implies $w_{i\nu}DS_i^{\mu\nu}/d\lambda_i = 0$. Thus Eq. (25) reduces to $DS_i^{\mu\nu}/d\lambda_i = 0$ and the linear momentum p_i^μ to $M_i v_i^\mu$. Furthermore in the absence of gravitation, the last term of Eq. (26) vanishes because the space is flat; in the presence of gravitation this is not the case, but the last term still vanishes, as a consequence of the form of $S_i^{\mu\nu}$ implied by the condition imposed on it and of the properties of the Newtonian curvature tensor[20]. Thus in the Newtonian case with or without gravitation, Eqs. (25) and (26) imply the vanishing of the covariant derivatives of the linear and of the angular momentum of each particle, and, regardless of its dipole moment, it still moves along a geodesic.

Determinantal Method in Perturbation Theory*

TATUYA SASAKAWA†
Department of Physics, University of California, Los Angeles, California
(Received 30 July 1963)

The determinantal method of deriving fundamental relations of a new perturbation theory, which the author has presented recently, is demonstrated. This method is simpler than the Green's function method which has been adopted in the previous paper. The Brillouin–Wigner perturbation theory is discussed for comparison.

I. INTRODUCTION

IN a recent issue of the *Journal of Mathematical Physics*, the author has presented a new method for solving eigenvalue problems.[1] In that paper, which will be referred to as I, the fundamental equations are derived on the basis of the Green's function method.

To summarize: The secular determinantal equation

$$\det(E - H) = 0 \qquad \text{[I.1.1]}, \tag{1}$$

has been transformed into the dispersion relation

$$\sum_i \frac{C_i}{E - \omega_i} = 1 \qquad \text{[I.1.4]}. \tag{2}$$

Here ω_i stands for the first-order energy in the conventional perturbation theory,

$$\omega_i = \epsilon_i + V_{ii}.$$

C_i was obtained by the Green's function method as

$$C_i = \sum_{j\neq i} \frac{g_{ij}}{\omega_i - \omega_j} + \sum_{j\neq i} \sum_{\substack{k>j \\ \neq i}} \frac{g_{ijk}}{(\omega_i - \omega_j)(\omega_i - \omega_k)} + \sum_{i\neq i} \sum_{\substack{k>j \\ \neq i}} \sum_{\substack{l>k \\ \neq i}} \frac{g_{ijkl}}{(\omega_i - \omega_j)(\omega_i - \omega_k)(\omega_i - \omega_l)} + \cdots \qquad \text{[I.2.34]}. \tag{3}$$

Here

$$g_{ij} = |V_{ij}|^2,$$
$$g_{ijk} = S_{ijk}(V_{ik}V_{kj}V_{ji}),$$
$$g_{ijkl} = S_{ikl}[S_{kl}(V_{il}V_{lk}V_{ki})V_{ji} - |V_{kl}|^2\,|V_{ij}|^2], \quad \text{etc.} \qquad \text{[I.2.31]}, \tag{4}$$

where

$$S_{ik}f_{ik} = f_{ik} + f_{ki},$$
$$S_{ikl}f_{ikl} = f_{ikl} + f_{kli} + f_{lik}, \quad \text{etc.} \qquad \text{[I.2.12]}. \tag{5}$$

In I, three methods for solving Eq. (2) were suggested. Several examples were demonstrated. All these examples show that this approach gives more rapid convergence than other conventional perturbation theories.

The transformation coefficient $(a|\lambda)$ between the perturbed wavefunction $|\lambda)$ and the unperturbed one $|a)$ is given by

$$|(a \mid \lambda)|^2 = \mathfrak{U}_a(E_\lambda) \frac{\prod_{a'\neq a} (E_\lambda - \omega_{a'})}{\prod_{\lambda'\neq\lambda} (E_\lambda - E_{\lambda'})} \qquad \text{[I.2.4]}, \tag{6}$$

and

$$(b \mid \lambda) = [\mathfrak{V}_{ba}(E_\lambda)/\mathfrak{U}_a(E_\lambda)](a \mid \lambda) \qquad \text{[I.2.7]}, \tag{7}$$

where

$$\mathfrak{U}_a(E) = 1 - \sum_{i\neq a} \sum_{\substack{j>i \\ \neq a}} g_{ij}/(E - \omega_i)(E - \omega_j) - \sum_{i\neq a} \sum_{\substack{j>i \\ \neq a}} \sum_{\substack{k>j \\ \neq a}} g_{ijk}/(E - \omega_i)(E - \omega_j) \times (E - \omega_k) - \cdots \qquad \text{[I.2.35]}, \tag{8}$$

and

$$\mathfrak{V}_{ba}(E) = \frac{V_{ba}}{E - \omega_b} \times \left[1 - \sum_{i\neq ab} \sum_{\substack{j>i \\ \neq ab}} g_{ij}/(E - \omega_i)(E - \omega_j) - \sum_{i\neq ab} \sum_{\substack{j>i \\ \neq ab}} \sum_{\substack{k>j \\ \neq ab}} g_{ijk}/(E - \omega_i) \times (E - \omega_j)(E - \omega_k) - \cdots\right] + \sum_{i\neq ab} h_{bia}/(E - \omega_b)(E - \omega_i) + \sum_{i\neq ab} \sum_{\substack{j>i \\ \neq ab}} h_{bija}/(E - \omega_b)(E - \omega_i)(E - \omega_j) + \cdots \qquad \text{[I.2.44]}. \tag{9}$$

* Supported by the National Science Foundation and the Army Research Office.
† On leave of absence from Kyoto University as a Fulbright Visiting Research Scholar. Now at Faculty of Science, Kyoto University, Kyoto, Japan.
[1] T. Sasakawa, J. Math. Phys. 4, 970 (1963).

Here

$$h_{bia} = V_{bi}V_{ia},$$
$$h_{bija} = S_{ij}(V_{bi}V_{ji}V_{ia}),$$
$$h_{bijka} = S_{ijk}[S_{jk}(V_{bk}V_{kj}V_{ji})V_{ia} - |V_{jk}|^2 V_{bi}V_{ia}], \quad \text{etc.} \quad \text{[I.2.45].} \tag{10}$$

Equations (6) and (7) yield the orthonormalized set of $|\lambda)$, if $|a)$ are orthonormalized.

The primary purpose of the present paper is to derive these formulas by the determinantal method. This will be demonstrated in Sec. II. For comparison, the determinantal derivation of the Brillouin–Wigner perturbation theory[2,3] will be given in Sec. III.

II. DERIVATION OF THE FORMULA

Derivation of (2), (3), and (4) is very simple if we use the determinantal method.[4] Let us multiply the nondiagonal matrix elements by a parameter s, which is, of course, set equal to 1 in the final results. The secular determinantal equation (1) is then

$$\begin{vmatrix} E-\omega_1 & -sV_{12} & \cdots & -sV_{1n} \\ -sV_{21} & E-\omega_2 & \cdots & -sV_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -sV_{n1} & -sV_{n2} & \cdots & E-\omega_n \end{vmatrix} = 0. \tag{11}$$

The left-hand side is expanded as power series in s

$$\prod_{i=1}^{n}(E-\omega_i) + s^2 \sum_{i=1}^{n}\sum_{j>i}^{n} \begin{vmatrix} 0 & V_{ij} \\ V_{ji} & 0 \end{vmatrix} \prod_{n'\neq ij}'' (E-\omega_{n'}) - s^3 \sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{k>j}^{n} \begin{vmatrix} 0 & V_{ij} & V_{ik} \\ V_{ji} & 0 & V_{jk} \\ V_{ki} & V_{kj} & 0 \end{vmatrix} \times \prod_{n'\neq ijk}''' (E-\omega_{n'}) + \cdots = 0. \tag{12}$$

Dividing each term by $\prod_{i=1}^{n}(E-\omega_i)$, one obtains the dispersion relation (2), if there is no degeneracy with respect to ω_i. g_{ij}, g_{ijk} etc., take on the form

$$g_{ij} = -\begin{vmatrix} 0 & V_{ij} \\ V_{ji} & 0 \end{vmatrix},$$
$$g_{ijk} = \begin{vmatrix} 0 & V_{ij} & V_{ik} \\ V_{ji} & 0 & V_{jk} \\ V_{ki} & V_{kj} & 0 \end{vmatrix},$$
$$g_{ijkl} = -\begin{vmatrix} 0 & V_{ij} & V_{ik} & V_{il} \\ V_{ji} & 0 & V_{jk} & V_{jl} \\ V_{ki} & V_{kj} & 0 & V_{kl} \\ V_{li} & V_{lj} & V_{lk} & 0 \end{vmatrix}, \tag{13}$$

etc. It is clear that these can be expressed as (4).

From the expression (13) one can easily prove that g_{ij}, g_{ijk}, etc., are all real. Let these quantities be expressed as $\det|V_{ij}|$. Then $\det|V_{ij}| = (\det|V_{ij}|)^t \equiv A + iB$. ($t$; transposed. A and B are real.) Accordingly, $(\det|V_{ij}|)^\dagger = A - iB$ (†; Hermitian conjugate). From the Hermitian nature of the matrix elements $(\det|V_{ij}|)^\dagger = \det|V_{ij}|$. Thus $B = 0$. Once we have proved that g_{ij}, g_{ijk}, etc. are real, it directly follows that C_i's are real, and thus the solution of Eq. (2) is real.

Next, we shall derive (6) and (7). The expectation value of the Green's function $(E-H)^{-1}$ is

$$\sum_\lambda \frac{(a\mid\lambda)(\lambda\mid a)}{E-E_\lambda} = (a|\frac{1}{E-H}|a),$$
$$\sum_\lambda \frac{(b\mid\lambda)(\lambda\mid a)}{E-E_\lambda} = (b|\frac{1}{E-H}|a). \tag{14}$$

Thus $|(a|\lambda)|^2$ and $(b|\lambda)(\lambda|a)$ are residues of the functions $(a|\ (E-H)^{-1}\ |a)$ and $(b|\ (E-H)^{-1}\ |a)$, respectively:

$$|(a\mid\lambda)|^2 = \lim_{E\to E_\lambda}(E-E_\lambda)(a|\frac{1}{E-H}|a),$$
$$(b\mid\lambda)(\lambda\mid a) = \lim_{E\to E_\lambda}(E-E_\lambda)(b|\frac{1}{E-H}|a). \tag{15}$$

The calculation of these functions is performed most simply as follows: The unperturbed wave-function $|a)$ is expressed as

$$|a) = (E-H)\frac{1}{E-H}|a)$$
$$= (E-H)\,|a)(a|\frac{1}{E-H}|a)$$
$$+ \sum_{b\neq a}(E-H)\,|b)(b|\frac{1}{E-H}|a). \tag{16}$$

Multiplying $(a|$ and $(b|$, $(b \neq a)$ from the left, we obtain a set of coupled equations,

$$\begin{vmatrix} E-\omega_a & -sV_{ab} & -sV_{ac} & \cdots \\ -sV_{ba} & E-\omega_b & -sV_{bc} & \cdots \\ -sV_{ca} & -sV_{cb} & E-\omega_c & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{vmatrix}$$

[2] L. Brillouin, J. Phys. Radium 3, 373 (1932).
[3] E. P. Wigner, Math. Naturw. Anz. Ungar. Akad. Wiss. 53, 475 (1935).
[4] See also M. Baker, Ann. Phys. (N.Y.) 4, 271 (1958).

$$\times \begin{vmatrix} (a|\frac{1}{E-H}|a) \\ (b|\frac{1}{E-H}|a) \\ (c|\frac{1}{E-H}|a) \\ \cdots \end{vmatrix} = \begin{vmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{vmatrix}. \quad (17)$$

The solution of this set of coupled equation is

$$(a|\frac{1}{E-H}|a) = \frac{D_{aa}(E)}{D(E)}, \qquad (b|\frac{1}{E-H}|a) = \frac{D_{ab}(E)}{D(E)} \quad (b \neq a), \quad (18)$$

where

$$D(E) \equiv \det(E - H). \quad (19)$$

D_{ab} is the cofactor of the (a, b) element of $D(E)$. Since (1) is transformed into (2), $D(E)$ can be expressed as

$$D(E) = \prod_{a=1}^{n}(E-\omega_a)\left(1 - \sum_{i=1}^{n}\frac{C_i}{E-\omega_i}\right). \quad (20)$$

$D_{aa}(E)$ is equal to the determinant $D(E)$, whose a row and a column are missing. Thus,

$$D_{aa}(E) = \prod_{\substack{a'=1 \\ (a'\neq a)}}^{n}{}'(E-\omega_{a'})\left[1 - \sum_{\substack{i=1 \\ (i\neq a)}}^{n}{}'\frac{C_i(a)}{E-\omega_i}\right]. \quad (21)$$

Here

$$C_i(a) = \sum_{j\neq i,a}{}''\frac{g_{ij}}{\omega_i-\omega_j} + \sum_{j\neq i,a}{}''\sum_{\substack{k>j \\ \neq i,a}}{}''\frac{g_{ijk}}{(\omega_i-\omega_j)(\omega_i-\omega_k)} + \sum_{j\neq i,a}{}''\sum_{\substack{k>j \\ \neq i,a}}{}''\sum_{\substack{l>k \\ \neq i,a}}{}''\frac{g_{ijkl}}{(\omega_i-\omega_j)(\omega_i-\omega_k)(\omega_i-\omega_l)} + \cdots. \quad (22)$$

We can easily check that the last factor in the right-hand side of Eq. (21) is equal to (8):

$$D_{aa}(E) = \prod_{\substack{a'=1 \\ (a'\neq a)}}^{n}{}'(E-\omega_{a'})\mathfrak{U}_a(E). \quad (23)$$

Since the solutions of (1) are E_λ ($\lambda = 1, \cdots, n$), $D(E)$ should be factorized as

$$D(E) = \prod_{\lambda=1}^{n}(E-E_\lambda). \quad (24)$$

Finally, from (15), (18), (21), and (23), we obtain (6).

A similar consideration as in the Fredholm theory of the integral equation[5] leads to

$$D_{ab}(E) = \prod_{a'\neq a}{}'(E-\omega_{a'})\left[s\frac{V_{ba}}{E-\omega_b} - s^2\sum_{\substack{i=1 \\ (i\neq a,b)}}^{n}{}''\begin{vmatrix} V_{ba} & V_{bi} \\ V_{ia} & 0 \end{vmatrix}\frac{1}{(E-\omega_b)(E-\omega_i)} + s^3\sum_{\substack{i=1 \\ (i\neq a,b)}}^{n}{}''\sum_{\substack{j>i \\ (j\neq a,b)}}^{n}{}''\begin{vmatrix} V_{ba} & V_{bi} & V_{bj} \\ V_{ia} & 0 & V_{ij} \\ V_{ja} & V_{ji} & 0 \end{vmatrix} \times \frac{1}{(E-\omega_b)(E-\omega_i)(E-\omega_j)} + \cdots\right]. \quad (25)$$

Then it is an easy matter to show that

$$D_{ab}(E) = \prod_{a'\neq a}{}'(E-\omega_{a'})\mathfrak{V}_{ba}(E), \quad (26)$$

$\mathfrak{V}_{ba}(E)$ being given by (9).

From equations (15), (18), (23), and (26), we obtain

$$\frac{(b|\lambda)}{(a|\lambda)} = \frac{D_{ab}(E_\lambda)}{D_{aa}(E_\lambda)} = \frac{\mathfrak{V}_{ba}(E_\lambda)}{\mathfrak{U}_a(E_\lambda)}. \quad (27)$$

This is Eq. (7). Thus we have derived all the fundamental equations of the new perturbation theory by the determinantal method.

III. COMPARISON WITH THE BRILLOUIN-WIGNER PERTURBATION THEORY

In the previous section, we have seen that fundamental formulas of the new perturbation theory is derived in a close connection with the Fredholm theory of the integral equation.

Let us examine the determinantal structure of the Brillouin–Wigner perturbation theory. Let us multiply all the matrix elements by a parameter s. The secular determinantal equation is then

$$\begin{vmatrix} E-\epsilon_1-sV_{11} & -sV_{12} & \cdots & -sV_{1n} \\ -sV_{21} & E-\epsilon_2-sV_{22} & \cdots & -sV_{2n} \\ \vdots & \vdots & & \vdots \\ -sV_{n1} & -sV_{n2} & \cdots & E-\epsilon_n-sV_{nn} \end{vmatrix} = 0. \quad (28)$$

[5] See, for example, E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, England, 1935), Chap. 11.

We obtain the zero-order equation as

$$\prod_{i=1}^{n} (E - \epsilon_i) = 0. \tag{29}$$

The zero-order energy is then $E^{(0)} = \epsilon_a$. The first-order equation is

$$\prod_{i=1}^{n} (E - \epsilon_i) - s \prod_{i=1}^{n} V_{ii} \prod_{j\neq i}{}' (E - \epsilon_j) = 0. \tag{30}$$

The first-order energy is then $E^{(1)} = \epsilon_a + V_{aa}$. The second-order equation is

$$\prod_{i=1}^{n} (E - \epsilon_i) - s \sum_{i=1}^{n} V_{ii} \prod_{j\neq i}{}' (E - \epsilon_j) + s^2 \sum_{i=1}^{n} \sum_{j>i}^{n} \begin{vmatrix} V_{ii} & V_{ij} \\ V_{ji} & V_{jj} \end{vmatrix} \prod_{k\neq ij}{}'' (E - \epsilon_k) = 0. \tag{31}$$

Dividing each term by $\prod'_{i\neq a} (E - \epsilon_i)$, we get

$$E - \epsilon_a - sV_{aa} - s \sum_{i\neq a}{}' \frac{V_{ii}(E - \epsilon_a)}{E - \epsilon_i} + s^2 \sum_{i\neq a} \begin{vmatrix} V_{aa} & V_{ai} \\ V_{ia} & V_{ii} \end{vmatrix} \frac{1}{E - \epsilon_i} + s^2 \sum_{i\neq a} \sum_{j>i} \begin{vmatrix} V_{ii} & V_{ij} \\ V_{ji} & V_{jj} \end{vmatrix} \frac{E - \epsilon_a}{(E - \epsilon_i)(E - \epsilon_j)} = 0. \tag{32}$$

After putting the first-order energy $E^{(1)}$ in the numerator of the first-order term, and the zero-order energy $E^{(0)}$ in the numerator of the second-order term of this equation, we obtain the second-order Brillouin–Wigner equation

$$E = \epsilon_a + sV_{aa} + s^2 \sum_{i\neq a}{}' \frac{V_{ai}V_{ia}}{E - \epsilon_i}. \tag{33}$$

A similar procedure will be adopted in higher-order terms: First, we write out nth-order Fredholm expansion. Then we divide each term by $\prod'_{i\neq a}(E - \epsilon_i)$. In the resulting equation, we put $(n - i)$th-order energy in the numerator of ith-order term. This is the determinantal derivation of the Brillouin–Wigner perturbation theory. Note that in the course of deriving the dispersion relation [Eq. (2)], we do not need any such approximate solution of E.

ACKNOWLEDGMENTS

The author should like to thank Professor S. A. Moszkowski for research facilities extended to him and interest in this work. The author is indebted to Dr. K. T. Mahanthappa, Dr. M. Yokota, and Dr. K. Maki for bringing the determinantal method to his notice.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Energy Sharing and Equilibrium for Nonlinear Systems

ROBERT S. NORTHCOTE,* AND RENFREY B. POTTS
Department of Mathematics, University of Adelaide, Adelaide, South Australia
(Received 22 October 1963)

A study is made of a one-dimensional system of identical particles in which the forces between neighbors are linear. The system is nonlinear because it is assumed that collisions occur between adjacent particles, which each have an effective diameter d. The energies E_i in the linear normal modes are computed numerically to show that energy is freely exchanged between all the modes in the system, as predicted by the theory. Furthermore, the time averages $\langle E_i \rangle$ of these energies show a strong tendency towards equipartition of energy among the modes. This is in distinct contrast to the computations of Ulam, Fermi, and Pasta, which showed that some nonlinear systems appear to be nonergodic. An equation of state and an expression for the total energy of the system as a function of thermodynamic coordinates are derived via statistical mechanics. Expected values for the pressure and temperature of the assembly may then be computed. A comparison of these with the numerical values of those variables arising from the computations shows that the nonlinear system approaches equilibrium.

I. INTRODUCTION

CONSIDERABLE interest has been shown in the surprising results obtained by Fermi, Pasta, and Ulam,[1] who carried out numerical computations on a one-dimensional nonlinear system and found, rather unexpectedly, very little tendency towards equipartition of energy among the linear normal modes. They considered an assembly of point particles coupled by forces which were linear except for small nonlinear terms. The interaction potential between two adjacent particles a distance r apart was assumed to be of the form

$$\phi(r) = \tfrac{1}{2}\gamma(r - a)^2 + \phi_1(r),$$

where a and γ are constants, and $\phi_1(r)$ is a perturbation term which gives rise to small nonlinear forces. In the cases studied, these forces were quadratic, cubic, or broken linear. FPU found, for example, that energy was shared periodically by only the first few modes when the assembly was set in motion in the first linear mode. There was no evidence of ergodicity in the nonlinear systems studied.

An explanation of these observations has been provided by Ford,[2,3] who investigated the general form of the Kryloff and Bogoliuboff series solution to the equations of motion. He explained this apparent nonergodicity in terms of a lack of internal resonance among the uncoupled frequencies. Jackson,[4,5] while agreeing that Ford's observation is correct in the case of weak coupling, shows that the behavior of such systems is also a function of the initial conditions of the assembly and the particular form of mode interaction.

In this paper we consider a one-dimensional assembly of particles in which nonlinearity is introduced by assuming that collisions occur between adjacent particles which each have an effective diameter d. The model is described in detail in the next section. The advantages of the model are that it has physical realism, nonlinear effects can be made small or large merely by altering the mean energy per particle of the system, and the computation is facilitated because the assembly behaves linearly between collisions.

If the system were linear, the energy E_i in the ith normal mode would be a constant of the motion. In the nonlinear system, E_i is constant only between collisions, having a finite jump discontinuity at each collision. Expressions for the coupling and energy sharing between the modes at a collision are derived exactly. It is shown that energy is shared among all the modes, irrespective of whether or not the system is set in motion in the first, or lowest, mode. However the rate of energy sharing is found to be dependent on the initial conditions. Partitioning of the total energy among the degrees of freedom is demonstrated by numerical computation of the time average $\langle E_i \rangle$ of the energy in each mode.

An equation of state, and an expression for the total energy of the system as a function of thermodynamic variables, may be simply derived via statistical mechanics by the use of an isobaric grand partition function. Since the energy is known,

* Present address: Digital Computer Laboratory, University of Illinois, Urbana, Illinois.

[1] E. Fermi, J. Pasta, and S. Ulam, Los Alamos Scientific Laboratory Rept. LA-1940 (1955), hereafter referred to as FPU.

[2] J. Ford, J. Math. Phys. 2, 387 (1961).

[3] J. Ford and J. Waters, J. Math. Phys. 4, 1293 (1963).

[4] E. A. Jackson, J. Math. Phys. 4, 551 (1963).

[5] E. A. Jackson, J. Math. Phys. 4, 686 (1963).

expected values for the temperature and pressure of the assembly may be computed. These provide a check for the numerical values of those variables arising from the computations. The estimates for the pressure and the temperature are derived from the forces acting on the end particles, which are held fixed, and from the total kinetic energy, respectively. Fluctuations in these estimates are computed in order to demonstrate that the assembly tends towards a state of equilibrium. Finally, the velocity autocorrelation function is discussed briefly.

II. THE MODEL

A. Definition of the Model

By the one-dimensional model, we mean a system of $N + 2$ identical particles,[6] each of mass M and effective diameter d, constrained to move on a straight line. The end particles, held fixed at a distance L apart, behave like rigid walls. If the particles are equally spaced, the distance between an adjacent pair is l, where

$$(N + 1)l = L. \tag{1}$$

It is supposed that adjacent particles are connected to each other by identical springs of spring constant γ and natural length $a - d$.

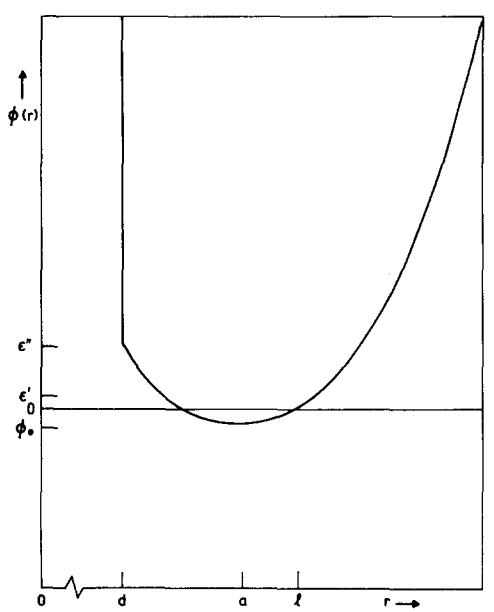

FIG. 1. The potential $\phi(r) = \phi_0 + \frac{1}{2}\gamma(r - a)^2$, where $\phi_0 = -\frac{1}{2}\gamma(l - a)^2$ if $\phi(l) = 0$, between particles of diameter d with centers r apart. The spring has constant γ, natural length $a - d$. The systems studied have a mean energy/particle given by ϵ' or ϵ''.

[6] Some authors use N degrees of freedom while others use $N - 1$, leading to confusion in the literature. In this paper we choose $N + 2$ particles, in which case there are N degrees of freedom if the end two particles are held fixed.

The force of interaction between two adjacent particles, with centers a distance r apart, is

$$f = -\gamma(r - a), \tag{2}$$

a linear function of r. Nonlinearity is introduced into the system by supposing that adjacent particles suffer elastic collisions when they meet, their centers then being a distance d apart. The system is thus a conservative one, with a potential energy of interaction given by

$$\begin{aligned} &\phi(r) = \infty, && r \le d, \\ \text{and}\quad &\phi(r) = \phi_0 + \tfrac{1}{2}\gamma(r - a)^2 && r > d. \end{aligned} \tag{3}$$

A graph of this function is sketched in Fig. 1. If the potential energy is taken to be zero when the particles are equally spaced, then $\phi(l) = 0$, and $\phi_0 = -\frac{1}{2}\gamma(l - a)^2$.

If the displacement of the jth particle from its equilibrium position is x_j at time t, then the equations of motion of the system, between collisions, are[7]

$$\begin{aligned} &M\ddot{x}_j - \gamma(x_{j+1} - 2x_j + x_{j-1}) = 0, \\ \text{with}\quad &x_0 = x_{N+1} \equiv 0 \quad \text{for all} \quad t. \end{aligned} \tag{4}$$

A collision occurs between particles $j - 1$ and j when any of the equalities

$$x_j - x_{j-1} + l - d = 0 \qquad (j = 1, 2, \cdots, N + 1) \tag{5}$$

is satisfied, where we have made use of the boundary conditions. The equality signs in (5) are replaced by inequality signs between collisions, i.e., between collisions,

$$x_j - x_{j-1} + l - d > 0 \qquad (j = 1, 2, \cdots, N + 1). \tag{6}$$

When a collision occurs, the positions of the particles are unaltered, but there is a discontinuity in at least one of the velocities. For a collision between particles $j - 1$ and j at time t, the velocities of the colliding particles immediately after the collision are determined according to the scheme:

collision index j	velocity changes
1	$\dot{x}_1(t+) = -\dot{x}_1(t-)$
$2, \cdots, N$	$\dot{x}_{j-1}(t+) = \dot{x}_j(t-)$, $\dot{x}_j(t+) = \dot{x}_{j-1}(t-)$
$N + 1$	$\dot{x}_N(t+) = -\dot{x}_N(t-)$.

(7)

[7] It will be assumed throughout this paper that the range of the suffices i and j is $(1, 2, \cdots, N)$ except where another range is specified, or the context indicates a different interpretation.

The collision indices $j = 1$ and $j = N + 1$ correspond to collisions on the left and right walls, respectively. All other collisions are internal. Equations (4) together with the conditions (6) and (7) completely define the motion of the assembly of particles.

B. The Mathematical Solution

Equation (4) may be rewritten, in matrix notation, in the form

$$\ddot{\mathbf{x}} + \mathbf{R}\mathbf{x} = \mathbf{0}, \tag{8}$$

where $\mathbf{x}$ is the position vector at any instant, and $\mathbf{R}$ is the positive definite matrix given, respectively, by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix},$$

and

$$\mathbf{R} = \frac{\gamma}{M}\begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & & & \\ 0 & -1 & 2 & -1 & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ 0 & & & & -1 & 2 & -1 \\ 0 & 0 & & \cdots & 0 & -1 & 2 \end{bmatrix}. \tag{9}$$

The vector $\mathbf{x}$ is of order N; $\mathbf{R}$ is of order $N \times N$.

The elements of the orthogonal symmetric matrix $\mathbf{S}$ which transforms $\mathbf{R}$ into diagonal form are given by

$$S_{ij} = [2/(N + 1)]^{\frac{1}{2}} \sin [ij\pi/(N + 1)]. \tag{10}$$

In the linear system, normal coordinates are defined by the transformation

$$\mathbf{q} = \mathbf{S}\mathbf{x}, \tag{11}$$

and the normal modes vibrate with frequencies $\omega_i/2\pi$, where

$$\omega_i = 2(\gamma/M)^{\frac{1}{2}} \sin [\tfrac{1}{2}i\pi/(N + 1)]. \tag{12}$$

The solution of (8) may be written down immediately in the form

$$\mathbf{x}(t) = \cos [\mathbf{R}^{\frac{1}{2}}(t - t_0)]\mathbf{x}(t_0) + \mathbf{R}^{-\frac{1}{2}} \sin [\mathbf{R}^{\frac{1}{2}}(t - t_0)]\dot{\mathbf{x}}(t_0),$$

from which it follows that

$$\dot{\mathbf{x}}(t) = -\mathbf{R}^{\frac{1}{2}} \sin [\mathbf{R}^{\frac{1}{2}}(t - t_0)]\mathbf{x}(t_0) + \cos [\mathbf{R}^{\frac{1}{2}}(t - t_0)]\dot{\mathbf{x}}(t_0).$$

These last two equations may be written together in the form

$$\begin{bmatrix} \mathbf{x}(t) \\ \dot{\mathbf{x}}(t) \end{bmatrix} = \mathbf{A}(t - t_0)\begin{bmatrix} \mathbf{x}(t_0) \\ \dot{\mathbf{x}}(t_0) \end{bmatrix}, \tag{13}$$

where

$$\mathbf{A}(t) = \begin{bmatrix} \mathbf{A}_1(t) & \mathbf{A}_2(t) \\ \mathbf{A}_3(t) & \mathbf{A}_1(t) \end{bmatrix}, \tag{14}$$

and

$$\mathbf{A}_2(t) = \mathbf{R}^{-\frac{1}{2}} \sin (\mathbf{R}^{\frac{1}{2}}t), \tag{15a}$$

$$\mathbf{A}_1(t) = \dot{\mathbf{A}}_2(t), \tag{15b}$$

$$\mathbf{A}_3(t) = -\mathbf{R}\mathbf{A}_2(t). \tag{15c}$$

The transformation (13), which is exact between collisions, is the basis for computations on the model. From it the values of the x coordinates and their derivatives at any time t may be obtained from their values at time t_0, provided that no collisions have occurred in the time interval (t, t_0).

The system is defined to be symmetric if $q_{2k}(t) \equiv 0 \equiv \dot{q}_{2k}(t) (k = 1, 2, \cdots, [\frac{1}{2}(N + 1)])$, and antisymmetric if $q_{2k-1}(t) \equiv 0 \equiv \dot{q}_{2k-1}(t) (k = 1, 2, \cdots, [\frac{1}{2}N])$. It is well known that symmetry and antisymmetry are conserved in the linear system. In the nonlinear system, antisymmetry is conserved, but symmetry is destroyed. This is easily demonstrated by considering a system of two free particles. If the system is symmetric, and $\dot{q}_1 > 0$ for example, a collision will occur on the right wall at some time t. Since $\dot{x}_2(t+) = -\dot{x}_2(t-)$, and $\dot{x}_1(t+) = \dot{x}_1(t-) = \dot{x}_2(t-)$, then $\dot{q}_2(t+) = \sqrt{2}\dot{x}_2(t-) \neq 0$. Symmetry has obviously been destroyed. On the other hand, if the system is antisymmetric, collisions will always occur either when the two particles meet midway between the walls, or when they hit the walls simultaneously. In each case it can be shown that antisymmetry is conserved. We shall always avoid this special case when choosing the initial conditions for numerical computations on the model.

III. THE NUMERICAL METHOD OF COMPUTATION

Suppose that a collision has occurred at time t_n, and that $\mathbf{x}(t_n)$ and $\dot{\mathbf{x}}(t_n+)$ are the position and velocity vectors immediately after the collision. After a subsequent time interval h, the position and velocity vectors can be found from

$$\begin{bmatrix} \mathbf{x}(t_n + h) \\ \dot{\mathbf{x}}(t_n + h) \end{bmatrix} = \mathbf{A}(h)\begin{bmatrix} \mathbf{x}(t_n) \\ \dot{\mathbf{x}}(t_n) \end{bmatrix}.$$

This iteration is then repeated, the $m + 1$st iteration being given by

$$\begin{bmatrix} \mathbf{x}[t_n + (m+1)h] \\ \dot{\mathbf{x}}[t_n + (m+1)h] \end{bmatrix} = \mathbf{A}(h) \begin{bmatrix} \mathbf{x}(t_n + mh) \\ \dot{\mathbf{x}}(t_n + mh) \end{bmatrix}.$$

After each iteration the inequalities (6) are tested. If no inequality is violated, further iterations are performed until one of the inequalities, say the jth, is not satisfied. This indicates that a collision between particles $j - 1$ and j has occurred in the previous time interval h. This procedure establishes the time t_{n+1} of the next collision to within h; in fact, if the first inequality to be unsatisfied is found after the $m + 1$st iteration, a first estimate of t_{n+1} is taken as

$$t'_{n+1} = t_n + mh.$$

The corresponding estimates of the position and velocity vectors immediately before this collision are

$$\mathbf{x}(t'_{n+1}) = \mathbf{x}(t_n + mh),$$

and

$$\dot{\mathbf{x}}(t'_{n+1}) = \dot{\mathbf{x}}(t_n + mh).$$

To enable a more accurate estimate of t_{n+1} to be made, K matrices $\mathbf{A}(h_i)$ $(i = 1, 2, \cdots, K)$ are evaluated, where h_1 is some convenient starting interval, and $h_{i+1} = p^{-1}h_i$, p being an integer greater than one. If $m_i + 1$ iterations have been computed with $\mathbf{A}(h_i)$, then an estimate of t_{n+1} is

$$t'_{n+1} = t_n + \sum_{i=1}^{K} m_i h_i,$$

and

$$t_{n+1} - t'_{n+1} < h_K.$$

A final estimate of t_{n+1} is obtained by making use of the known values of the position and velocity vectors at time t'_{n+1}. If h_K is sufficiently small, an extrapolation to $O[(\delta t)^3]$ by Newton's method will establish t_{n+1} to the accuracy of the computer. The final estimate of the collision time is then

$$t_{n+1} = t'_{n+1} + \Delta t,$$

where Δt is the small time correction established by the third-order approximation. The values of the position and velocity vectors at time t_{n+1} are computed by extrapolation from the values at time t'_{n+1}.

Having determined the time t_{n+1} and the values of the position and velocity vectors immediately before the collision, the velocities of the colliding particles are altered according to (7) to determine the velocity vector $\dot{\mathbf{x}}(t_{n+1}+)$ immediately after the collision. The whole process is then repeated to determine the time of the next collision, and so on. A method for numerically evaluating the elements of the iteration matrices $\mathbf{A}(h_i)$ $(i = 1, \cdots, K)$ will now be derived.

If an approximation to the function $\sin\theta$ can be represented by the economized polynomial of degree $2n + 1$,

$$\sin\theta \approx \sum_{r=0}^{n} c_{2r+1}\theta^{2r+1} \quad |\theta| \leq \tfrac{1}{2}\pi, \tag{16}$$

then

$$\mathbf{R}^{-\frac{1}{2}} \sin(\mathbf{R}^{\frac{1}{2}}t) \approx \sum_{r=0}^{n} c_{2r+1}\mathbf{R}^r t^{2r+1}. \tag{17}$$

For the range of θ in (16), the corresponding range of t in (17) is

$$\begin{aligned} |t| &\leq \tfrac{1}{4}\pi(\gamma/M)^{-\frac{1}{2}}, \\ &= \tfrac{1}{4}\pi \text{ units}, \end{aligned} \tag{18}$$

if we choose $(\gamma/M)^{-\frac{1}{2}}$ as the unit of time. The partitioned blocks of $\mathbf{A}(t)$ in (15) are now given to order n by

$$\mathbf{A}_1(t) = \sum_{r=0}^{n} c_{2r}\mathbf{R}^r t^{2r}, \tag{19a}$$

$$\mathbf{A}_2(t) = \sum_{r=0}^{n} c_{2r+1}\mathbf{R}^r t^{2r+1}, \tag{19b}$$

and

$$\mathbf{A}_3(t) = -\mathbf{R}\mathbf{A}_2(t), \tag{19c}$$

with

$$c_{2r} = (2r + 1)c_{2r+1}. \tag{20}$$

In order to evaluate the elements of the matrices in (19), it is necessary to obtain an expression for the elements of $\mathbf{R}^r$ for positive integers r. The matrix $\mathbf{R}$, which is given by (9), is a band matrix of order $N \times N$. It may be shown, by mathematical induction for example, that the elements of $\mathbf{R}^r$, for $r > 0$, are given by the expression

$$[\mathbf{R}^r]_{i,j} = (\gamma/M)^r \sum_{s=0}^{2r} (-)^{r+s} \times \left[\binom{2r}{s} - \binom{2r}{r-m} - \binom{2r}{r-m'}\right] \delta_{i+r,j+s}, \tag{21}$$

where $\delta_{i,j}$ is the Kronecker delta function,

$$m = i + j, \tag{22}$$

and

$$m' = 2(N + 1) - m. \tag{23}$$

It is also assumed that

$$\binom{2r}{k} = 0 \quad \text{for} \quad k < 0. \tag{24}$$

The elements of $\mathbf{A}(t)$ are evaluated from (19) and (21). However the structure of $\mathbf{A}(t)$ makes possible a considerable reduction in the number of elements requiring evaluation. It is obvious that the elements in the ith row of $\mathbf{A}_1(t)$, for example, are symmetric about the ith element in that row,i.e.,

$$[\mathbf{A}_1(t)]_{i,i+k} = [\mathbf{A}_1(t)]_{i,i-k},$$

providing the column subscripts are valid. In general, if N is sufficiently large, i may be chosen so that both $m > n$ and $m' > n$ are satisfied for every element in the ith row. This amounts to satisfying the condition

$$n \leq i \leq N + 1 - n.$$

With i in this range, the second and third terms in (21) are each zero for all $r(r = 1, 2, \cdots, n)$. The elements in the ith row of $\mathbf{A}_1(t)$ are then

$$[\mathbf{A}_1(t)]_{i,i+k} = (-)^k \sum_{r=0}^{n} \binom{2r}{r-k} c_{2r} z^{2r} \qquad (k = 0, 1, \cdots, n), \tag{25}$$

where

$$z = (\gamma/M)^{\frac{1}{4}} t. \tag{26}$$

The variable z is then a measure of the time in terms of the unit $(\gamma/M)^{-\frac{1}{4}}$.

When $i < n$ there are elements $[\mathbf{A}_1(t)]_{i,j}$ for which $m \leq n$. It is evident from (21) that for such elements a correction must be subtracted from the appropriate value derived from (25). The corrections, which are obtained from (19a) and (21), are

$$B_m = (-)^m \sum_{r=0}^{n} \binom{2r}{r-m} c_{2r} z^{2r},$$

where m is defined by (22). This term is just the element $[\mathbf{A}_1(t)]_{i,i+m}$ given by (25). Consideration of the cases for which $m' \leq n$ when $i > N + 1 - n$ shows that the correction terms are then $B_{m'}$.

In order to compute each of the N^2 elements of $\mathbf{A}_1(t)$ to the nth order, it is apparent that only the $n + 1$ quantities

$$B_k = (-)^k \sum_{r=0}^{n} \binom{2r}{r-k} c_{2r} z^{2r} \qquad (k = 0, 1, \cdots, n) \tag{27}$$

need be evaluated. The c_{2r} are the coefficients in any nth-order economized polynomial approximation to $\cos\theta$, and z is defined by (26). Since B_k is zero for $k > n$, every element of $\mathbf{A}_1(t)$ may now be expressed in the form

$$[\mathbf{A}_1(t)]_{i,j} = B_k - B_m - B_{m'}, \tag{28}$$

where

$$k = |i - j|,$$

and m and m' are given by (22) and (23), respectively. Similar expressions for the elements of $\mathbf{A}_2(t)$ and $\mathbf{A}_3(t)$ may be derived in the same way, using the coefficients c_{2r+1} in the polynomial approximation to $\sin\theta$ given by (16).

The coefficients c_{2r} and c_{2r+1} may be derived from the Chebyshev series expansions[8] for $\cos\frac{1}{2}\pi x$ and $\sin\frac{1}{2}\pi x$, respectively. If single precision arithmetic (8 decimal digits) is used for the computations on a computer, the numerical error is a minimum if only the first seven terms in the Chebyshev series are used. The initial time interval for the iteration procedure should be made as large as possible to increase the speed of the computation, yet should be small enough to avoid missing any collisions. It was found, experimentally, that suitable starting intervals are approximately $h_1 = 0.4$ units for $N < 10$, and $h_1 = 0.2$ units for $10 < N < 40$. The final extrapolation to determine the collision times is valid when $|\delta t| < 2(5)^{\frac{1}{2}} \times 10^{-3}$ units. If succeeding iteration intervals are bisected when a collision is indicated, it is apparent that only 8 matrices $\mathbf{A}(h_i)$ are required. The computer storage requirement for these matrices is less than 150 words, although there are over 4000 elements in each matrix when $N = 32$.

The mathematical analysis in this section emphasizes one of the main advantages of the model; that the computation may be reduced to a very simple form. Numerical computations have been carried out on IBM 1620 and IBM 7090 computers with assemblies for which N was 3, 4, 7, 8, 15, 16, 31, and 32. These values for N were chosen because Ford suggested that the nonlinear systems studied by FPU would not show energy sharing when $N + 1$ is a prime or a power of two. In such cases, the normal frequencies ω_i are linearly independent, and there is very little resonance amongst the ω_i.

The speed of the computation, measured in terms of the average time taken to compute a collision, is approximately proportional to N. For $N = 31$, the 7090 calculated 100 collisions per particle, and carried out all the additional computations required,

[8] See for example, C. W. Clenshaw, *Mathematical Tables*, (Her Majesty's Stationery Office, London, 1962), Vol. 5, p. 20.

in about 70 minutes. Ample means for checking the accuracy of the computer program are available since the total energy of any system is constant, the motion is reversible, and antisymmetry is conserved. For instance, the velocities $\dot{x}_i$ in a system of $N = 31$ particles which was started in the first mode were reversed after two hundred collisions had occurred. After a further two hundred collisions, the system returned to the initial configuration, in the sense that all the energy returned to the first mode to within an error of less than 0.001%.

In the numerical computations, realistic values were assigned to the parameters M, γ, l, a, and d of the system. The values so assigned are given in Table I.

IV. ENERGY SHARING

The total energy of the system is

$$E = \tfrac{1}{2}M(\mathbf{x}'\mathbf{R}\mathbf{x} + \dot{\mathbf{x}}'\dot{\mathbf{x}}) \tag{29}$$

or, in terms of the normal coordinates of the linear system,

$$E = \tfrac{1}{2}M \sum_{i=1}^{N} (\omega_i^2 q_i^2 + \dot{q}_i^2). \tag{30}$$

Since the system is conservative, E is a constant of the motion.

If the system were linear,

$$E_i = \tfrac{1}{2}M(\omega_i^2 q_i^2 + \dot{q}_i^2), \tag{31}$$

the energy associated with the ith normal mode, would also be a constant of the motion. However in the nonlinear system, E_i is constant only between collisions, having a finite jump discontinuity at each collision. For a collision at time t, the changes in these energies are

$$\begin{aligned}\Delta E_i(t) &= \tfrac{1}{2}M[\dot{q}_i(t+)^2 - \dot{q}_i(t-)^2] \\ &= \tfrac{1}{2}M\Delta\dot{q}_i(t)[2\dot{q}_i(t-) + \Delta\dot{q}_i(t)],\end{aligned} \tag{32}$$

where

$$\Delta\dot{q}_i(t) = \dot{q}_i(t+) - \dot{q}_i(t-). \tag{33}$$

From (7) and (11) it may be shown that, when the two particles $j - 1$ and j collide,

$$\Delta\dot{q}_i(t) = [\gamma(N+1)/2M]^{-\frac{1}{2}}\omega_i C_{ij}\,|\Delta\dot{x}_j|, \tag{34}$$

where ω_i is given by (12), and

$$C_{ij} = \cos[\tfrac{1}{2}i(2j-1)\pi/(N+1)]. \tag{35}$$

The matrix $\mathbf{C}$ is of order $N \times (N+1)$. Equation (34) shows that there is "coupling" between the modes whenever a collision occurs, while (32) gives the corresponding changes in the energies E_i.

A. Coupling between the Modes

The increment $\Delta\dot{q}_i(t)$ for a collision j at time t is the product of three factors: the magnitude of the impulse at the collision; the coupling term C_{ij}; and a modulating term which is directly proportional to the frequency ω_i. The impulsive force is determined by the velocities of the colliding particles, and therefore increases, on the average, as the total energy of the system increases. The absolute value of the coupling term is a maximum ($|C_{ij}| = 1$) when

$$i(2j-1) = 2m(N+1), \tag{36}$$

for some integer m. Of all the cases for which the condition (36) is approximately satisfied, there are two which will be of particular interest. If $N \gg 1$, i.e., $N^{-1} < 0.1$, say, $|C_{ij}| \approx 1$ for the first few modes ($i \approx 1$) when collisions occur between particles which are close to the left wall ($j \approx 1$), or the right wall ($j \approx N+1$). The coupling term is also large in magnitude for the modes $i = 2k$ and $i = N - 2(k-1)$, ($k \approx 1$), when collisions occur near the center of the system ($j \approx [\frac{1}{2}(N+1)]$). As the integer approximations become worse, the numerical value of $|C_{ij}|$ decreases. Although $|C_{ij}|$ can be large for the low modes, in general the increment $\Delta\dot{q}_i$ will be larger in magnitude for the high modes because of the modulating term, which is proportional to ω_i. The sign of $\Delta\dot{q}_i$, which is determined by C_{ij}, is positive if

$$4m - 1 < i(2j-1)/(N+1) < 4m + 1 \tag{37}$$

for some integer m, and negative otherwise.

B. Energy Sharing

The energy sharing is due entirely to the discontinuities in the kinetic energy associated with each mode when collisions occur. It is obvious that the increments ΔE_i in the energy of the ith mode are strongly dependent on the increments $\Delta\dot{q}_i$. Because the latter are proportional to the frequency ω_i, it follows that the rate of energy sharing for the high modes should be much greater than that for the low modes, if N is large.

Since $C_{ij}\Delta\dot{q}_i > 0$ for all i and j, the energy increment $\Delta E_i(t)$ is positive if

$$C_{ij}\dot{q}_i(t-) > 0. \tag{38}$$

When

$$C_{ij}\dot{q}_i(t-) < 0, \tag{39}$$

$\Delta E_i(t)$ is positive or negative according as

$$|\Delta\dot{q}_i(t)| \gtrless 2\,|\dot{q}_i(t-)|. \tag{40}$$

The potential energies of each mode are continuous functions of time throughout the motion. However, as the system behaves linearly during the time interval between collisions, there is energy sharing between the potential and kinetic energies of each mode. This form of energy sharing is periodic with period $T_i = 2\pi/\omega_i$ for the ith mode, and is therefore much more rapid for high modes than for low modes. It follows that the kinetic energies of the modes, and hence the discontinuities in these at each collision, are dependent on the frequencies ω_i. This dependence is negligible unless the time between any two collisions is of the order of $(\gamma/M)^{-\frac{1}{2}}$.

C. A Dominant Mode

Suppose that at some time during the motion the kth mode becomes dominant, in the sense that the energy E_k is a large fraction of the total energy E of the system. If the first mode is dominant, with $\dot{q}_1 > 0$ and $|\dot{q}_1| \gg |\dot{q}_i|$ $(i \neq 1)$, then the next collision will most probably occur near the right wall. When $j > \frac{1}{2}(N+1)$, $C_{1j} < 0$, and in this case $C_{1j}\dot{q}_1 < 0$. Similar considerations when the kth mode is dominant will show that the collisions j most likely to occur subsequently are those for which $C_{kj}\dot{q}_k < 0$. From (39) and (40) it is seen that this condition will usually lead to a decrease in E_k.

The above argument does not preclude the possibility that during the motion some mode may become dominant. Indeed, if the model is to exhibit ergodicity, this is to be expected. Such a condition could occur, for example, if the potential energy $\frac{1}{2}M\omega_k^2 q_k^2$ of the kth mode is much larger than the corresponding kinetic energy. At subsequent collisions there would be an increase in the kinetic energy and in the total energy E_k of the mode if either $|\Delta\dot{q}_k(t)| > 2\,|\dot{q}_k(t-)|$, or $C_{kj}\dot{q}_k(t-) > 0$. However, if E_k is a large fraction of E after any collision, there is a greater probability that at the next collision ΔE_k will be negative, rather than positive.

D. Numerical Computations

The above analysis does not yield an exact expression for the amount of energy sharing to be expected over any time interval. However it does indicate qualitatively that energy is shared between all modes, that the rate of energy sharing is greater for the higher modes, and that the probability of any mode becoming, or remaining, dominant is small. For quantitative results we turn to the computer solutions of the motion for different numbers of particles.

For any given N, the initial configuration of the system may be taken to be

$$E_i = \delta_{i,k}E,$$

say, where k is odd if the special cases of antisymmetry are to be avoided. This corresponds to the kth normal mode in the linear system. Since collisions must occur if the model is to be nonlinear, there is a lower bound on the total energy E. In the numerical computations we always chose $k = 1$, or $k = N$ when N was odd. We also investigated weakly coupled and strongly coupled systems, the strength of the coupling being dependent on the total energy of the system. As the energy is increased, so too is the collision rate, and the average magnitude of the impulse at each collision. The initial behavior of any system is strongly dependent on the starting configuration, the number of particles, and the mean energy per particle.

1. Case $k = 1$

When a system is started in the first mode, there is a gradual loss of energy from that mode with an increase in the energies of all other modes. The energy loss from the first mode is explained by the fact that most of the collisions occur initially in the neighborhood of one of the walls [the right wall if $\dot{q}_1(0) > 0$], and, as we noted above, the condition $C_{1j}\dot{q}_1 < 0$ is usually satisfied. It was also shown that for such collisions $|C_{1j}| \approx 1$, if $N \gg 1$. The loss of energy is then a maximum, even though the actual loss is small due to the modulating term. The second and third modes increase more rapidly than the others, initially. This is because $C_{2j} > 0$ if $\frac{1}{4}N > j$, or $j > \frac{3}{4}N$, and there is an almost continuous stepwise increase in E_2 as $C_{2j}\dot{q}_2 > 0$ at almost every collision. Although the coupling at each collision is often stronger for the third mode, E_2 increases more rapidly than E_3 as $C_{3j}\dot{q}_3$ is often negative. As the increments E_i for the higher modes are negative almost as often as they are positive, there is no appreciable increase in the energies E_i initially. Eventually, while the first mode is still dominant, the sign of $\dot{q}_1$ is reversed. While the magnitude of $\dot{q}_1$ is small, the first inequality in (40) is satisfied for some collisions, and there are small positive increments in E_1. There is then a period without collisions during which the entire system of particles "swings" in the opposite direction to the original motion. Collisions then occur predominantly in the neighborhood of the other wall, leading to further losses of energy from the first mode. After some time, the configuration of the system bears little resemblance to the initial

TABLE I. The numerical values assigned to the parameters in the computation. The particles each have mass M, effective diameter d, and a mean separation of $l - d$. Adjacent particles are connected by springs which have spring constant γ and natural length $a - d$.

Parameter	Numerical Value
M	3.0×10^{-23} g
γ	400 dyne/cm
l	4.000×10^{-8} cm
a	3.995×10^{-8} cm
d	3.400×10^{-8} cm

state, and the modes begin to exchange energy freely in a more random manner. Frequently one mode is found to have much more energy than any of the others. There is, however, no evidence of the periodic behavior found in the systems studied by FPU and others. The only significant difference in the energy sharing properties of weakly and strongly coupled systems started in the first mode appeared to be the rate at which energy was shared among all the modes.

For the values of the parameters used in the computations (see Table I), the potential energy of interaction between two particles about to collide is $\phi(d+) = 0.708 \times 10^{-14}$ ergs, and $\phi_0 = -0.5 \times 10^{-18}$ ergs. If we denote the mean energy per particle (actually the mean energy per degree of freedom) by

$$\epsilon = E/N,$$

then the energies of the systems studied were given by $\epsilon' = 0.4 \times 10^{-14}$ ergs, and $\epsilon'' = 0.7 \times 10^{-14}$ ergs. These energy levels are illustrated in Fig. 1. It is evident that systems with energies $N\epsilon'$ will be weakly nonlinear, while those with energies $N\epsilon''$ will be much more strongly nonlinear.

In Fig. 2 we plot E_1 and E_2 as a percentage of

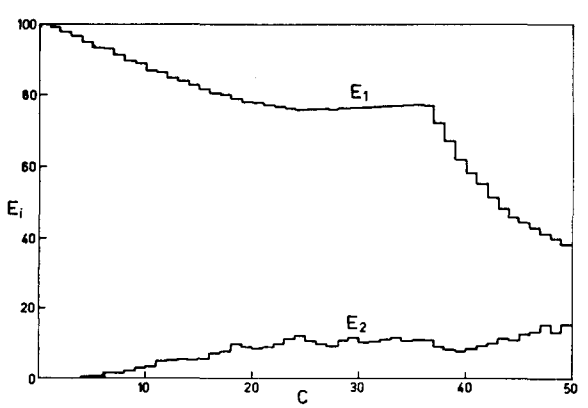

FIG. 2. The energy in the first two modes after C collisions for $N = 15$ particles, mean energy/particle $\epsilon = \epsilon''$, and $E_i = \delta_{ik}E$ at $t = 0$, with $k = 1$.

the total energy E for the system $N = 15$, which was started in the first mode with $\epsilon = \epsilon''$. The "curve" is a histogram because of the constancy of the energies between collisions. E_1 decreased initially until, after about 20 collisions, $|\dot{q}_1|$ was small. The increments ΔE_1 were then negligible, and for several collisions E_1 remained almost constant. The explanation for this behavior is that $T_1 = 2\pi/\omega_1$ is long compared with the average time between collisions, so that $\dot{q}_1$ does not alter appreciably from one collision to the next. The whole system of particles then swung in the opposite direction until the 37th collision occurred on the left wall. E_1 then decreased more rapidly again. The partition of the total energy among all the modes at intervals up to 1500 collisions is shown in Table II. Once the first mode has ceased to be dominant, it is apparent that energy is freely exchanged between all the modes. This is further illustrated by computation of the time averages $\langle E_i \rangle$ of the energies, given by

$$\langle E_i \rangle = t^{-1} \int_0^t E_i(s)\, ds. \tag{41}$$

Numerical values for these averages are given in Table III. The approach to equipartition of the energy among the modes, in the time average, is strongly evident.

In a similar system with the same starting configuration except that $\epsilon = \epsilon'$, the decay of energy from the first mode is much slower. The collision rate in this system was not much more than 25% of the rate in the more strongly nonlinear system. Furthermore, the impulsive force at each collision was less than in the first case. However an approach to equipartition was still evident.

2. Case $k = N$, N odd

In systems for which the initial configuration is $E_N = E$, there is a very rapid loss of energy from the Nth mode. Most of the collisions occur near the center of the system initially, and the first few have indices j which are even if $\dot{q}_N > 0$. This explains the loss of energy from the Nth mode, since the condition $C_{Nj}\dot{q}_N < 0$ is satisfied when such collisions occur. The energy loss is large because at each collision $|C_{Nj}| \approx 1$, and the modulating term is a maximum. Furthermore, the increments $|\Delta\dot{q}_N|$ are so large that $\dot{q}_N$ quickly becomes negative. Any further collisions for which j is even will lead to small increases in E_N. However when $\dot{q}_N < 0$, collisions for which j is odd soon begin to occur, leading to further decreases in E_N. After a short

TABLE II. The energies E_i, as percentages of E, after C collisions and time z units for the case $N = 15$, $\epsilon = \epsilon''$, and $k = 1$. $[z = (\gamma/M)^{\frac{1}{2}}t]$

C	z	$i=1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.6	95.2	0.4	0.6	0.4	0.2	0.0	0.0	0.0	0.0	0.1	0.4	0.8	1.0	0.7	0.2
10	0.8	88.8	2.1	1.8	0.4	0.1	0.6	0.4	0.0	0.6	1.8	2.0	1.1	0.3	0.0	0.0
15	1.2	82.9	4.9	2.4	0.3	0.5	0.1	0.7	3.1	3.2	0.8	0.0	0.5	0.4	0.1	0.0
30	2.9	76.4	11.3	2.0	2.3	0.6	1.5	0.9	0.1	0.7	1.9	0.8	0.2	0.9	0.3	0.3
45	13.2	46.4	10.9	1.1	1.6	5.3	4.7	3.5	2.4	0.6	0.7	1.4	0.7	1.8	13.4	5.6
60	15.7	32.5	12.0	0.9	1.0	0.7	4.2	6.1	2.8	0.4	0.1	11.0	14.8	2.6	5.6	5.2
75	21.0	33.3	9.2	3.5	4.0	0.1	2.2	1.9	3.6	2.8	8.1	3.7	8.7	1.9	4.6	12.5
90	28.4	17.2	5.8	6.6	1.8	5.5	3.5	2.6	2.3	6.0	13.4	14.1	9.9	9.4	0.4	1.6
105	32.7	21.2	15.4	0.5	3.8	1.6	1.5	14.5	1.6	2.1	1.1	4.0	2.1	2.2	19.4	9.2
120	39.2	9.9	13.8	1.2	3.1	8.1	8.6	10.5	9.0	0.6	3.7	8.0	0.1	12.9	7.8	3.0
135	47.2	11.9	6.0	5.0	4.0	5.7	1.6	2.6	4.3	3.7	2.5	10.0	8.9	0.4	19.1	14.3
150	57.6	12.0	3.7	2.5	1.0	5.5	12.7	14.5	8.5	4.1	4.5	2.1	12.5	4.8	2.9	8.6
300	123.4	6.2	1.0	28.5	15.1	0.3	7.2	3.8	1.8	3.2	23.1	0.4	0.1	2.3	6.5	0.6
450	196.3	7.3	11.4	7.6	40.6	4.2	0.7	4.0	6.6	1.0	2.0	4.1	1.9	4.9	0.1	3.6
600	262.5	4.3	1.7	24.4	2.2	23.4	20.1	3.2	2.6	0.5	4.5	0.8	2.4	0.5	6.1	3.3
750	326.2	6.3	10.5	5.6	4.5	0.5	0.7	8.7	1.5	5.7	29.3	7.2	2.9	6.1	6.4	4.1
900	395.8	5.7	6.0	1.3	5.3	3.1	9.3	16.8	11.3	4.6	4.3	4.7	2.7	2.1	13.2	9.6
1050	470.0	0.9	3.5	9.2	3.3	6.8	3.3	7.8	6.1	5.9	7.4	1.0	0.5	7.4	32.2	4.8
1200	528.9	11.2	12.9	1.1	21.2	10.3	10.4	3.0	0.3	4.9	0.6	0.7	11.6	1.9	4.9	5.0
1350	593.9	7.1	1.5	0.1	1.2	0.5	1.1	9.8	5.2	26.8	9.0	13.3	10.0	2.9	7.3	4.2
1500	668.3	11.5	5.8	3.4	6.1	8.1	2.4	1.5	0.8	5.9	34.3	11.6	1.6	1.1	1.4	4.5

time, E_N is almost negligible. The energies E_{N-2}, E_{N-4}, $\cdots$ increase rapidly initially. It has been shown that $|C_{ij}| \approx 1$ for $i = N-2, N-4, \cdots$, and $i = 2, 4, \cdots$, when collisions occur in the neighborhood of $j=[\frac{1}{2}(N+1)]$. When the modulating term is taken into account, it is obvious that rapid increases in E_{N-2}, E_{N-4}, $\cdots$ are to be expected. An interesting feature is that E_2 increases slowly but fairly steadily. Indeed, for large N, it is not long before $\langle E_2 \rangle$ becomes greater than $\langle E_N \rangle$. The coupling term C_{2j} is negative for a large number of collisions initially. While $\dot{q}_2$ remains negative, the increment ΔE_2 at each collision, although small, is always positive, leading to a slow but steady increase in E_2.

In Fig. 3 we plot E_{13} and E_{15} for the first 50 collisions in a system which was started in the Nth mode with $N = 15$ and $\epsilon = \epsilon''$. The rapid exchange of energy between these two modes is clearly shown. Values for the energies E_i and the time averages $\langle E_i \rangle$ for all the modes are given in Tables IV and V, respectively. The approach of the time averages to equipartition is again evident. Comparison with Fig. 2 and Tables II and III shows that this approach is more rapid for the system started in the Nth mode, because of the more rapid exchange of energy initially.

The general behavior of every system studied was very similar, once the initial conditions had ceased to be dominant. In particular, a slight variation in the number of particles from $N = 31$ to $N = 32$, say, did not yield essentially different results, provided antisymmetric systems were excluded from consideration. That is, the nonlinear systems studied here do not show less energy sharing when $N + 1$ is a prime or a power of two, which was a reason given by Ford for the lack of energy sharing in the systems studied by FPU. In our systems, if all the energy was initially in one normal mode, it became freely distributed among all the modes, provided there was sufficient energy for at least one collision to occur.

V. THE STATISTICAL MECHANICS OF THE MODEL

In statistical mechanics, an isobaric grand partition function[9] of an assembly of N particles confined to a volume V may be defined by

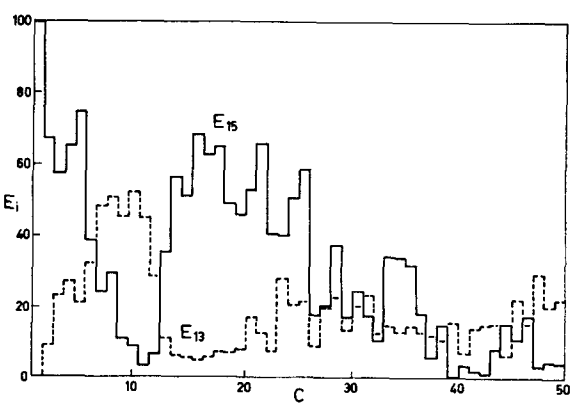

FIG. 3. The energy in the 13th and 15th modes after C collisions for the case $N = 15$, $\epsilon = \epsilon''$, $k = 15$.

[9] L. van Hove, Physica **16**, 137 (1950).

TABLE III. The percentage time averages $\langle E_i \rangle$ after C collisions and time z for $N = 15$, $\epsilon = \epsilon''$, and $k = 1$.

C	z	$i = 1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.6	99.0	0.1	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.8	96.8	0.5	0.6	0.5	0.3	0.2	0.1	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1
15	1.2	93.1	1.7	1.5	0.7	0.3	0.3	0.4	0.6	0.5	0.3	0.2	0.1	0.2	0.2	0.1
30	2.9	83.9	6.5	2.1	1.0	0.8	0.9	0.8	0.8	0.6	0.5	0.3	0.3	0.4	0.4	0.5
45	13.2	76.1	9.6	2.4	1.3	1.2	1.0	2.2	1.3	1.0	0.5	0.9	0.3	0.7	0.6	0.7
60	15.7	69.7	10.2	2.2	1.3	1.3	1.3	3.4	1.4	1.2	0.6	1.6	1.0	1.0	1.8	2.0
75	21.0	60.6	9.7	2.2	1.7	1.1	1.3	3.4	2.8	2.7	1.4	1.6	2.8	1.4	2.4	4.7
90	28.4	52.4	9.2	3.0	1.9	1.4	1.6	2.8	3.1	3.0	3.3	2.4	4.4	2.3	2.6	6.6
105	32.7	47.9	9.8	3.1	2.5	1.5	2.3	3.7	3.0	3.3	3.2	2.6	4.7	2.9	3.5	6.1
120	39.2	43.4	11.2	2.8	2.6	1.7	2.5	4.3	3.2	3.4	3.1	2.8	4.6	3.0	3.8	7.5
135	47.2	37.9	10.8	3.1	2.8	2.2	2.7	4.1	4.9	3.3	3.6	3.3	5.9	3.0	4.7	7.8
150	57.6	33.4	10.3	2.8	2.6	4.0	3.2	4.6	4.5	3.0	4.6	3.0	6.5	3.1	4.7	9.8
300	123.4	19.6	7.8	5.9	4.5	5.5	6.2	6.6	5.5	4.1	5.6	5.3	5.7	4.5	4.6	8.6
450	196.3	13.5	8.7	7.4	6.1	5.6	6.2	7.4	5.4	5.2	6.9	5.2	5.9	5.2	4.1	7.1
600	262.5	11.3	9.2	8.8	6.7	5.6	6.1	6.6	6.0	5.3	6.5	5.4	5.8	5.2	4.9	6.7
750	326.2	9.9	9.1	9.2	6.7	5.7	6.2	6.8	6.0	5.7	7.1	5.4	5.8	5.4	4.8	6.3
900	395.8	8.7	9.5	8.6	6.9	5.7	6.1	6.7	5.9	6.1	6.8	5.9	5.6	6.1	4.8	6.7
1050	470.0	7.9	9.1	8.2	6.4	6.3	6.0	6.5	5.9	6.1	6.8	5.9	5.6	6.5	5.3	7.3
1200	528.9	7.5	8.5	8.2	6.5	6.6	6.2	6.3	5.9	6.0	6.6	6.1	5.8	6.7	5.4	7.7
1350	593.9	7.3	7.9	8.1	6.8	7.0	6.2	6.3	5.8	6.4	6.4	5.9	5.9	6.7	5.7	7.5
1500	668.3	7.2	7.5	7.8	6.6	6.6	6.3	6.5	6.3	6.5	6.6	6.4	6.1	6.7	5.6	7.3

$$Z_i(T, N, P) = \sum_{(V)} Z(T, N, V) \exp(-PV/kT), \quad (42)$$

where $Z(T, N, V)$ is the ordinary partition function. T is the absolute temperature of the assembly, P is the total pressure, and k is Boltzmann's constant. It may be shown that

$$E = kT[T(\partial/\partial T)(\ln Z_i) + P(\partial/\partial P)(\ln Z_i)]. \quad (43)$$

This equation expresses the total energy of the assembly as a function of the temperature and the pressure if the number of particles in the assembly is fixed. Also, if the volume V is given, then P is derived implicitly from

$$V = -kT(\partial/\partial P) \ln Z_i(T, N, P). \quad (44)$$

On replacing the variable V by the variable L of our assembly, we obtain, in terms of the configurational partition function Z_C,

$$Z_i(T, N, P) = \left(\frac{MkT}{2\pi\hbar^2}\right)^{\frac{1}{2}N} \times \int_0^\infty \exp(-PL/kT) Z_C(T, N, L)\, dL = \left(\frac{MkT}{2\pi\hbar^2}\right)^{\frac{1}{2}N} \bar{Z}_C(T, N, P/kT). \quad (45)$$

$\bar{Z}_C$ is the Laplace transform of Z_C, and $\hbar$ is Planck's constant. It should be noted that there are N degrees of freedom in the assembly. The configurational partition function of the model is given by

$$Z_C = \int \cdots \int \exp\left[-\sum_{i=1}^{N+1} \phi(y_i - y_{i-1})/kT\right] d\mathbf{y},$$

integrated over $0 = y_0 < y_1 < \cdots < y_{N+1} = L$, where we have written $y_i = jl + x_i$.

Repeated use of the convolution theorem gives

$$\bar{Z}_C(T, N, P/kT) = (\mathfrak{L}\{\exp[-\phi(r)/kT]\})^{N+1},$$

where $\mathfrak{L}$ is the Laplace transform operator

$$\mathfrak{L} = \int_0^\infty \exp(-Pr/kT)\, dr.$$

Since $\exp[-\phi(r)/kT] = 0$ for $r \leq d$,

$$\mathfrak{L}\{\exp[-\phi(r)/kT]\} = \exp(-Pd/kT)\mathfrak{L}\{\exp[-\phi(r + d)/kT]\},$$

which reduces to a standard Laplace transform when the functional form of ϕ given by (3) is introduced. The final result is

$$\bar{Z}_C = ((kT/\gamma)^{\frac{1}{2}} F(y) \times \exp\{-[Pd + \phi_0 + \tfrac{1}{2}\gamma(a - d)^2]/kT\})^{N+1}, \quad (46)$$

where

$$y = [P - \gamma(a - d)](\gamma kT)^{-\frac{1}{2}}, \quad (47)$$

and

$$F(y) = \exp(\tfrac{1}{2}y^2) \int_y^\infty \exp(-\tfrac{1}{2}s^2)\, ds \quad (48)$$

is an error function which has been tabulated.[10]

An equation of state for the model is derived

[10] See, for example, *The British Association for the Advancement of Science, Mathematical Tables*, (University Press, Cambridge, England, 1939), Vol. 7.

TABLE IV. The percentage energies E_i; $N = 15$, $\epsilon = \epsilon''$, $k = 15$.

C	z	$i=1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
5	1.0	0.0	1.9	0.0	0.7	0.0	0.3	0.0	0.3	0.7	0.1	0.4	0.1	21.0	0.1	74.6
10	3.4	0.0	9.9	0.1	6.2	0.3	0.6	1.3	0.4	5.9	0.1	12.3	0.5	52.5	0.8	9.1
15	8.7	0.0	10.6	0.1	0.8	0.1	2.8	0.1	1.9	4.9	0.1	19.8	0.6	5.9	1.2	51.0
30	13.7	0.1	7.1	1.4	2.7	1.6	1.2	1.8	8.2	0.5	20.1	16.8	7.5	13.3	0.3	17.3
45	27.4	0.5	9.7	3.3	6.0	8.1	7.8	0.0	4.4	3.5	23.0	5.1	0.1	6.7	6.0	15.8
60	35.2	3.8	7.2	1.5	0.4	0.3	2.7	0.1	2.8	10.2	11.9	16.4	5.3	31.4	2.7	3.2
75	43.2	4.3	9.5	6.3	11.7	7.1	9.9	5.3	6.2	9.7	2.0	0.2	8.2	6.8	6.5	6.3
90	50.6	0.3	16.3	22.6	6.8	11.5	0.4	3.1	0.5	11.8	3.4	0.5	1.2	0.4	17.5	3.7
105	58.1	0.6	4.7	14.4	9.6	3.7	8.0	3.2	1.8	1.7	0.6	14.7	13.3	10.3	2.3	11.2
120	64.2	3.8	0.4	15.6	0.7	13.0	5.8	10.6	8.9	0.6	3.4	1.4	9.2	11.7	12.2	2.6
135	70.1	3.9	0.1	3.3	1.0	14.6	2.4	1.1	6.4	16.3	7.2	1.5	12.4	1.6	19.7	8.5
150	77.2	2.4	5.0	3.9	3.9	19.0	7.6	5.7	4.1	1.0	13.5	10.5	5.8	2.2	13.6	1.8
300	151.7	1.4	1.2	0.1	9.0	0.8	11.9	4.7	25.1	2.8	5.6	4.3	2.8	2.5	17.3	10.7
450	228.4	0.4	24.8	7.8	4.0	5.4	2.9	14.5	5.4	11.6	6.3	1.9	3.5	7.0	1.3	3.0
600	296.2	5.1	15.3	3.8	3.0	6.8	2.8	17.0	4.5	6.0	0.5	6.3	4.7	17.5	2.9	3.8
750	353.3	8.5	6.9	3.7	4.7	1.2	2.5	4.1	13.9	4.1	3.9	9.7	1.9	14.4	2.7	17.8
900	415.7	21.9	2.4	2.7	1.8	5.8	8.7	2.4	5.8	10.3	11.0	0.7	9.7	3.0	11.3	2.5
1050	499.6	6.9	0.1	8.5	1.5	8.5	4.2	17.8	9.4	20.9	0.5	1.0	0.6	8.8	2.2	9.2
1200	553.8	0.8	7.3	5.2	15.0	15.2	7.9	5.9	4.4	12.8	4.6	2.0	1.2	0.3	2.1	15.4
1350	611.9	24.7	0.2	2.2	4.5	1.0	2.2	1.3	21.5	10.6	9.6	0.2	1.1	0.6	12.8	7.5
1500	677.6	7.9	22.5	3.8	3.7	10.8	10.5	2.1	0.1	0.3	2.1	10.8	2.5	0.2	3.0	19.7

from (44), (45), and (46) in the form

$$P + \gamma(l - a) = (\gamma kT)^{\frac{1}{2}}/F(y). \tag{49}$$

An equation similar to this has been derived by Koppell.[11] We have shown elsewhere[12] that the evaluation of the isobaric grand partition function Z_i above is much simpler than the difficult saddle-point method which Koppell used to evaluate Z_C. Our equation clearly shows that the difference between the pressure of the nonlinear system and the pressure due to the tension in the springs is given by a correction term, which is a function of the temperature, and the function $F(y)$ defined by (47) and (48). Nagamiya[13] has considered a model in which the mutual interaction energy between particles was taken to be

$$\phi(r) = \phi_0 + \tfrac{1}{2}\gamma(r - a)^2 \qquad -\infty < r < \infty.$$

Here, no cutoff was assumed to exist at $r = d$. The model was therefore linear, and the equation

TABLE V. The percentage time averages $\langle E_i \rangle$; $N = 15$, $\epsilon = \epsilon''$, $k = 15$.

C	z	$i=1$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
5	1.0	0.0	1.5	0.0	0.8	0.1	0.3	0.2	0.4	0.7	0.2	1.0	0.3	18.8	0.3	75.4
10	3.4	0.0	6.1	0.1	3.8	0.2	0.4	0.8	0.6	3.5	0.3	7.0	0.5	40.8	0.7	35.3
15	8.7	0.0	9.3	0.1	3.0	0.4	1.4	1.8	1.3	7.1	0.7	14.5	0.7	30.1	0.7	28.8
30	13.7	0.0	7.5	0.1	2.8	0.4	1.2	2.0	1.6	5.4	1.4	12.6	0.8	23.5	1.6	39.1
45	27.4	0.1	11.0	0.6	3.2	1.7	2.7	3.3	2.9	7.0	3.8	10.7	2.5	18.9	4.6	26.9
60	35.2	0.5	9.8	1.4	2.7	1.4	2.4	3.7	3.6	7.3	6.1	11.1	3.7	18.6	4.7	22.8
75	43.2	1.1	9.2	2.0	2.9	2.1	2.8	4.7	4.0	7.6	6.7	11.4	3.9	16.3	5.1	20.3
90	50.6	1.2	10.0	3.6	4.3	3.1	2.9	5.3	3.8	7.5	6.1	10.2	4.3	14.4	5.1	18.1
105	58.1	1.1	10.3	7.0	4.5	2.8	3.7	5.1	3.7	7.4	6.1	9.6	4.4	13.0	5.1	16.1
120	64.2	1.2	9.6	7.8	4.6	2.9	3.6	5.4	3.8	7.4	5.9	9.3	5.1	12.8	5.8	14.9
135	70.1	1.4	8.9	8.0	4.3	3.8	3.6	5.5	4.1	7.3	5.7	8.6	5.4	12.3	6.7	14.4
150	77.2	1.5	8.2	7.5	4.1	4.7	3.7	5.5	4.4	7.6	5.8	8.8	5.8	11.4	7.1	13.9
300	151.7	2.8	7.2	6.5	3.9	5.9	6.4	6.2	5.4	6.9	6.7	7.4	5.6	8.8	9.0	11.4
450	228.4	2.9	6.0	5.7	5.3	6.6	5.9	6.4	5.4	7.3	6.5	7.4	7.2	9.1	8.5	9.8
600	296.2	4.5	7.5	5.4	5.4	7.8	6.2	6.3	5.6	6.8	6.2	7.2	6.8	8.1	7.6	8.6
750	353.3	5.8	7.8	5.4	5.4	7.5	6.1	5.9	5.5	6.8	6.3	6.7	6.7	8.1	7.6	8.5
900	415.7	7.0	7.4	5.2	6.0	7.0	6.6	6.0	5.6	7.0	6.1	6.9	6.3	7.9	7.2	7.9
1050	488.6	8.0	7.0	5.1	6.0	7.2	6.8	6.0	5.6	6.5	6.2	6.7	6.0	7.9	7.2	7.8
1200	553.8	7.9	6.8	5.2	6.0	7.0	6.8	6.4	5.9	6.5	6.1	6.5	6.2	7.8	7.3	7.5
1350	611.9	8.2	6.7	5.1	6.2	6.9	6.9	6.6	5.8	6.5	6.1	6.3	6.0	7.9	7.2	7.5
1500	677.6	8.2	7.0	5.3	6.0	6.7	6.8	6.5	5.8	6.6	6.2	6.4	6.0	7.7	7.0	7.7

[11] D. Koppell, Phys. Fluids **6**, 609 (1963).
[12] R. Northcote and R. Potts, Phys. Fluids (to be published).
[13] T. Nagamiya, Proc. Phys. Math. Soc. Japan **22**, 705 (1940).

of state derived was

$$P + \gamma(l - a) = 0. \tag{50}$$

This is just our equation of state (49) with the correction term replaced by zero. The correction term, then, represents the difference between the linear and the nonlinear cases. The result in the linear case is, of course, formulated in Hooke's Law.

For further comparison, it is possible to recast (49) in the form

$$\gamma(l - d)(\gamma kT)^{-\frac{1}{2}} = -y + F(y)^{-1}. \tag{51}$$

One could assume that there is a minimum distance between molecular centers, say d, in Nagamiya's model. By adding an appropriate term in each side of (50) and effecting some minor rearrangement we have

$$\gamma(l - d)(\gamma kT)^{-\frac{1}{2}} = -y. \tag{52}$$

Equations (51) and (52) are represented graphically in Fig. 4.

Combining together (43), (45), and (46), the total energy of the assembly may be obtained in the form

$$E = (N + \tfrac{1}{2})kT + \tfrac{1}{2}(N + 1)\gamma^{-1}\{P^2 - \gamma^2(l - a)^2 - (\gamma kT)^{\frac{1}{2}}F(y)^{-1}[P + \gamma(a - d)]\}, \tag{53}$$

where we have used $\phi_0 = -\frac{1}{2}\gamma(l - a)^2$. On using (49) to eliminate $F(y)$, the mean energy per degree of freedom $\epsilon = E/N$, where N is assumed to be infinite, is found to be given by

$$\epsilon = kT - \tfrac{1}{2}(l - d)[P + \gamma(l - a)]. \tag{54}$$

For given values of the parameters γ, l, a, and d, (47), (49), and (54) is a set of three independent equations in the four variables ϵ, P, T, and y. Any one of these variables may therefore be expressed as a function of any one of the others. Theoretical values for the expected temperature and pressure of any assembly may therefore be computed. Sketches of the curves $T = T(\epsilon)$ and $P = P(\epsilon)$, for the numerical values of the parameters given in Table I, have been drawn in Figs. 5 and 6, respectively.

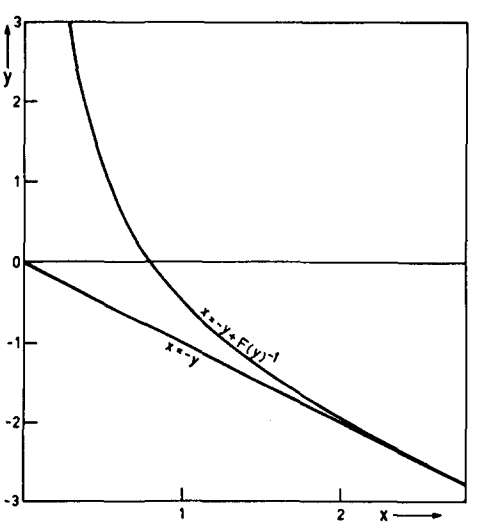

FIG. 4. Curves for the equation of state, in dimensionless form, for both the linear and nonlinear systems. $x = \gamma(l - d)(\gamma kT)^{-\frac{1}{2}}$, and $y = [P - \gamma(a - d)](\gamma kT)^{-\frac{1}{2}}$.

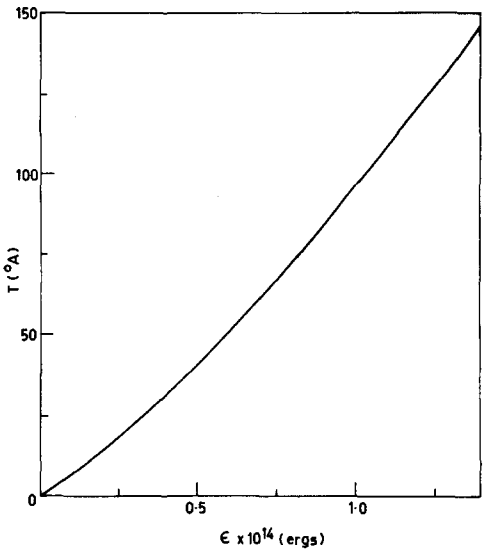

FIG. 5. The thermodynamic temperature T as a function of the mean energy/particle ϵ, for the specific values of the parameters given in Table I.

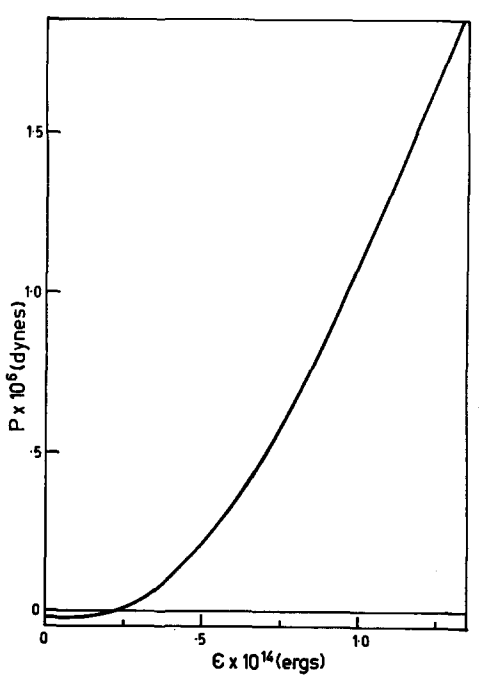

FIG. 6. The thermodynamic pressure P as a function of ϵ, for the values of the parameters given in Table I.

Fluctuations in the temperature and the pressure are expected in an isolated assembly. The relative fluctuations in the temperature T are given by

$$\sigma_T^2 = \overline{(T - \bar{T})^2}/\bar{T}^2. \tag{55}$$

However, because we are using ideal walls for the calculation of the pressure, we cannot consider fluctuations in the actual pressure. This is because any estimate of the pressure requires a finite time for computation. We therefore consider fluctuations in the time average of the pressure,[14] defined by

$$\sigma_{\langle P\rangle}^2 = \overline{(\langle P\rangle - \overline{\langle P\rangle})^2}/\overline{\langle P\rangle}^2. \tag{56}$$

Computation of these fluctuations will yield further evidence on whether or not the assembly of particles under consideration attains a state of equilibrium with respect to temperature and pressure.

VI. COMPUTATION OF THE THERMODYNAMIC VARIABLES

In the preceding section we developed the statistical mechanics of the model, i.e., we found expressions for the thermodynamic variables when N was assumed to be infinite. We proceed now to find mathematical expressions for the corresponding variables of the finite model.

A. The Mean Temperature

Since there are N degrees of freedom in the model, it follows from the equipartition principle that

$$\langle \mathrm{KE}\rangle = \tfrac{1}{2}Nk\langle T\rangle,$$

where $\langle \mathrm{KE}\rangle$ denotes the time average of the kinetic energy of the system. The time average of the temperature is therefore given by

$$\langle T\rangle = 2\langle \mathrm{KE}\rangle/(Nk). \tag{57}$$

A measure of the instantaneous temperature is taken as $T = 2 \times \mathrm{KE}/(Nk)$.

The time average of the kinetic energy over any time interval $(0, \tau)$ is

$$\langle \mathrm{KE}\rangle = \tfrac{1}{2}M\tau^{-1}\int_0^\tau \dot{\mathbf{x}}'\dot{\mathbf{x}}\,dt. \tag{58}$$

If there are no collisions in this interval then, on integrating by parts and rearranging, we have

$$\tfrac{1}{2}M\int_0^\tau (\dot{\mathbf{x}}'\dot{\mathbf{x}} - \mathbf{x}'\mathbf{R}\mathbf{x})\,dt = \tfrac{1}{2}M[\mathbf{x}'\dot{\mathbf{x}}]_0^\tau.$$

But from (29), since the system is conservative,

$$\tfrac{1}{2}M\int_0^\tau (\dot{\mathbf{x}}'\dot{\mathbf{x}} + \mathbf{x}'\mathbf{R}\mathbf{x})\,dt = \tau E.$$

Addition of the last two equations yields the time average of the kinetic energy over the interval $(0, \tau)$, if no collisions occur:

$$\langle \mathrm{KE}\rangle = \tfrac{1}{2}E + \tfrac{1}{4}M\tau^{-1}[\mathbf{x}'\dot{\mathbf{x}}]_0^\tau. \tag{59}$$

The usual result for the linear system is obtained in the limit as $\tau \to \infty$.

However, if there are collisions during the interval $(0, \tau)$, there is a correction term C in (59) so that

$$\langle \mathrm{KE}\rangle = \tfrac{1}{2}E + \tfrac{1}{4}M\tau^{-1}[\mathbf{x}'\dot{\mathbf{x}}]_0^\tau + C, \tag{60}$$

where

$$C = -\tfrac{1}{4}M\tau^{-1}\sum_i [\mathbf{x}'\dot{\mathbf{x}}]_{t_i-}^{t_i+},$$

and t_i $(i = 1, 2, \cdots)$ are the times at which the collisions occur. This correction term for the nonlinear case in which there are collisions is

$$C = \tfrac{1}{4}M\tau^{-1}(l - d)\sum_i |\Delta\dot{x}_j(t_i)|, \tag{61}$$

where the summation is over all collisions between particles $j - 1$ and j at times t_i during the time interval. The time average $\langle T\rangle$, which we shall call the mean temperature, is computed form (57), (60), and (61).

B. The Mean Pressure

Between collisions, the forces exerted by the constraints holding particles 0 and $N + 1$ fixed are

$$f_0 = -\gamma(x_1 + l - a),$$

and

$$f_{N+1} = \gamma(-x_N + l - a),$$

respectively. These forces are the instantaneous pressures acting on the left and right walls, respectively, provided there are no collisions on the walls. Thus, in the linear system, the time average of the mean pressure acting on the two walls taken over an interval $(0, \tau)$ is

$$\langle P\rangle = -\gamma(l - a) - \tfrac{1}{2}\gamma\tau^{-1}\int_0^\tau (x_1 - x_N)\,dt. \tag{62}$$

From (8) it follows that

$$\langle P\rangle = -\gamma(l - a) + \tfrac{1}{2}\gamma\tau^{-1}(\boldsymbol{\alpha}_1' - \boldsymbol{\alpha}_N')[\dot{\mathbf{x}}]_0^\tau,$$

where $\boldsymbol{\alpha}_1'$ and $\boldsymbol{\alpha}_N'$ are the first and last rows, respectively, of the matrix $\mathbf{R}^{-1}$. In the limit as $\tau \to \infty$ we obtain the mean pressure of the linear system

$$\langle P\rangle = -\gamma(l - a), \tag{63}$$

which is the result predicted in (50).

When there are collisions, however, there are additional contributions to the pressure. These are due to impulses on the walls when end collisions

[14] See for example, D. ter Haar, *Elements of Statistical Mechanics*, (Rinehart and Company, New York, 1954), p. 113.

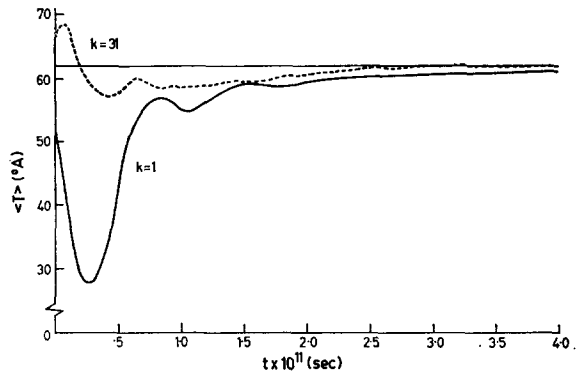

FIG. 7. The mean temperature computed for the systems $N = 31$, $\epsilon = \epsilon''$, and $E_i = \delta_{ik}E$ at $t = 0$. The predicted thermodynamic temperature is $T'' = 62.07°\text{A}$.

occur. Thus a correction term must be added in the right side of (62) to obtain an estimate of the nonlinear pressure. When the definite integral in (62) is also evaluated for this case, a measure of the time average of the pressure over the interval $(0, \tau)$ in the nonlinear system is found to be given by

$$\langle P \rangle = -\gamma(l - a) + \tfrac{1}{2}\gamma\tau^{-1}(\alpha_1' - \alpha_N')[\dot{\mathbf{x}}]_0^\tau + C'. \quad (64)$$

The term C' may be shown to be

$$C' = M(N + 1)^{-1}\tau^{-1} \sum_i |\Delta\dot{x}_i(t_i)|, \quad (65)$$

where the summation is again taken over all collisions in the time interval. Provided the time interval $(0, \tau)$ is large, we shall refer to the pressure $\langle P \rangle$ given by (64) as the mean pressure of the system.

C. Results of the Computations

The numerical computations for different systems showed that the mean temperatures and pressures approach the values predicted theoretically, when $N \gg 1$. That a small number of particles gives a good approximation to the infinite assembly seems a characteristic of molecular dynamics. Even when we chose N as small as 7 or 8, the mean temperature approached to within 4 % of the predicted temperature for the infinite assembly, while the error in the mean pressure was of the order of 30%. However it is realistic to discuss only the results for the largest systems on which computations were carried out.

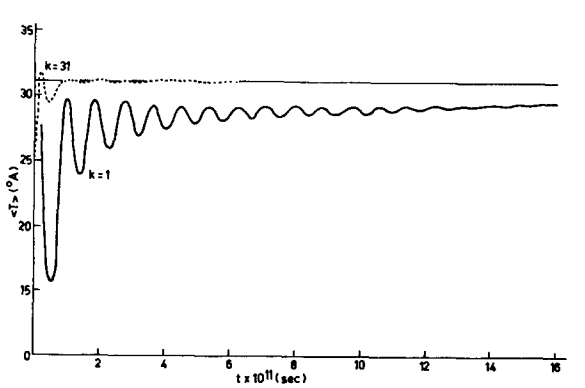

FIG. 8. The mean temperature for the weakly nonlinear systems $N = 31$, $\epsilon = \epsilon'$, $E_i = \delta_{ik}E$ at $t = 0$. $T' = 31.7°\text{A}$. The distinctly linear behavior of the system for $k = 1$ is apparent.

In Fig. 7 we plot $\langle T \rangle$ for the two systems in which $N = 31$, $\epsilon = 0.7 \times 10^{-14}$ ergs, and the initial configurations were $E_1 = E$ and $E_N = E$, respectively. It is quite clear that $\langle T \rangle$ approaches near the predicted value in each case. Figure 8 shows the behavior of $\langle T \rangle$ in the two similar systems where $\epsilon = 0.4 \times 10^{-14}$ ergs. In a linear system in which all the energy is in the ith normal mode, say, the curve of $\langle T \rangle$ as a function of the time t will oscillate periodically about the mean $E/(Nk)$ with a period $2\pi/\omega_i$. The oscillations about the mean will decrease to zero as $t \to \infty$. This explains the periodic oscillations in the curve labeled $k = 1$. In the weakly nonlinear system, the first mode remained dominant for a considerable time as there was little energy sharing with the other modes, and the oscillations in $\langle T \rangle$ are therefore to be expected. The period of these oscillations is approximately $2\pi/\omega_1 = 1.753 \times 10^{-11}$ sec. However, as the energy becomes distributed among all the modes, $\langle T \rangle$ begins to approach near the predicted value of 31.7°A. The approach is slow because of the dominance of the first mode for such a long time. The curve $k = 31$ approaches the equilibrium value much more quickly because of the rapid exchange of energy among all the modes when the initial configuration is $E_{31} = E$.

Some actual values for $\langle T \rangle$ and $\langle P \rangle$, together with the relative fluctuations σ_T^2 and $\sigma_{\langle P \rangle}^2$ computed from (55) and (56), respectively, are given in Table VI. The approach to the predicted temperature is very good. Indeed the error in the first two systems is about 2%, while in the more strongly nonlinear systems it is only about 1%. The computed values for σ_T^2 indicate that the systems tend towards a state of equilibrium with respect to temperature. The computed values for the mean pressure, although not as accurate as those for the mean temperature, also show an approach near to the predicted pressures. Curves for the pressure $\langle P \rangle$ are illustrated in Figs. 9 and 10. Of course the fluctuations $\sigma_{\langle P \rangle}^2$ should tend to zero if $\langle P \rangle$ approaches an equilibrium value. The given estimates for $\sigma_{\langle P \rangle}^2$ therefore support the theory that the mean pressures approach equilibrium values. It should be pointed

TABLE VI. Time averages of the temperature $\langle T \rangle$ and pressure $\langle P \rangle$, together with the relative fluctuations σ_T^2 and $\sigma_{\langle P \rangle}^2$, respectively, over intervals of C collisions and time t sec. $N = 31$, $E_k = E$ at $t = 0$, and $\epsilon = E/N$. The theoretically predicted values are $T' = 31.72°A$, $T'' = 62.07°A$, $P' = 0.1060 \times 10^{-6}$ dyn, $P'' = 0.5031 \times 10^{-6}$ dyn.

ϵ	k	$C \times 10^{-2}$	$t \times 10^{10}$	$\langle T \rangle$	σ_T^2	$\langle P \rangle \times 10^6$	$\sigma_{\langle P \rangle}^2$
ϵ'	1	0–15	3.31	30.6	0.078	0.0605	0.7578
		15–30	2.66	31.5	0.018	0.0875	0.0048
ϵ'	31	0–15	2.84	31.3	0.023	0.0813	0.0080
		15–30	2.65	31.5	0.022	0.0988	0.0008
ϵ''	1	0–15	0.77	61.4	0.016	0.5101	0.0042
		15–30	0.85	61.5	0.015	0.4476	0.0004
ϵ''	31	0–15	0.85	61.6	0.019	0.4695	0.0025
		15–30	0.87	61.4	0.017	0.4494	0.0003

out that no contributions to the fluctuations σ_T^2 and $\sigma_{\langle P \rangle}^2$ were computed for the first 150 collisions when $\epsilon = \epsilon'$, and for the first 300 collisions when $\epsilon = \epsilon''$.

VII. THE AUTOCORRELATION FUNCTION

In a system of N particles, the autocorrelation function associated with the ith particle is defined by

$$\rho_{N,i}(t) = \frac{\lim_{T\to\infty} \frac{1}{T} \int_0^T p_i(t+\tau) p_i(\tau)\, d\tau}{\lim_{T\to\infty} \frac{1}{T} \int_0^T p_i(\tau)^2\, d\tau},$$

where p_i is the momentum of the particle. For the purposes of numerical computation, an estimate of this function is given by

$$\rho_{N,i}(t, T) = \frac{\int_0^T \dot{x}_i(t+\tau)\dot{x}_i(\tau)\, d\tau}{\int_0^T \dot{x}_i(\tau)^2\, d\tau},$$

provided $T \gg t$.

It has been shown by Mazur and Montroll[15] that a knowledge of the behavior of the autocorrelation function can show whether or not equilibrium is established in the system. Preliminary computations have been made in which $\rho_{N,i}(t)$ was evaluated for several (200) values of t. In Fig. 11 the curve $\rho_{N,1}(t)$ for $T = 0.6 \times 10^{-9}$ sec is drawn. It is interesting to compare this with the hypothetical curve contained in the paper by Mazur and Montroll.

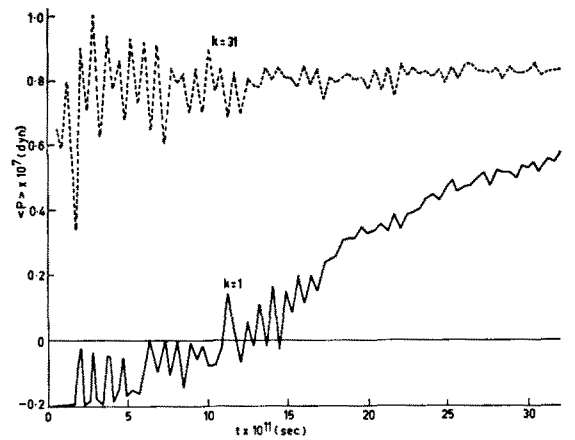

FIG. 10. The mean pressure for $N = 31$, $\epsilon = \epsilon'$, $E_i = \delta_{ik}E$ at $t = 0$. $P' = 0.106 \times 10^{-6}$ dyn.

VIII. CONCLUSIONS

We have shown that, in contrast to the systems studied by FPU and others, our nonlinear systems show a behavior which can reasonably be described as ergodic. This is evidenced both by the equipartition of the energy of each system among all the modes, in the time average, and by the rapid approach of the temperature and pressure to their equilibrium values. It is hoped to use our model to verify numerically some of the other results of statistical mechanics. In particular, calculation

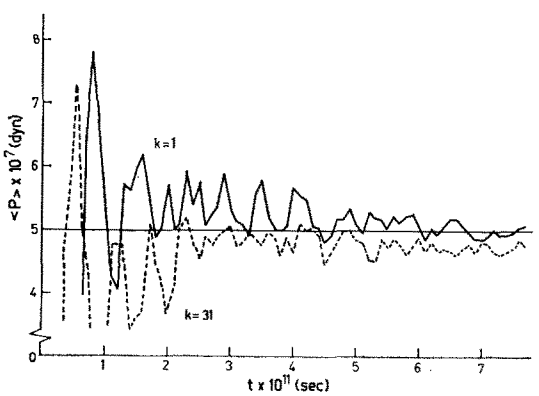

FIG. 9. The mean pressure for the systems $N = 31$, $\epsilon = \epsilon''$, $E_i = \delta_{ik}E$ at $t = 0$. The predicted thermodynamic pressure is $P'' = 0.503 \times 10^{-6}$ dyn.

[15] P. Mazur and E. Montroll, J. Math. Phys. 1, 70 (1960).

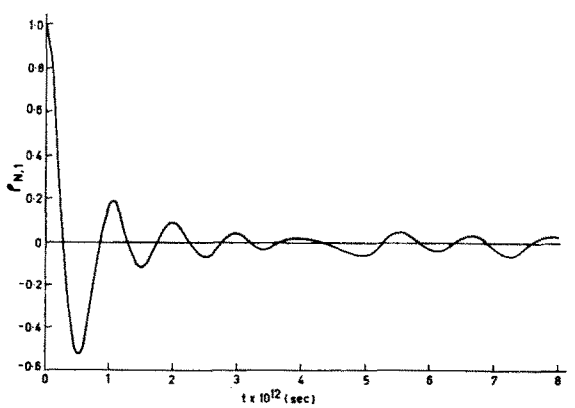

FIG. 11. The velocity autocorrelation function $\rho_{N,i}$ for the system $N = 31$, $\epsilon = \epsilon'$, $k = 1$.

of the entropy function and its fluctuations for both a system and a subsystem is proposed. This should lead to a numerical evaluation of the H-theorem.

It should be possible to derive the macroscopic properties of the model by the generation of a stochastic process. The time intervals between two collisions could be chosen at random from a probability distribution. The trajectory of any system in phase space may then be followed by repeated application of the transformation (13) for the selected time intervals, together with a random choice of a collision after each iteration. We hope to carry out computations to show that this method yields systems which are statistically equivalent to the ones which have been discussed in this paper.

ACKNOWLEDGMENTS

In the early stages of this work, helpful discussions on the problem were held with Dr. Robert Hermann and Mr. Richard Rothery of the General Motors Research Laboratories, Detroit. The authors also wish to thank Dr. C. A. Hurst and Mr. J. G. Sanderson of the University of Adelaide for several fruitful discussions.

Some of the computations were carried out on a 7090 while one of the authors (RBP) was a visitor at the IBM Research Center, Yorktown Heights. One of us (RSN) gratefully acknowledges the support of the Australian Government for an award under the British Commonwealth Scholarship and Fellowship Plan.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Spectral Analysis of the Anisotropic Neutron Transport Kernel in Slab Geometry with Applications*

A. LEONARD AND T. W. MULLIKIN
The RAND Corporation, Santa Monica, California
(Received 15 October 1963)

A spectral analysis of the transport kernel for anisotropic scattering in finite slabs is achieved by first solving a type of generalized scattering problem for a subcritical slab. Initially, the scattering problem is stated as an inhomogeneous integral transport equation with a complex-valued source function. This is readily transformed to singular integral equations and linear constraints in which the space and angle variables enter as parameters. Dual singular equations appear in applications of Case's method to transport problems, but we cannot yet completely explain this duality. The singular equations are transformed to Fredholm equations by an extension of Muskhelishvili's standard method and by analytic continuation. It is shown that, for a wide class of scattering functions, this particular Fredholm reduction yields equations which converge rapidly under iteration for all neutron productions and slab thicknesses. The ultimate solution of the singular equations contains arbitrary constants which, when evaluated by the aforementioned linear constraints, display explicitly the Fredholm determinant and the eigenfunctions of the transport kernel. An immediate consequence of this result is the criticality condition and the associated neutron distribution. Specific applications to linear anisotropic and isotropic scattering in slab geometry are discussed. In addition, it is seen that the case of isotropic scattering in spheres can be treated with this method, and, in fact, the spectral analysis of the kernel for the slab problem immediately applies to the sphere kernel.

1. INTRODUCTION

IN this paper, we consider an integral operator A, derivable from the steady-state transport equation for anisotropic scattering in a one-dimensional slab of finite thickness. By means of the theory of singular integral equations and analytic continuation, we reduce a certain scattering problem to a form suitable for computation. From this solution we obtain both the eigenvalues and eigenfunctions of the operator A. For isotropic scattering, these results give, by a familiar observation,[1] a spectral analysis of the isotropic transport kernel for a finite sphere.

The spectral analysis is important in the determination of critical dimensions of nuclear reactors in which the energy of neutrons is considered constant. By the standard method of taking a Laplace transform in the time variable,[1] these results can also be applied to determine the discrete spectrum of the time-dependent transport equation.

Recently, various transport problems for plane geometry have been treated by Case's singular eigenfunction method.[2-5] This method is one of determining singular solutions to the integrodifferential form of the homogeneous transport equation by separation of variables in infinite space. Solutions to boundary-value problems are expanded in terms of these singular solutions. Determination of the expansion coefficients is achieved by solving singular integral equations.

The method of this paper also leads to a study of singular integral equations, which are adjoint to those derived in the applications of Case's method. For this reason our method seems to be dual to that of Case. In deriving these equations we do not use singular solutions to the integrodifferential equation form of the transport problem. Rather, we consider the familiar associated Fredholm equation which incorporates the boundary conditions for any given problem. In another paper we have shown by operational methods that this integral equation can be readily replaced by singular integral equations and linear constraints. The linear constraints come from properties of the analytic continuation of a certain function.

Our transformation to singular equations is achieved by a method which differs from the Fourier transform methods of the Weiner–Hopf analysis[6] of integral operators with difference kernels on a half-line. Our method can be applied to the study of a wide class of such operators defined over an arbitrary interval. We plan to pursue this in a later paper.

We wish to emphasize one other point of our

* This research is sponsored by the United States Air Force under Project RAND–Contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development Headquarters, United States Air Force.

[1] B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957).
[2] K. M. Case, Ann. Phys. (N. Y.) 9, 1 (1960).
[3] J. R. Mika, Nucl. Sci. Eng. 11, 415 (1961).
[4] G. J. Mitsis, Argonne National Laboratory, ANL-6459 (1961).
[5] R. Zelazny, J. Math. Phys. 2, 538 (1961).
[6] N. Wiener and E. Hopf, Sitzber. Preuss. Akad. Wiss. Physik. Math. Kl. 1931, 696. (1931).

analysis. Because of analytic properties of certain functions, we are able to extend the standard results of singular equation theory[7] by means of analytic continuation. This leads to certain Fredholm equations which are simpler than those given by the standard theory.

We first consider in Sec. 2 the problem of anisotropic scattering in a one-dimensional slab which is of subcritical thickness τ. If the slab is exposed on one face to a monodirectional beam of neutrons and a steady state is reached, the neutron source function J, a function of two variables which gives the distribution of neutrons in space and angle emergent from collisions, satisfies a familiar linear integral equation.[1,8-10] For a subcritical reactor this integral equation can be solved by a Neumann series in an integral operator Λ. From this series it is difficult to obtain any information about the spectrum of the operator Λ.

With the assumption that the scattering function $f(\Omega, \Omega')$ has a finite Legendre polynomial expansion, it has been shown[11] that the integral equation for J can be replaced by a functional equation which relates J at the symmetric depths x and $\tau - x$. In Sec. 3 this equation is shown to lead to linear singular integral equations and a finite number of linear constraints.

In Sec. 4, we use singular equation theory and analytic continuation to convert the singular equations to Fredholm equations with continuous kernels. The solution to these Fredholm equations contain arbitrary constants which are determined by the linear constraints mentioned above. The determinant of this system of linear algebraic equations contains the Fredholm determinant of the operator Λ. For given f and $c > 1$, the critical dimension is determined by the smallest τ which makes this determinant vanish. By a limiting process we obtain the critical source function, which is an eigenfunction of the operator Λ. In fact, all eigenvalues and eigenfunctions of Λ can be so determined. In certain cases, we also obtain solutions to the Milne problem for $\tau = \infty$.[1,8,10]

In Sec. 5, specific results for the slab problem are obtained in the case of scattering functions linear in $(\Omega \cdot \Omega')$, and in the case of isotropic scattering. The results for isotropic scattering are related to previous results obtained by Mitsis[4] and Zelazny,[5] who constructed their solutions from Case's[2] singular solutions to the transport equation.

In Sec. 6 it is shown that the sphere criticality problem for isotropic scattering is also easily handled by the method of this paper. In fact, a slight modification of the slab reactor results immediately gives estimates for the sphere.

2. THE TRANSPORT EQUATION

We shall consider a homogeneous plane-parallel reactor of finite thickness τ. Depth into the slab is measured by the variable x, $0 \leq x \leq \tau$, in a direction normal to a face of the slab and in units of a mean free path. We let $\Psi(x, \mu)$ denote the angular neutron density at depth x in a direction specified by θ, the inclination to the positive x axis. Here $\mu = \cos\theta$.

We consider an axially symmetric beam of neutrons having unit flux incident on the slab at $x = 0$ from some direction $\mu_0 \geq 0$. The steady-state transport equation for the angular density of *scattered* neutrons is then[8]

$$\mu[\partial\Psi(x, \mu, \mu_0)/\partial x] + \Psi(x, \mu, \mu_0) = J(x, \mu, \mu_0), \quad (2.1)$$

where the source function J is given by

$$J(x, \mu, \mu_0) = \frac{1}{2}\int_{-1}^{1} f(\mu, \nu)\Psi(x, \nu, \mu_0)\, d\nu + f(\mu, \mu_0)\exp(-x/\mu_0). \quad (2.2)$$

Boundary conditions result from the fact that no scattered neutrons enter the faces of the slab, and are given by

$$\Psi(0, \mu, \mu_0) = \Psi(\tau, -\mu, \mu_0) = 0, \quad 0 \leq \mu \leq 1. \quad (2.3)$$

The last term on the right-hand side of (2.2) is the resultant source due to the reduced incident flux.

We assume that the positive scattering function f in (2.2) is given by a finite expansion in Legendre polynomials,

$$\text{(i)} \quad f \geq \epsilon > 0, \qquad \text{(ii)} \quad f(\mu, \nu) = c\sum_{i=0}^{N} b_i P_i(\mu)P_i(\nu), \quad b_0 = 1. \quad (2.4)$$

For simplicity, we have assumed azimuthal symmetry. Mullikin[11] has shown that the general case of azimuthal dependence leads to additional independent equations which may be handled in the same way with very little additional effort.

[7] N. I. Muskhelishvili, *Singular Integral Equations*, (P. Noordhoff Ltd., Groningen, The Netherlands, 1953).

[8] K. M. Case, F. DeHoffmann, and G. Placzek, *Introduction to the Theory of Neutron Diffusion* (U. S. Government Printing Office, 1953), Vol. 1.

[9] S. Chandrasekhar, *Radiative Transfer* (Dover Publications, Inc., New York, 1960; Oxford University Press, London, 1950).

[10] E. Hopf, "Mathematical Problems of Radiative Equilibrium" (Cambridge Tracts, No. 31, 1934).

[11] T. W. Mullikin, The RAND Corporation, RM-3558-PR, (1963) (to appear in Astrophys. J.).

From (2.1), (2.2), and (2.3) we obtain expressions for Ψ in terms of J by

$$\Psi(x, \mu, \mu_0) = \frac{1}{\mu}\int_0^x \exp[(y - x)/\mu] \times J(y, \mu, \mu_0)\, dy \qquad (2.5)$$

for $0 < \mu \leq 1$, and

$$\Psi(x, \mu, \mu_0) = -\frac{1}{\mu}\int_x^\tau \exp[(y - x)/\mu] \times J(y, \mu, \mu_0)\, dy \qquad (2.6)$$

for $-1 \leq \mu < 0$. Combined with (2.2), this gives the integral equation

$$J(x, \mu, \mu_0) = f(\mu, \mu_0)\exp(-x/\mu_0) + \frac{1}{2}\int_0^1 f(\mu, \sigma)\int_0^x \exp[(y - x)/\sigma]J(y, \sigma, \mu_0)\, dy\, \frac{d\sigma}{\sigma} - \frac{1}{2}\int_{-1}^0 f(\mu, \sigma)\int_x^\tau \exp[(y - x)/\sigma]J(y, \sigma, \mu_0)\, dy \frac{d\sigma}{\sigma}. \qquad (2.7)$$

We abbreviate this equation by

$$J = c\Lambda(J) + B. \qquad (2.8)$$

It is easily shown that because of (2.4) the operator Λ is strictly positive.[12] This implies that Λ has a real positive eigenvalue $\bar{\lambda}$ of unit multiplicity, which dominates the modulus of all other eigenvalues. This eigenvalue is a function of the parameters $\tau, b_1, \cdots, b_N$. The eigenfunction ρ associated with $\bar{\lambda}$ can be chosen to be uniformly positive.[12]

We will need the following result (see Hopf[10]). Let the constants $b_1, \cdots, b_N$ be restricted by (2.4). Then the maximum eigenvalue $\bar{\lambda}$ of Λ satisfies

$$\bar{\lambda}(\tau, b_1, \cdots, b_N) < 1, \qquad 0 \leq \tau < \infty. \qquad (2.9)$$

If the constant c is restricted by

$$c\bar{\lambda} < 1, \qquad (2.10)$$

Eq. (2.8) has the unique solution given by the Neumann series

$$J = \sum_{n=0}^{\infty} c^n \Lambda^n(B). \qquad (2.11)$$

The solution is an analytic function of the complex variable μ_0 for all $0 < |\mu_0| < \infty$ with a pole of order N at ∞, and real valued and nonnegative for real μ_0, $\mu_0 \neq 0$.

We observe from (2.4) that

$$f(\mu, -\nu) = f(-\mu, \nu). \qquad (2.12)$$

[12] S. Karlin, J. Math. Mech. 8, 907 (1959).

It then follows easily from (2.7) that

$$J(\tau - x, -\mu, \mu_0) = \exp(-\tau/\mu_0)J(x, \mu, -\mu_0). \qquad (2.13)$$

The solution given by (2.11) is computationally unsatisfactory for reactors which are near criticality because of the slow rate of convergence of the series. We shall solve (2.8) for subcritical reactors by replacing (2.8) by singular linear integral equations which give a more satisfactory procedure of numerical solution. From the solution of these equations we shall obtain a method for computing $\bar{\lambda}(\tau, b_1, \cdots, b_N)$ and hence the criticality condition

$$c\bar{\lambda}(\tau, b_1, \cdots, b_N) = 1. \qquad (2.14)$$

If we then take the solution J to the subscritical problem and let c increase to criticality we shall find, by dominance of $\bar{\lambda}$,

$$\lim_{c\bar{\lambda}\to 1}(1 - c\bar{\lambda})J = \rho. \qquad (2.15)$$

This function ρ is an eigenfunction of the operator Λ and satisfies

$$\rho = c\Lambda(\rho). \qquad (2.16)$$

It represents the angular density of source neutrons in space and direction in a critical reactor.

It is easily shown that Λ, as defined by (2.7), has a kernel which is square integrable in the variables x, y, μ, and ν for $0 \leq x, y \leq \tau$, $-1 \leq \mu, \nu \leq 1$. Therefore the eigenvalues of Λ are discrete, although not necessarily real, since Λ is not Hermitian in general. From our solution to singular integral equations, we can determine all the eigenvalues and eigenfunctions of Λ.

3. LINEAR SINGULAR EQUATIONS

In the previous section, we arrived at an integral equation (2.7) whose solution is the source function J for a scattering problem. Mullikin[11] has shown that (2.7) can be replaced by another equation in which the roles of the independent variables x and μ and the parameter μ_0 are interchanged. We merely outline this derivation.

In (2.7) the variable μ_0 is usually restricted by $0 \leq \mu_0 \leq 1$ as the cosine of an angle in the interval $[0, \frac{1}{2}\pi]$. In the term $f(\mu, \mu_0)\exp(-x/\mu_0)$ let us replace μ_0 in the exponential by a complex number z, $z \neq 0$. We then consider the equation

$$K(x, \mu, \mu_0, z) = f(\mu, \mu_0)\exp(-x/z) + c\Lambda(K)(x, \mu, \mu_0, z). \qquad (3.1)$$

It is easily shown that this equation has a solution which is analytic in z for $|z| > 0$. For $z = \mu_0$ this solution agrees with the solution J to (2.7).

We now apply the integral operator Λ to both sides of (3.1). This can readily be seen to give an equation for $\Lambda(K)$ which is similar to (3.1) except that the forcing term consists of integrals over the parameters μ_0 and z. Applying $(I - c\Lambda)^{-1}$ to this equation, we obtain a representation of $\Lambda(K)$ in terms of K and J. When this is combined with (3.1) we obtain

$$\begin{aligned} &K(x, \mu, \nu, z) \\ &- \frac{cz}{2} \int_{-1}^{1} \frac{K(x, \mu, \sigma, z) - K(x, \mu, z, z)}{z - \sigma} f(\nu, \sigma)\, d\sigma \\ &= f(\mu, \nu) \exp(-x/z) \\ &+ \frac{cz}{2} \int_{0}^{1} \frac{J(x, \mu, z) - J(x, \mu, \sigma)}{z - \tau} f(\nu, \sigma)\, d\sigma \\ &+ \frac{cz}{2} \int_{-1}^{0} \frac{J(x, \mu, z) - \exp(-\tau/z) J(\tau - x, -\mu, -\sigma)}{z - \sigma} \\ &\times f(\nu, \sigma)\, d\sigma. \end{aligned} \tag{3.2}$$

If we now consider fixed values of the variables x, μ, and z, then the left-hand side of (3.2) contains a linear operator L_z which acts only on the third argument of the function K. The variable ν enters the right-hand side of (3.2) only in the function $f(\nu, \sigma)$. If we assume that f has a finite expansion in Legendre polynomials, then it can be shown[11] that the operator $(I - L_z)$ has an inverse and that (3.2) can be solved for K in terms of J.

After solving (3.2) for K, we again set ν equal to z to obtain

$$\begin{aligned} &J(x, \mu, z) = k(\mu, z) \exp(-x/z) \\ &- \frac{z}{2} \int_{0}^{1} \frac{J(x, \mu, z) - J(x, \mu, \sigma)}{z - \sigma} k(\sigma, z)\, d\sigma \\ &+ \frac{z}{2} \int_{-1}^{0} \frac{J(x, \mu, z) - \exp(-\tau/z) J(\tau - x, -\mu, -\sigma)}{z - \sigma} \\ &\times k(\sigma, z)\, d\sigma. \end{aligned} \tag{3.3}$$

Notice that now x and μ are merely parameters. The function k is given by

$$k(\mu, z) \equiv c \sum b_i P_i(\mu) h_i(z), \tag{3.4}$$

where h_i are defined recursively by

$$\begin{aligned} &h_0(z) = 1, \qquad h_1(z) = (1 - c)z, \\ &(j + 1)h_{j+1}(z) + jh_{j-1}(z) \\ &\qquad = z[(2j + 1) - cb_j]h_j(z). \end{aligned} \tag{3.5}$$

This formula for the h_i is obtained by Mika[3] by a different method than that used by Mullikin.[11]

We are primarily interested in $J(x, \mu, z)$ for z in the interval [0, 1]. Once this is determined, we have the analytic continuation of J to complex z given by (3.3). For the moment we restrict z to complex values outside the interval $[-1, 1]$ and rearrange (3.3) as

$$\begin{aligned} &\lambda(z) J(x, \mu, z) = k(\mu, z) \exp(-x/z) \\ &\quad - \frac{z}{2} \int_{0}^{1} \frac{k(\sigma, z) J(x, \mu, \sigma)}{z - \sigma}\, d\sigma \\ &\quad - \frac{z}{2} \exp(-\tau/z) \int_{-1}^{0} \frac{k(\sigma, z) J(\tau - x, -\mu, -\sigma)}{z - \sigma}\, d\sigma. \end{aligned} \tag{3.6}$$

The function λ is defined by

$$\lambda(z) = 1 + z \int_{-1}^{1} \frac{\gamma(\sigma)\, d\sigma}{\sigma - z}, \qquad z \notin [-1, 1], \tag{3.7}$$

where

$$\gamma(\sigma) = \tfrac{1}{2} k(\sigma, \sigma). \tag{3.8}$$

To obtain (3.7) from (3.4) we have used the result given by Mullikin[11] that

$$\int_{-1}^{1} \frac{k(\sigma, z) - k(\sigma, \sigma)}{\sigma - z}\, d\sigma = 0. \tag{3.9}$$

It is also shown by Mullikin[11] that

$$k(\mu, z) = k(-\mu, -z), \tag{3.10}$$

and hence that

$$\gamma(\sigma) = \gamma(-\sigma). \tag{3.11}$$

For special functions f, the function γ has been called the "characteristic function" by Chandrasekhar.[9]

The expression (3.4) for k shows that Eq. (3.6) can be expanded in terms of Legendre polynomials in μ to give $(N + 1)$ independent equations. Since this expansion can be done at any stage of our treatment, we simply study the single equation (3.6) with μ as a parameter.

The first piece of information to be gained from the functional equation (3.6) is a set of linear constraints. We want a solution to (3.6) which is analytic in z for $0 < |z| < \infty$. If the function λ vanishes at some point, or set of points, z_i with z_i finite and outside the interval $[-1, 1]$, then for J to be analytic the right-hand side of (3.6) must vanish to the same order as λ at these points. This gives linear constraints such as

$$\begin{aligned} &k(\mu, z_i) \exp(-x/z_i) \\ &= \frac{z_i}{2} \int_{0}^{1} \frac{k(\sigma, z_i) J(x, \mu, \sigma)}{z_i - \sigma}\, d\sigma \\ &\quad + \exp(-\tau/z_i) \frac{z_i}{2} \int_{-1}^{0} \frac{k(\sigma, z_i) J(\tau - x, -\mu, -\sigma)}{z_i - \sigma}\, d\sigma, \end{aligned} \tag{3.12}$$

for z_i satisfying

$$\lambda(z_i) = 0, \qquad \lambda'(z_i) \neq 0. \tag{3.13}$$

Complications arise if λ vanishes at infinity. It follows easily from (3.11), by induction on n, that for $n \geq 1$

$$h_n(z) = P_n(z) \prod_{i=0}^{n-1} \left(1 - \frac{cb_i}{2i+1}\right) + \text{lower order } P_i. \tag{3.14}$$

Then from (3.7), (3.4), and (3.14), we have

$$\lambda(\infty) = (1 - c) - c \sum_{i=1}^{N} \frac{b_i}{2i+1} \prod_{j=0}^{i-1} \left(1 - \frac{cb_j}{2j+1}\right). \tag{3.15}$$

Again, by induction on N, it readily follows that

$$\lambda(\infty) = \prod_{i=0}^{N} \left(1 - \frac{cb_i}{2i+1}\right). \tag{3.16}$$

Now formula (3.6) gives the analytic continuation of $J(x, \mu, z)$ off the interval $-1 \leq z \leq 1$, and J must be so chosen on the interval $[-1, 1]$ that this function is analytic in $0 < |z| < \infty$ with a pole of order N at infinity. From (3.14) and (3.16) we see that if $\lambda(\infty) \neq 0$, then $h_N(z)$ is of degree N, and hence also $k(\mu, z)$ and J are of degree N in z, giving the correct-order pole at infinity. However, if $cb_N = 2N + 1$, then k is of degree N but $\lambda(\infty) = 0$, and conditions have to be imposed to prevent J having a higher-order pole than N at infinity. Also, if some $cb_i = 2i + 1$, $i < N$, it follows easily that $k(\mu, z)$ is at most of degree $(N - 1)$ in z and that λ vanishes at infinity at least to order two. Then conditions again have to be imposed to give J the correct behavior at infinity.

It should be remembered that the constants b_i are not completely arbitrary but are restricted by the condition (2.4) of positivity of the scattering function f. However, it seems difficult to relate these restrictions in general to possibilities discussed above.

Further calculations will require the frequent use of Plemelj's formulas.[11] The function φ, defined by

$$\varphi(z) = \frac{1}{2\pi i} \int_a^b \frac{f(\sigma)\, d\sigma}{\sigma - z}, \tag{3.17}$$

is analytic in the complex plane cut along $[a, b]$. Provided f satisfies a Hölder condition on (a, b), Plemelj's formulas give the limits φ^+ and φ^- of φ from the upper and lower half-planes, respectively, as z tends to ν, $a < \nu < b$:

$$\varphi^+(\nu) = \tfrac{1}{2} f(\nu) + \frac{1}{2\pi i} \int_a^b \frac{f(\sigma)\, d\sigma}{\sigma - \nu}, \qquad \varphi^-(\nu) = -\tfrac{1}{2} f(\nu) + \frac{1}{2\pi i} \int_a^b \frac{f(\sigma)\, d\sigma}{\sigma - \nu}. \tag{3.18}$$

The integrals are computed as Cauchy principal values.

Remembering that J is to be analytic in $0 < |z| < \infty$, we now apply Plemelj's formulas to compute the limits in (3.6) as z tends to ν with $-1 < \nu < 1$, $\nu \neq 0$. Adding the limits from the lower and upper half-planes, we obtain a singular integral equation for $J(x, \mu, \nu)$ on the interval $-1 < \nu < 1$:

$$\lambda_0(\nu) J(x, \mu, \nu) = k(\mu, \nu) \exp(-x/\nu) - \frac{\nu}{2} \int_0^1 \frac{k(\sigma, \nu) J(x, \mu, \sigma)\, d\sigma}{\nu - \sigma} - \exp(-\tau/\nu) \frac{\nu}{2} \int_{-1}^0 \frac{k(\sigma, \nu)}{\nu - \sigma} J(\tau - x, -\mu, -\sigma)\, d\sigma. \tag{3.19}$$

The singular integrals are computed as Cauchy principal values, and λ_0 is given by

$$\lambda_0(\nu) = 1 + \nu \int_{-1}^1 \frac{\gamma(\sigma) - \gamma(\nu)}{\sigma - \nu}\, d\sigma - \nu\gamma(\nu) \ln \frac{1 + \nu}{1 - \nu}. \tag{3.20}$$

To simplify the subsequent analysis, we choose to write (3.19) as two uncoupled equations on the interval (0, 1). We do this by defining two functions by

$$P(x, \mu, \nu) = J(x, \mu, \nu) + J(\tau - x, -\mu, \nu), \qquad Q(x, \mu, \nu) = J(x, \mu, \nu) - J(\tau - x, -\mu, \nu), \qquad 0 < \nu < 1. \tag{3.21}$$

The equations for P and Q are then

$$\lambda_0 P = l_1 + V(P) - U(P), \qquad \lambda_0 Q = l_2 + V(Q) + U(Q). \tag{3.22}$$

For brevity, we have defined two integral operators, V and U, by

$$V(J)(x, \mu, \nu) = \frac{\nu}{2} \int_0^1 \frac{k(\sigma, \nu) J(x, \mu, \sigma)}{\sigma - \nu}\, d\sigma, \qquad U(J)(x, \mu, \nu) = \exp(-\tau/\nu) \frac{\nu}{2} \int_0^1 \frac{k(-\sigma, \nu)}{\sigma + \nu} \times J(x, \mu, \sigma)\, d\sigma \tag{3.23}$$

for $0 < \nu < 1$, and we have introduced the known functions l_1 and l_2 given by

$$l_1(x, \mu, \nu) \equiv k(\mu, \nu) \exp(-x/\nu) + k(\mu, -\nu) \exp[(x - \tau)/\nu],$$
$$l_2(x, \mu, \nu) \equiv k(\mu, \nu) \exp(-x/\nu) - k(\mu, -\nu) \exp[(x - \tau)/\nu]. \quad (3.24)$$

In the next section we use techniques of analytic function theory to convert (3.22) to Fredholm equations with continuous kernels. Although Mika,[3] by using Muskhelishvili's standard method,[7] has already treated singular equations of this type for a semi-infinite medium, which are adjoint in the singular part of (3.22), we choose to employ a slight modification of the standard method. In this way, we take advantage of certain simplifications resulting from the analytic properties of the various functions.

4. FREDHOLM EQUATIONS

In this section, then, we consider the singular integral equation

$$\lambda_0(\nu)R(\nu) = F(\nu) + \frac{\nu}{2}\int_0^1 \frac{k(\sigma, \nu)R(\sigma)\, d\sigma}{\sigma - \nu}, \quad (4.1)$$

which is Eq. (3.22) for P (or Q), with l_1 (or l_2) and the bounded operator U included in the definition of F. That is, if (4.1) is to represent (3.22) for P, then

$$F(\nu) = l_1(x, \mu, \nu) - U(P)(x, \mu, \nu), \quad (4.2)$$

or, if it is to represent (3.22) for Q, we have

$$F(\nu) = l_2(x, \mu, \nu) + U(Q)(x, \mu, \nu). \quad (4.3)$$

Since x and μ are merely parameters in our equations, we will not display them. We will treat F as a known function and invert the singular operator by a slightly modified form of the theory of singular integral equations.[7] Finally, we will obtain a regular Fredholm equation for F.

Since k is a polynomial, a sectionally holomorphic function $\varphi(z)$ is defined by

$$\varphi(z) = \frac{1}{2\pi i}\int_0^1 \frac{k(\sigma, z)R(\sigma)\, d\sigma}{\sigma - z}. \quad (4.4)$$

By Plemelj's formulas it is easily verified that we can write (4.1) as

$$\lambda^-(\nu)\varphi^+(\nu) = \lambda^+(\nu)\varphi^-(\nu) + 2\gamma(\nu)F(\nu), \quad 0 \le \nu \le 1, \quad (4.5)$$

where

$$\lambda^+(\nu) = \lambda_0(\nu) + \pi i\nu\gamma(\nu),$$
$$\lambda^-(\nu) = \lambda_0(\nu) - \pi i\nu\gamma(\nu). \quad (4.6)$$

We divide (4.5) by λ^- and rewrite it as

$$\varphi^+(\nu) - [\lambda^+(\nu)/\lambda^-(\nu)]\varphi^-(\nu) = 2\gamma(\nu)F(\nu)/\lambda^-(\nu). \quad (4.7)$$

By standard methods,[7] we now construct a nonvanishing sectionally holomorphic function $E(z)$ ($z \notin [0, 1]$), which solves the homogeneous equation

$$E^+(\nu)/E^-(\nu) = \lambda^+(\nu)/\lambda^-(\nu). \quad (4.8)$$

We find that

$$E(z) = \exp[\Gamma(z)]/(1 - z)^\alpha \quad (4.9)$$

is the required solution, where the function Γ is defined by

$$\Gamma(z) = \int_0^1 \frac{\theta(t)\, dt}{t - z}, \quad (4.10)$$

and

$$\theta(t) = \frac{1}{\pi}\tan^{-1}\frac{\pi t\gamma(t)}{\lambda_0(t)}, \qquad 0 \le t \le 1. \quad (4.11)$$

It can be shown[3,11] that the index, α, is a nonnegative integer given by

$$\alpha = \theta(1) \le N + 1. \quad (4.12)$$

We will henceforth assume that

$$\lambda_0^2(\nu) + [\nu\gamma(\nu)]^2 \neq 0, \qquad 0 \le \nu \le 1. \quad (4.13)$$

If (4.13) does not hold, the index α is increased and is accompanied by an equal increase in the number of constraints on the equation for J [see (3.12)], where the additional z_i are just the zeros of the left-hand side of (4.13).

Inserting (4.8) into (4.7) gives

$$\varphi^+(\nu) - [E^+(\nu)/E^-(\nu)]\varphi^-(\nu) = 2\gamma(\nu)F(\nu)/\lambda^-(\nu), \quad (4.14)$$

or

$$\frac{\varphi^+(\nu)}{E^+(\nu)} - \frac{\varphi^-(\nu)}{E^-(\nu)} = \frac{2\gamma(\nu)F(\nu)}{\lambda^-(\nu)E^+(\nu)}. \quad (4.15)$$

Now consider the function G

$$G(z) = \frac{\varphi(z)}{E(z)} - \frac{1}{2\pi i}\int_0^1 \frac{2\gamma(\sigma)F(\sigma)\, d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma - z)}. \quad (4.16)$$

G is clearly analytic in the finite cut plane and, from (4.15), we find that G is continuous across the cut and hence is analytic in the entire finite plane. Finally, if

$$\beta \equiv \text{degree of } z \text{ in } k(\sigma, z), \quad (4.17)$$

then, by considering the behavior of the right-hand side of (4.16) at infinity, we conclude that G is a polynomial of degree $\alpha + \beta - 1$ at most, which

we choose to write as

$$G(z) = \frac{p_{\alpha+\beta-1}(z)}{\pi i} = \frac{1}{\pi i}\sum_{i=0}^{\alpha+\beta-1} a_i z^i. \tag{4.18}$$

Therefore, (4.16) becomes after multiplication by $E(z)$

$$\varphi(z) = \frac{E(z)}{\pi i}\int_0^1 \frac{\gamma(\sigma)F(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma - z)} + \frac{1}{\pi i} \times E(z)p_{\alpha+\beta-1}(z). \tag{4.19}$$

The standard singular integral equation theory concludes at this point by equating $\varphi^+ - \varphi^-$ as computed from (4.4) and from (4.19). If this is done for R set equal to P (or Q) and for F replaced by its definition in terms of P (or Q), then we obtain a Fredholm equation for P (or Q) on the interval [0, 1]. The kernel is such an equation is complicated. Fortunately we can use the analytic properties of the functions P and Q and the polynomial nature of the kernel k to obtain simpler Fredholm equations for the function F given by (4.2) [or by (4.3)]. This will also lead to an evaluation of the coefficients a_i in the polynomial $p_{\alpha+\beta-1}$.

We, therefore, specialize to the case where $R = P$, and F is defined by (4.2). By an obvious modification we can apply our results to the function Q.

Using the definition (3.23) of U and V, we see that $F(z)$, defined by (4.2), is simply related to $\varphi(-z)$ in (4.4). But $\varphi(-z)$ can also be computed by (4.19), and this gives us the following equation for F:

$$F(z) = l_1(z) - ze^{-\tau/z}E(-z) \times \left[\int_0^1 \frac{\gamma(\sigma)F(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + z)} + \sum_{i=0}^{\alpha+\beta-1} a_i(-z)^i\right]. \tag{4.20}$$

We define the linear operator L by

$$L(F)(z) = ze^{-\tau/z}E(-z)\int_0^1 \frac{\gamma(\sigma)F(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + z)}, \tag{4.21}$$

and choose to write (4.20) as the following set of integral equations:

$$\begin{aligned} \tilde{g}(z) &= l_1(z) - L(\tilde{g})(z), \\ g_i(z) &= (-z)^{i+1}e^{-\tau/z}E(-z) - L(g_i)(z), \\ &\qquad i = 0, \cdots, \alpha + \beta - 1. \end{aligned} \tag{4.22}$$

Because of the linearity of L, we clearly have

$$F(z) = \tilde{g}(z) + \sum_{i=0}^{\alpha+\beta-1} a_i g_i(z). \tag{4.23}$$

For z in the interval [0, 1], the Fredholm operator L has a continuous kernel. If the norm of L satisfies $||L|| < 1$ in some Banach space, then Eqs. (4.22) may be solved by iteration. These iterations will converge uniformly if

$$\max_{0\le\nu\le1}\left[\nu e^{-\tau/\nu}E(-\nu)\int_0^1 \frac{|\gamma(\sigma)|\,d\sigma}{E^+(\sigma)\lambda^-(\sigma)(\sigma + \nu)}\right] < 1. \tag{4.24}$$

For any given γ this inequality will certainly hold for all sufficiently large τ. We will show below that for certain γ this holds for all $\tau \geq 0$. The convergence of the iteration scheme for all γ and all $\tau \geq 0$ is an open question.

If we let $\mathfrak{C}$ denote a circle centered at the origin and enclosing the cut along [0, 1] and the point z, we can show[13] by contour integration that

$$\begin{aligned} &\int_0^1 \frac{\gamma(\sigma)\,d\sigma}{E^+(\sigma)\lambda^-(\sigma)(\sigma - z)} \\ &= \frac{1}{zE(0)} - \frac{1}{zE(z)} - \frac{1}{2\pi i}\int_{\mathfrak{C}} \frac{dw}{E(w)w(w - z)} \\ &= \frac{1}{zE(0)} - \frac{(1 - z)^\alpha \exp[-\Gamma(z)]}{z} \\ &\quad - \sum_{i=0}^{\alpha-1} \frac{z^i}{2\pi i}\int_{\mathfrak{C}} \frac{\exp[-\Gamma(w)](1 - w)^\alpha\,dw}{w^{i+2}}. \end{aligned} \tag{4.25}$$

The series in z terminates since coefficients of higher-order terms can be shown to vanish by choosing $\mathfrak{C}$ with arbitrarily large radius.

Now the left-hand side of (4.25) vanishes at infinity, and so must the right-hand side. The principle part of the Laurent expansion about infinity of $\exp[-\Gamma(z)](1 - z)^\alpha/z$ can be used to evaluate the various integrals around $\mathfrak{C}$ in terms of such quantities as $\int_0^1 t^n\theta(t)\,dt$. We shall abbreviate the polynomial part of (4.25) by $q(z)_{\alpha-1}$, a polynomial of degree $\alpha - 1$. Then we have

$$\nu E(-\nu)\int_0^1 \frac{\gamma(\sigma)\,d\sigma}{E^+(\sigma)\lambda^-(\sigma)(\sigma + \nu)} = 1 - \frac{E(-\nu)}{E(0)} - \nu E(-\nu)q(-\nu)_{\alpha-1}. \tag{4.26}$$

As a first application of this identity, we suppose that γ is positive in the interval [0, 1]. It follows readily[13] that $\alpha = 1$ and that, by (4.26) and the facts that $E(-\nu)$ and $E^+(\nu)\lambda^-(\nu)$ are positive for all γ,

$$0 \leq \nu E(-\nu)\int_0^1 \frac{\gamma(\sigma)\,d\sigma}{E^+(\sigma)\lambda^-(\sigma)(\sigma + \nu)} < 1, \qquad 0 \leq \nu \leq 1. \tag{4.27}$$

[13] T. W. Mullikin, The RAND Corporation, RM-3376-JPL (1962) (to appear in Trans. Am. Math. Soc.).

In this case of positive γ on [0, 1], condition (4.24) is valid, the operator L on [0, 1] is contracting for all $\tau \geq 0$, and Eqs. (4.22) can be solved by iteration for all $\tau \geq 0$.

For the remainder of this paper we assume that the Fredholm equations (4.22) have been solved for the functions $\tilde{g}$ and g_i on the interval [0, 1]. Once these functions are computed on [0, 1], their analytic continuation to complex z outside the cut $[-1, 0]$ is given again by (4.22).

We now turn our attention to the evaluation of the $\alpha + \beta$ constants a_i, which is needed to determine the function F.

In Sec. 3 we noted that linear constraints are obtained by demanding regularity of J at the zeros of λ in the plane cut along $[-1, 1]$. By an application of the argument principle to λ, it can be shown[3,11] that the number of zeros of λ equals 2α. Also by (3.5) and the fact that γ is an even function, we see that the zeros of λ occur in symmetric pairs $\pm z_i$. Now from (2.13) and (3.21), we see that we seek solutions P and Q which satisfy

$$\begin{aligned} P(x, \mu, \nu) &= \exp(-\tau/\nu)P(x, \mu, -\nu), \\ Q(x, \mu, \nu) &= -\exp(-\tau/\nu)Q(x, \mu, -\nu). \end{aligned} \tag{4.28}$$

Therefore, independent conditions are obtained by requiring regularity of P and Q at only one of each of the pairs of roots $\pm z_i$, and this is sufficient to obtain regularity of $J(x, \mu, z)$ at all the zeros $\pm z_i$.

We assume, for simplicity, that λ has only simple zeros which are finite and outside the cut $[-1, 1]$, i.e., we assume in particular that the constant c is such that $\lambda(\infty) \neq 0$ as given by (3.16). Then from (3.12), (3.21), (3.23), and (4.2), we see that for the function F we obtain α linear constraints

$$F(z_i) + \exp(-\tau/z_i)[F(-z_i) - l_1(-z_i)] = 0 \tag{4.29}$$

for $i = 1, \cdots, \alpha$. Using (4.22) and (4.23), we may write this as

$$\sum_{i=0}^{\alpha+\beta-1} D_{ij}(\gamma, c, b_1, \cdots, b_N)a_j = l_1(z_i)$$

$$- L(\tilde{g})(z_i) - e^{-\tau/z_i}L(\tilde{g})(-z_i), \quad i = 1, \cdots, \alpha, \tag{4.30}$$

where

$$\begin{aligned} D_{ij}(\tau, c, b_1, \cdots, b_N) &= z_i^{j+1}E(z_i) + (-z_i)^{j+1}e^{-\tau/z_i}E(-z_i) \\ &\quad - L(g_j)(z_i) - e^{-\tau/z_i}L(g_j)(-z_i). \end{aligned} \tag{4.31}$$

We also require β additional equations involving the a_i to completely determine F. These equations appear to be obtained most easily by investigating (4.19) at the zeros of γ. Assume that there are η pair of zeros, $\pm\zeta_i$, i.e.,

$$\gamma(\pm\zeta_i) = 0 \qquad i = 1, \cdots, \eta. \tag{4.32}$$

Using (4.4) for φ, we multiply (4.19) by πi and write it for $\nu = \pm\zeta_i$,

$$\begin{aligned} &\frac{1}{2}\int_0^1 \frac{k(\sigma, \pm\zeta_i)P(\sigma)\,d\sigma}{\sigma \mp \zeta_i} \\ &= E(\pm\zeta_i)\int_0^1 \frac{\gamma(\sigma)f(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma \mp \zeta_i)} \\ &\quad + E(\pm\zeta_i)p_{\alpha+\beta-1}(\pm\zeta_i), \quad i = 1, \cdots, \eta. \end{aligned} \tag{4.33}$$

Since $\gamma(\sigma) = \frac{1}{2}k(\sigma, \sigma)$, we see that $k(\sigma, \zeta_i)$ is a polynomial in σ which vanishes at ζ_i. We may then write

$$\frac{k(\sigma, \pm\zeta_i)}{\sigma \mp \zeta_i} = \sum_{j=0}^{\delta-1} w_{ij}(\zeta_i)(\pm 1)^{j+1}\sigma^j, \tag{4.34}$$

where the w_{ij} are polynomials and δ is the degree of σ in $k(\sigma, \nu)$. Denoting the moments of P by

$$m_j(P) = \int_0^1 \sigma^j P(\sigma)\,d\sigma, \tag{4.35}$$

and using the definition (4.21) of L, we find that (4.33) becomes

$$\begin{aligned} \frac{1}{2}\sum_{j=0}^{\delta-1} w_{ij}(\zeta_i)(\pm 1)^{j+1}m_j(P) &= \mp\frac{e^{-\tau/\zeta_i}}{\zeta_i}L(F)(-\zeta_i) \\ &\quad + E(\pm\zeta_i)p_{\alpha+\beta-1}(\pm\zeta_i). \end{aligned} \tag{4.36}$$

Equations (4.36) are 2η in number. From the results (3.4) and (3.5) we can show that, except in special cases,

$$\delta + \beta = 2\eta. \tag{4.37}$$

For a given value of c, there is at most one value b_{N-1} such that $2\eta < \delta + \beta$. In such a case, one must obtain additional equations for the constants. This is done, for example, by expanding (4.19) in a Laurent series about infinity and equating the coefficients of z^n. Or one may approach the special value of b_{N-1} as a limit of a sequence of cases where (4.37) holds.

In general, therefore, the δ quantities $m_j(P)$ may be eliminated from the $\delta + \beta$ equations (4.36), leaving β algebraic equations for the a_i. These equations combined with the α equations (4.30) will uniquely determine (at least for the subcritical reactor) the constants $a_i (i = 0, \cdots, \alpha + \beta - 1)$. We write these equations in matrix notation,

$$\mathbf{M}(\gamma, c, b_1, \cdots, b_N)\mathbf{a} = \mathbf{v}(x, \mu, \gamma). \tag{4.38}$$

(We suppress the dependence of the vector $\mathbf{v}$ upon $c, b_1, \cdots, b_N$).

To gain further insight into the problem, we choose to solve (4.38) by expanding $\mathbf{v}$ in terms of the eigenvectors of $\mathbf{M}$. For simplicity, we will assume that $\mathbf{M}$ has $\alpha + \beta$ distinct eigenvalues so that this expansion is possible. However, the following results for the critical thickness and flux hold even if $\mathbf{M}$ has eigenvalues with multiplicity greater than one. We therefore proceed by writing

$$\mathbf{v} = \sum_{i=0}^{\alpha+\beta-1} \tilde{v}_i(x, \mu, \gamma)\xi_i(\tau, c, b_1, \cdots, b_N), \tag{4.39}$$

where the ξ_i are the normalized eigenvectors of $\mathbf{M}$ which solve the following eigenvalue problem:

$$\mathbf{M}\xi_i = \epsilon_i\xi_i. \tag{4.40}$$

We choose the eigenvalues ϵ_i to be so arranged that

$$|\epsilon_{i+1}| \geq |\epsilon_i|. \tag{4.41}$$

The $\tilde{v}_i$ are then given by

$$\tilde{v}_i = (\mathbf{v}, \xi_i^\dagger). \tag{4.42}$$

The "$\dagger$" denotes eigenvectors of the Hermitian conjugate matrix which are defined as follows:

$$\mathbf{M}^\dagger\xi_i^\dagger = \epsilon_i^\dagger\xi_i^\dagger. \tag{4.43}$$

Using the expansion (4.39), we write the solution to (4.37) as

$$\mathbf{a} = \sum_{i=0}^{\alpha+\beta-1} \frac{\tilde{v}_i(x, \mu, \gamma)}{\epsilon_i(\tau_i, c, b_1, \cdots, b_N)} \xi_i(\tau, c, b_1, \cdots, b_N). \tag{4.44}$$

This is well defined since we have been restricting our attention to singular equations among whose solutions is the bounded analytic function J which satisfies (2.8) for scattering in a subcritical slab. This combined with (4.22) and (4.23) completes the determination of F.

From (3.6) and (3.21) we can easily show that a representation of P which is analytic in $0 < |z| < \infty$ is given by

$$P(z) = [F(z) + \pi i z\phi(z)]/\lambda(z), \tag{4.45}$$

with ϕ as defined by (4.4), and $R = P$. We have suppressed the parameters x and μ. The numerator has been so chosen as to vanish at the zeros of λ, and Plemelj's formulas readily show (4.45) to be analytic across the cut $[-1, 1]$. By using (4.19) as the computation of ϕ and the identity (4.26), we can express P in terms of F, for z outside $[-1, 0]$, by

$$P(z) = \frac{E(z)}{\lambda(z)}\left\{F(z)(zq_{\alpha-1}(z) - 1/E(0)) + zp_{\alpha+\beta-1}(z) + z\int_0^1 \frac{\gamma(\sigma)[F(\sigma) - F(z)]}{\lambda^-(\sigma)E^+(\sigma)(\sigma - z)}\, d\sigma\right\}. \tag{4.46}$$

A similar expression can be obtained for the Q function with a different F function and a different polynomial $p_{\alpha+\beta-1}$. This has reduced the scattering problem to the computation of F by the solving of the Fredholm equations (4.22) and the linear algebraic equations (4.38).

As is standard in the Wiener–Hopf analysis[6] we observe that, by (4.8), $\lambda(z)/E(z)E(-z)$ is bounded and analytic in the finite plane with polynomial growth at infinity, and hence a polynomial. If $\lambda(\infty) \neq 0$, we equate leading terms and zeros to write

$$\frac{\lambda(z)}{E(z)E(-z)} = \lambda(\infty)\prod_{i=1}^{\alpha}(z_i^2 - z^2). \tag{4.47}$$

By evaluating this at $z = 0$, we find an expression which is valid even when $\lambda(\infty) = 0$, namely

$$\frac{E(z)}{\lambda(z)} = \frac{\exp\left[\int_0^1 \frac{(t+2z)\theta(t)}{t(t+z)}\, dt\right]}{\prod_{i=1}^{\alpha}(1 - z/z_i)^2}. \tag{4.48}$$

For computational purposes this is to be used in (4.46).

We wish now to extract criticality conditions from our solution for a subcritical slab. Since J, and hence P, must become infinite as criticality is reached, we see from (4.46) that this can only happen when F becomes infinite. For a given set of the constants $c, b_1, \cdots, b_N$ this could conceivably happen in either of two ways. It may happen that the Fredholm equations (4.22) have solutions which become infinite as τ approaches certain critical values—we have only shown this is impossible if $\gamma(\nu) > 0$ for $0 \leq \nu \leq 1$. This clearly will not be true if $c - 1$ is positive and sufficiently small, for then the critical thickness will be large, and (4.24) will be satisfied for τ around the critical value.

We will restrict our attention to those cases where the Fredholm equations (4.22) have bounded solutions for all τ of interest. Then F, as given by (4.23), can become infinite only by having the coefficients a_i become infinite. By (4.44), this means that the eigenvalue ϵ_0 of $\mathbf{M}$ of minimum modulus approaches zero as criticality is attained. To avoid the eigenvalue analysis of $\mathbf{M}$, we can define the

critical dimension τ_c as that minimum τ such that the determinant of $\mathbf{M}$ vanishes,

$$\det \mathbf{M}(\tau, c, b_1, \cdots, b_N) = 0. \tag{4.49}$$

From (4.23), (4.44), and (4.46), we also obtain the result that

$$\lim_{\tau \to \tau_c} \epsilon_0(\tau, c, b_1, \cdots, b_N) P(x, \mu, z) = C(z, \tau_c, b_1, \cdots, b_N) \tilde{v}_0(x, \mu, \tau_c), \tag{4.50}$$

where C is a function of z which is not identically zero. From (2.7) we see that

$$P(x, \mu, z) = c\Lambda(P)(x, \mu, z) + f(\mu, z) \exp(-x/z) + f(-\mu, z) \exp[(x - \tau)/z]. \tag{4.51}$$

We multiply this by $\epsilon_0(\tau, c, b_1, \cdots, b_N)$ and let τ approach τ_c to find

$$\tilde{v}_0 = c\Lambda(\tilde{v}_0), \quad \text{for} \quad \tau = \tau_c. \tag{4.52}$$

The function $C(z)$ can be canceled as a constant of proportionality. This shows that Eq. (4.38) contains both the criticality condition and the neutron source function $\tilde{v}_0$ for a critical reactor.

For the spectral analysis of Λ for a given $b_1, \cdots, b_N$ and τ, one determines all values of c for which (4.49) holds. An equivalent analysis which is easier computationally is to fix c and $b_1, \cdots, b_N$ and search for all $\tau_i > \tau_c$ such that

$$\det \mathbf{M}(\tau_i, c, b_1, \cdots, b_N) = 0 \quad i = 1, \cdots. \tag{4.53}$$

This defines the condition for existence of higher-order *even* eigenfunctions of Λ for each of the τ_i. Again, by the ordering (4.41) of the eigenvalues of $\mathbf{M}$, the corresponding eigenfunction is given by $v_0(x, \mu, \tau_i)$. By a trivial modification of the above results, we may obtain the equations for $Q(x, \mu, z)$ which would then lead to a different set of τ_i for which odd eigenfunctions of Λ exist.

Finally, we note that the above results should yield a method for solving the homogeneous Milne problem for $c > 1$. That is, if for any subsequence (i_j) the limit exists in

$$\rho(x, \mu) = \lim_{j \to \infty} v_0(x, \mu, \tau_{i_j}), \tag{4.54}$$

then ρ solves the equation

$$\rho = c\Lambda(\rho), \qquad \tau = \infty, \tag{4.55}$$

and, hence, is the solution to the Milne problem. Further discussion of this procedure appears in the next section.

We will now apply the results of this section to specific cases, namely linear anisotropic scattering and isotropic scattering.

5. EXAMPLES IN SLAB GEOMETRY

A. Linear Anisotropic Scattering

In this section we will study the criticality problem assuming that the scattering function is given by

$$f(\mu, \mu') = c[1 + b_1 \mu \mu']. \tag{5.1}$$

Demanding positivity of f leads easily to the following inequality for b_1:

$$|b_1| < 1. \tag{5.2}$$

Using the results of Sec. 3, we find that an elementary calculation for k and γ gives

$$\begin{aligned} k(\sigma, \nu) &= c[1 - (c - 1)b_1 \sigma \nu], \\ \gamma(\sigma) &= \tfrac{1}{2} k(\sigma, \sigma) \\ &= \tfrac{1}{2}(c - 1)cb_1[1/(c - 1)b_1 - \sigma^2]. \end{aligned} \tag{5.3}$$

The above equation immediately reveals the zeros of γ as

$$\pm\zeta_0 = \pm[(c - 1)b_1]^{-\frac{1}{2}}. \tag{5.4}$$

The function λ may be calculated from (3.7). The result is

$$\lambda(z) = 1 - c[z - (c - 1)b_1 z^3] \tanh^{-1}(1/z) - (c - 1)cb_1 z^2. \tag{5.5}$$

It is easily verified that, if either $c < 1 + 1/b_1$ and $1 > b_1 > 0$, or $-1 < b_1 \leq 0$, then λ has only one pair of zeros, $\pm z_0$, in the complex plane cut from -1 to 1. They are pure imaginary for $c > 1$ and real for $c < 1$. In this case, $\alpha = 1$ and $\alpha + \beta = 2$ so that two arbitrary constants a_0, a_1 arise. By combining (4.30) and (4.36) Eqs. (4.32) for a_0 and a_1 may be written as follows:

$$\mathbf{M}(\tau, c, b_1, \cdots, b_N)\mathbf{a} = \mathbf{v}(x, \mu, \tau), \tag{5.6}$$

where

$$\begin{aligned} M_{11} &= z_0 E(z_0) - z_0 e^{-\tau/z_0} E(-z_0) - L(g_0)(z_0) - e^{-\tau/z_0} L(g_0)(-z_0), \\ M_{12} &= z_0^2 E(z_0) + z_0^2 e^{-\tau/z_0} E(-z_0) - L(g_1)(z_0) - e^{-\tau/z_0} L(g_1)(-z_0), \\ M_{21} &= \zeta_0 E(\zeta_0) + \zeta_0 E(-\zeta_0) + e^{\tau/\zeta_0} L(g_0)(\zeta_0) - e^{-\tau/\zeta_0} L(g_0)(-\zeta_0), \\ M_{22} &= \zeta_0^2 E(\zeta_0) - \zeta_0^2 E(-\zeta_0) + e^{\tau/\zeta_0} L(g_1)(\zeta_0) - e^{-\tau/\zeta_0} L(g_1)(-\zeta_0), \end{aligned} \tag{5.7}$$

and

$$v_1 = l_1(x, \mu, z_0) - L(\tilde{g})(x, \mu, z_0) - e^{-\tau/z_0}L(\tilde{g})(x, \mu, -z_0),$$
$$v_2 = -e^{\tau/\zeta_0}L(\tilde{g})(x, \mu, \zeta_0) + e^{-\tau/\zeta_0}L(\tilde{g})(x, \mu, -\zeta_0). \quad (5.8)$$

The functions $\tilde{g}$, g_0, and g_1 are computed from (4.22). The quantities $L(g_i)$ and $L(\tilde{g})$ are then computed by (4.21).

To find the critical thickness, we let τ increase, with c and b_1 fixed, until

$$\det \mathbf{M}(\tau_c, c, b_1) = 0. \quad (5.9)$$

From (4.50) the corresponding critical source function is given by

$$J(x, \mu) = M_{21}v_1(x, \mu, \tau_c) - M_{11}v_2(x, \mu, \tau_c). \quad (5.10)$$

Since the determinant of M depends on the functions $\tilde{g}$, g_1, and g_0, the solution of (5.9) for τ_c implies a search for τ_c by repeated calculations of the g_0, g_1 functions for various values of τ.

Many physically interesting cases are in the range $c - 1 \ll 1$. In this case, τ_c is large and approaches infinity as c decreases to 1. Then from (4.22) we approximate g_0 and g_1 by 0 to obtain an approximation for τ_c from (5.9), as

$$\left[1 - \frac{z_0}{\zeta_0}\frac{E(\zeta_0) + E(-\zeta_0)}{E(\zeta_0) - E(-\zeta_0)}\right]E(z_0) = \left[1 + \frac{z_0}{\zeta_0}\frac{E(\zeta_0) + E(-\zeta_0)}{E(\zeta_0) - E(-\zeta_0)}\right]E(-z_0)\exp(-\tau_c/z_0). \quad (5.11)$$

The flux, which is derived from the source function J by integration, can be shown to reduce to

$$\rho(x, \tau_c) \equiv \int_{-1}^{1} J(x, \mu, \tau_c)\, d\mu = A\cos\left(\frac{\tau_c - 2x}{2|z_0|}\right), \quad (5.12)$$

for some constant A.

The quantities appearing in (5.11) can be related to the solution of the Milne problem. For fixed c and b_1, we readily see that there is a sequence $\{\tau_i\}$ tending to infinity for which

$$\det \mathbf{M}(\tau_i, c, b_1) = 0. \quad (5.13)$$

Also we conclude from (5.7) that, even though z_0 is pure imaginary,

$$\lim_{i\to\infty} \exp(-\tau_i/z_0) = \frac{E(z_0)}{E(-z_0)} \times \frac{\zeta_0[E(\zeta_0) - E(-\zeta_0)] - z_0[E(\zeta_0) + E(-\zeta_0)]}{\zeta_0[E(\zeta_0) - E(-\zeta_0)] + z_0[E(\zeta_0) + E(-\zeta_0)]}. \quad (5.14)$$

For each such τ_i, we have an eigenfunction $J(x, \mu, \tau_i)$ of Λ given by (5.10) with τ_c replaced by τ_i. We see then that

$$\lim_{i\to\infty} J(x, \mu, \tau_i) = \lim_{i\to\infty}[\zeta_0 E(\zeta_0) + \zeta_0 E(-\zeta_0)] \times [k(\mu, z_0)\exp(-x/z) + k(-\mu, z_0) \times \exp[(x - \tau_i)/z_0]] + O[\exp(-x)]. \quad (5.15)$$

This function, part of which we have merely abbreviated by $O[\exp(-x)]$, gives a solution to the Milne problem.

Defining ρ_i by

$$\rho_i(x) = \int_{-1}^{1} J(x, \mu, \tau_i)\, d\mu, \quad (5.16)$$

we see that

$$\lim_{i\to\infty} \rho_i(x) = A[\exp(-x/z_0) + \exp(x/z_0) \times \lim_{i\to\infty} \exp(-\tau_i/z_0)] + O[\exp(-x)]. \quad (5.17)$$

If we neglect the term $O[\exp(-x)]$, we see that the asymptotic part of (5.17) leads naturally to the concept of an extrapolation distance τ_a, at which the asymptotic solution vanishes. Then from (5.17) and (5.14), we can relate τ_a and τ_c by

$$\exp(-\tau_c/z_0) = -\exp(2\tau_a/z_0), \quad (5.18)$$

or

$$\tau_c = \pi|z_0| - 2\tau_a. \quad (5.19)$$

Higher-order approximations to the criticality condition may be achieved by iterating the Fredholm equations (4.22) and taking these iterates into account in the determinant of $\mathbf{M}$.

If $c > 1 + 1/b_1$ and $1 > b_1 > 0$, the function λ has two pairs of zeros in the cut plane.[3] Both pairs are pure imaginary if $cb_1 > 3$, and one pair is imaginary and the other is real if $cb_1 < 3$. This extra pair of zeros increases α to two and increases the size of the matrix $\mathbf{M}$ to three by three. The additional required matrix elements may be easily found. Since in these cases c is restricted by $c > 2$, the critical dimension will not be large.

In the case $cb_1 < 3$, it is again possible to find a solution to the Milne problem as the limit of eigenfunctions for a sequence of values of τ. However, in the case $cb_1 > 3$ it is not clear that this can be done. The trouble is due to the fact that the determinant of $\mathbf{M}$ as well as the eigenfunctions contain two oscillatory functions of τ, $\exp(-\tau/z_0)$ and $\exp(-\tau/z_1)$ for pure imaginary numbers z_0 and z_1,

and it is not clear that the $J(x, \mu, \tau_i)$ converge as the τ_i tend to infinity.

B. Isotropic Scattering

We now specialize to the case where

$$f(\mu, \mu') = c. \tag{5.20}$$

The functions k, γ, and λ are now simply given by

$$k(\sigma, \nu) = c, \qquad \gamma(\sigma) = \tfrac{1}{2}c, \qquad \lambda(z) = 1 - cz \tanh^{-1}(1/z). \tag{5.21}$$

The matrix $\mathbf{M}$ degenerates into a single element and we find that

$$a_0 = \frac{l_1(x, z_0) - L(\tilde{g})(x, z_0) - \exp(-\tau/z_0)L(\tilde{g})(x, -z_0)}{z_0E(z_0) - z_0 \exp(-\tau/z_0)E(-z_0) - L(g_0)(z_0) - \exp(-\tau/z_0)L(g_0)(-z_0)}. \tag{5.22}$$

We can immediately write down the criticality condition[14]

$$z_0E(z_0) - z_0e^{-\tau_c/z_0}E(-z_0) - L(g_0)(z_0) - \exp(-\tau_c/z_0)L(g_0)(-z_0) = 0, \tag{5.23}$$

and the flux

$$\rho(x) = l_1(x, z_0) - L(\tilde{g})(x, z_0) - \exp(-\tau/z_0)L(\tilde{g})(x, -z_0). \tag{5.24}$$

As before, the ratio $E(z_0)/E(-z_0)$ may be related to the extrapolation distance for the Milne problem. If this is done we may rewrite the criticality condition (5.23) as

$$\frac{1 + \frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + z_0)}}{1 + \frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + z_0)}} = -\exp[-(\tau + 2\tau_a)/z_0]. \tag{5.25}$$

Since both sides of this equation are complex numbers on the unit circle, we may equate arguments to get

$$\tan^{-1}\left[\frac{-|z_0|\frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}{1 + \frac{c}{2}\int_0^1 \frac{\sigma g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}\right] = \frac{1}{2}\left[\pi - \frac{(\tau + 2\tau_a)}{|z_0|}\right]. \tag{5.26}$$

From Sec. 4, the function g_0 is the solution to the following Fredholm integral equation:

$$g_0(\nu) = -\nu \exp(-\tau/\nu)E(-\nu) \times \left[1 - \frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + \nu)}\right]. \tag{5.27}$$

Since γ is positive ($=\frac{1}{2}c$), we see by the results of Sec. 4 that the following iterative scheme for approximating g_0 will converge *uniformly to the solution for any c and* τ:

$$g_0^{(0)}(\nu) = 0,$$
$$g_0^{(n+1)}(\nu) = -\nu \exp(-\tau/\nu)E(-\nu) \times \left[1 - \frac{c}{2}\int_0^1 \frac{g_0^{(n)}(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma + \nu)}\right]. \tag{5.28}$$

Hence, using the above precedure, the critical slab thickness for the entire range of values of $c(1 < c < \infty)$ may be determined. An analogous iteration scheme may be written for $\tilde{g}(x, \nu)$.

Mitsis[4] has also formulated the slab criticality problem for isotropic scattering using Case's singular eigenfunction expansion. As expected, his resulting Fredholm kernel is adjoint to the kernel in Eq. (5.27). Corresponding iterations of the two adjoint equations, however, give precisely the same results for the neutron flux and critical thickness.

6. SPHERICAL REACTOR: ISOTROPIC SCATTERING

The neutron density $\rho_0(x)$, in a homogeneous critical sphere with isotropic fission production and scattering, satisfies the equation

$$x\rho_0(x) = \frac{c}{2}\int_{-\tau/2}^{+\tau/2} y\rho_0(y) \times \left[\int_0^1 \frac{\exp(-|x - y|/\mu)}{\mu}\,d\mu\right] dy, \tag{6.1}$$

where τ is the sphere diameter.[1] With the substitution

$$\rho(x) = (x - \tfrac{1}{2}\tau)\rho_0(x - \tfrac{1}{2}\tau), \tag{6.2}$$

Eq. (6.1) becomes identical to the integral equation for the neutron density in a slab reactor of thickness τ. However, from (6.2) we have the following symmetry condition on ρ:

$$\rho(x) = -\rho(\tau - x). \tag{6.3}$$

We require therefore the first odd eigenfunction of Λ.

[14] A. Leonard, and T. W. Mullikin, The RAND Corporation, RM-3256-PR (1962).

According to the discussion in Sec. 4, the odd eigenfunctions are found by considering Eq. (4.1) for Q. By a few simple sign changes in the analysis of Sec. 5B, we find that the constant a_0 is given by

$$a_0 = \frac{-l_2(x, z_0) + \exp(-\tau/z_0)L(\tilde{g})(x, -z_0) - L(\tilde{g})(x, z_0)}{z_0 \exp(-\tau/z_0)E(-z_0) + z_0E(z_0) + L(g_0)(z_0) - \exp(-\tau/z_0)L(g_0)(-z_0)}, \tag{6.4}$$

with the functions $\tilde{g}$ and $\tilde{g}_0$ now given by the solution to the following Fredholm equations:

$$\tilde{g}(x, \nu) = l_2(x, \nu) + L(\tilde{g})(x, \nu), \tag{6.5}$$

$$g_0(\nu) = \nu \exp(-\tau/\nu)E(-\nu) + L(g_0)(\nu). \tag{6.6}$$

Equation (6.4) leads to the following *sphere criticality condition*[14]:

$$\tan^{-1}\left[\frac{|z_0|\frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}{1 + \frac{c}{2}\int_0^1 \frac{\sigma g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}\right] = \left[\frac{\tau + 2\tau_a}{2|z_0|} - \pi\right]. \tag{6.7}$$

Again iterative schemes completely analogous to (5.26) may be defined which converge uniformly to the solution of (6.5) and (6.6) for every c and τ.

In the criticality condition for the sphere (6.7) and the slab (5.26), the multiple-valued function "arctan" appears. The multiple roots of these equations, τ_i, correspond to the τ_i of (4.53) and, therefore, define the existence of higher-order eigenfunctions of the operator Λ for isotropic scattering.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Some Exact Radial Integrals for Dirac–Coulomb Functions*

J. T. Reynolds, D. S. Onley, and L. C. Biedenharn
Duke University, Durham, North Carolina
(Received 14 August 1963)

The zero energy loss Dirac–Coulomb integrals are evaluated using the technique of contour integration. The expressions obtained have a closed analytic form, showing that these integrals are formally similar to the corresponding classical and nonrelativistic quantum mechanical, zero energy loss integrals which also have exact elementary solutions.

Application of the zero energy loss Dirac–Coulomb integrals occurs in inelastic electron scattering and similar problems. The investigation of the finite energy loss Dirac–Coulomb integrals requires a study of the zero energy loss integrals as a preliminary.

I. INTRODUCTION

By using the technique of multipole field expansion, the treatment of problems involving energy transfer from charged particles in the Coulomb field, in cases for which retardation effects are negligible, can often be reduced essentially to the evaluation of integrals containing, in the integrand, factors of the form, $r^{-L-1}Y_L^M$. Such is the case, for example, in the processes of Coulomb excitation, bremsstrahlung, pair production, etc. The evaluation of these integrals is not a completely straightforward task since the integrals with which we are concerned are more general than the Sonine–Schafheitlin integral[1] and in a similar way display discontinuous behavior in the zero energy loss limit.

The treatment of such integrals in the classical[2–5] and nonrelativistic quantum mechanical[2,3,6,7] cases has been discussed in several places. The authors were led to the problem of extending this work to the relativistic case during the course of calculating, accurately, the differential cross section for

* Supported in part by the Army Research Office (Durham), and the National Science Foundation.

[1] W. Magnus and F. Oberhettinger, *Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), p. 35.

[2] L. C. Biedenharn and P. J. Brussaard, *Coulomb Excitation* (to be published by Oxford University Press).

[3] K. Alder, A. Bohr, T. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).

[4] K. Alder and A. Winther, Kgl. Danske Videnskab. Selskab, Mat. Fys. Medd. **31**, No. 1 (1956).

[5] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields* (Pergamon Press, Ltd., London, 1959).

[6] L. C. Biedenharn, J. L. McHale, and R. M. Thaler, Phys. Rev. **100**, 376 (1955).

[7] M. Goldstein, R. M. Thaler, and L. C. Biedenharn, Los Alamos Scientific Laboratory Rept. LA-2106 (1957) (unpublished).

According to the discussion in Sec. 4, the odd eigenfunctions are found by considering Eq. (4.1) for Q. By a few simple sign changes in the analysis of Sec. 5B, we find that the constant a_0 is given by

$$a_0 = \frac{-l_2(x, z_0) + \exp(-\tau/z_0)L(\tilde{g})(x, -z_0) - L(\tilde{g})(x, z_0)}{z_0 \exp(-\tau/z_0)E(-z_0) + z_0E(z_0) + L(g_0)(z_0) - \exp(-\tau/z_0)L(g_0)(-z_0)}, \tag{6.4}$$

with the functions $\tilde{g}$ and $\tilde{g}_0$ now given by the solution to the following Fredholm equations:

$$\tilde{g}(x, \nu) = l_2(x, \nu) + L(\tilde{g})(x, \nu), \tag{6.5}$$

$$g_0(\nu) = \nu \exp(-\tau/\nu)E(-\nu) + L(g_0)(\nu). \tag{6.6}$$

Equation (6.4) leads to the following *sphere criticality condition*[14]:

$$\tan^{-1}\left[\frac{|z_0|\frac{c}{2}\int_0^1 \frac{g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}{1 + \frac{c}{2}\int_0^1 \frac{\sigma g_0(\sigma)\,d\sigma}{\lambda^-(\sigma)E^+(\sigma)(\sigma^2 + |z_0|^2)}}\right] = \left[\frac{\tau + 2\tau_a}{2|z_0|} - \pi\right]. \tag{6.7}$$

Again iterative schemes completely analogous to (5.26) may be defined which converge uniformly to the solution of (6.5) and (6.6) for every c and τ.

In the criticality condition for the sphere (6.7) and the slab (5.26), the multiple-valued function "arctan" appears. The multiple roots of these equations, τ_i, correspond to the τ_i of (4.53) and, therefore, define the existence of higher-order eigenfunctions of the operator Λ for isotropic scattering.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

Some Exact Radial Integrals for Dirac–Coulomb Functions*

J. T. Reynolds, D. S. Onley, and L. C. Biedenharn
Duke University, Durham, North Carolina
(Received 14 August 1963)

The zero energy loss Dirac–Coulomb integrals are evaluated using the technique of contour integration. The expressions obtained have a closed analytic form, showing that these integrals are formally similar to the corresponding classical and nonrelativistic quantum mechanical, zero energy loss integrals which also have exact elementary solutions.

Application of the zero energy loss Dirac–Coulomb integrals occurs in inelastic electron scattering and similar problems. The investigation of the finite energy loss Dirac–Coulomb integrals requires a study of the zero energy loss integrals as a preliminary.

I. INTRODUCTION

By using the technique of multipole field expansion, the treatment of problems involving energy transfer from charged particles in the Coulomb field, in cases for which retardation effects are negligible, can often be reduced essentially to the evaluation of integrals containing, in the integrand, factors of the form, $r^{-L-1}Y_L^M$. Such is the case, for example, in the processes of Coulomb excitation, bremsstrahlung, pair production, etc. The evaluation of these integrals is not a completely straightforward task since the integrals with which we are concerned are more general than the Sonine–Schafheitlin integral[1] and in a similar way display discontinuous behavior in the zero energy loss limit.

The treatment of such integrals in the classical[2–5] and nonrelativistic quantum mechanical[2,3,6,7] cases has been discussed in several places. The authors were led to the problem of extending this work to the relativistic case during the course of calculating, accurately, the differential cross section for

* Supported in part by the Army Research Office (Durham), and the National Science Foundation.

[1] W. Magnus and F. Oberhettinger, *Special Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1949), p. 35.

[2] L. C. Biedenharn and P. J. Brussaard, *Coulomb Excitation* (to be published by Oxford University Press).

[3] K. Alder, A. Bohr, T. Huus, B. Mottelson, and A. Winther, Rev. Mod. Phys. **28**, 432 (1956).

[4] K. Alder and A. Winther, Kgl. Danske Videnskab. Selskab, Mat. Fys. Medd. **31**, No. 1 (1956).

[5] L. D. Landau and E. M. Lifschitz, *The Classical Theory of Fields* (Pergamon Press, Ltd., London, 1959).

[6] L. C. Biedenharn, J. L. McHale, and R. M. Thaler, Phys. Rev. **100**, 376 (1955).

[7] M. Goldstein, R. M. Thaler, and L. C. Biedenharn, Los Alamos Scientific Laboratory Rept. LA-2106 (1957) (unpublished).

the inelastic scattering of high-energy electrons by nuclei.[8]

The existence of relatively simple forms for the exact integrals for arbitrary energy transfer in the classical and nonrelativistic cases is a consequence of the unusual and special symmetry of the nonrelativistic Coulomb field.[2] Since the relativistic Kepler problem is characterized by a group with fewer symmetry operators than the nonrelativistic problem, it is to be expected that the integrals will be considerably more difficult to handle. It is the essential point of the present paper to demonstrate—by construction—that the zero energy loss Dirac–Coulomb integrals are formally similar to the classical and nonrelativistic zero energy loss integrals and the entire system of integrals for this case have exact elementary solutions.

The fact that there exists a simple, exact result for the zero energy loss Dirac–Coulomb integrals is of some intrinsic interest. Such a result, for example, necessarily contains all previous similar results by taking the appropriate limits $\hbar \to 0$ and /or $c \to \infty$. Also, the possibility of close examination of such limiting processes is one of the great attractions connected with "exact" solutions.

Since the classical and nonrelativistic integrals are intimately connected with the present results we shall summarize previous results for convenience in Secs. II and III. Section IV is devoted to the derivation of the explicit forms for the zero energy loss Dirac–Coulomb integrals. In this derivation different initial and final energies are carried as far as possible to indicate exactly where the zero energy loss limit must be taken and to provide a starting point for the investigation of the effect of such energy loss. A short discussion of the calculations that have been made is presented in Sec. V.

II. CLASSICAL ZERO ENERGY LOSS INTEGRALS

In the classical zero energy loss case the relevant integrals that occur, for example, in Coulomb excitation, are of the form[2,3]

$$I_{L,M}(\theta) \equiv (\epsilon^2 - 1)^{-L+\frac{1}{2}} \int_{-\varphi_0}^{\varphi_0} e^{iM\varphi}(\epsilon \cos\varphi - 1)^{L-1}\, d\varphi$$
$$= (2\pi)^{\frac{1}{2}}(L-1)!\, \epsilon^{-\frac{1}{2}}(\epsilon^2-1)^{-(L/2)+\frac{1}{4}} P^{-L+\frac{1}{2}}_{M-\frac{1}{2}}(1/\epsilon), \quad (1)$$

where ϵ and φ_0 are defined in terms of the deflection angle, θ, as $\epsilon = 1/\sin\frac{1}{2}\theta$ and $\varphi_0 = \frac{1}{2}(\pi - \theta)$. These integrals come directly from the function, $r^{-L-1}Y^M_L$, of the projectile coordinates as measured from the nuclear center of mass, integrated along the Kepler orbit of the projectile.

The essential simplicity of this result can be seen from the fact that the integrand is essentially a polynomial in cos φ and sin φ. The integral is easily evaluated for any multipolarity, L, and for the lowest multipole orders, one obtains the following explicit expressions:

$$\begin{aligned}
I_{1,\pm1}(\theta) &= 2\sin\tfrac{1}{2}\theta,\\
I_{2,\pm2}(\theta) &= \tfrac{2}{3}\sin^2\tfrac{1}{2}\theta,\\
I_{2,0}(\theta) &= 2\tan^2\tfrac{1}{2}\theta[1 - \tfrac{1}{2}(\pi-\theta)\tan\tfrac{1}{2}\theta],\\
I_{3,\pm3}(\theta) &= \tfrac{4}{15}\sin^3\tfrac{1}{2}\theta,\\
I_{3,\pm1}(\theta) &= 2(\sin^3\tfrac{1}{2}\theta/\cos^4\tfrac{1}{2}\theta)\\
&\quad\times\left[\frac{2+\sin^2\frac{1}{2}\theta}{3} - \frac{(\pi-\theta)}{2}\tan\tfrac{1}{2}\theta\right].
\end{aligned} \quad (2)$$

III. NONRELATIVISTIC QUANTUM-MECHANICAL ZERO ENERGY LOSS COULOMB INTEGRALS

The usual nonrelativistic Coulomb wavefunctions for a particle of charge Z_1 and spin $\frac{1}{2}$ in the field of a nucleus with charge Z_2 are given by[2,9]

$$|\eta,\ \kappa,\ \mu\rangle = (kr)^{-1}F_{\eta,l(\kappa)}(kr)\chi^\mu_\kappa, \quad (3)$$

where κ is the Dirac quantum number, k is the wave number of the projectile and $\eta \equiv Z_1Z_2e^2/\hbar v$. The spin-angle function, χ^μ_κ, is defined by

$$\chi^\mu_\kappa \equiv \sum_\tau \langle l\ \mu-\tau\ \tfrac{1}{2}\tau \mid j\mu\rangle Y^{\mu-\tau}_l(\theta,\varphi)\chi^\tau_{\frac{1}{2}}, \quad (4)$$

and $F_{\eta,l}$ is defined in terms of the confluent hypergeometric function by

$$F_{\eta,l}(kr) = \frac{2^l e^{-\frac{1}{2}\pi\eta}\,|\Gamma(l+1+i\eta)|}{\Gamma(2l+2)}(kr)^{l+1}e^{-ikr} \times {}_1F_1(l+1-i\eta,\ 2l+2;\ 2ikr). \quad (5)$$

The relevant integrals that occur reduce to radial integrals of the form

$$\langle \eta,\ l'\ |r^{-L-1}|\ \eta,\ l\rangle \equiv \int_0^\infty F_{\eta,l'}(kr)F_{\eta,l}(kr)r^{-L-1}\, dr. \quad (6)$$

Expressions for these integrals exist in closed form and are surprisingly simple. Thus, for example, one obtains[3]

$$\langle \eta,\ l+L\ |r^{-L-1}|\ \eta,\ l\rangle = \langle \eta,\ l\ |r^{-L-1}|\ \eta,\ l+L\rangle = 2^{L-2}k^L\frac{[\Gamma(L)]^2}{\Gamma(2L)}\left|\frac{\Gamma(l+1+i\eta)}{\Gamma(l+L+1+i\eta)}\right|. \quad (7)$$

[8] T. A. Griffy, D. S. Onley, J. T. Reynolds, and L. C. Biedenharn, Phys. Rev. **128**, 833 (1962); D. S. Onley, T. A. Griffy, and J. T. Reynolds, *ibid.* **129**, 1689 (1963).

[9] G. Breit and M. H. Hull in *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1959), Vol. 41.

The integrals are most easily obtained by means of recursion relations to be discussed below. For the lowest multipole orders, one obtains the following expressions[3]:

$$\langle \eta, l+1 \,|r^{-2}|\, \eta, l\rangle = \langle \eta, l \,|r^{-2}|\, \eta, l+1\rangle = \frac{k}{2}\frac{1}{|l+1+i\eta|},$$

$$\langle \eta, l+2 \,|r^{-3}|\, \eta, l\rangle = \langle \eta, l \,|r^{-3}|\, \eta, l+2\rangle = \frac{k^2}{6}\frac{1}{|l+1+i\eta|\,|l+2+i\eta|},$$

$$\langle \eta, l \,|r^{-3}|\, \eta, l\rangle = \frac{k^2}{2l(l+1)(2l+1)} \times [2l+1-\pi\eta+2\eta \operatorname{Im}\Psi(l+1+i\eta)], \quad (8)$$

$$\langle \eta, l+3 \,|r^{-4}|\, \eta, l\rangle = \langle \eta, l \,|r^{-4}|\, \eta, l+3\rangle = \frac{k^3}{15}\frac{1}{|l+1+i\eta|\,|l+2+i\eta|\,|l+3+i\eta|},$$

$$\langle \eta, l+1 \,|r^{-4}|\, \eta, l\rangle = \langle \eta, l \,|r^{-4}|\, \eta, l+1\rangle = \frac{k^3}{3l(l+1)(l+2)(2l+1)(2l+3)\,|l+1+i\eta|} \times \{3\,|l+1+i\eta|^2\,[2l+1-\pi\eta+2\eta \times \operatorname{Im}\Psi(l+1+i\eta)] - l(l+1)(2l+1)\},$$

where $\Psi(z)$ is the logarithmic derivative of the gamma function. Unlike the classical case, for which Eq. (1) is general, no general result for

$$\langle \eta, l+M \,|r^{-L-1}|\, \eta, l\rangle$$

(with *arbitrary* L, M) has been given, although it surely exists.

The symmetry properties of the Coulomb field may be utilized to derive all of the preceding results by operator techniques.[2,10] Since these techniques may serve as a guide for the application of similar techniques to the relativistic problem, they will be discussed briefly here. It may be shown directly that the Lenz–Runge vector $\mathbf{A}$, defined by

$$\mathbf{A} \equiv Z_1Z_2e^2m\hbar^{-1}\,|2mH|^{-\frac{1}{2}} \times [\hat{\mathbf{r}} - \hbar(2Z_1Z_2e^2m)^{-1}(\mathbf{L}\times\mathbf{p} - \mathbf{p}\times\mathbf{L})], \quad (9)$$

is a constant of motion, where H, $\mathbf{L}$, and $\mathbf{p}$ are the Hamiltonian, the orbital angular momentum, and the linear momentum operators for the projectile. This operator has the property that

$$\boldsymbol{\sigma}\cdot\mathbf{A}\,|\eta, \kappa, \mu\rangle = (-1)(\kappa^2+\eta^2)^{\frac{1}{2}}\,|\eta, -\kappa, \mu\rangle, \quad (10)$$

where $(\frac{1}{2}\hbar)\boldsymbol{\sigma}$ is the spin operator, which implies

$$\left[1 + \frac{\kappa\hbar^2}{Z_1Z_2e^2m}\left(\frac{d}{dr} + \frac{\kappa+1}{r}\right)\right]\frac{F_{\eta,l(\kappa)}(kr)}{r} = \left(1+\frac{\kappa^2}{\eta^2}\right)^{\frac{1}{2}}\frac{F_{\eta,l(-\kappa)}(kr)}{r}. \quad (11)$$

[10] We hope to return to this subject in a later paper.

This last equation defines a raising (κ negative) and lowering (κ positive) operator for l, and it is, in fact, *a concise, and complete expression for the recursion relations for the nonrelativistic Coulomb wavefunctions.* Further, one can solve Eq. (11) directly to obtain Eq. (5)—without appeal to analysis—by first solving the equivalent relation to (11) for the discrete wavefunctions and making a simple transformation back to the continuum case.[2]

One may also show that the operator $\boldsymbol{\sigma}\times\mathbf{L}$ has the property, when operating on $|\eta, \kappa, \mu\rangle$, of changing κ to $-\kappa-1$. Thus, by using suitable combinations of $\boldsymbol{\sigma}\times\mathbf{L}$ and $\boldsymbol{\sigma}\cdot\mathbf{A}$, one may find a relation between any two Coulomb wavefunctions. In addition, by deriving operator identities involving the operators $\boldsymbol{\sigma}\cdot\mathbf{A}$ and $\boldsymbol{\sigma}\times\mathbf{L}$ one obtains, by taking matrix elements, the recursion relations from which follow Eqs. (7) and (8). The details of this procedure will be presented elsewhere,[10] and we shall content ourselves in the present paper to a single example. Upon defining the operator η by

$$\tilde{\eta} \equiv Z_1Z_2e^2m\hbar^{-1}[2mH]^{-\frac{1}{2}}, \quad (12)$$

one can verify directly the identity

$$\begin{aligned}4(K+1)\tilde{\eta}^{-1}\boldsymbol{\sigma}\cdot\mathbf{A}\{&(\boldsymbol{\sigma}\times\mathbf{L})_z(\boldsymbol{\sigma}\times\mathbf{L})_z r^{-3}(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})\\ &\times(\boldsymbol{\sigma}\times\mathbf{L})_z(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})\tilde{\eta}^{-1}\boldsymbol{\sigma}\cdot\mathbf{A} - (\boldsymbol{\sigma}\times\mathbf{L})_z\tilde{\eta}^{-1}(\boldsymbol{\sigma}\cdot\mathbf{A})r^{-3}\\ &\times(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})(\boldsymbol{\sigma}\times\mathbf{L})_z(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})(\boldsymbol{\sigma}\times\mathbf{L})_z\} = (Z_1Z_2e^2m/\hbar^2)^2\\ &\times(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})(\boldsymbol{\sigma}\times\mathbf{L})_z(\boldsymbol{\sigma}\times\mathbf{L})_z(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})(\boldsymbol{\sigma}\times\mathbf{L})_z\\ &\times\{-2[\tilde{\eta}^{-2}, r^{-1}] + K[\tilde{\eta}^{-2}, [\tilde{\eta}^{-2}, r^{-1}]]\\ &- 2K^2[\tilde{\eta}^{-2}, r^{-1}]\tilde{\eta}^{-2} + 2\tilde{\eta}^{-1}(\boldsymbol{\sigma}\cdot\mathbf{A})\\ &\times[\tilde{\eta}^{-2}, r^{-1}]\tilde{\eta}^{-1}(\boldsymbol{\sigma}\cdot\mathbf{A})\}. \end{aligned} \quad (13)$$

Since $(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})$ has the effect, when operating on χ^{μ}_{κ}, of changing the sign of κ, taking the appropriate matrix element of (13) yields the difference equation

$$|l+3+i\eta|\,\langle \eta, l+3\,|r^{-3}|\,\eta, l+1\rangle - |l+1+i\eta|\,\langle \eta, l+2\,|r^{-3}|\,\eta, l\rangle = 0, \quad (14)$$

of which a solution, as shown in Eq. (8), is

$$\langle \eta, l+2\,|r^{-3}|\,\eta, l\rangle = k^2/(6\,|l+1+i\eta|\,|l+2+i\eta|). \quad (15)$$

As might be expected, there is a correspondence between the classical integrals, Eq. (1), and the nonrelativistic integrals, Eq. (6). In fact, it can be shown directly[6] that, by simultaneously taking the limits $\hbar\to 0$, l_i, l_f, $\eta\to\infty$, with $l_i - l_f$ remaining finite, there results the limit

$$\lim \langle \eta, l_f\,|r^{-L-1}|\,\eta, l_i\rangle = \tfrac{1}{4}(k/\eta)^L I_{L, l_f-l_i}(2\tan^{-1}\eta/l_i), \quad (16)$$

which also demonstrates the correspondence

$$\eta/l \rightleftarrows \tan \tfrac{1}{2}\theta.$$

IV. ZERO ENERGY LOSS DIRAC-COULOMB INTEGRALS

The well known spherical coordinate solutions of the Dirac equation for a projectile of mass m_0, charge $(-Z_1)$, and spin $\frac{1}{2}$ in the Coulomb field of a point nucleus of charge Z_2, are[11]

$$|\kappa\mu\rangle = \begin{bmatrix} F_\kappa(kr)\Phi^\mu_{-\kappa} \\ G_\kappa(kr)\Phi^\mu_\kappa \end{bmatrix}. \tag{17}$$

The index, κ, is the Dirac quantum number and the spin angle functions Φ^μ_κ are given explicitly by

$$\Phi^\mu_\kappa = \sum_\tau \langle l(\kappa)\, \mu-\tau\, \tfrac{1}{2}\tau \mid j(\kappa)\mu\rangle i^{l(\kappa)} Y^{\mu-\tau}_{l(\kappa)}\chi^\tau_{\frac{1}{2}}. \tag{18}$$

The radial functions are defined in terms of confluent hypergeometric functions to be (with $c = \hbar = 1$)

$$\begin{Bmatrix} F_\kappa(kr) \\ G_\kappa(kr) \end{Bmatrix} = \begin{Bmatrix} -S(\kappa)[(E - m_0)/(E + m_0)]^{\frac{1}{2}} \\ 1 \end{Bmatrix} \frac{(kr)^{\gamma-1}2^{\gamma-1}e^{\pi\eta/2}\,|\Gamma(\gamma + i\eta)|}{\Gamma(2\gamma + 1)}$$

$$\times \begin{Bmatrix} -i(\gamma + i\eta)e^{i\phi}e^{-ikr}\ {}_1F_1(\gamma + 1 + i\eta,\, 2\gamma + 1;\, 2ikr) + i(\gamma - i\eta)e^{-i\phi}e^{-ikr}\ {}_1F_1(\gamma + i\eta,\, 2\gamma + 1;\, 2ikr) \\ (\gamma + i\eta)e^{i\phi}e^{-ikr}\ {}_1F_1(\gamma + 1 + i\eta,\, 2\gamma + 1;\, 2ikr) + (\gamma - i\eta)e^{-i\phi}e^{-ikr}\ {}_1F_1(\gamma + i\eta,\, 2\gamma + 1;\, 2ikr) \end{Bmatrix}, \tag{19}$$

where E is the energy of the projectile, k is the wavenumber, $S(\kappa)$ is the sign of κ, and

$$\gamma = [\kappa^2 - (\alpha Z_1 Z_2)^2]^{\frac{1}{2}},$$

$$\eta = \alpha Z_1 Z_2 E/k, \tag{20}$$

$$e^{2i\phi(\kappa)} = e^{-i\pi}\left[\frac{\kappa - i(\alpha Z_1 Z_2 m_0/k)}{\gamma + i\eta}\right], \quad -\pi \leq \phi \leq 0.$$

α is the fine-structure constant.

In calculations of processes involving energy transfer in the field of the nucleus leading to a 2^L electric-multipole transition, the radial functions occur in integrals of the form

$$I(\kappa_1, L, \kappa_2) \equiv \int_0^\infty [S(\kappa_1)S(\kappa_2)F_{\kappa_1}(k_1r)F_{\kappa_2}(k_2r) + G_{\kappa_1}(k_1r)G_{\kappa_2}(k_2r)]r^{-L+1}\, dr, \tag{21}$$

where k_1 and k_2 are, respectively, the wavenumbers of the incoming and outgoing projectiles. (The limit $k_2 \rightarrow k_1$ will be taken in the course of the derivation.)

In order to transform these integrals into more useful forms, it is necessary to introduce the convergence factor, e^{-sr}, into the integrand where s is a small positive number. The limit $s \rightarrow 0$ will eventually be taken. Employing integral representations for both confluent hypergeometric functions,[12] changing the order of integration, and integrating over r yields

$$\int_0^\infty \frac{1}{r^{L-1}} \begin{Bmatrix} S(\kappa_1)S(\kappa_2)F_{\kappa_1}(k_1r)F_{\kappa_2}(k_2r) \\ G_{\kappa_1}(k_1r)G_{\kappa_2}(k_2r) \end{Bmatrix} e^{-sr}\, dr = C(\kappa_1, \kappa_2, L) \begin{Bmatrix} (E - m_0)/(E + m_0) \\ 1 \end{Bmatrix}$$

$$\times \left[\begin{Bmatrix} - \\ + \end{Bmatrix} (\gamma_1 + i\eta_1)(\gamma_2 + i\eta_2)e^{i(\phi_1+\phi_2)}F_2^t(\gamma_1 + \gamma_2 - L,\, \gamma_1 + 1 + i\eta_1,\, \gamma_2 + 1 + i\eta_2,\, 2\gamma_1 + 1,\, 2\gamma_2 + 1;\, x,\, y) \right.$$

$$+ \begin{Bmatrix} - \\ + \end{Bmatrix} (\gamma_1 - i\eta_1)(\gamma_2 - i\eta_2)e^{-i(\phi_1+\phi_2)}F_2^t(\gamma_1 + \gamma_2 - L,\, \gamma_1 + i\eta_1,\, \gamma_2 + i\eta_2,\, 2\gamma_1 + 1,\, 2\gamma_2 + 1;\, x,\, y)$$

$$+ (\gamma_1 + i\eta_1)(\gamma_2 - i\eta_2)e^{i(\phi_1-\phi_2)}F_2^t(\gamma_1 + \gamma_2 - L,\, \gamma_1 + 1 + i\eta_1,\, \gamma_2 + i\eta_2,\, 2\gamma_1 + 1,\, 2\gamma_2 + 1;\, x,\, y)$$

$$\left. + (\gamma_1 - i\eta_1)(\gamma_2 + i\eta_2)e^{-i(\phi_1-\phi_2)}F_2^t(\gamma_1 + \gamma_2 - L,\, \gamma_1 + i\eta_1,\, \gamma_2 + 1 + i\eta_2,\, 2\gamma_1 + 1,\, 2\gamma_2 + 1;\, x,\, y) \right], \tag{22}$$

where

$$C(\kappa_1, \kappa_2, L) = \frac{2^{\gamma_1+\gamma_2-2}e^{\pi(\eta_1+\eta_2)/2}\,|\Gamma(\gamma_1 + i\eta_1)\Gamma(\gamma_2 + i\eta_2)|\, k_1^{\gamma_1-1}k_2^{\gamma_2-1}\Gamma(\gamma_1 + \gamma_2 - L)(ik_1 + ik_2)^{-\gamma_1-\gamma_2+L}}{\Gamma(2\gamma_1 + 1)\Gamma(2\gamma_2 + 1)}, \tag{23}$$

[11] R. C. Young, Phys. Rev. 115, 577 (1959). Further references are given in this paper.

[12] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 1, p. 255. We shall hereafter denote this reference as *HTF*-1.

and

$$F_2^t(\alpha, a, \bar{a}, b, \bar{b}; x, y) = \{\Gamma(b)/[\Gamma(a)\Gamma(b-a)]\} \times \int_0^1 du\, u^{a-1}(1-u)^{b-a-1}(1-ux-it)^{-\alpha} \times {}_2F_1(\alpha, \bar{a}, \bar{b}; y/(1-ux-it)), \quad (24)$$

with $t \equiv s/(k_1+k_2)$, $y = 2k_2/(k_1+k_2)$, and $x = 2k_1/(k_1+k_2)$. Here, ${}_2F_1(\cdots)$ denotes the integral representation of this hypergeometric function since the series is divergent.[13] It is seen that for $t = 0$, $F_2^t(\alpha, a, \bar{a}, b, \bar{b}; x, y)$ is just the integral representation for Appell's double hypergeometric series.[14] It will be noted that the singular points of the integrand are at $u = 0$, 1, $(1-it)/x$, $(1-y-it)/x$, and ∞.

The radial integrals are still not in a form that can be easily evaluated, so the null contours C_1 and C_2, shown in Figs. 1 and 2, are introduced. Both contours contain a loop of large but not infinite radius R. As the radii of the circuits about 0 and 1 approach zero, both of these contours contain the path of integration of expression (24). On these contours the function $I(u)$ is defined by

$$I(u) \equiv \{\Gamma(b)/[\Gamma(a)\Gamma(b-a)]\}u^{a-1} \times (1-u)^{b-a-1}(1-ux-it)^{-\alpha} \times {}_2F_1(\alpha, \bar{a}, \bar{b}; y/(1-ux-it)), \quad (25)$$

with the phase of $I(u)$ determined on both C_1 and C_2 by defining arg (u) and arg $(1-u)$ to be zero at $u = A$, and

$$\lim_{t\to 0} \arg\{(1-ux-it)_{u=A}\} = 0. \quad (26)$$

If the radii of the loops about 0 and 1 are considered to be infinitesimal, then $F_2^t(\alpha, a, \bar{a}, b, \bar{b}; x, y)$ may be expressed in terms of more-easily handled integrals by considering the identity

$$\int_{C_2} I(u)\, du - e^{2\pi i(b-\alpha)} \int_{C_1} I(u)\, du = 0. \quad (27)$$

This combination of the two integrals is just that needed in order that the integrations about $(1-y-it)/x$ and $(1-it)/x$ in the two contours cancel, yielding

$$(1-e^{2\pi i(b-\alpha)})F_2^t(\alpha, a, \bar{a}, b, \bar{b}; x, y) = \int_1^{(\infty+)} I(u)\, du - e^{2\pi i(b-\alpha)} \int_0^{(\infty+)} I(u)\, du. \quad (28)$$

To evaluate the first integral on the right-hand side of (28), we make use of the identity[15]

$${}_2F_1(\alpha, \bar{a}, \bar{b}; 1/(1-z)) = (1-z)^{\alpha}(-z)^{-\alpha}\, {}_2F_1(\alpha, \bar{b}-\bar{a}, \bar{b}; 1/z) \quad (29)$$

to obtain

$$I(u) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} u^{a-1} \times (1-u)^{b-a-1}(1-y-ux-it)^{-\alpha} \times {}_2F_1\left(\alpha, \bar{b}-\bar{a}, \bar{b}; \frac{y}{y-1+ux+it}\right). \quad (30)$$

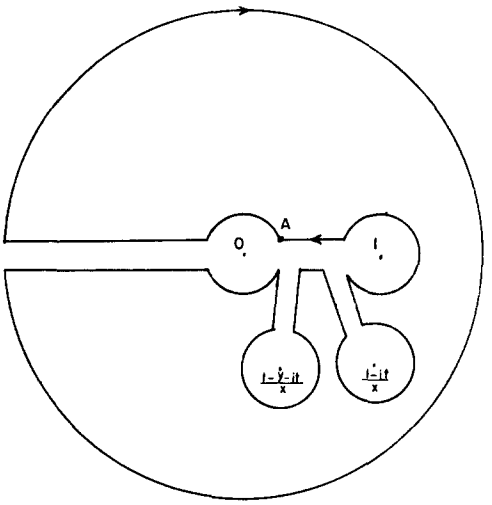

FIG. 1. The first null contour, C_1, on which $I(u)$ is defined. This contour contains an arc of large but not infinite radius R about the origin, and arcs of small radius about the four finite singular points. All adjacent parallel lines are supposed to coincide with the horizontal lines on the real axis.

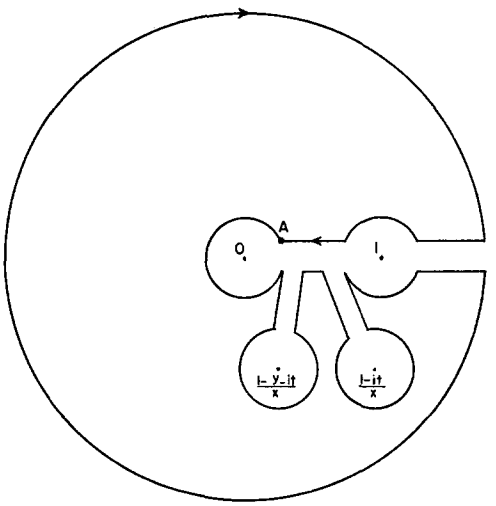

FIG. 2. The second null contour, C_2, on which $I(u)$ is defined. The phases of $I(u)$ on the two null contours are equal at $u = A$.

[13] *HTF*-1, p. 59.
[14] *HTF*-1, p. 224.
[15] *HTF*-1, p. 64.

To obtain an explicit solution to the integral, the zero energy loss limit must now be taken, so that $x = y = 1$. Then, making a series expansion of the hypergeometric function and changing the order of integration and summation yields

$$\lim_{\substack{t\to 0\\ R\to\infty}} \int_1^{(\infty+)} I(u)\,du = \frac{\Gamma(b)\Gamma(\alpha-b+1)}{\Gamma(a)\Gamma(\alpha-a+1)} \times e^{-i\pi(b-a-\alpha)}(1-e^{2\pi i(b-\alpha)}) \times {}_3F_2(\alpha, \bar{b}-\bar{a}, \alpha-b+1; \bar{b}, \alpha-a+1; 1). \quad (31)$$

Here, use has been made of the identity

$$\lim_{\substack{t\to 0\\ R\to\infty}} \int_1^{(\infty+)} u^{a-1}(1-u)^{b-a-1}(u+it)^{-\alpha-n}\,du = \frac{\Gamma(\alpha+n-b+1)\Gamma(b-a)}{\Gamma(\alpha+n+1-a)} \times e^{i\pi(a-\alpha)}[e^{-i\pi(b-\alpha)} - e^{i\pi(b-\alpha)}], \quad (32)$$

valid for $\mathrm{Re}(b-a)>0$, and all values of $\alpha+n-b+1$ not zero or a negative integer. This may be more easily seen by making the transformation $\omega = 1/u$.

To obtain a solution of the second integral on the right-hand side of (28), the hypergeometric function is expanded into its series, integration and summation are interchanged, and the change of variables $\omega = u/(u-1)$ is made to yield

$$\int_0^{(\infty+)} I(u)\,du = e^{i\pi a} \sum_{n=0}^{\infty} \frac{(\alpha)_n(\bar{a})_n}{(\bar{b})_n n!}\, y^n \times \int_0^{(1+)} d\omega\, \omega^{a-1}(1-\omega)^{\alpha+n-b} \times [1-\omega(1-x)-it(1-\omega)]^{-n-\alpha}, \quad (33)$$

where $(l)_m \equiv \Gamma(l+m)/\Gamma(l)$, and where the contour of the integration is shown in Fig. 3. To evaluate this expression, the zero energy loss limit, $x = y = 1$, must be taken. Then, by using the identity

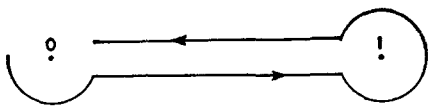

FIG. 3. The contour of integration used in right-hand side of Eq. (33). The radii of the arcs about 0 and 1 approach zero as $R \to \infty$ and $t \to 0$.

$$\lim_{\substack{t\to 0\\ R\to\infty}} \int_0^{(1+)} d\omega\, \omega^{a-1}(1-\omega)^{\alpha+n-b}[1-it(\omega-1)]^{-n-\alpha} = [1-e^{2\pi i(\alpha-b)}]\frac{\Gamma(a)\Gamma(\alpha+n+1-b)}{\Gamma(\alpha+n+1+a-b)}, \quad (34)$$

valid for $\mathrm{Re}\,(a) > 0$, and for all values of

$$\alpha + n + 1 - b$$

except zero and negative integers, there results

$$\lim_{\substack{t\to 0\\ R\to\infty}} \int_0^{(\infty+)} I(u)\,du = e^{i\pi a}\frac{\Gamma(b)\Gamma(\alpha+1-b)}{\Gamma(b-a)\Gamma(\alpha+a+1-b)}[1-e^{2\pi i(\alpha-b)}] \times {}_3F_2(\alpha, \bar{a}, \alpha+1-b; \bar{b}, \alpha+a+1-b; 1). \quad (35)$$

Upon noting that, in all four cases in Eq. (22) both $(\bar{b}-\bar{a}-\alpha+a-1)$ and $(\bar{a}-\alpha-a-1+b)$ are positive integers (zero energy loss means $\eta_1 = \eta_2$), use may be made of the very useful identity (Saalschütz)

$${}_3F_2(\alpha, a+k, \beta; \gamma, a; 1) = \frac{\Gamma(a)\Gamma(a+k-\beta)\Gamma(\gamma)\Gamma(\gamma-\beta-\alpha)}{\Gamma(a+k)\Gamma(a-\beta)\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \times \sum_m \frac{(-k)_m(\beta)_m(1-\gamma+\beta)_m}{m!\,(1+\beta-a-k)_m(1-\gamma+a+\beta)_m}, \quad (36)$$

where k is a positive integer. This expression is valid whenever both sides of the equation are defined. Hence, there results the expression

$$F_2^0(\alpha, a, \bar{a}, b, \bar{b}; 1, 1) = e^{-i\pi(b-a-\alpha)}\frac{\Gamma(b)\Gamma(\alpha-b+1)\Gamma(\bar{b}-\bar{a}-\alpha+b-1)\Gamma(\bar{b})\Gamma(\bar{b}-2\alpha+b-1)}{\Gamma(a)\Gamma(\bar{b}-\bar{a})\Gamma(b-a)\Gamma(\bar{b}-\alpha)\Gamma(\bar{b}-\alpha+b-1)} \times \sum_m \frac{(\alpha-a+1-\bar{b}+\bar{a})_m(\alpha-b+1)_m(2-\bar{b}+\alpha-b)_m}{m!\,(2+\alpha+\bar{a}-b-\bar{b})_m(2+2\alpha-b-\bar{b})_m} + e^{i\pi a}\frac{\Gamma(b)\Gamma(\alpha-b+1)\Gamma(\bar{a}-\alpha-1+b)\Gamma(\bar{b})\Gamma(\bar{b}+b-2\alpha-1)}{\Gamma(b-a)\Gamma(\bar{a})\Gamma(a)\Gamma(\bar{b}-\alpha)\Gamma(\bar{b}+b-\alpha-1)} \times \sum_m \frac{(\alpha+a+1-b-\bar{a})_m(\alpha+1-b)_m(2+\alpha-b-\bar{b})_m}{m!\,(2+\alpha-b-\bar{a})_m(2+2\alpha-b-\bar{b})_m}, \quad (37)$$

in which, for the case at hand, all sums are finite.

The zero energy loss Dirac–Coulomb integrals for $\gamma_1 \neq \gamma_2$ may now be written as

$$\int_0^\infty \begin{Bmatrix} S(\kappa_1)S(\kappa_2)F_{\kappa_1}(kr)F_{\kappa_2}(kr) \\ G_{\kappa_1}(kr)G_{\kappa_2}(kr) \end{Bmatrix} \frac{1}{r^{L-1}}\,dr$$

$$= \frac{2^{L-1}\Gamma(2L+1)k^{L-2}\pi\Gamma(\gamma_1+\gamma_2-L)}{\Gamma(\gamma_1+\gamma_2+L+1)\Gamma(L+1+\gamma_1-\gamma_2)\Gamma(L+1+\gamma_2-\gamma_1)\sin\pi(\gamma_2-\gamma_1-L)}$$

$$\times \begin{Bmatrix} (E-m_0)/(E+m_0) \\ 1 \end{Bmatrix} \mathrm{Re}\left\{\frac{\Gamma(\gamma_1+L+i\eta)\Gamma(\gamma_2-i\eta)}{|\Gamma(\gamma_1+i\eta)\Gamma(\gamma_2+i\eta)|}\exp i[\phi_1+\phi_2-\tfrac{1}{2}\pi(\gamma_2-\gamma_1)+\tfrac{1}{2}\pi(L+2)]\right.$$

$$\times \left[(\gamma_1+L+i\eta)\sum_m\left(\begin{Bmatrix}-\\+\end{Bmatrix}(-L)_m - e^{-2i\phi_1}(-L-1)_m\right)\frac{(\gamma_2-\gamma_1-L)_m(-\gamma_1-\gamma_2-L)_m}{m!\,(-\gamma_1-L-i\eta)_m(-2L)_m}\right.$$

$$\left.\left. + (\gamma_2-i\eta)e^{-2i\phi_2}\sum_m\left((-L+1)_m - \begin{Bmatrix}-\\+\end{Bmatrix}e^{-2i\phi_1}(-L)_m\right)\frac{(\gamma_2-\gamma_1-L)_m(-\gamma_1-\gamma_2-L)_m}{m!\,(1-\gamma_1-L-i\eta)_m(-2L)_m}\right]\right\}, \tag{38}$$

where Re { } means the real part of { }. It will be noted that no sum in this expression contains more than $L+2$ terms.

Because of the factor $\sin\pi(\gamma_2-\gamma_1-L)$ in the denominator of (38), the expression is not defined for $\gamma_1=\gamma_2$, and a limiting process must be carried out to include this case. A separate limit is required for each multipolarity. These limits have been performed for multipolarities one through four and are given by

$$\int_0^\infty \begin{Bmatrix} -F_\kappa(kr)F_{-\kappa}(kr) \\ G_\kappa(kr)G_{-\kappa}(kr) \end{Bmatrix} dr$$

$$= \frac{\Gamma(2\gamma-1)}{k(\gamma^2+\eta^2)^{\frac{1}{2}}\Gamma(2\gamma+2)}\begin{Bmatrix} (E-m_0)/(E+m_0) \\ 1 \end{Bmatrix}$$

$$\times \left[\begin{Bmatrix}-\\+\end{Bmatrix}(\gamma^2+\eta^2)\left(4\gamma+2\eta[\pi+i\Psi(\gamma-i\eta)\right.\right.$$

$$\left. - i\Psi(\gamma+i\eta)] - \frac{\gamma+2\eta^2}{\gamma^2+\eta^2}\right)$$

$$+ \frac{\alpha Z m_0}{k}(2\eta - 4\eta\gamma - (2\gamma^2+2\eta^2)$$

$$\left.\times [\pi + i\Psi(\gamma-i\eta) - i\Psi(\gamma+i\eta)])\right],$$

$$\int_0^\infty \begin{Bmatrix} F_\kappa(kr)^2 \\ G_\kappa(kr)^2 \end{Bmatrix}\frac{1}{r}\,dr$$

$$= \frac{4\Gamma(2\gamma-2)}{\Gamma(2\gamma+3)}\begin{Bmatrix} (E-m_0)/(E+m_0) \\ 1 \end{Bmatrix}$$

$$\times \left[\begin{Bmatrix}-\\+\end{Bmatrix}\left(-3\kappa\gamma - \frac{6\eta^2\gamma m_0}{E}\right.\right.$$

$$- \frac{\kappa}{\gamma^2+\eta^2}(\gamma^3-\gamma+3\eta^2\gamma-3\eta^2)$$

$$+ [\pi\eta + i\eta\Psi(\gamma-i\eta) - i\eta\Psi(\gamma+i\eta)]$$

$$\times \left[-3\kappa - \frac{m_0}{E}(\gamma^2-1+3\eta^2)\right]$$

$$\left.+ \frac{\eta^2 m_0}{E(\gamma^2+\eta^2)}(3\gamma+\gamma^2-1+3\eta^2)\right)$$

$$+ \{3[\pi\eta + i\eta\Psi(\gamma-i\eta) - i\eta\Psi(\gamma+i\eta)]$$

$$\times (\gamma^2+\eta^2) + \gamma(2\gamma+1)$$

$$\left.\times (2\gamma-1) + 3\eta^2(2\gamma-1)\}\right],$$

$$\int_0^\infty \begin{Bmatrix} -F_\kappa(kr)F_{-\kappa}(kr) \\ G_\kappa(kr)G_{-\kappa}(kr) \end{Bmatrix}\frac{1}{r^2}\,dr$$

$$= \frac{8k\Gamma(2\gamma-3)(\gamma^2+\eta^2)^{\frac{1}{2}}}{\Gamma(2\gamma+4)}\begin{Bmatrix} (E-m_0)/(E+m_0) \\ 1 \end{Bmatrix}$$

$$\times \left[\begin{Bmatrix}-\\+\end{Bmatrix}\left(\gamma\left(\frac{16}{3}\gamma^2 - \frac{31}{3} + 20\eta^2\right)\right.\right.$$

$$+ \eta[\pi + i\Psi(\gamma-i\eta) - i\Psi(\gamma+i\eta)]$$

$$\times (6\gamma^2 - 11 + 10\eta^2) + \frac{1}{\gamma^2+\eta^2}(-3\gamma^3+3\gamma$$

$$\left.- 15\eta^2\gamma - 6\eta^2\gamma^2 + 11\eta^2 - 10\eta^4)\right)$$

$$+ \frac{\alpha Z m_0}{k}\cdot\frac{1}{\gamma^2+\eta^2}\left(\eta\left(-\frac{52}{3}\gamma^3 + 2\gamma^2\right.\right.$$

$$- 2 + 10\eta^2 - 20\eta^2\gamma + \frac{22}{3}\gamma\Big)$$

$$+ [\pi + i\Psi(\gamma-i\eta) - i\Psi(\gamma+i\eta)]$$

$$\left.\left.\times (-2\gamma^4 + 2\gamma^2 - 12\eta^2\gamma^2 + 2\eta^2 - 10\eta^4)\right)\right],$$

$$\int_0^\infty \begin{Bmatrix} F_\kappa(kr)^2 \\ G_\kappa(kr)^2 \end{Bmatrix} \frac{1}{r^3}\, dr$$

$$= \frac{280k^2\Gamma(2\gamma-4)}{\Gamma(2\gamma+5)} \begin{Bmatrix} (E-m_0)/(E+m_0) \\ 1 \end{Bmatrix}$$

$$\times \left[\begin{Bmatrix} - \\ + \end{Bmatrix} \left(\kappa \left(-\frac{22}{21}\gamma^3 + \frac{46}{21}\gamma - 6\eta^2\gamma \right) \right.\right.$$

$$+ \frac{\alpha Z m_0}{k} \left(-\frac{44}{21}\eta\gamma^3 + \frac{134}{21}\eta\gamma - 4\eta^3\gamma \right)$$

$$+ \frac{\kappa}{\gamma^2+\eta^2} \left(-\frac{6}{35}\gamma^5 + \frac{6}{7}\gamma^3 - \frac{24}{35}\gamma - \frac{12}{7}\eta^2\gamma^3 \right.$$

$$\left. + \frac{34}{7}\eta^2\gamma - 2\eta^4\gamma + \frac{12}{7}\eta^2\gamma^2 - \frac{20}{7}\eta^2 + 4\eta^4 \right)$$

$$+ \frac{\alpha Z m_0}{k(\gamma^2+\eta^2)} \left(\frac{12}{7}\eta\gamma^3 - \frac{20}{7}\eta\gamma + 4\eta^3\gamma + \frac{6}{35}\eta\gamma^4 \right.$$

$$\left. - \frac{6}{7}\eta\gamma^2 + \frac{24}{35}\eta + \frac{12}{7}\eta^3\gamma^2 - \frac{34}{7}\eta^3 + 2\eta^5 \right)$$

$$+ [\pi + i\Psi(\gamma - i\eta) - i\Psi(\gamma + i\eta)]$$

$$\times \left[\kappa \left(-\frac{12}{7}\eta\gamma^2 + \frac{20}{7}\eta - 4\eta^3 \right) \right.$$

$$+ \frac{\alpha Z m_0}{k} \left(-\frac{6}{35}\gamma^4 + \frac{6}{7}\gamma^2 - \frac{24}{35} \right.$$

$$\left. \left. - \frac{12}{7}\eta^2\gamma^2 + \frac{34}{7}\eta^2 - 2\eta^4 \right) \right]$$

$$+ 2 \left\{ \frac{32}{105}\gamma^5 - \frac{16}{21}\gamma^3 + \frac{6}{35}\gamma - \frac{3}{7}\eta^2\gamma^2 \right.$$

$$+ \frac{46}{21}\eta^2\gamma^3 - \frac{37}{21}\eta^2\gamma + \frac{5}{7}\eta^2 + 2\eta^4\gamma - \eta^4$$

$$+ [\pi\eta + i\eta\Psi(\gamma - i\eta) - i\eta\Psi(\gamma + i\eta)]$$

$$\left. \times \left(\frac{3}{7}\gamma^4 - \frac{5}{7}\gamma^2 + \frac{10}{7}\eta^2\gamma^2 - \frac{5}{7}\eta^2 + \eta^4 \right) \right\} \left. \right]. \quad (39)$$

Now in calculations of processes that lead to a 2^L magnetic-multipole transition, the radial functions occur in integrals of the form (where the zero energy loss limit has already been taken)

$$M(\kappa_1, L, \kappa_2) \equiv \int_0^\infty [S(\kappa_1)F_{\kappa_1}(kr)G_{\kappa_2}(kr) + S(\kappa_2)G_{\kappa_1}(kr)F_{\kappa_2}(kr)]r^{-L+1}\, dr. \quad (40)$$

These integrals may be evaluated by use of the expression, valid for $\gamma_1 \neq \gamma_2$,

$$\int_0^\infty \frac{1}{r^{L-1}} S(\kappa_1)F_{\kappa_1}(kr)G_{\kappa_2}(kr)\, dr$$

$$= \frac{[(E-m_0)/(E+m_0)]^{\frac{1}{2}}2^{L-1}\Gamma(2L+1)k^{L-2}\pi\Gamma(\gamma_1+\gamma_2-L)}{\Gamma(\gamma_1+\gamma_2+L+1)\Gamma(L+1+\gamma_1-\gamma_2)\Gamma(L+1+\gamma_2-\gamma_1)\sin\pi(\gamma_2-\gamma_1-L)}$$

$$\times \operatorname{Im} \left\{ \frac{\Gamma(\gamma_1+L+i\eta)\Gamma(\gamma_2-i\eta)}{|\Gamma(\gamma_1+i\eta)\Gamma(\gamma_2+i\eta)|} \exp i[\phi_1+\phi_2-\tfrac{1}{2}\pi(\gamma_2-\gamma_1)+\tfrac{1}{2}\pi(L+2)] \right.$$

$$\times \left((\gamma_1+L+i\eta) \sum_m [-(-L)_m - e^{-2i\phi_1}(-L-1)_m] \frac{(\gamma_2-\gamma_1-L)_m(-\gamma_1-\gamma_2-L)_m}{m!\,(-\gamma_1-L-i\eta)_m(-2L)_m} \right.$$

$$\left. \left. + (\gamma_2-i\eta)e^{-2i\phi_2} \sum_m [-(-L+1)_m - e^{-2i\phi_1}(-L)_m] \frac{(\gamma_2-\gamma_1-L)_m(-\gamma_2-\gamma_1-L)_m}{m!\,(1-\gamma_1-L-i\eta)_m(-2L)_m} \right) \right\}, \quad (41)$$

where Im { } means the imaginary part of { }. This formula was derived using exactly the same techniques used to derive the electric-multipole integrals. Again, because of the factor

$$\sin\pi(\gamma_2-\gamma_1-L)$$

in the denominator, in order to include the case $\gamma_1 = \gamma_2$, a limit must be taken for each multipolarity. For lowest multipole orders, these limits have been performed and are given by

$$\int_0^\infty S(\kappa)F_\kappa(kr)G_\kappa(kr)\, dr$$

$$= \frac{\Gamma(2\gamma-1)}{(E+m_0)\Gamma(2\gamma+2)} \left\{ \kappa \left[4\gamma + 2\eta[\pi + i\Psi(\gamma - i\eta) - i\Psi(\gamma+i\eta)] - \frac{\gamma+2\eta^2}{\gamma^2+\eta^2} \right] \right.$$

$$\left. + \frac{\alpha Z m_0}{k} \left[-[\pi + i\Psi(\gamma - i\eta) - i\Psi(\gamma+i\eta)] - \frac{2\eta\gamma-\eta}{\gamma^2+\eta^2} \right] \right\}, \quad (42)$$

$$\int_0^\infty S(\kappa)F_\kappa(kr)G_{-\kappa}(kr)\frac{1}{r}\,dr$$
$$= 4\left(\frac{E-m_0}{E+m_0}\right)^{\frac{1}{2}}\frac{\Gamma(2\gamma-2)}{\Gamma(2\gamma+3)(\gamma^2+\eta^2)^{\frac{1}{2}}}$$
$$\times\{3\eta(\kappa+1)(\gamma^2+\eta^2)[\pi+i\Psi(\gamma-i\eta)$$
$$-i\Psi(\gamma+i\eta)]+4\gamma^3-\gamma+6\eta^2\gamma$$
$$-3\eta^2+\kappa(2\gamma-1)[\gamma(2\gamma+1)+3\eta^2]\},$$

$$\int_0^\infty S(\kappa)F_\kappa(kr)G_\kappa(kr)\frac{1}{r^2}\,dr=\frac{8(E-m_0)\Gamma(2\gamma-3)}{\Gamma(2\gamma+4)}$$
$$\times\left\{\kappa\left[\frac{16}{3}\gamma^3-\frac{31}{3}\gamma+20\eta^2\gamma+[\pi+i\Psi(\gamma-i\eta)\right.\right.$$
$$-i\Psi(\gamma+i\eta)](6\eta\gamma^2-11\eta+10\eta^3)+\frac{1}{\gamma^2+\eta^2}$$
$$\left.\times(-3\gamma^3+3\gamma-15\eta^2\gamma-6\eta^2\gamma^2+11\eta^2-10\eta^4)\right]$$
$$+\frac{\alpha Zm_0}{k}\left[-20\eta\gamma+[\pi+i\Psi(\gamma-i\eta)\right.$$
$$-i\Psi(\gamma+i\eta)](-3\gamma^2+3-15\eta^2)$$
$$+\frac{1}{\gamma^2+\eta^2}(-6\eta\gamma^3+11\eta\gamma$$
$$\left.\left.-10\eta^3\gamma+3\eta\gamma^2-3\eta+15\eta^3)\right]\right\}.$$

Just as in the nonrelativistic case, there exists an operator that changes the sign of κ in the relativistic Coulomb wavefunctions given in (17). This is the operator[16]

$$\mathcal{R}=-S(\kappa)p^{-1}[K^2-(\alpha Z)^2+\eta^2]^{-\frac{1}{2}}$$
$$\times\{i\alpha Zm_0(\boldsymbol{\sigma}\cdot\hat{\mathbf{r}})+K\rho_1(H+m_0\rho_3)\},\qquad(43)$$

where

$$S(\kappa)\equiv K[K^2]^{-\frac{1}{2}},\qquad p\equiv[H^2-m_0^2]^{\frac{1}{2}}.\qquad(44)$$

(Note that this is not given quite correctly in Ref. 16.) By using operator techniques similar to those used in the nonrelativistic case, one is able to derive recursion relations for the relativistic radial integrals. However, the work on this subject to date is quite incomplete and the topic will not be pursued here.

V. CALCULATIONS

Extensive calculations have been made for multipolarities one through four, using electrons as projectiles, in order to investigate the dependence of the integrals on the multipolarity L, the charge of the nucleus Z, and the energy of the projectile E. The electric and magnetic integrals have similar characteristics as functions of L, Z, and E. It should be noted that, aside from an overall energy-dependent factor, the corresponding nonrelativistic integrals contain the dependence on Z and E in a single parameter η, while the relativistic integrals depend on Z and E separately.

For small values of the parameters κ_1 and κ_2 and for all L, the integrals have appreciable dependence on Z, the integrals increasing monotonically with Z (decreasing monotonically when the projectiles are positrons). This dependence on Z decreases as κ_1 and κ_2 are increased. The higher the multipolarity L, the more rapidly the integrals fall off as a function of κ_1 and κ_2. Except for an overall energy-dependent factor there is very little energy dependence in the radial integrals for projectile energy greater than 10 MeV.

In the complete relativistic limit ($E\to\infty$), $I(\kappa_1, L, \kappa_2)$ and $I(-\kappa_1, L, -\kappa_2)$ become equal, whereas the nonrelativistic limits of these two integrals are different [being $p^{-2}\langle l(\kappa_1),\ \eta\ |r^{-L-1}|\ l(\kappa_2),\ \eta\rangle$, and $p^{-2}\langle l(-\kappa_1),\ \eta\ |r^{-L-1}|\ l(-\kappa_2),\ \eta\rangle$, respectively]. On the other hand, for large E and small values of κ_1 and κ_2, $I(|\kappa_1|, L, |\kappa_2|-1)$ and $I(|\kappa_1|, L, -|\kappa_2|)$ are quite different but their nonrelativistic limits

$$[p^{-2}\langle l(|\kappa_1|),\ \eta\ |r^{-L-1}|\ l(|\kappa_2|-1),\ \eta\rangle$$

and

$$p^{-2}\langle l(|\kappa_1|),\ \eta\ |r^{-L-1}|\ l(-|\kappa_2|),\ \eta\rangle]$$

are exactly equal. These two facts demonstrate that in no straightforward manner can the relativistic integrals be approximated by the nonrelativistic formulas.

ACKNOWLEDGMENTS

The computations reported here were performed at the Duke University Computing Center on an IBM 7072 digital computer. We are indebted to Dr. Gallie and his staff for their cooperation. We would also like to acknowledge the assistance of Dr. J. N. Huffaker in some of the preliminary calculations.

[16] L. C. Biedenharn, Phys. Rev. 126, 845 (1962).

Integration of the Partial Differential Equations for the Hypergeometric Functions F_1 and F_D of Two and More Variables*

PER O. M. OLSSON
Institute of Theoretical Physics, University of Stockholm, Stockholm, Sweden
(Received 19 June 1963)

From Kummer's 24 solutions of the ordinary hypergeometric equation and their connections, the general solution of the partial differential equations for the Appell function F_1 in the neighborhood of the singular points is found. Connections between the solutions are given to the extent required to continue the F_1 function to the neighborhood of any of its singular points.

The corresponding problem for the hypergeometric function F_D of more than two variables is briefly indicated.

THE integration of the partial differential equations

$$\begin{aligned} x(1-x)\frac{\partial^2 F}{\partial x^2} + y(1-x)\frac{\partial^2 F}{\partial x\,\partial y} \\ + [c-(a+b_1+1)x]\frac{\partial F}{\partial x} \\ - b_1 y\frac{\partial F}{\partial y} - ab_1 F = 0, \\ y(1-y)\frac{\partial^2 F}{\partial y^2} + x(1-y)\frac{\partial^2 F}{\partial x\,\partial y} \\ + [c-(a+b_2+1)y]\frac{\partial F}{\partial y} \\ - b_2 x\frac{\partial F}{\partial x} - ab_2 F = 0, \end{aligned} \tag{1}$$

satisfied by the Appell function $F_1(a, b_1, b_2, c, x, y)$ which is defined in the neighborhood of the origin by the expansion

$$F_1(a, b_1, b_2, c, x, y) = \sum_{\substack{m=0\\n=0}}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n x^m y^n}{(c)_{m+n} m!\, n!}, \qquad |x|<1, \quad |y|<1, \tag{2}$$

has been investigated by many writers. Further solutions can be obtained in form of an integral proposed by Picard:

$$F(x, y) = \int_c u^{b_1+b_2-c} \times (u-1)^{c-a-1}(u-x)^{-b_1}(u-y)^{-b_2}\, du. \tag{3}$$

Here the path of integration is either closed on the Riemann surface of the integrand or an open path ending at zeros of the integrand.

From this integral, 60 solutions in terms of the Appell function F_1 have been found. A table of these sixty integrals can be found in Appell and Kampé de Fériet.[1] This set of solutions corresponds, in a sense, to Kummer's 24 solutions of the hypergeometric equation since it contains all known solutions of (1) expressible in terms of F_1 functions. There is, however, an important difference between the two tables. The set of F_1 solutions does not give complete information concerning the behavior of the general solution of (1) near its singular points, whereas the Kummer set gives complete information in this respect concerning the solutions of the hypergeometric equation.

Erdélyi[2] pointed out the deficiency of the set of F_1 solutions in this respect and also showed how further solutions could be obtained to complete the table. These additional solutions are not expressed in F_1 functions but in terms of G_2 functions, defined by the expansion

$$G_2(a_1, a_2, b_1, b_2, x, y) = \sum_{\substack{m=0\\n=0}}^{\infty} (a_1)_m(a_2)_n \times (b_1)_{n-m}(b_2)_{m-n}\frac{x^m}{m!}\frac{y^n}{n!}, \quad |x|<1, \quad |y|<1. \tag{4}$$

This function, as well as F_1, belongs to a class of 34 hypergeometric functions of two variables and of order two in Horn's classification.

Erdélyi uses the method of contour integration on (3) but with more general contours than had hitherto been attempted. The solutions in terms of G_2 functions are not explicitly given but can be obtained from his paper. Erdélyi, however, does not give the connections between the various solutions.

In this paper we shall derive explicitly the G_2 solutions of Erdélyi, and also give the connections between various solutions to the extent required

* This work has been carried out under the auspices of the Swedish Atomic Research Council.

[1] P. Appell et J. Kampé de Fériet, *Fonctions hypergéométriques et hypersphériques* (Gauthier-Villars, Paris, 1926), p. 62.

[2] A. Erdélyi, Acta Mat. **83**, 131 (1950).

to obtain the analytic continuations of the F_1 function (2) to the neighborhood of any of its singular points. We shall use the following simple method of integrating. (1) The F_1 function is expressed as a sum, or integral, over the ordinary hypergeometric ${}_2F_1$ functions, which are then replaced by their well-known analytic continuations. It is not difficult to select the correct continuation of the ${}_2F_1$ function which, after simple transformations, leads to continuations of the F_1 function in terms of other hypergeometric functions, the expansions of which are convergent near any desired singular point of the system (1), and which are, in addition, particular integrals of the system. In this way, as we shall see, we can obtain a set of solutions expressed in F_1 and G_2 functions such that, for at least every real point (x, y) which is not a singular point of (1), the set contains three linearly independent solutions whose expansions are convergent at the point. Since the general solution is a linear combination of three independent solutions, its analytic properties are known from such a set, which we will call a complete set of solutions.

From the thus completed set of solutions and their connections, one can derive solutions of the differential equations for the hypergeometric function F_D of three variables introduced by Lauricella:

$$F_D(a, b_1, b_2, b_3, x, y, z) = \sum \frac{(a)_{n_1+n_2+n_3}(b_1)_{n_1}(b_2)_{n_2}(b_3)_{n_3}x^{n_1}y^{n_2}z^{n_3}}{(c)_{n_1+n_2+n_3}n_1!\,n_2!\,n_3!}. \quad (5)$$

These solutions, as far as we know, have not been obtained earlier. The method of deriving these solutions is the same as in deriving the complete set of solutions for the system of equations for F_1. The F_D function is expressed as a sum, or as an integral, over F_1 functions, for which we now know the continuations. There seems to be no reason why a complete set of solutions of the equations for the function F_D should not be obtainable in this way. One can of course then proceed and derive the solutions of the differential equations for the Lauricella function F_D of more than three variables.

The knowledge of the analytic continuations of the F_1 and F_D functions obviously solves the problem of continuing analytically other hypergeometric functions in cases when they can be transformed or reduced to them. Cases where hypergeometric functions can be reduced to Appell or Lauricella functions for certain values of the parameters are special but not unimportant, since the reducible cases may be reached after use of suitable transformations or recurrence relations.

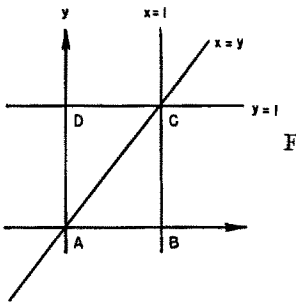

Fig. 1

It is instructive to consider the singular manifolds of the system (1) which are

$$(0, y) \quad (1, y) \quad (\infty, y) \quad (x, x = y)$$
$$(x, 0) \quad (x, 1) \quad (x, \infty). \quad (6)$$

Here, e.g., $(0, y)$ means $x = 0$ and y arbitrary. Figure 1 gives a picture of these singular manifolds for real variables. The points A and C are intersections of three singular manifolds; namely $x = 0$, $y = 0$, and $x = y$ for A and $x = 1$, $y = 1$, and $x = y$ for C. The point $x = y = \infty$ has the same character, whereas the points B and D are intersections of two singular manifolds only. The points which are intersections of three singular manifolds are actually more complicated than the others since no continuations of the F_1 function in terms of hypergeometric series, convergent in the entire neighborhood of the singular point, seem to be possible.

It is possible to express in many ways the Appell function F_1 in terms of the ordinary hypergeometric function ${}_2F_1$ in which the variable is either x or y. We can then take advantage of the known continuations of this function in order to obtain continuations of the F_1 function. A table of these connections between the Kummer solutions is available[3]. What we have in mind is best illustrated by the following example:

The series expansion

$${}_2F_1(a, b, c, y) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n n!} y^n$$

allows us to sum over n in (2) and we obtain

$$F_1(a, b_1, b_2, c, x, y) = \sum_{m=0}^{\infty} \frac{(a)_m(b_1)_m x^m}{(c)_m m!} {}_2F_1(a + m, b_2, c + m, y). \quad (7)$$

Suppose now that we are interested in the behavior of the F_1 function near the point (0, 1). We shall

[3] Bateman Manuscript Project, *Higher Transcendental Functions*, edited by A. Erdélyi (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 106.

then obviously replace the ${}_2F_1$ function by its continuation to the vicinity of $y = 1$, which is

$$
\begin{aligned}
{}_2F_1(a, b, c, y) &= \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \\
&\times {}_2F_1(a, b, a+b-c+1, 1-y) \\
&+ \frac{\Gamma(c)\Gamma(a+b-c)}{\Gamma(a)\Gamma(b)}(1-y)^{c-a-b} \\
&\times {}_2F_1(c-a, c-b, c-a-b+1, 1-y), \quad (8)
\end{aligned}
$$

where $|\arg(1-y)| < \pi$.

Inserting this expression for the ${}_2F_1$ function in (7), after the appropriate change of the parameters we obtain

$$
\begin{aligned}
F_1(a, b_1, b_2, c, x, y) &= \frac{\Gamma(c)\Gamma(c-a-b_2)}{\Gamma(c-a)\Gamma(c-b_2)} \\
&\times \sum_{\substack{m=0\\n=0}}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n x^m(1-y)^n}{(c-b_2)_m(a+b_2-c+1)_n m!\, n!} \\
&+ \frac{\Gamma(c)\Gamma(a+b_2-c)}{\Gamma(a)\Gamma(b_2)}(1-y)^{c-a-b_2} \\
&\times \sum_{\substack{m=0\\n=0}}^{\infty} \frac{(c-b_2)_{m+n}(b_1)_m(c-a)_n x^m(1-y)^n}{(c-b_2)_m(c-b_2-a+1)_n m!\, n!}, \quad (9)
\end{aligned}
$$

using $\Gamma(a+n) = \Gamma(a)(a)_n$ and returning to the double sums. The two series are recognized as expansions of the Appell function F_2,

$$
F_2(a, b_1, c_1, c_2, x, y) = \sum_{\substack{m=0\\n=0}}^{\infty} \frac{(a)_{m+n}(b_1)_m(b_2)_n x^m y^n}{(c_1)_m(c_2)_n m!\, n!}, \quad |x| + |y| < 1. \quad (10)
$$

Hence

$$
\begin{aligned}
F_1(a, b_1, b_2, c, x, y) &= \frac{\Gamma(c)\Gamma(c-a-b_2)}{\Gamma(c-a)\Gamma(c-b_2)} \\
&\times F_2(a, b_1, b_2, c-b_2, a+b_2-c+1, x, 1-y) \\
&+ \frac{\Gamma(c)\Gamma(a+b_2-c)}{\Gamma(a)\Gamma(b_2)}(1-y)^{c-a-b_2} \\
&\times F_2(c-b_2, b_1, c-a, c-b_2, c-b_2-a+1, x, 1-y), \quad (11)
\end{aligned}
$$

which is the desired analytic continuation of the F_1 function to the neighborhood of (0, 1).

Each term in (11) is a solution of (1), as can be inferred from the way the result was derived. In the second F_2 function, the first parameter is equal to the fourth, in which case the F_2 function is known to reduce to a F_1 function according to

$$
\begin{aligned}
F_2(a, b_1, b_2, a, c_2, x, y) &= (1-x)^{-b_1} \\
&\times F_1\left(b_2, b_1, a-b_1, c_2, \frac{y}{1-x}, y\right). \quad (12)
\end{aligned}
$$

The first F_2 function can be reduced to a G_2 function with the aid of

$$
\begin{aligned}
&F_2(c_1+c_2-1, b_1, b_2, c_1, c_2, x, y) \\
&= (1-x)^{-b_1}(1-y)^{-b_2} \\
&\times G_2\left(b_1, b_2, 1-c_1, 1-c_2, \frac{x}{1-x}, \frac{y}{1-y}\right), \quad (13)
\end{aligned}
$$

a result that is derived in the paper by Erdélyi.[4]

By choosing the proper continuation among the different connections between the Kummer solutions ${}_2F_1$, we could have obtained the continuation (11) directly in F_1 and G_2 functions, but that choice is less obvious and we prefer here the choice that leads most directly to continuations of the F_1 function to the desired domain.

From the series expansions of the functions, we see that

$$
\begin{aligned}
&F_1(a, b_1, b_2, c, x, y) = F_1(a, b_2, b_1, c, y, x), \\
&F_2(a, b_1, b_2, c_1, c_2, x, y) \\
&\qquad = F_2(a, b_2, b_1, c_2\, c_1, y, x), \quad (14) \\
&G_2(a_1, a_2, b_1, b_2, x, y) = G_2(a_2, a_1, b_2, b_1, y, x).
\end{aligned}
$$

Then in (11) we can interchange b_1 and b_2, and x and y, and obtain the continuation to the neighborhood of (1,0). Using (12), (13), and (14), we obtain the following continuations expressed in terms of F_1 and G_2 functions:

$$
\begin{aligned}
&F_1(a, b_1, b_2, c, x, y) \\
&= \frac{\Gamma(c)\Gamma(a+b_2-c)}{\Gamma(a)\Gamma(b_2)}(1-x)^{-b_1}(1-y)^{c-a-b_2}F_1\left(c-a, b_1, c-b_1-b_2, c-a-b_2+1, \frac{1-y}{1-x}, 1-y\right) \\
&+ \frac{\Gamma(c)\Gamma(c-a-b_2)}{\Gamma(c-a)\Gamma(c-b_2)}(1-x)^{-b_1}y^{-b_2}G_2\left(b_1, b_2, b_2-c+1, c-a-b_2, \frac{x}{1-x}, \frac{1-y}{y}\right) \\
&= \frac{\Gamma(c)\Gamma(a+b_1-c)}{\Gamma(a)\Gamma(b_1)}(1-x)^{c-a-b_1}(1-y)^{-b_2}F_1\left(c-a, c-b_1-b_2, b_2, c-a-b_1+1, 1-x, \frac{1-x}{1-y}\right) \\
&+ \frac{\Gamma(c)\Gamma(c-a-b_1)}{\Gamma(c-a)\Gamma(c-b_1)}x^{-b_1}(1-y)^{-b_2}G_2\left(b_1, b_2, c-a-b_1, b_1-c+1, \frac{1-x}{x}, \frac{y}{1-y}\right). \quad (15)
\end{aligned}
$$

[4] Ref. 2, p. 149.

The two F_1 functions on the right-hand side (rhs) of (15) are found in the aforementioned table of F_1 solutions[3] where they are denoted z_{16} and z_{17}. The G_2 solutions are two functions from a set of 60 solutions in terms of this function. This set can be deduced from the paper by Erdélyi[4]. In the form of F_2 functions, the two G_2 solutions above were, however, first found by Borngässer,[5] who gave the solutions in the neighborhood of the intersections of two singular manifolds of (1). The remaining solutions which are convergent near intersections of three singular manifolds were first given by Erdélyi.[2]

So far we have obtained two obviously independent solutions near each of the two points (0, 1) and (1, 0). Since (1) has three linearly independent solutions, one convergent solution near each of the two points is missing. These solutions, however, will be obtained simply by changing parameters in results that we shall obtain as we proceed.

Considering next the neighborhood of $(0, \infty)$, it is obvious that in (7) we shall use

$$
\begin{aligned}
{}_2F_1(a, b, c, y) &= \frac{\Gamma(c)\Gamma(b-a)}{\Gamma(b)\Gamma(c-a)}(-y)^{-a} \\
&\times {}_2F_1\left(a, a-c+1, a-b+1, \frac{1}{y}\right) \\
&+ \frac{\Gamma(c)\Gamma(a-b)}{\Gamma(a)\Gamma(c-b)}(-y)^{-b} \\
&\times {}_2F_1\left(b, b-c+1, b-a+1, \frac{1}{y}\right), \\
&\qquad |\arg(-y)| < \pi, \qquad (16)
\end{aligned}
$$

which, after a change of the parameters, allows us to write

$$
\begin{aligned}
F_1(a, b_1, b_2, c, x, y) &= \frac{\Gamma(c)\Gamma(b_2-a)}{\Gamma(b_2)\Gamma(c-a)}(-y)^{-a}\sum_{\substack{m=0\\n=0}}^{\infty}\frac{(a)_{m+n}(b_1)_m(b_2-a)_{-m}(1+a-c)_n(x/-y)^m(1/y)^n}{(1+a-b_2+m)_n m!\,n!} \\
&+ \frac{\Gamma(c)\Gamma(a-b_2)}{\Gamma(a)\Gamma(c-b_2)}(-y)^{-b_2}\sum_{\substack{m=0\\n=0}}^{\infty}\frac{(b_1)_m(a-b_2)_m(b_2)_n(1+b_2-c-m)_n x^m(1/y)^n}{(c-b_2)_m(1+b_2-a-m)_n m!\,n!}.
\end{aligned}
$$

From the properties of the Γ function, it follows that

$$(a)_{-m} = (-1)^m/(1-a)_m,$$

which applied to the double sums above, leads directly to the expansions for the F_1 and G_2 functions, so that we may write

$$
\begin{aligned}
F_1(a, b_1, b_2, c, x, y) &= \frac{\Gamma(c)\Gamma(b_2-a)}{\Gamma(b_2)\Gamma(c-a)}(-y)^{-a} \\
&\times F_1\left(a, b_1, 1+a-c, a-b_2+1, \frac{x}{y}, \frac{1}{y}\right) \\
&+ \frac{\Gamma(c)\Gamma(a-b_2)}{\Gamma(a)\Gamma(c-b_2)}(-y)^{-b_2} \\
&\times G_2\left(b_1, b_2, 1+b_2-c, a-b_2, -x, -\frac{1}{y}\right) \\
&= \frac{\Gamma(c)\Gamma(b_1-a)}{\Gamma(b_1)\Gamma(c-a)}(-x)^{-a} \\
&\times F_1\left(a, 1+a-c, b_2, 1+a-b_1, \frac{1}{x}, \frac{y}{x}\right) \\
&+ \frac{\Gamma(c)\Gamma(a-b_1)}{\Gamma(a)\Gamma(c-b_1)}(-x)^{-b_1} \\
&\times G_2\left(b_1, b_2, a-b_1, 1+b_1-c, -\frac{1}{x}, -y\right). \qquad (17)
\end{aligned}
$$

In order to obtain the last equality, which determines the behavior of the F_1 function near $(\infty, 0)$, we have used the symmetries (14). Two more F_1 solutions are found here, namely z_8 and z_9 in the table of F_1 solutions, and two G_2 solutions, both derived by Borngässer.[5]

In order to investigate the neighborhood of (1, 1) we take, e.g., the first F_2 solution appearing in (11),

$$
\begin{aligned}
&F_2(a, b_1, b_2, c-b_2, a+b_2-c+1, x, 1-y) \\
&= \sum \frac{(a)_n(b_2)_n(1-y)^n}{(a+b_2-c+1)_n n!}\,{}_2F_1(a+n, b_1, c-b_2, x),
\end{aligned}
$$

and use (8) in order to get power series in $1-x$. After the previously indicated manipulations with the Γ functions and the symbols $(a)_n$, we obtain

[5] L. Borngässer, "Uber hypergeometrische Functionen zweier Veränderlichen" (Dissertation, Darmstadt, Germany, 1932).

$$F_2(a, b_1, b_2, c - b_2, a + b_2 - c + 1, x, 1 - y) = \frac{\Gamma(c - b_2)\Gamma(c - a - b_1 - b_2)}{\Gamma(c - a - b_2)\Gamma(c - b_1 - b_2)}$$
$$\times F_1(a, b_1, b_2, 1 + a + b_1 + b_2 - c, 1 - x, 1 - y) + \frac{\Gamma(c - b_2)\Gamma(a + b_1 + b_2 - c)}{\Gamma(a)\Gamma(b_1)}$$
$$\times (1 - x)^{c-a-b_1-b_2} G_2\left(c - b_1 - b_2, b_2, a + b_1 + b_2 - c, c - a - b_2, x - 1, \frac{y - 1}{1 - x}\right). \quad (18)$$

In dealing with the second F_2 function in (11) in the same way we obtain the reduction formula (12) and a F_1 solution already found in (15). Thus it is sufficient to introduce (18) into (11) and reduce the second F_2 function according to (12). This gives

$$F_1(a, b_1, b_2, c, x, y) = \frac{\Gamma(c)\Gamma(c - a - b_1 - b_2)}{\Gamma(c - a)\Gamma(c - b_1 - b_2)} F_1(a, b_1, b_2, 1 + a + b_1 + b_2 - c, 1 - x, 1 - y)$$
$$+ \frac{\Gamma(c)\Gamma(a + b_2 - c)}{\Gamma(a)\Gamma(b_2)} (1 - x)^{-b_1}(1 - y)^{c-a-b_2} F_1\left(c - a, b_1, c - b_1 - b_2, c - a - b_2 + 1, \frac{1 - y}{1 - x}, 1 - y\right)$$
$$+ \frac{\Gamma(c)\Gamma(c - a - b_2)\Gamma(a + b_1 + b_2 - c)}{\Gamma(a)\Gamma(b_1)\Gamma(c - a)} (1 - x)^{c-a-b_1-b_2}$$
$$\times G_2\left(c - b_1 - b_2, b_2, a + b_1 + b_2 - c, c - a - b_2, x - 1, \frac{1 - y}{x - 1}\right)$$
$$= \frac{\Gamma(c)\Gamma(c - a - b_1 - b_2)}{\Gamma(c - a)\Gamma(c - b_1 - b_2)} F_1(a, b_1, b_2, 1 + a + b_1 + b_2 - c, 1 - x, 1 - y) + \frac{\Gamma(c)\Gamma(a + b_1 - c)}{\Gamma(a)\Gamma(b_1)}$$
$$\times (1 - x)^{c-a-b_1}(1 - y)^{-b_2} F_1\left(c - a, c - b_1 - b_2, b_2, c - a - b_1 + 1, 1 - x, \frac{1 - x}{1 - y}\right)$$
$$+ \frac{\Gamma(c)\Gamma(c - a - b_1)\Gamma(a + b_1 + b_2 - c)}{\Gamma(a)\Gamma(b_2)\Gamma(c - a)} (1 - y)^{c-a-b_1-b_2}$$
$$\times G_2\left(b_1, c - b_1 - b_2, c - a - b_1, a + b_1 + b_2 - c, \frac{1 - x}{y - 1}, y - 1\right). \quad (19)$$

Here again we have used the symmetries of the functions to obtain the second equality. The first F_1 function, which is convergent in the entire neighborhood of (1, 1), is the solution z_2 in the table of F_1 solutions. The other solutions are not convergent in the entire neighborhood of (1, 1), since, in order that the series be convergent independent of the values of the parameters, we must have $|1 - y| < |1 - x|$ for the first equality (19), and $|1 - y| > |1 - x|$ for the second. The expansions of the functions are thus, in general, not convergent if $x = y$. This is a singular manifold of Eqs. (1) as important as $x = 1$ or $x = 0$, and further solutions, defined near (1, 1) also when $x = y$, should be obtainable from expansions of the F_1 function that contain powers of $x - y$. Before dealing with the domain $x = y$, we investigate the remaining points $(1, \infty)$ and (∞, ∞). We take again the first F_2 solution in (11) and expand it in ${}_2F_1$ functions which are then replaced with (16). This gives

$$F_2(a, b_1, b_2, a + b_1 - c + 1, c - b_1, 1 - x, y)$$
$$= \frac{\Gamma(c - b_1)\Gamma(b_2 - a)}{\Gamma(b_2)\Gamma(c - a - b_1)} (1 - y)^{-a} F_1\left(a, b_1, c - b_1 - b_2, 1 + a - b_2, \frac{1 - x}{1 - y}, \frac{1}{1 - y}\right)$$
$$+ \frac{\Gamma(c - b_1)\Gamma(a - b_2)}{\Gamma(a)\Gamma(c - b_1 - b_2)} (1 - y)^{-b_2} G_2\left(b_1, b_2, c - a - b_1, a - b_2, x - 1, \frac{1}{y - 1}\right), \quad (20)$$

and a similar result by interchange of b_1 and b_2 and the arguments x and y. The second F_2 function in (11) can be maintained in the form in which it appears in (19) since it is then convergent near $(1, \infty)$ as are the functions in (20). Inserting into (11) we obtain

$$F_1(a, b_1, b_2, c, x, y) = \frac{\Gamma(c)\Gamma(b_2 - a)}{\Gamma(c - a)\Gamma(b_2)} (1 - y)^{-a} F_1\left(a, b_1, c - b_1 - b_2, 1 + a - b_2, \frac{1 - x}{1 - y}, \frac{1}{1 - y}\right)$$
$$+ \frac{\Gamma(c)\Gamma(a + b_1 - c)}{\Gamma(a)\Gamma(b_1)} (1 - x)^{c-a-b_1}(1 - y)^{-b_2} F_1\left(c - a, b_2, c - b_1 - b_2, c - a - b_1 + 1, \frac{1 - x}{1 - y}, 1 - x\right)$$
$$+ \frac{\Gamma(c)\Gamma(a - b_2)\Gamma(c - a - b_1)}{\Gamma(a)\Gamma(c - a)\Gamma(c - b_1 - b_2)} (1 - y)^{-b_2} G_2\left(b_1, b_2, c - a - b_1, a - b_2, x - 1, \frac{1}{y - 1}\right)$$
$$= \frac{\Gamma(c)\Gamma(b_1 - a)}{\Gamma(c - a)\Gamma(b_1)} (1 - x)^{-a} F_1\left(a, c - b_1 - b_2, b_2, 1 + a - b_1, \frac{1}{1 - x}, \frac{1 - y}{1 - x}\right) + \frac{\Gamma(c)\Gamma(a + b_2 - c)}{\Gamma(a)\Gamma(b_2)}$$
$$\times (1 - x)^{-b_1}(1 - y)^{c-a-b_2} F_1\left(c - a, b_1, c - b_1 - b_2, c - a - b_2 + 1, \frac{1 - y}{1 - x}, 1 - y\right)$$
$$+ \frac{\Gamma(c)\Gamma(a - b_1)\Gamma(c - a - b_2)}{\Gamma(a)\Gamma(c - a)\Gamma(c - b_1 - b_2)} (1 - x)^{-b_1} G_2\left(b_1, b_2, a - b_1, c - a - b_2, \frac{1}{x - 1}, y - 1\right). \tag{21}$$

Here we encounter two more F_1 solutions that we have not obtained earlier, namely the first function on the rhs in each of the equalities (21). These two functions, which are denoted z_{28} and z_{39} in the table of F_1 solutions,[3] can be transformed into those denoted z_8 and z_9 respectively. The latter solutions occur already in (17) and so the last two F_1 solutions obtained are proportional to F_1 solutions found previously. Actually there exist five known transformations that transform a F_1 function into another F_1 function.[6] A set of ten F_1 function would then suffice to derive the 60 known solutions since 50 additional solutions can be obtained from the five transformations. Among the ten solutions to start with, there must not be any solutions proportional to each other, since if this were the case we would obtain of course less than 60 solutions. As we shall see, we shall have obtained exactly ten nonproportional F_1 solutions when we have completed the task of continuing the F_1 function.

The derivation of the result near (∞, ∞) should now offer no difficulties for the reader, and we write down the result directly:

$$F_1(a, b_1, b_2, c, x, y) = \frac{\Gamma(c)\Gamma(b_2 - a)}{\Gamma(b_2)\Gamma(c - a)} (-y)^{-a} F_1\left(a, b_1, 1 + a - c, 1 + a - b_2, \frac{x}{y}, \frac{1}{y}\right)$$
$$+ \frac{\Gamma(c)\Gamma(a - b_1 - b_2)}{\Gamma(a)\Gamma(c - b_1 - b_2)} (-x)^{-b_1}(-y)^{-b_2} F_1\left(1 + b_1 + b_2 - c, b_1, b_2, 1 + b_1 + b_2 - a, \frac{1}{x}, \frac{1}{y}\right)$$
$$+ \frac{\Gamma(c)\Gamma(a - b_2)\Gamma(b_1 + b_2 - a)}{\Gamma(a)\Gamma(b_1)\Gamma(c - a)} (-x)^{b_2-a}(-y)^{-b_2} G_2\left(1 + a - c, b_2, b_1 + b_2 - a, a - b_2, -\frac{1}{x}, -\frac{x}{y}\right)$$
$$= \frac{\Gamma(c)\Gamma(b_1 - a)}{\Gamma(b_1)\Gamma(c - a)} (-x)^{-a} F_1\left(a, 1 + a - c, b_2, 1 + a - b_1, \frac{1}{x}, \frac{y}{x}\right)$$
$$+ \frac{\Gamma(c)\Gamma(c - b_1 - b_2)}{\Gamma(a)\Gamma(c - b_1 - b_2)} (-x)^{-b_1}(-y)^{-b_2} F_1\left(1 + b_1 + b_2 - c, b_1, b_2, 1 + b_1 + b_2 - a, \frac{1}{x}, \frac{1}{y}\right)$$
$$+ \frac{\Gamma(c)\Gamma(a - b_1)\Gamma(b_1 + b_2 - a)}{\Gamma(a)\Gamma(b_2)\Gamma(c - a)} (-x)^{-b_1}(-y)^{b_1-a} G_2\left(b_1, 1 + a - c, a - b_1, b_1 + b_2 - a, -\frac{y}{x}, -\frac{1}{y}\right). \tag{22}$$

The point (∞, ∞) is an intersection of three singular manifolds, and here again we have to use different solutions on different sides of $x = y$. As was the case near (1, 1), the series are generally not convergent when $|x| = |y|$ since one of the arguments is either x/y or y/x. There exist, however, solutions that are convergent when $x = y$. Following the general idea underlying the derivations in this paper we look for an expansion of the F_1 function that contains powers of $x - y$. One of the well-known transformations supplies such an expansion:

$$F_1(a, b_1, b_2, c, x, y) = (1 - y)^{-a} \times F_1\left(a, b_1, c - b_1 - b_2, c, \frac{y - x}{y - 1}, \frac{y}{y - 1}\right).$$

To the F_1 function thus transformed we apply (17) and obtain

[6] Ref. 3, p. 239.

$$F_1(a, b_1, b_2, c, x, y)$$

$$= \frac{\Gamma(c)\Gamma(c-a-b_1-b_2)}{\Gamma(c-a)\Gamma(c-b_1-b_2)} x^{-a} F_1\left(a, 1+a-c, b_2, 1+a+b_1+b_2-c, \frac{x-1}{x}, \frac{x-y}{x}\right)$$

$$+ \frac{\Gamma(c)\Gamma(a+b_1+b_2-c)}{\Gamma(a)\Gamma(b_1+b_2)} x^{b_1+b_2-c}(1-x)^{c-a-b_1-b_2}$$

$$\times G_2\left(c-b_1-b_2, b_2, a+b_1+b_2-c, 1-b_1-b_2, \frac{1-x}{x}, \frac{x-y}{1-x}\right)$$

$$= \frac{\Gamma(c)\Gamma(c-a-b_1-b_2)}{\Gamma(c-a)\Gamma(c-b_1-b_2)} y^{-a} F_1\left(a, b_1, 1+a-c, 1+a+b_1+b_2-c, \frac{y-x}{y}, \frac{y-1}{y}\right)$$

$$+ \frac{\Gamma(c)\Gamma(a+b_1+b_2-c)}{\Gamma(a)\Gamma(b_1+b_2)} y^{b_1+b_2-c}(1-y)^{c-a-b_1-b_2}$$

$$\times G_2\left(b_1, c-b_1-b_2, 1-b_1-b_2, a+b_1+b_2-c, \frac{y-x}{1-y}, \frac{1-y}{y}\right). \tag{23}$$

This continuation can be used when (19) fails but there are points in the neighborhood of (1, 1) for which neither (19) nor (23) is convergent. Such points are in the domain

$$|x-y| > |1-y| = |1-x| \tag{24}$$

since for such points the expansions of the G_2 functions in (19) and (23) are not convergent. An obvious remedy is to invert the argument $x - y/1 - x$ of the first G_2 function in (23) by expanding it in ${}_2F_1$ functions of this argument and then use (16). This gives

$$G_2(a_1, a_2, b_1, b_2, x, y)$$

$$= \frac{\Gamma(1-b_2)\Gamma(a_2-b_1)}{\Gamma(a_2)\Gamma(1-b_1-b_2)} y^{-b_1}$$

$$\times G_2\left(a_1, b_1+b_2, b_1, a_2-b_1, -xy, \frac{1}{y}\right)$$

$$+ \frac{\Gamma(1-b_2)\Gamma(b_1-a_2)}{\Gamma(b_1)\Gamma(1-b_2-a_2)} y^{-a_2}$$

$$\times F_1\left(a_2+b_2, a_1, a_2, 1+a_2-b_1, -x, -\frac{1}{y}\right), \tag{25}$$

with the aid of which the G_2 function in (23) is defined also in the domain (24). But there are still elusive points, namely points in the domain

$$|x-y| = |1-x| = |1-y|.$$

Writing

$$x = 1 - z_1, \qquad y = 1 - z_2,$$

we have

$$|z_1 - z_2| = |z_1| = |z_2|,$$

or the relation

$$z_1 = e^{\pm\frac{1}{3}i\pi} z_2. \tag{26}$$

Similar difficulties are encountered also in the analytic continuations of the function ${}_2F_1(a, b, c, z)$ to the neighborhood of $z = e^{\pm\frac{1}{3}i\pi}$. The question of convergence for those exceptional points requires a special investigation which, however, we shall not discuss in the present paper.

The two F_1 functions found on the rhs of (23) can be transformed into each other with the aid of one of the five transformations transforming F_1 functions into F_1 functions. The rhs of (23) then contains only one F_1 solution. This solution and the first F_1 function on the rhs of (19) are convergent in the entire neighborhood of (1, 1). They are, however, identical and more than one solution, convergent in the entire neighborhood, cannot be found from the methods we have used here. This statement holds for all intersections of three singular manifolds.

If we apply another transformation,

$$F_1(a, b_1, b_2, c, x, y) = (1-x)^{-b_1}(1-y)^{c-a-b_2}$$

$$\times F_1\left(c-a, b_1, c-b_1-b_2, c, \frac{x-y}{x-1}, y\right),$$

and then use (17) in order to investigate the neighborhood of (∞, ∞), we obtain

$$
\begin{aligned}
&F_1(a, b_1, b_2, c, x, y)\\
&= \frac{\Gamma(c)\Gamma(a-b_1-b_2)}{\Gamma(a)\Gamma(c-b_1-b_2)}(-y)^{a-c}(1-x)^{-b_1}(1-y)^{c-a-b_2}F_1\Big(c-a, b_1, 1-a, 1+b_1+b_2-a, \frac{x-y}{y(x-1)}, \frac{1}{y}\Big)\\
&+ \frac{\Gamma(c)\Gamma(b_1+b_2-a)}{\Gamma(c-a)\Gamma(b_1+b_2)}(-y)^{b_1+b_2-c}(1-x)^{-b_1}(1-y)^{c-a-b_2}\\
&\times G_2\Big(b_1, c-b_1-b_2, 1-b_1-b_2, b_1+b_2-a, \frac{x-y}{1-x}, -\frac{1}{y}\Big)\\
&= \frac{\Gamma(c)\Gamma(a-b_1-b_2)}{\Gamma(a)\Gamma(c-b_1-b_2)}(-x)^{a-c}(1-x)^{c-a-b_1}(1-y)^{-b_2}\\
&\times F_1\Big(c-a, 1-a, b_2, 1+b_1+b_2-a, \frac{1}{x}, \frac{y-x}{x(y-1)}\Big) + \frac{\Gamma(c)\Gamma(b_1+b_2-a)}{\Gamma(c-a)\Gamma(b_1+b_2)}(-x)^{b_1+b_2-c}\\
&\times (1-x)^{c-a-b_1}(1-y)^{-b_2}G_2\Big(c-b_1-b_2, b_2, b_1+b_2-a, 1-b_1-b_2, -\frac{1}{x}, \frac{y-x}{1-y}\Big). \qquad (27)
\end{aligned}
$$

The discussion of convergence here is similar to the one for the neighborhood of (1, 1). The elusive points are

$$x = e^{\pm\frac{1}{3}i\pi}y.$$

This concludes the task of continuing the F_1 function in terms of F_1 and G_2 functions. Actually we have obtained now three independent solutions, convergent for any point in the neighborhood of singular intersections except for the neighborhood of (0, 0), provided (x, y) is not in the exceptional domains discussed above. In order to obtain the general solution near (0, 0) we have only to continue the solutions near, e.g., (1, 1) to near (0, 0) with the aid of (19). Thus the solution

$$F_1(a, b_1, b_2, 1+a+b_1+b_2-c, 1-x, 1-y)$$

gives the two F_1 solutions

$$
\begin{aligned}
&x^{-b_1}y^{b_1-c+1}F_1\Big(1+b_1+b_2-c, b_1, a-c+1, b_1-c+2, \frac{y}{x}, y\Big),\\
&x^{b_2-c+1}y^{-b_2}F_1\Big(1+b_1+b_2-c, a-c+1, b_2, b_2-c+2, x, \frac{x}{y}\Big), \qquad (28)
\end{aligned}
$$

and the two G_2 solutions

$$
\begin{aligned}
&x^{1-c}G_2(a-c+1, b_2, c-1, b_1-c+1, -x, -(y/x)),\\
&y^{1-c}G_2(b_1, a-c+1, b_2-c+1, c-1, -(x/y), -y) \qquad (29)
\end{aligned}
$$

whereas the two other F_1 solutions in (19) give the F_1 solutions not previously obtained:

$$
\begin{aligned}
&(y-x)^{1-b_1-b_2}y^{b_1+b_2-c}(1-x)^{b_2-1}(1-y)^{c-a-b_2}\\
&\qquad\times F_1\Big(1-b_2, 1-a, c-b_1-b_2, 2-b_1-b_2, \frac{y-x}{1-x}, \frac{y-x}{y(1-x)}\Big),\\
&(x-y)^{1-b_1-b_2}x^{b_1+b_2-c}(1-x)^{c-a-b_1}(1-y)^{b_1-1}\\
&\qquad\times F_1\Big(1-b_1, c-b_1-b_2, 1-a, 2-b_1-b_2, \frac{x-y}{x(1-y)}, \frac{x-y}{1-y}\Big). \qquad (30)
\end{aligned}
$$

and the four G_2 solutions

$$
\begin{aligned}
&(1-x)^{-b_1}(1-y)^{c-a-b_2}y^{1-c}G_2\Big(b_1, 1-a, 1-b_1-b_2, c-1, \frac{x-y}{y(1-x)}, -y\Big),\\
&(1-x)^{c-a-b_1}(1-y)^{-b_2}x^{1-c}G_2\Big(1-a, b_2, c-1, 1-b_1-b_2, -x, \frac{y-x}{x(1-y)}\Big),\\
&(1-x)^{c-a-b_1}(1-y)^{c-b_2-1}(x-y)^{1-c}G_2\Big(c-b_1-b_2, 1-a, b_2-c+1, c-1, \frac{x(1-y)}{y-x}, \frac{y-x}{1-y}\Big),\\
&(1-x)^{c-b_1-1}(1-y)^{c-a-b_2}(y-x)^{1-c}G_2\Big(1-a, c-b_1-b_2, c-1, b_1-c+1, \frac{x-y}{1-x}, \frac{y(1-x)}{x-y}\Big). \qquad (31)
\end{aligned}
$$

We have here the same complication as that discussed for the solutions in the vicinity of (1, 1) and (∞, ∞). Thus there is in (28)–(31) only one solution convergent when $|x - y| > |x| = |y|$. The missing solutions can be obtained in terms of F_1 functions if (25) is used on the G_2 functions in (31). Since we obtain then F_1 solutions that are easily obtained from a simple transformation of F_1 solutions we have already derived, it is not necessary to write them down.

Actually we have so far derived ten nonproportional F_1 solutions and fifteen nonproportional G_2 solutions. With the aid of the five known transformations of F_1 functions into F_1 functions we can obtain a total of 60 solutions in terms of F_1 functions. Erdélyi[7] has given three transformation formulas that transform a G_2 function into a G_2 function, and thus we can from the fifteen G_2 solutions obtain another 45 solutions. These 120 solutions are all solutions known at present that are expressible in terms of F_1 and G_2 functions. With the aid of this set of solutions, the general solution can be obtained in terms of convergent power series, apart from the exceptional domains discussed above and for which we believe no convergent hypergeometric series are possible. In this respect the set constitutes what we have termed a complete set of solutions. We present here the ten F_1 solutions obtained. The number found after each solution refers to the formula in the present text in which the functions can be found.

Table of the ten distinct F_1 solutions

$$F_1(a, b_1, b_2, c, x, y), \tag{2}$$

$$(1-x)^{-b_1}(1-y)^{c-a-b_2}F_1\left(c-a, b_1, c-b_1-b_2, c-a-b_2+1, \frac{1-y}{1-x}, 1-y\right), \tag{15}$$

$$(1-x)^{c-a-b_1}(1-y)^{-b_2}F_1\left(c-a, c-b_1-b_2, b_2, c-a-b_1+1, 1-x, \frac{1-x}{1-y}\right), \tag{15}$$

$$(-y)^{-a}F_1\left(a, b_1, 1+a-c, 1+a-b_2, \frac{x}{y}, \frac{1}{y}\right), \tag{17}$$

$$(-x)^{-a}F_1\left(a, 1+a-c, b_2, 1+a-b_1, \frac{1}{x}, \frac{y}{x}\right), \tag{17}$$

$$F_1(a, b_1, b_2, 1+a+b_1+b_2-c, 1-x, 1-y), \tag{19}$$

$$(-x)^{-b_1}(-y)^{-b_2}F_1\left(1+b_1+b_2-c, b_1, b_2, 1+b_1+b_2-a, \frac{1}{x}, \frac{1}{y}\right), \tag{22}$$

$$x^{-b_1}y^{b_1-c+1}F_1\left(1+b_1+b_2-c, b_1, a-c+1, b_1-c+2, \frac{y}{x}, y\right), \tag{28}$$

$$x^{b_2-c+1}y^{-b_2}F_1\left(1+b_1+b_2-c, a-c+1, b_2, b_2-c+2, x, \frac{x}{y}\right), \tag{28}$$

$$(x-y)^{1-b_1-b_2}y^{b_1+b_2-c}(1-x)^{b_2-1}(1-y)^{c-a-b_2} \times F_1\left(1-b_2, 1-a, c-b_1-b_2, 2-b_1-b_2, \frac{y-x}{1-x}, \frac{y-x}{y(1-x)}\right). \tag{30}$$

Similarly, there are among the G_2 solutions that we have derived fifteen functions which can not be transformed into each other by expansions in ${}_2F_1$ functions and use of the Kummer connections.

[7] Ref. 2, p. 148.

These fifteen solutions are:

Table of the fifteen distinct G_2 solutions

$$x^{-b_1}(1-y)^{-b_2}G_2\left(b_1,\ b_2,\ c-a-b_1,\ b_1-c+1,\ \frac{1-x}{x},\ \frac{y}{1-y}\right), \tag{15}$$

$$(1-x)^{-b_1}y^{-b_2}G_2\left(b_1,\ b_2,\ b_2-c+1,\ c-a-b_2,\ \frac{x}{1-x},\ \frac{1-y}{y}\right), \tag{15}$$

$$x^{-b_1}G_2\left(b_1,\ b_2,\ a-b_1\ 1+b_1-c,\ -\frac{1}{x},\ -y\right), \tag{17}$$

$$y^{-b_2}G_2\left(b_1,\ b_2,\ 1+b_2-c,\ a-b_2,\ -x,\ -\frac{1}{y}\right), \tag{17}$$

$$(1-x)^{c-a-b_1-b_2}G_2\left(c-b_1-b_2,\ b_2,\ a+b_1+b_2-c,\ c-a-b_2,\ x-1,\ \frac{1-y}{x-1}\right), \tag{19}$$

$$(1-y)^{c-a-b_1-b_2}G_2\left(b_1,\ c-b_1-b_2,\ c-a-b_1,\ a+b_1+b_2-c,\ \frac{1-x}{y-1},\ y-1\right), \tag{19}$$

$$(1-x)^{-b_1}G_2\left(b_1,\ b_2,\ a-b_1,\ c-a-b_2,\ \frac{1}{x-1},\ y-1\right), \tag{21}$$

$$(1-y)^{-b_2}G_2\left(b_1,\ b_2,\ c-a-b_1\ a-b_2,\ x-1,\ \frac{1}{y-1}\right), \tag{21}$$

$$x^{b_2-a}y^{-b_2}G_2\left(a-c+1,\ b_2,\ b_1+b_2-a,\ a-b_2,\ -\frac{1}{x},\ -\frac{x}{y}\right), \tag{22}$$

$$x^{-b_1}y^{b_1-a}G_2\left(b_1,\ a-c+1,\ a-b_1,\ b_1+b_2-a,\ -\frac{y}{x},\ -\frac{1}{y}\right), \tag{22}$$

$$x^{b_1+b_2-c}(1-x)^{c-a-b_1-b_2}G_2\left(c-b_1-b_2,\ b_2,\ a+b_1+b_2-c,\ 1-b_1-b_2,\ \frac{1-x}{x},\ \frac{x-y}{1-x}\right), \tag{23}$$

$$x^{b_1+b_2-c}\ (1-x)^{c-a-b_1}(1-y)^{-b_2}G_2\left(c-b_1-b_2,\ b_2,\ b_1+b_2-a,\ 1-b_1-b_2,\ -\frac{1}{x},\ \frac{y-x}{1-y}\right), \tag{27}$$

$$x^{1-c}G_2\left(a-c+1,\ b_2,\ c-1,\ b_1-c+1,\ -x,\ -\frac{y}{x}\right), \tag{29}$$

$$y^{1-c}G_2\left(b_1,\ a-c+1,\ b_2-c+1,\ c-1,\ -\frac{x}{y},\ -y\right), \tag{29}$$

$$(1-x)^{-b_1}(1-y)^{c-a-b_2}y^{1-c}G_2\left(b_1,\ 1-a,\ 1-b_1-b_2,\ c-1,\ \frac{x-y}{y(1-x)},\ -y\right). \tag{31}$$

Any other G_2 solution found in the text can be transformed into one of the fifteen functions in the table by means of the transformation formulas[7]

$$G_2(a_1,\ a_2,\ b_1,\ b_2,\ x,\ y) = (1+x)^{-b_2}(1-xy)^{-a_2}G_2\left(1-a_1-b_1,\ a_2,\ b_1,\ b_2,\ \frac{-x}{1+x},\ \frac{y(1+x)}{1-xy}\right)$$

$$= (1+y)^{-b_1}(1-xy)^{-a_1}G_2\left(1-a_2-b_2,\ a_1,\ b_2,\ b_1,\ \frac{-y}{1+y},\ \frac{x(1+y)}{1-xy}\right)$$

$$= (1+x)^{-b_2}(1+y)^{-b_1}(1-xy)^{1-a_1-a_2}G_2\left(1-a_1-b_1,\ 1-a_2-b_2,\ b_1,\ b_2,\ -\frac{x(1+y)}{1+x},\ -\frac{y(1+x)}{1+y}\right), \tag{32}$$

which, applied to the fifteen solutions in the table, give us the aforementioned additional 45 solutions in terms of G_2 functions.

By expressing the F_1 function in ${}_2F_1$ functions, and using systematically the analytic continuations and the transformation theory of the latter functions with the purpose of continuing analytically the F_1 function to the neighborhoods of its singular points, we have obtained what we believe to be all solutions expressible in terms of F_2 and G_2 functions, an extensive transformation theory of the solutions and connections between the various functions. From these connections the analytic properties of the functions can be obtained for the whole complex domain of the variables x and y, except possibly for certain rather special domains, which were not investigated.

We have already mentioned that the results obtained here can be used to derive solutions of the system of partial differential equations associated with the Lauricella function F_D of more than two variables, and the possibility that the properties of the hypergeometric functions of two and more variables can be derived from the properties of the ${}_2F_1$ function suggests itself. An attempt to obtain the analytic continuations of the Appell function F_2 has shown, not unexpectedly, that hypergeometric functions other than the ${}_2F_1$ functions, particularly ${}_3F_2$ functions, must be taken into account. It is possible, however, that these functions can be studied conveniently in terms of ${}_2F_1$ functions as an intermediate stage, and that the properties of the more complex hypergeometric functions can be derived from less complex functions using a technique similar to the one outlined in this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 5, NUMBER 3 MARCH 1964

The Complete High-Energy Behavior of Ladder Diagrams in Perturbation Theory*

J. C. POLKINGHORNE
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 11 October 1963)

The Mellin-transform method for obtaining the high-energy behavior of Feynman integrals is modified and applied to the set of ladder diagrams. The *complete* set of terms of the form $s^{-1}(\ln s)^n$ is summed, and gives an equation for the trajectory function which is analogous to that obtained by Fredholm methods for a Yukawa potential. A perturbation expansion for $\alpha(t)$ valid for t large is given, and the threshold behavior investigated. The results confirm the reliability of the perturbation-theory method of investigation. They also exhibit directly the connection between high-energy behavior and the poles of the scattering amplitude.

1. INTRODUCTION

INVESTIGATIONS into the high-energy behavior of perturbation theory have been used to exhibit Regge poles,[1] Reggeise elementary particles,[2] exhibit Regge cuts,[3,4] determine singularities of Regge trajectories,[5] and investigate the high-energy behavior of production processes.[6] In all this work the method used has been to take a class of Feynman diagrams, evaluate the leading asymptotic behavior of each diagram, and then sum these contributions. Of course this is not a rigorous method, but where comparison with rigorous results is possible, it is found not to be misleading. Consequently we believe it to be a reliable heuristic method.

In a recent paper, Trueman and Yao[7] have shown how to evaluate more than just the leading term in the asymptotic behavior of individual diagrams. They sum up what might at first sight seem the most important corrections for the set of ladder diagrams with the extraordinary result of turning a single Regge pole into a double pole! They also point out that, if further corrections are added in the double pole, it again becomes a single pole. Such capricious behavior is scarcely acceptable, and it is clearly necessary for the vindication of the heuristic value of this fruitful method of investigation that *all* terms in the asymptotic behavior should be summed and be shown to give a reasonable Regge behavior.

Before this can be done it is necessary to modify the Mellin-transform method[4,7] since it turns out that its original form does not give the correct reality properties. This is done in Sec. 2. In Sec. 3, the series of terms of the form $s^{-1}(\ln s)^n$ is completely summed for ladder diagrams composed of bosons of unit mass interacting through a ϕ^3 term. The result is an implicit equation for the trajectory function $\alpha(t)$ which is analogous to the equation obtained by Cassandro, Cini, Jona-Lasinio, and Sertorio[8] for interaction through a Yukawa potential. A series expansion for the leading (single) Regge pole, valid for large s, is given in Sec. 4. In Sec. 5 the behavior in the neighborhood of the elastic threshold is investigated.

We conclude that a complete summation confirms the value of the method though it remains, of course, without a strictly rigorous justification.

2. MELLIN-TRANSFORM METHOD

The use of Mellin transforms for evaluating high-energy behavior is due to Bjorken and Wu[4] and has been further exploited by Trueman and Yao.[7] However, the form in which they develop the theory gives a contribution

$$f \sim \Gamma(-\alpha)e^{-i\frac{1}{2}\pi\alpha}s^{\alpha} \qquad (1)$$

from a Regge pole of trajectory α. This form makes f real along the imaginary axis for real α. We know,

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research, OAR, through the European Office, Aerospace Research, United States Air Force.

[1] J. C. Polkinghorne, J. Math. Phys. 4, 503 (1963); P. G. Federbush and M. T. Grisaru, Ann. Phys. (N. Y.) 22, 263, 299 (1963); I. G. Halliday, Nuovo Cimento, 30, 177 (1963).

[2] M. Gell-Mann and M. L. Goldberger, Phys. Rev. Letters 9, 275 (1962); M. Gell-Mann, M. L. Goldberger, F. E. Low, and F. Zachariasen, Phys. Letters 4, 265 (1963), and to be published.

[3] J. C. Polkinghorne, J. Math. Phys. 4, 1396 (1963).

[4] J. D. Bjorken and T. T. Wu, Phys. Rev. 130, 2566 (1963).

[5] J. C. Polkinghorne, J. Math. Phys. 4, 1393 (1963).

[6] I. G. Halliday and J. C. Polkinghorne, Phys. Rev. 132, 852 (1963).

[7] T. L. Trueman and T. Yao, Phys. Rev. 132, 2741 (1963).

[8] M. Cassandro, M. Cini, G. Jona-Lasinio, and L. Sertorio, Nuovo Cimento 28, 1351 (1963), referred to as C³JS.

however, that the reality properties of the correct theory of ladder diagrams must give a real scattering amplitude for real α when s is real and negative. It is necessary, therefore, to modify the approach slightly.

In using Mellin transforms it is necessary to avoid cuts in the variable being transformed. We therefore evaluate the limit with $s \to -\infty$, and for convenience we also take $t < 4$, that is below its threshold. The scattering amplitude corresponding to the ladder with $(n+1)$ rungs is of the form

$$f_n(s,t) = g^2\left(\frac{-g^2}{16\pi^2}\right)^n \Gamma(n+1) \times \int_0^1 d\alpha_i\, d\beta_i \frac{\delta(\sum\alpha + \sum\beta - 1)C^{n-1}}{D^{n+1}}, \quad (2)$$

where the α's are the Feynman parameters of the rungs and the β's the Feynman parameters of the remaining lines. C and D are the well-known Feynman numerator and denominator functions. C is positive definite, and for $s < (n+1)^2$, $t < 4$, D is negative definite. Therefore, in evaluating our limit, we may rewrite (2) in the form

$$f_n(s,t) = (-1)^{n+1} g^2\left(\frac{-g^2}{16\pi^2}\right)^n \times \int_0^\infty dx_1 \cdots dx_{n+1}\, dy_i \Delta^{-2} e^{Q}, \quad (3)$$

where x_i are parameters associated with the rungs, and y_i parameters associated with the remaining lines of the ladder, Δ is the same function of the x_i and y_i that C is of the α_i and β_i, and Q is similarly derived from D/C. (3) may be transformed into (2) by writing

$$x_i = \rho\alpha_i\,, \qquad y_i = \rho\beta_i\,, \quad (4)$$

and performing the ρ integration.

Q may be written in the form

$$Q = -\left[\sigma \prod_{i=1}^{n+1} x_i/\Delta + J(x,y,t)\right], \quad (5)$$

where $\sigma = -s$. We are to evaluate the limit $\sigma \to \infty$. Accordingly, we take the Mellin transform

$$L_n(\alpha,t) = \int_0^\infty d\sigma\, \sigma^{-\alpha-1} f_n(-\sigma,t) = \Gamma(-\alpha)(-1)^{n+1} g^2\left(\frac{-g^2}{16\pi^2}\right)^{n+1} \times \int_0^\infty dx_1 \cdots dx_n\, dy_i \Delta^{-2-\alpha} e^{-J} \prod_{i=1}^{n+1} x_i^{\alpha}. \quad (6)$$

Following Trueman and Yao,[7] we extend the region of definition of (6) to exhibit the multiple pole at $\alpha = -1$ by integrating by parts and obtain

$$L_n(\alpha,t) = \frac{\Gamma(-\alpha)}{(\alpha+1)^{n+1}} g^2\left(\frac{-g^2}{16\pi^2}\right)^{n+1} \int_0^\infty dx_1 \cdots dx_n\, dy_i \times \prod_{i=1}^{n+1} x_i^{\alpha+1} \frac{\partial^{n+1}}{\partial x_1 \cdots \partial x_{n+1}} [\Delta^{-2-\alpha} e^{-J}]. \quad (7)$$

The coefficient of $(\alpha+1)^{-m-1}$ in (7) is

$$L_n(m) = \Gamma(-\alpha) g^2\left(\frac{-g^2}{16\pi^2}\right)^{n+1} \times \sum_{l=0}^{n-m} (-1)^l \int_0^\infty dx_1 \cdots dx_n\, dy_i \times \frac{\left(\ln \prod_{i=1}^{n+1} x_i\right)^{n-m-l}}{(n-m-l)!} \frac{\partial^{n+1}}{\partial x_1 \cdots \partial x_{n+1}} \left[\frac{e^{-J}(\ln\Delta)^l}{\Delta\, l!}\right]. \quad (8)$$

The approximation of taking only leading asymptotic behavior corresponds to summing $L_n(n)$.

The form in which we have set up the Mellin-transform method in this section gives contributions of the form

$$f \sim \Gamma(-\alpha)\,(-s)^{\alpha}, \quad (9)$$

which has the correct reality properties. When α passes through a nonnegative integer, the poles of the gamma function give poles of the scattering amplitude. This establishes directly in perturbation theory the connection between high-energy behavior and bound states.

3. SUMMATION OF THE SERIES

The first term in the integrand of (8) may be expanded by the multinomial theorem to give a sum of terms of the form

$$\prod_{i=1}^{n+1} \frac{(\ln x_i)^{s_i}}{s_i!}, \qquad \sum s_i = n - m - l. \quad (10)$$

Any x_i for which $s_i = 0$ corresponds to an integration which can be performed immediately to give x_i put equal to zero in the expression in square brackets and the corresponding differential operator removed. Putting $x_i = 0$ is equivalent to contracting the corresponding rung of the ladder. Both J and Δ then factorize into the product of two terms corresponding to the subdiagrams on either side of the contracted line. It is this factorization which enables the series to be summed. When Δ factorizes, the logarithm in the square brackets may also be expanded using the binomial theorem. Thus the basic units which will occur in the summation are

either of the form

$$F_k(s_1, \cdots, s_k; \beta) \equiv -\left(\frac{-g^2}{16\pi^2}\right)^{k+1} \int_0^\infty dx_1 \cdots dx_k\, dy_i \times \prod_{i=1}^{k} \frac{(\ln x_i)^{s_i}}{s_i!} \frac{\partial^k}{\partial x_1 \cdots \partial x_k} \left[\frac{(-\ln \Delta_k)^\beta}{\beta!} \frac{e^{-J_k}}{\Delta_k}\right], \quad (11)$$

with $s_i > 0$ and Δ_k and J_k corresponding to Fig. 1, or of the form

$$G_k(s_1, \cdots, s_k; \beta) \equiv -g\left(\frac{-g^2}{16\pi^2}\right)^{k} \int_0^\infty dx_1 \cdots dx_k\, dy_i \times \prod_{i=1}^{k} \frac{(\ln x_i)^{s_i}}{s_i!} \frac{\partial^k}{\partial x_1 \cdots \partial x_k} \left[\frac{(-\ln \bar{\Delta}_k)^\beta}{\beta!} \frac{\rho^{-\bar{J}_k}}{\bar{\Delta}_k}\right], \quad (12)$$

with $s_i > 0$ and $\bar{\Delta}_k$ and $\bar{J}_k$ corresponding to Fig. 2. It is convenient to allow the index k to take the value zero, corresponding to contributions associated with the simple bubble diagram.

We now evaluate

$$L = \sum_{n,m} L_n(m)(1 + \alpha)^{-m-1}. \quad (13)$$

The simplest way to do this is as follows. Each term in the sum has a coefficient formed of the product of a G, a number of F's, and another G.

These terms are formed by contracting lines in ladder diagrams of appropriate length. It follows that a term involving a given number of F factors of given types may be formed in just the number of distinct ways in which the F factors can be arranged in order. This number is a multinomial coefficient. When the set of terms corresponding to f F factors is summed, disregarding the powers of $(1 + \alpha)^{-1}$, the result is $G(\sum F)^f G$. However, the power of $(1 + \alpha)^{-1}$ corresponding to a given term is in fact $1 + \sum n_i$ where the n associated with an F factor is given by

$$n = 1 - \sum (s_i - 1) - \beta, \quad (14)$$

and the n associated with a G factor is given by

$$n = -\sum (s_i - 1) - \beta. \quad (15)$$

The sum could easily be performed if the power of $(1 + \alpha)^{-1}$ associated with a term with f F factors were $f + 1$, for then the coefficient of $(1 + \alpha)^{-(f+1)}$ would be just

$$(\sum G)\cdot(\sum F)^f \cdot(\sum G), \quad (16)$$

giving a geometric series. It is convenient therefore to introduce the functions

$$F_k'(\alpha; s_1, \cdots, s_k; \beta) \equiv -\left(\frac{-g^2}{16\pi^2}\right)^{k+1} \int_0^\infty dx_1 \cdots dx_k\, dy_i$$

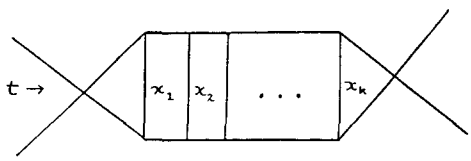

FIG. 1. A contracted interior subdiagram.

$$\times \prod_{i=1}^{k} \frac{(\ln x_i)^{s_i}(1 + \alpha)^{s_i - 1}}{s_i!} \times \frac{\partial^k}{\partial x_1 \cdots \partial x_k} \left[\frac{(-\ln \Delta_k)^\beta (1 + \alpha)^\beta e^{-J_k}}{\Delta_k \beta!}\right], \quad (17)$$

and $G_k'(\alpha; s_1, \cdots s_k; \beta)$ similarly defined. These functions will appear as coefficients of the desired powers of $(1 + \alpha)^{-1}$. Then

$$L = \mathcal{G}\mathcal{F}^{-1}\mathcal{G}, \quad (18)$$

where

$$\mathcal{F}(\alpha) \equiv \alpha + 1 - \sum_{k=0}^{\infty} \bar{F}_k, \quad (19)$$

$$\mathcal{G}(\alpha) \equiv g + \sum_{k=1}^{\infty} \bar{G}_k, \quad (20)$$

$$\bar{F}_k(\alpha) \equiv \sum_{s_i,\beta} F_k'(\alpha; s_1, \cdots s_n, \beta) = -\left(\frac{-g^2}{16\pi^2}\right)^{k+1} \int_0^\infty dx_1 \cdots dx_k\, dy_i \times \prod_{i=1}^{k} \frac{x_i^{\alpha+1} - 1}{\alpha + 1} \frac{\partial^k}{\partial x_1 \cdots \partial x_k} [\Delta_k^{-2-\alpha} e^{-J_k}], \quad (21)$$

$$\bar{G}_k(\alpha) \equiv \sum_{s_i,\beta} G_k'(\alpha; s_1, \cdots s_k; \beta), \quad (22)$$

and has an integral expression similar to (21) in terms of $\bar{\Delta}_k$ and $\bar{J}_k$.

The Regge poles of the theory are given by the solutions of

$$\mathcal{F}(\alpha) = 0. \quad (23)$$

This equation is the analogue of Eq. (11) of C²JS, to which it is very similar in form. The functions $\bar{F}_k$ and $\bar{G}_k$ are, of course, functions of t. For $t < 4$, Eq. (23) is a real equation for α. The integrals in (21) can be rewritten in a Feynman-like form which can then be used to continue (23) above $t = 4$. An example of this is given in Sec. 5.

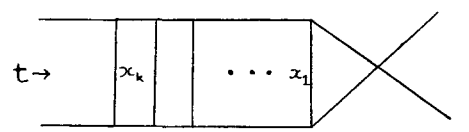

FIG. 2. A contracted end subdiagram.

4. PERTURBATION SOLUTION

A perturbative solution of Eq. (23) is easily found. The first few terms are given by

$$\alpha + 1 = F_0(0) + [F_1(1;0) + F_0(1)F_0(0)] + [F_2(1,1;0) + F_1(1;1)F_0(0) + F_0(2)F_0(0)^2 + F_0(1)F_1(1;0) + F_0^2(1)F_0(0)] + O(g^8). \quad (24)$$

The first term on the right of Eq. (24) is the term obtained from taking the leading asymptotic behavior only.[1] All the functions appearing in (24) are, of course, functions of t, the momentum transfer. We may expect the series (18) to give the correct solution when t is large, but it cannot converge when t is near threshold because $F_0(n)$ is infinite at $t = 4$. The threshold behavior requires, therefore, a separate discussion which is given in the next section.

5. THRESHOLD BEHAVIOR

The reason for the term-by-term divergence of the perturbation solution (24) near threshold is the phenomenon of the condensation of poles near Re $\alpha = -\frac{1}{2}$ at threshold.[9] Consequently, a different form of approximation to Eq. (23) is required to investigate the threshold behavior. The discussion we give here closely parallels that given in C^{2}JS for the Yukawa potential.

One may expect that the appropriate approximation results from working with the $\bar{F}_k$ functions rather than the F_k functions. The first approximation is given by

$$\alpha + 1 - \bar{F}_0(\alpha; t) = 0, \quad (25)$$

where we now exhibit explicitly the t dependence of $\bar{F}_0$. The expression for $\bar{F}_0$ is most easily discussed if the integral corresponding to (21) is rewritten in the more familiar Feynman representation. This gives

$$\bar{F}_0(\alpha; t) = \frac{g^2}{16\pi^2}\Gamma(-\alpha)\int_0^1 \frac{du}{[\epsilon u(1-u) + (1-2u)^2]^{-\alpha}}, \quad (26)$$

with

$$\epsilon = 4 - t. \quad (27)$$

The asymptotic form of (26) near $\epsilon = 0$ may be evaluated by elementary means to be

$$\bar{F}_0 \sim \frac{g^2}{16\pi^2}\left[\frac{\Gamma(-\alpha)}{2\alpha+1} - \frac{\pi\Gamma(\frac{1}{2})\Gamma(\alpha+\frac{3}{2})}{2\sin\pi(\alpha+\frac{1}{2})}(\tfrac{1}{4}\epsilon)^{\alpha+\frac{1}{2}}\right]. \quad (28)$$

The threshold behavior[9] of the Regge poles arises from the poles of (28) at $\alpha = -\frac{1}{2}$. There is a trajectory given by

$$\alpha \sim -\tfrac{1}{2} + \frac{g^2\pi^{\frac{1}{2}}}{16\pi^2}\left[1 - (\tfrac{1}{4}\epsilon)^{g^2\pi^{\frac{1}{2}}/16\pi^2}\right]. \quad (29)$$

In addition, since (25) is of the form of Eq. (53) of C^{2}JS, it will have other solutions in which Re $\alpha \to -\frac{1}{2}$ as $\epsilon \to 0$.

6. CONCLUSION

The principal result of the paper is the derivation of Eq. (23) for the trajectory function $\alpha(t)$. Its similarity to the form of the corresponding equation for a Yukawa potential exhibits the close connection between interaction through a Yukawa potential and interaction through the Bethe–Salpeter ladder approximation. The methods developed here are capable of application to more complicated sets of diagrams, and also to evaluating terms of the form $s^{-n}(\ln s)^m$ $(n > 1)$ for ladder diagrams. We believe that the results so far obtained confirm the value of this method of investigation. An agreeable feature is that the connection between high-energy behavior and bound-state poles is directly exhibited by Eq. (9) without recourse to complex angular momentum.

[9] V. N. Gribov and I. Ya. Pomeranchuk, Phys. Rev. Letters 9, 238 (1962); B. R. Desai and R. G. Newton (to be published).