JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 2, NUMBER 6 NOVEMBER–DECEMBER, 1961

A Criterion for the Free Character of Fields*

JAN LOPUSZANSKI†
Institute of Mathematical Sciences, New York University, New York, New York
(Received April 3, 1961)

It is shown that a field is a free field as soon as the states generated by the Heisenberg field and the incoming field operators operating on the vacuum coincide [statement (i)]. Several conclusions are drawn from statement (i) concerning the strong convergence of the field for t tending to infinity [statement (ii)], the uselessness of the local clothed particle representation [statement (iii)], and the diagonalization of the Hamiltonian [statement (iv)], as well as the time behavior of the mathematical vacuum [statement (v)].

I. INTRODUCTION

IN this note we prove [statement (i)], under pretty weak conditions, that a field is a free field satisfying canonical commutation relations provided that the states generated by the field operator and the incoming field operator operating on the vacuum coincide. We use throughout the formalism of Lehmann, Symanzik, and Zimmermann. Our statement (i) is closely related to that made recently by Federbush and Johnson.[1] They gave a rigorous proof of it using powerful mathematical tools. Our proof is an elementary one although not so general as theirs (e.g., we assume the asymptotic conditions to be valid). Our argument is applicable also to some nonrelativistic field theories. Our statement is also related to the work of Greenberg[2] on clothed particle representation in relativistic field theories as well as to the theorem due to Haag.[3]

From the basic statement (i) some conclusions can be drawn. It follows for instance that, for each physically reasonable theory, the field renormalization constant must be necessarily different from one [statement (ii)]. It follows also [statement (iii)] that the concept of local clothing operation (for definition see Sec. III B) becomes useless in the relativistic theory even if one uses the nonorthogonal set of asymptotic stationary states (for definition see Van Hove[4]). We show also [statement (iv)] that it may be sometimes sufficient to find the vacuum and one particle eigenstates of the Hamiltonian expressed in terms of the Heisenberg operators at $t=0$ (in the Schrödinger picture) to achieve complete diagonalization of this Hamiltonian. This diagonalization of the Hamiltonian has, however, a kinematic significance only, it does not solve the dynamics of the problem. Finally we show [statement (v)] that in a physically reasonable theory with interaction the mathematical (bare) vacuum for finite time t belongs to a different Hilbert space than the physical vacuum.

II. THE BASIC STATEMENT

Let us consider a relativistic real scalar renormalized field $A(x)$ given by the formula[5]

$$A(x)\equiv\frac{1}{(2\pi)^{\frac{3}{2}}}\int\frac{d^3k}{(2k_0)^{\frac{1}{2}}}\{e^{ikx}a(\mathbf{k},x_0)+e^{-ikx}a^+(\mathbf{k},x_0)\}$$
$$=A_{\text{in}}(x)-\int\Delta^{\text{ret}}(x-y)j(y)dy, \quad (1)$$

* This work was supported by the National Sciences Foundation while the author was a temporary member at the Institute.
† On leave of absence from the Institute of Theoretical Physics, University of Wroclaw, Wroclaw, Poland.
[1] P. G. Federbush and K. A. Johnson, Phys. Rev. **120**, 1926 (1960). See also R. Jost, Proc. Kiev Conference 1959, as well as B. Schroer, unpublished thesis, Hamburg, 1958.
[2] O. W. Greenberg, Phys. Rev. **115**, 706 (1959).
[3] R. Haag, Kgl. Danske. Videnskab. Selskab, Mat-fys. Medd. **29**, No. 12 (1955).
[4] L. Van Hove, Physica **21**, 901 (1955); **22**, 343 (1956).
[5] We assume Z_3^{-1} to be finite in
$$\langle 0|[a(\mathbf{k},x_0),\ a^+(\mathbf{k}',x_0)]|0\rangle=(1/Z_3)\delta(\mathbf{k}-\mathbf{k}').$$

where

$$A_{\text{in}}(x)\equiv A_{\text{in}}^{(+)}(x)+A_{\text{in}}^{(-)}(x)=\frac{1}{(2\pi)^{\frac{3}{2}}}\int\frac{d^3k}{(2k_0)^{\frac{1}{2}}}$$
$$\times\{e^{ikx}a_{\text{in}}(\mathbf{k})+e^{-ikx}a_{\text{in}}^{+}(\mathbf{k})\}\quad(2)$$

with

$$[a_{\text{in}}(\mathbf{k}),a_{\text{in}}^{+}(\mathbf{k}')]=\delta(\mathbf{k}-\mathbf{k}'),\text{ otherwise }0.$$
$$k_0=+(\mathbf{k}^2+m^2)^{\frac{1}{2}};\quad kx=\mathbf{kx}-k_0x_0.$$

The operator $j(y)$ accounts for the interaction of the field $A(x)$ with itself or with another field.

We assume[6]

$$[j(\mathbf{y},t),A(\mathbf{x},t)]=0.\quad(3)$$

We assume in addition that $a(\mathbf{k},t)$ converges[7] weakly to $a_{\text{in}}(\mathbf{k})$ for $t\to-\infty$ and that

$$a_{\text{in}}^{+}(\mathbf{k})\cdots a_{\text{in}}^{+}(\mathbf{k}_n)\,|0\rangle\equiv|\mathbf{k}_1,\cdots\mathbf{k}_2\rangle,\ n=0,1,2,\cdots$$

form a complete set of eigenfunctions of the energy momentum four-vector $\mathbf{P}$, H. The normalizable physical vacuum state $|0\rangle$ is defined by

$$a_{\text{in}}(\mathbf{k})\,|0\rangle=0.\quad(4)$$

We assume that the physical vacuum and the one particle states are not degenerate. The metric is positive definite.

Under these assumptions we have

Statement (i): If

$$A(x)\,|0\rangle=A_{\text{in}}(x)\,|0\rangle,\quad(5)$$

then $A(x)=A_{\text{in}}(x)$, i.e., the field under consideration is a free field.

Proof: From

$$A(x)\,|0\rangle=A_{\text{in}}(x)\,|0\rangle-\int\Delta^{\text{ret}}(x-y)j(y)\,|0\rangle dy$$

and (5) it follows that

$$j(x)\,|0\rangle=0.\quad(6)$$

On account of (3) and (5) we have

$$A(\mathbf{y},t)j(\mathbf{x},t)\,|0\rangle=j(\mathbf{x},t)A_{\text{in}}(\mathbf{y},t)\,|0\rangle$$

which vanishes because of (6). We insert now for $A_{\text{in}}(\mathbf{y},t)$ the expression (2) and taking into account (4) we get

$$\int d^3k\frac{1}{(2k_0)^{\frac{1}{2}}}e^{-i(\mathbf{ky}-k_0t)}j(\mathbf{x},t)a_{\text{in}}^{+}(\mathbf{k})\,|0\rangle=0.$$

The Fourier transform with respect to $\mathbf{y}$ must vanish too, viz.,

$$j(x)a_{\text{in}}^{+}(\mathbf{k})\,|0\rangle\equiv j(x)\,|\mathbf{k}\rangle=0.\quad(7)$$

From

$$A(x)\,|\mathbf{k}\rangle=A_{\text{in}}(x)\,|\mathbf{k}\rangle-\int\Delta^{\text{ret}}(x-y)j(y)\,|\mathbf{k}\rangle dy$$

and (7) it follows that

$$A(x)\,|\mathbf{k}\rangle=A_{\text{in}}(x)\,|\mathbf{k}\rangle.\quad(8)$$

Using (3) and (8) as well as (7) we get

$$A(\mathbf{x},t)j(\mathbf{y},t)\,|\mathbf{k}\rangle=j(\mathbf{y},t)\{A_{\text{in}}^{(+)}(\mathbf{x},t)+A_{\text{in}}^{(-)}(\mathbf{x},t)\}\,|\mathbf{k}\rangle=0.$$

The term $A_{\text{in}}^{(+)}|\mathbf{k}\rangle$ is either the vacuum state or it vanishes; so we recall (6) and we are left with

$$j(\mathbf{y},t)A_{\text{in}}^{(-)}(\mathbf{x},t)\,|\mathbf{k}\rangle=0.$$

The Fourier transform with respect to $\mathbf{x}$ yields

$$j(y)\,|\mathbf{q},\mathbf{k}\rangle=0.$$

Proceeding in this way we are able to show that

$$j(y)\,|\mathbf{k}_1,\cdots\mathbf{k}_n\rangle=0\quad n=1,2,\cdots.$$

Since the set is complete it follows that:

$$j(y)=0$$

and the proof is accomplished on account of (1).

III. CONCLUSIONS

The following few conclusions can be drawn from statement (i). From now on we confine ourselves to the case of the interaction of the field with itself. The results can be easily extended to more complex cases.

A. The Case of Strong Convergence

Statement (ii):

Let $a(\mathbf{k},t)$ and $a^{+}(\mathbf{k},t)$ converge *strongly* to $a_{\text{in}}(\mathbf{k})$ and $a_{\text{in}}^{+}(\mathbf{k})$, respectively, then the theory is a free field theory.

Proof: The assumption implies

$$Z_3=1,\quad(9)$$

where[8]

$$Z_3^{-1}=1+\int_0^\infty\sigma(x^2)d(x^2);\quad\sigma(x^2)\geq0.\quad(10)$$

From (9) and (10),

$$\sigma(x^2)=0\quad(11)$$

follows, which is equivalent with (5) or (6). From here the statement (*ii*) follows in virtue of statement (*i*).

Thus the renormalization of the field operators is an essential feature of each physically reasonable theory. In other words the theory becomes trivial if the cloud effects of virtual quanta are absent.

[6] Notice that if

$$[A(\mathbf{x},t),A(\mathbf{y},t)]=[\dot{A}(\mathbf{x},t),\dot{A}(\mathbf{y},t)]=0$$

and $[A(\mathbf{x},t),\dot{A}(\mathbf{y},t)]$ is independent of t, then (3) follows.

[7] To be more rigorous we should consider the convergence of integrals over the $\mathbf{k}$ space involving test functions and linear in the $a(\mathbf{k},t)$ rather than $a(\mathbf{k},t)$ itself.

[8] H. Lehmann, Nuovo cimento **2**, 342 (1954); H. Umezawa and S. Kamefuchi, Progr. Theoret. Phys. (Kyoto) **6**, 543 (1951); see also G. Källén, Helv. Phys. Acta **25**, 417 (1952) and H. Lehmann, Nuovo cimento **11**, 342 (1954).

B. Uselessness of the Local Clothing Operation

The clothing operation consists in finding (in addition to the physical vacuum) such operators[9] $c^+(\mathbf{k},0)$ and $c(\mathbf{k},0)$ which operating on the physical vacuum generate the physical one-particle states

$$Hc^+(\mathbf{k},0)|0\rangle=k_0c^+(\mathbf{k},0)|0\rangle \tag{12}$$

and destroy the vacuum

$$c(\mathbf{k},0)|0\rangle=0 \text{ respectively.}^{10} \tag{13}$$

They satisfy also

$$[c^+(\mathbf{k},0),c^+(\mathbf{q},0)]=[c(\mathbf{k},0),c(\mathbf{q},0)]=0.$$

All this is performed using the Schrödinger picture (i.e., $t=0$). In addition $c^+(\mathbf{k},t)$ has to converge weakly to $a_{\substack{\text{in}\\\text{out}}}{}^+(\mathbf{k})$ for $t\to\mp\infty$. It is clear that the states

$$c^+(\mathbf{k},0)\cdots c^+(\mathbf{k}_n,0)|0\rangle \quad n=2,3,\cdots \tag{14}$$

do not have to be eigenfunctions of the Hamiltonian and even do not have to be orthogonal; at least it is so for each finite t.[4]

Statement (i) discloses the trivial nature of a certain class of the clothed particle representations, called here for convenience the local one.[11] The local clothed field is supposed to satisfy the locality condition (3). After we replace the A field by the C field in (3) as well as in other formulas (*mutatis mutandis*) we have

Statement (iii):

Let us assume that the clothed particle field satisfies the assumptions listed at the beginning of this note, in particular (3), then the clothed particle field is a free field. In other words: a physically reasonable theory of a clothed field cannot satisfy canonical commutation relations; moreover, it cannot be local [in the sense of (3)] for finite time t.[11a]

C. Diagonalization of the Hamiltonian

We are going now to present a statement slightly different from (iii), viz.,

Statement (iv):

Let us assume that the field C satisfies assumptions listed at the beginning of this note, [in particular (3)] and let the field $c(\mathbf{k},0)$ and $c^+(\mathbf{k},0)$ satisfy the relation

$$c(\mathbf{k},0)c^+(\mathbf{k}',0)|0\rangle=\frac{1}{Z_3}\delta(\mathbf{k}-\mathbf{k}')|0\rangle, \tag{15}$$

and assume that the interaction Hamiltonian affects neither the vacuum nor the one-particle states, then the theory is a free field theory.

Proof: In Schrödinger picture the Hamiltonian reads

$$H\{c(0)\}=Z_3H_0\{c(0)\}+H_1\{c(0)\}=H_0\{c_{\text{in}}\}.$$

Because of our assumption, we have

$$Z_3H_0\{c(0)\}|0\rangle=0 \tag{16}$$

$$Z_3H_0\{c(0)\}c_{\text{in}}{}^+(\mathbf{k})|0\rangle=k_0c_{\text{in}}{}^+(\mathbf{k})|0\rangle. \tag{17}$$

On the other hand, from (15) it follows that

$$Z_3H_0\{c(0)\}c^+(\mathbf{k},0)|0\rangle=k_0c^+(\mathbf{k},0)|0\rangle. \tag{18}$$

Taking into account that the one-particle spectrum is not degenerate and Eqs. (17) and (18), we get

$$c_{\text{in}}{}^+(\mathbf{k})|0\rangle=c^+(\mathbf{k},0)|0\rangle \tag{19a}$$

and from (16) it follows that

$$c(\mathbf{k},0)|0\rangle=c_{\text{in}}(\mathbf{k})|0\rangle; \tag{19b}$$

since

$$e^{ik_0t}c^+(\mathbf{k},t)=e^{iHt}c^+(\mathbf{k},0)e^{-iHt} \quad (t\text{-finite}) \tag{20}$$

holds it follows from (19) that

$$C(x)|0\rangle=C_{\text{in}}(x)|0\rangle; \tag{5}$$

then statement (iv) follows from statement (i) [provided that (3) holds].

Let us exhibit a certain aspect of the statement (iv). Let us assume that we start working with the Hamiltonian in the Schrödinger picture expressed in terms of the Heisenberg physical field $a(\mathbf{k},t)$. The physical field must not necessarily satisfy the canonical commutation rules for equal time. Let us further assume that we succeed in finding the physical vacuum and the one particle state. This allows us to diagonalize the Hamiltonian in the sector of Hilbert space containing the vacuum and the one particle states. Let us then construct in a reasonable but otherwise arbitrary manner a complete orthonormal set $|\mathbf{k}_1,\cdots\mathbf{k}_n,t\rangle_c$ in the Hilbert space in such a way that the vacuum and one particle states are the constituants of this set. After this is accomplished we are able to define the creation and destruction operators $c^+(\mathbf{k},0)$ and $c(\mathbf{k},0)$, respectively, which in the $|\ \rangle_c$ representation have the form

$$\begin{bmatrix}0&0&0&\cdot\\ \sqrt{1}&0&0&\cdot\\ 0&\sqrt{2}&0&\cdot\\ 0&0&\sqrt{3}&\cdot\\ \cdots&\cdots&\cdots&\cdots\end{bmatrix}\begin{bmatrix}0&\sqrt{1}&0&0&\cdot\\ 0&0&\sqrt{2}&0&\cdot\\ 0&0&0&\sqrt{3}&\cdot\\ \cdots&\cdots&\cdots&\cdots&\cdots\end{bmatrix}. \tag{21}$$

We have then $|\mathbf{k}_1,\cdots\mathbf{k}_n,t\rangle_c=$combinatorial factor $\times c^+(\mathbf{k}_1,t)\cdots c^+(\mathbf{k}_n,t)|0\rangle$. Since the states $|\ \rangle_c$ are ortho-

[9] From now on we reserve the letter A and a for the Heisenberg field which represent really the physical situation and which secures the connection between A_{in} and A_{out} fields. We call this A field the physical field.

[10] L. Van Hove, see reference 4; O. W. Greenberg and S. S. Schweber, Nuovo cimento **8**, 378 (1958); J. Lopuszanski, Physica **25**, 745 (1959).

[11] The name "local clothed field" is somewhat confusing. We chose this word because we could not find any better name.

[11a] M. A. Braun and Yu. V. Novozhilov, [J. Exptl. Theor. Phys. **39**, 1317 (1960)] arrived at a similar conclusion. I am obliged to Dr. Schweber for calling it to my attention.

normal the operators c^+ and c will satisfy the canonical commutation relations.

Statement (iv) may enable us to diagonalize the Hamiltonian without having found c explicitly. After we found the vacuum and one particle states we just construct on their basis the complete orthonormal set. In terms of this set the Hamiltonian may be diagonal—this depends on whether the C field satisfies (3). This does not mean, however, that the scattering problem is also solved.

To make it clear we should like to emphasize the following: the Heisenberg a field is weakly convergent to a certain incoming field $a_{\text{in}}(\mathbf{k})$ for $t \to -\infty$. Although

$$a_{\text{in}}^+(\mathbf{k})|0\rangle = c^+(\mathbf{k},0)|0\rangle, \tag{22}$$

it is not necessarily true that

$$a_{\text{in}}^+(\mathbf{k}) = c^+(\mathbf{k},0).$$

The fields in x space built up from the a_{in} and c operators, respectively, are both free fields. Both operators generate two complete sets of eigenfunctions of the Hamiltonian, in general, independent from each other. They are related by an unitary transformation; say $D\{a_{\text{in}}\} = D\{c(0)\}$:

$$a_{\text{in}}^+(\mathbf{k}) = D^{-1}c(\mathbf{k},0)D.$$

D accounts for the degeneracy of the spectrum of H for more than one particle. For instance, the scattering operator S is an example of the D transformation.

D. On the Time Behavior of the Mathematical Vacuum[12]

The statement we are going to display in this paragraph is closely related to that of Haag,[13] more precisely to the first part of Haag's theorem concerning the vacuum.[14]

We shall show that the time behavior of the vacuum determines whether the field is free or not. To show it we do not have to assume either that the field $(Z_3)^{\frac{1}{2}}A(x)$ is related to the free field by a unitary transformation or that both of them satisfy canonical commutation relations for a fixed time.

We have

Statement (v):

If the state $|0(t)\rangle$ defined by

$$a(\mathbf{k},t)|0(t)\rangle = 0 \tag{23}$$

exists for each finite t and has the properties that it is (a) invariant with respect to 3-dimensional space translations, and that (b)

$$\langle \mathbf{k}_1 \cdots \mathbf{k}_n | 0(t)\rangle \neq 0 \quad n = 0, 1, 2, \cdots \tag{24a}$$

as well as

$$0 < \langle 0(t)|0(t)\rangle < \infty \tag{24b}$$

hold, then the theory is a free field theory.

Proof: Assumption (b) means that $|0(t)\rangle$ is a vector of the Hilbert space spanned on the incoming eigenstates of the energy-momentum 4-vector. There exists at least one operator $V(t)$ which transforms $|0\rangle$ into $|0(t)\rangle$; the matrix elements of $V(t)$, e.g.,

$$\langle \mathbf{k}_1 \cdots \mathbf{k}_n | V(t)|0\rangle,$$

do not vanish identically [by assumption (b)].

Assumption (a) implies that

$$\mathbf{P}|0(t)\rangle = 0. \tag{25}$$

But the physical vacuum is also an eigenstate of the vanishing 3-momentum; it is well known that there do not exist other normalizable eigenstates of vanishing 3-momentum. Because of (24b) and (25) it must be

$$|0(t)\rangle = N(t)|0\rangle \quad N(t) \text{ is a finite } c \text{ number}, \neq 0.$$

From (23) we get

$$a(\mathbf{k},t)|0\rangle = 0. \tag{26}$$

So far all this is well known [see, e.g., reference 3]. From (1) and (2) it follows that

$$a(\mathbf{k},t)|0\rangle = a_{\text{in}}(\mathbf{k})|0\rangle - [i/(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}] \times \int dy\, j(y)\theta(t-y_0)e^{-iky}|0\rangle. \tag{27}$$

In view of (4), (26) and (27) we get

$$\int_{-\infty}^{t} dy_0 e^{-ik_0 y_0} \int d^3y\, j(y) e^{-iky}|0\rangle = 0.$$

We differentiate with respect to t and get

$$\int d^3y\, j(\mathbf{y},t) e^{-iky}|0\rangle = 0 \tag{28}$$

or taking the Fourier transform of (28)

$$j(y)|0\rangle = 0. \tag{6}$$

Thus we succeeded in reducing the problem to the problem considered in statement (i) and this accomplishes the proof.

Conclusion

The necessary condition for a field theory with interaction to be a physically reasonable one is that the mathematical vacuum for finite time t shall be orthogonal to all physical incoming states including the physical vacuum. The same is true for all mathematical

[12] This paragraph is based on the suggestion made by Dr. Friedrichs in a conversation.

[13] See reference 3.

[14] D. Hall and A. S. Wightman, Mat. fys.-Medd. Dan. Vidselsk **31**, No. 5 (1957). L. Van Hove, Physica **18**, 145 (1952).

states (provided that Z_3 is finite). In other words, the mathematical vacuum has to belong to a different Hilbert space than that of incoming states, both Hilbert spaces belonging to a larger inseparable Hilbert space. No restrictions, however, are imposed on the field $(Z_3)^{\frac{1}{2}}A(x)$ to satisfy canonical commutation relations for fixed t; we have to use, at any rate, a myriotic representative of the commutation relations.

To carry through the proof of statement (v) we do not need as much relativistic invariance of the theory. What we need is that all quantities appearing in the theory are functions of $A(x)$; then the $a(\mathbf{k},t)$ and $a^+(\mathbf{k},t)$ operators appear symmetrically in the theory *mutatis mutandis*. In the nonrelativistic model theories like Lee's or Ruijgrok-Van Hove's model the pair production is not taken into account which causes that the theory is not a free field theory although $|0(t)\rangle = |0\rangle$. In the first mentioned model there is in addition a lack on crossing symmetry.

Note added in proof. A different, elegant proof of statement (v) was suggested to me by Haag. The outline of the proof is as follows. One starts with

$$a(\mathbf{k},t) = k_0 \int_{x_0=t} \frac{e^{-ikx}}{(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}} A(x) d^3x + i \int_{x_0=t} \frac{e^{-ikx}}{(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}} \frac{\partial A(x)}{\partial x_0} d^3x.$$

Because of (26) and

$$\partial A(x)/\partial x_0 = i[H, A(x)]$$

we get

$$0 = k_0 \int_{x_0=t} \frac{e^{-ikx}}{(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}} A(x) d^3x |0\rangle - H \int_{x_0=t} \frac{e^{-ikx}}{(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}} A(x) d^3x |0\rangle.$$

Since the one particle states are supposed not to be degenerate the vector

$$\int_{x_0=t} \frac{e^{-ikx}}{(2\pi)^{\frac{3}{2}}(2k_0)^{\frac{1}{2}}} A(x) d^3x |0\rangle$$

is a one particle eigenstate with momentum $-\mathbf{k}$. Since (5) holds we can apply statement (i).

ACKNOWLEDGMENTS

The author is extremely grateful to Dr. B. Zumino without whose constant assistance and criticism this note could not have been accomplished at all. The author is pleased to express his great indebtedness to Dr. K. O. Friedrichs for stimulating discussions and for the hospitality extended to him at the I. M. S. of N. Y. U. as well as to Dr. R. Haag for his searching criticism and to Mr. W. Garczynski for discussions.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Analyticity of Wightman Functions

CHRISTIAN FRONSDAL*
University of Pennsylvania, Philadelphia, Pennsylvania
(Received April 18, 1961)

The problem of mapping certain domains in the space of complex four-vectors onto the space of their inner products is solved by a novel method. The "primitive" domains of regularity of the three and four point Wightman functions are determined. The domains whose X-space characterization was obtained by Streater, for the holomorphy envelope of the union of several primitive domains (both for the three and four point function) are also determined. The corresponding problem in perturbation theory is examined, and the analytic form of the perturbation theory boundary of the three and four point functions is obtained by the same method. The problem of the four-point function is reduced to constructing the holomorphy envelope of the union of three domains. It is shown that the boundary of the domain of four-point singularitites in perturbation theory bears no resemblance to any part of the boundary of the primitive domains, or the domains found or conjectured by Streater.

INTRODUCTION

IN this paper we solve the algebraic problems of mapping certain domains in the space of n complex four vectors onto the complex space of their invariant inner products. The background for this problem is, very briefly, the following.

A theory of quantized fields $A(Y_1)$, $B(Y_2),\cdots$, satisfying a set of fundamental postulates (positive energy, causality, unitarity, relativistic invariance) is completely specified by the vacuum expectation values of products of field operators,[1] The latter are invariant functions of coordinates and are called Wightman functions, and the fundamental postulates reduce to postulated properties of these functions. Consider the three-point function

$$ABC=\langle A(Y_1)B(Y_2)C(Y_3)\rangle,$$

and its Fourier transform

$$\overline{ABC}=\int e^{-i(p_1X_1+p_2X_2)}ABCd^4X_1d^4X_2,$$

with $X_i^\mu=Y_i^\mu-Y_i^\mu+1$. The postulates of positive energy and relativistic invariance imply[2] that ABC is the limit of a function of the inner products (X_iX_j), regular when X_i^μ is in the "extended tube" $\mathcal{T}'$. This is defined as follows. Let $X_i^\mu\in\mathcal{T}$ mean that $\mathrm{Im}X_i^\mu$ is in the forward lightcone. Then $X_i^\mu\in\mathcal{T}'$ if there exists a complex Lorentz transformation Λ such that $(\Lambda X_i)^\mu\in\mathcal{T}$. Similarly, the permuted function ACB is the limit of a function regular in a permuted domain. The consequence of causality (local commutativity of the field operators) is that ABC and ACB are limits of the same analytic function.[1] Hence the function ABC may be continued into the union of the two domains. The envelope of holomorphy of these two domains may then be computed.

Streater's[3] theorem gives this holomorphy envelope in X space. In the first part of this paper we explain our method in detail by deriving the boundary, in the space of the invariants, of both the primitive domain and Streaters holomorphy envelope.

In the case of the three-point function it turned out[4] that perturbation theory examples gave a useful clue to the problem of finding the holomorphy envelope. We therefore rederive the form of the three-point boundary manifolds in perturbation theory. Perturbation theory leads to integral representations for the Wightman functions. We shall assume that the reader is familiar with these, since solving the perturbation theory problem is not our main task. Any background that may be needed may be found in reference 5.

In the second part of this paper we apply our methods to the four point function. Our methods may be used to obtain, in a simple way, the mapping in the space of invariants of all the domains which have been discussed in the literature. The holomorphy envelope of the union of 4 primitive domains (not related by TCP) is discussed in a preliminary way. It turns out that our methods are well suited to decide whether the manifolds thus obtained bear any resemblance to the perturbation theory boundary. Thus, the conjecture[5] that the perturbation theory boundary might be related through simple changes of sign to the primitive boundary is disproved. (This result has no bearing on the possibility of a similarity of the perturbation theory boundary to the complete 4-point function domain.)

I. THE THREE-POINT FUNCTION

In the following, a rather detailed derivation of the domain of analyticity of the three-point function is given. By so doing we shall be able to put the subsequent discussion of the four-point function in a better relief.

* Present address: University of California, Los Angeles, California.

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

[2] This is the Bargman-Hall-Wightman theorem. D. Hall and A. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **31**, No. 5 (1957).

[3] R. F. Streater, Proc. Roy. Soc. (London) **A256**, 39 (1960).

[4] G. Källén and A. Wightman, Danske Videnskab Selskab, Mat.-fys. Skr. **1**, No. 6 (1958). This paper gives the first derivation of the holomorphy envelope of the union of the three primitive domains.

[5] A. Chi-Tai Wu, University of Maryland, Physics Department Technical Report No. 186 (July, 1960).

The field theoretical primitive domain is derived[4] in Sec. A, and extended by means of Streater's theorem[3] in B. The perturbation theory boundary manifolds are found[6] in Sec. C, and compared with the field theoretical results in D.

As parts of the following discussion are applicable without added complication to the four-point function (and to the five-point function) as well, it is sometimes convenient to refrain from specialization.

A. The Primitive Domain

The algebraic problem may be described as follows. Let X_i^μ, $i=1, \cdots, n-1$, be $n-1$ complex four-vectors, and a_i^μ, b_i^μ the real and imaginary parts. That is

$$X_i^\mu \equiv a_i^\mu + ib_i^\mu, \quad i=1, \cdots, n-1. \tag{1}$$

The forward tube $\mathcal{T}$ is the domain defined by

$$X \in \mathcal{T}, \quad \text{if} \quad b_i^2 \geqslant 0, \quad b_i^0 > 0, \quad i=1, \cdots, n-1, \tag{2}$$

where the metric is such that

$$(b_i b_i) \equiv b_i^{0^2} - b_i^{1^2} - b_i^{2^2} - b_i^{3^2}. \tag{3}$$

Let $Z_{ij}=Z_{ji}$, $i, j=1, \cdots, n-1$, be a set of $\frac{1}{2}n(n-1)$ complex numbers. Points in X space are mapped on points in Z space by

$$Z_{ij} = (X_i X_j). \tag{4}$$

The points $X \in \mathcal{T}$ are mapped by (4) on a subset $M\mathcal{T}$ of Z space. The number of (real) dimensions of $M\mathcal{T}$ is the same as that of Z space if $n \leqslant 5$, and then equals $n(n-1)$. The n^2-n-1 dimensional boundary of $M\mathcal{T}$ will be denoted $BM\mathcal{T}$. The problem of this section is to be determine $BM\mathcal{T}$.

Let Z be a regular point on $BM\mathcal{T}$. By this is meant that there exists a n^2-n-1 dimensional plane which is tangent to $BM\mathcal{T}$ at Z. Then there exists a vector N, with components $N_{ij}=N_{ji}$, normal to $BM\mathcal{T}$ at Z, and such that $Z+\lambda N \in M\mathcal{T}$ for sufficiently small real positive λ.

If $Z \in BM\mathcal{T}$, and $Z' \in M\mathcal{T}$ is in a neighborhood of Z, then there exist complex infinitesimal four-vectors dX_i^μ such that

$$Z_{ij} = (X_i X_j), \quad Z_{ij}' = (X_i + dX_i,\ X_j + dX_j), \tag{5}$$

where

$$dX_i^\mu \equiv da_i^\mu + i db_i^\mu, \tag{6}$$

and

$$(b_i + db_i)^2 \geqslant 0, \quad b_i^0 + db_i^0 > 0. \tag{7}$$

The projection of $Z'-Z$ on N, with respect to a Euclidean metric in the $n(n-1)$ dimensional real Z space is

$$(Z'-Z, N) \equiv \sum_{i,j} \mathrm{Re}\{N_{ij}^*(Z_{ij}'-Z_{ij})\} = 2\sum_{ij} \mathrm{Re}\{N_{ij}^*(X_j dX_i)\}. \tag{8}$$

Here only the lowest order in dX_i^μ has been retained, because for any Z' not on $BM\mathcal{T}$, Z may be so chosen that (8) does not vanish. The definition of N requires that (8) be non-negative for any dX_i^μ that satisfy (7).

The only components of dX_i^μ not left arbitrary by (7) are $(b_i db_i)$ and $(f_i db_i)$, where f_i^μ is a real four-vector in the forward lightcone. Therefore, (8) must be of the form

$$(Z'-Z, N) = \sum_i \{\lambda_i (b_i db_i) + (f_i db_i)\} \geqslant 0. \tag{9}$$

Comparing coefficients of dX_i^μ in (8) and (9) we get

$$\sum_i N_{ij}^* X_j^\mu = -i\lambda_i b_i^\mu - i f_i^\mu. \tag{10}$$

Here $f_i^\mu \neq 0$ only in the case that db_i^μ is restricted by (7), that is when $b_i^\mu = 0$. Then $\mathrm{Im} Z_{ii} = 0$, which defines a subspace of the same dimension as $BM\mathcal{T}$. For this to be a measurable part of $BM\mathcal{T}$, all the X_j^μ, $j \neq i$, must be arbitrary.[7] Hence $N_{jk}=0$ except N_{ii}, and (10) reduces to $N_{ii}^* a_i^\mu = -i f_i^\mu$. Thus a_i^μ is timelike and $\mathrm{Re} Z_{ii} > 0$. That this is a cut rather than a boundary proper is seen from the fact that the sign of N_{ij} is undetermined. *A glance at (10) also shows that this is the only kind of cut in the domain.* The λ_i must vanish unless (b_i, db_i) is restricted by (7),[8] that is when $b_i^2 = 0$, and in that case λ_i must be non-negative. If the right-hand side of (10) vanishes for any i, then any variation of X_i^μ leads to a point Z' on $BM\mathcal{T}$, which means that the equations defining $BM\mathcal{T}$ do not involve X_i^μ. Hence, if any λ_i vanishes we get only the $n-1$ point parts of $BM\mathcal{T}$. The only nontrivial part of $BM\mathcal{T}$ is that on which all f_i^μ and no λ_i vanish. It is characterized by the existence of a unique (up to a positive factor) $N_{ij}(Z)$ satisfying

$$\sum_{j=1}^{n-1} N_{ij}^* X_j^\mu = -i\lambda_i b_i^\mu, \quad i=1, \cdots, n-1, \tag{11}$$

$$b_i^2 = 0, \quad b_i^0 > 0, \quad \lambda_i > 0, \quad i=1, \cdots, n-1. \tag{12}$$

If (11) is multiplied by X_i^μ and summed over i, the symmetry of N_{ij} gives

$$\sum_i \lambda_i (a_i^\mu b_i^\nu - a_i^\nu b_i^\mu) = 0. \tag{13}$$

Equation (11) can always be solved for N_{ij}^* but the solution will be symmetric iff (13) is satisfied. Hence (11) and (13) are equivalent. The result (13) was first obtained by Wightman.[9] Even though our method of the normal leads to a well-known result in this particular case, its very great generality makes it a valuable tool in more complicated situations.

[6] These were first determined in reference 4.

[7] Except, of course, that $X \in \mathcal{T}$.

[8] This simple observation (together with the similar remark above) constitutes a proof of the fact that of all $X \in \mathcal{T}$, only points on $B\mathcal{T}$ can be mapped on $BM\mathcal{T}$. This result was first obtained in reference 2.

[9] Appendix II of reference 4.

Before proceeding with the task of solving these equations in the three-point case, it is useful to discuss the general properties of (13). As a means of showing the close relationship of the methods of this and later sections with those of Wightman[9] and others,[10] we give another derivation, We have already noted that N_{ij} must depend on Z only; it is therefore invariant under the complex Lorentz group. From (11) we obtain

$$\sum_{j,l} N_{ij}{}^{*}N_{kl}{}^{*}Z_{jl}=-\lambda_i\lambda_k(b_ib_k), \quad i,k=1,\cdots,n-1. \tag{14}$$

The left-hand sides of these equations are invariant under any complex Lorentz transformation

$$X_i{}^{\mu}\rightarrow(\Lambda X_i)^{\mu}$$

such that $\Lambda X_i\in\mathcal{T}$. Therefore, the right-hand sides must also be invariant. In particular, taking $i=k$ in (14) we see that no Λ must exist such that $b_i\rightarrow b_i'$ with $b_i'^2>0$. Specializing to infinitesimal transformations this is expressed by the condition that no real matrix $\beta_{\mu\nu}$ exist such that

$$\sum_{\mu,\nu}\beta_{\mu\nu}(a_i{}^{\mu}b_i{}^{\nu}-a_i{}^{\nu}b_i{}^{\mu})>0, \tag{15}$$

and this is equivalent to (13).

As an aid in the practical task of solving (11) or (13), we shall prove that no more than $n-1$ among the $a_i{}^{\mu}$, $b_i{}^{\mu}$ can be linearly independent. Suppose that m of the $b_i{}^{\mu}$ are linearly independent, and let $c_p{}^{\mu}$, $p=1,\cdots 4-m$, be $4-m$ linearly independent vectors normal to $b_i{}^{\mu}$. Then (13) gives

$$\sum_i\lambda_i(c_pa_i)b_i{}^{\mu}=0, \quad p=1,\cdots,4-m. \tag{16}$$

Of these $4-m$ linear equations for the $b_i{}^{\mu}$, only $n-1-m$ can be independent. Hence there exist

$$(4-m)-(n-1-m)=5-n$$

relations of the form

$$\sum_p h_p{}^{\sigma}\lambda_i(c_pa_i)=0, \quad \sigma=1,\cdots,5-n. \tag{17}$$

Since $\lambda_i\neq 0$, this means that $5-n$ linearly independent vectors (note that $5-n\leqslant 4-m$) are normal to $a_i{}^{\mu}$ and to $b_i{}^{\mu}$. Therefore only $4-(5-n)=n-1$ among the $a_i{}^{\mu}$, $b_i{}^{\mu}$ can be linearly independent; Q.E.D.

We now return to the practical task of solving (11), (12). We could, instead, choose to solve (12), (13), but without gaining in simplicity.

There is always a set of real Lorentz transformations Λ_i such that

$$X_i{}^{\mu}=(\Lambda_i\bar{X}_i)^{\mu}, \tag{18}$$

$$\bar{X}_i{}^{\mu}=\{\tfrac{1}{2}(r_i+w_i),\ \tfrac{1}{2}(r_i-w_i),\ s_i,\ t_i\}, \tag{19}$$

with real s_i and t_i. Points in the forward tube are

$$X\in\mathcal{T} \text{ iff } \operatorname{Im}w_i>0, \quad \operatorname{Im}r_i>0, \tag{20}$$

and $b_i{}^2=0$ is the same as

$$\operatorname{Im}r_i=0. \tag{21}$$

In this way the $\frac{1}{2}n(n-1)$ inner products are expressible in terms of $n-1$ complex numbers and a set of real parameters. The complex numbers may conveniently be eliminated by the relations $Z_{ii}=(X_iX_i)=(\bar{X}_i\bar{X}_i)$, which give

$$w_i=r_i^{-1}(Z_{ii}+s_i{}^2+t_i{}^2). \tag{22}$$

The w_i are arbitrary except for (20), or

$$r_i\operatorname{Im}Z_{ii}>0. \tag{23}$$

It will turn out that (11) implies conditions on the real parameters only.

In the three-point case it is convenient to distinguish the case of $b_1{}^{\mu}$ and $b_2{}^{\mu}$ parallel (which gives the S curve) and $b_1{}^{\mu}$ not parallel to $b_2{}^{\mu}$ (which gives the F_{12} curve).

Parallel b_1 and b_2.

In this case a Lorentz frame exists in which, on the boundary,

$$X_i=\{\tfrac{1}{2}(r_i+w_i),\ \tfrac{1}{2}(r_i-w_i),\ 0,\ 0\}, \quad i=1,\ 2. \tag{24}$$

Putting the last two components equal to zero is possible because we have proved that only two among the four vectors $a_i{}^{\mu}$, $b_i{}^{\mu}$ are linearly independent. Equation (11) now reduces to four equations, of which two are homogeneous and give

$$N_{ij}{}^{*}=\rho\begin{pmatrix}1, & k\\ k, & k^2\end{pmatrix}, \quad k=-\frac{r_1}{r_2}, \tag{25}$$

and the two inhomogeneous equations are

$$\rho(w_1+kw_2)=-i\lambda_1\operatorname{Im}w_1=-i\lambda_2k^{-1}\operatorname{Im}w_2. \tag{26}$$

From (24) we find $Z_{12}=r_1w_2+r_2w_1$. Eliminating w_i by means of (22) [with $s_i=t_i=0$], we obtain

$$S: \quad Z_{11}+2kZ_{12}+k^2Z_{22}=0. \tag{27}$$

The restriction to positive λ_i in (26) gives the relevance condition

$$\operatorname{Im}Z_{11}\operatorname{Im}Z_{22}<0, \tag{28}$$

and when this is compared with (23);

$$k>0. \tag{29}$$

The results (27), (28), and (29) are identical to those of Källén and Wightman.[4]

Nonparallel b_1 and b_2.

In this case a Lorentz frame exists such that the space parts of $b_1{}^{\mu}$ and $b_2{}^{\mu}$ are antiparallel, and on the

[10] Unpublished work, especially by R. Jost.

boundary

$$X_i=\{\tfrac{1}{2}(r_i+w_i),\ (-)^i\tfrac{1}{2}(r_i-w_i),\ 0,\ 0\},\quad i=1,\ 2. \tag{30}$$

Instead of (25) and (26) we now find

$$N_{ij}{}^*=\rho\begin{pmatrix} r_2w_2, & -\gamma \\ -\gamma, & r_1w_1 \end{pmatrix};\quad \gamma=r_1r_2, \tag{31}$$

$$\gamma\rho(w_1w_2-\gamma)=-i\lambda_1r_1\,\mathrm{Im}w_1=-i\lambda_2r_2\,\mathrm{Im}w_2. \tag{32}$$

Eliminating w_i as above we get

$$F:\quad Z_{11}Z_{22}-2\gamma Z_{12}+\gamma^2=0. \tag{33}$$

Positive λ_1 and λ_2 in (32) means

$$\mathrm{Im}Z_{11}\,\mathrm{Im}Z_{22}>0; \tag{34}$$

and when this is compared with (23),

$$\gamma>0. \tag{35}$$

The results (33)–(35) are identical to those of Källén and Wightman.[4]

We end this section by giving some compact expressions for N_{ij}. If (27) and (33) be solved for k and γ, we may write

$$N_{ij}{}^*=-ik\epsilon(Z_{11})\partial k/\partial Z_{ij},\ k=-[Z_{12}\pm(-D)^{\frac{1}{2}}]/Z_{22} \tag{36}$$

on the S curve, and

$$N_{ij}{}^*=-i\epsilon(Z_{11})\partial\gamma/\partial Z_{ij},\quad \gamma=Z_{12}\pm(-D)^{\frac{1}{2}} \tag{37}$$

on the F_{12} curve. Here $\epsilon(Z_{11})$ is the sign of $\mathrm{Im}Z_{11}$, and D is the determinant of the matrix Z_{ij}. Introducing (36) into (8) we get

$$(Z'-Z,\ N)=\sum_{ij}\mathrm{Re}\{N_{ij}{}^*dZ_{ij}\}=k\epsilon(Z_{11})\,\mathrm{Im}dk, \tag{38}$$

which shows that the allowed side of the S curve is that on which

$$k\epsilon(Z_{11})\,\mathrm{Im}k>0. \tag{39}$$

Similarly, the allowed side of the F_{12} curve is

$$\epsilon(Z_{11})\,\mathrm{Im}\gamma>0. \tag{40}$$

B. Analytic Completion

Up to now we have explored the analyticity in x space which follows from the positiveness of the energy of a complete set of states. This domain may be considerably enlarged by using local commutativity of the field operators. The consequence of local commutativity is the following.[1] Let

$$X_i{}^\mu=Y_i{}^\mu-Y_{i+1}{}^\mu. \tag{41}$$

Then, to the primitive domain must be added all domains which may be obtained by applying a permutation to the n complex four-vectors $Y_i{}^\mu$.

It happens that the Z-space domain obtained in this way is not a natural domain of holomorphy. We are therefore confronted with the very difficult task of computing the holomorphy envelope. Fortunately, Streater[3] has found a shortcut to part of the answer. The Streater theorem, which is based on Dyson's integral representation[11] is this.

Theorem A.[3] The holomorphy envelope of the union of the two domains

$$\cdots,\quad Y_{n-1}-Y_n,\quad Y_n-Y_{n+1}\in\mathcal{T}' \tag{42}$$

and

$$\cdots,\quad Y_{n-1}-Y_{n+1},\quad Y_{n+1}-Y_n\in\mathcal{T}' \tag{43}$$

is the domain

$$\cdots,\quad Y_{n-1}-Y_n,\quad Y_{n-1}-Y_{n+1}\in\mathcal{T}', \tag{44}$$

less the cut

$$(Y_n-Y_{n+1})^2=\text{real positive}. \tag{45}$$

Here $\mathcal{T}'$ is the extended tube, or the inverse image of $M\mathcal{T}$.

The application of this to the three-point case is straightforward. The three primitive domains are

$$X_1,X_2\in\mathcal{T}',\quad -X_1,X_1+X_2\in\mathcal{T}',\quad X_1+X_2,-X_2\in\mathcal{T}'. \tag{46}$$

Streater's theorem says that the holomorphy envelope of the union of these domains is not smaller than the union of the domains

$$X_1,X_1+X_2\in\mathcal{T}',\quad -X_1,X_2\in\mathcal{T}',\quad X_1+X_2,X_2\in\mathcal{T}' \tag{47}$$

less the $X_1{}^2$, $X_2{}^2$ and $(X_1+X_2)^2$ cuts. Hence part of the analytic completion is accomplished by replacing each of the primitive domains by a "secondary domain." One such domain is

$$\{-X_1,\ X_2\in\mathcal{T}'\}\bigcap\{X_1+X_2\in\mathcal{T}'\}; \tag{48}$$

the others are obtained by applying a permutation of the $Y_i{}^\mu$ to (48).

The boundary of the secondary domain is easily determined. It contains, of course, the three cuts. To obtain the three-point part of the boundary we have only to reverse the sign of $\mathrm{Im}w_1$ in (20). Then (23) is replaced by

$$r_1\,\mathrm{Im}Z_{11}<0,\quad r_2\,\mathrm{Im}Z_{22}>0. \tag{49}$$

Now (23) was used to obtain (29) and (35). Using (49) instead we get $k<0$ and $\gamma<0$. With negative k and γ, (27) and (33) become the S' and F_{12}' curves of Wightman and Källén.[4]

If we construct the union of the three secondary domains we get a domain that is bounded by the cuts and by the three F_{ij}' curves:

$$Z_iZ_j+\gamma(Z_k-Z_i-Z_j)+\gamma^2=0,\quad 0>\gamma>-\infty. \tag{50}$$

Here i, j, k is a cyclic permutation of 1, 2, 3, and

$$Z_i=Z_{ii},\quad i=1,\ 2,\qquad Z_3=(X_1+X_2)^2=Z_{11}+2Z_{12}+Z_{22}. \tag{51}$$

[11] F. J. Dyson, Phys. Rev. **110**, 1460 (1958).

This domain is only slightly smaller than the holomorphy envelope.[12]

C. Perturbation Theory

The algebraic problem is the following[13]: Let Y_i^μ be n complex four-vectors,[14] satisfying

$$Y_i^2=a_i\geqslant 0, \quad i=1,\cdots,n, \tag{52}$$

$$|(Y_iY_j)|=0. \tag{53}$$

Let $n-1$ complex four-vectors be defined by

$$X_i^\mu=Y_i^\mu-Y_{i+1}^\mu, \quad i=1,\cdots,n-1, \tag{54}$$

and let

$$Z_{ij}=(X_iX_j). \tag{55}$$

Conditions (52), (53) define a manifold D in Y space, which is mapped on the domain MD in Z space by (54), (55). This latter domain is of $n(n-1)$ real dimensions. It is bounded by BMD, which is a manifold of n^2-n-1 real dimensions. We shall determine the analytical form of the equations which define BMD.

On a point on the boundary in Z space there exists a complex symmetric matrix $N_{ij}(Z)$, unique up to a real (positive except on cuts) factor, with the property that

$$0\geqslant \mathrm{Re}\{\sum_{i,j=1}^{n-1} N_{ij}{}^*dZ_{ij}\}=2\,\mathrm{Re}\{\sum_{i,j=1}^{n-1} N_{ij}{}^*(X_jdX_i)\}, \tag{56}$$

for all infinitesimal $dX_i^\mu=dY_i^\mu-dY_{i+1}^\mu$ such that (52) and (53) are preserved under the transformation $Y_i^\mu\to Y_i^\mu+dY_i^\mu$. This definition of N_{ij} is entirely analogous to that of Sec. A.

Because (52) and (53) are conditions on Y_i^μ rather than on X_i^μ, it is convenient to re-express (56) in terms of dY_i^μ. Then we get

$$0\geqslant \mathrm{Re}\{\sum_{i,j=1}^{n} M_{ij}{}^*(Y_idY_j)\}, \tag{57}$$

where M_{ij} is related to N_{ij} and satisfies

$$\sum_{i=1}^{n} M_{ij}=0. \tag{58}$$

Preservation of (53) does not impose any restriction on dY_i^μ that is relevant for the discussion of (57), but (52) implies that

$$(Y_idY_i) \text{ are real, and if } Y_i^2=0, \text{ positive.} \tag{59}$$

Because of the particular importance of the case $Y_i^2=0$, we shall introduce a positive integer m defined by

$$Y_i^2=a_i>0, \quad i\leqslant m, \tag{60}$$

$$Y_i^2=0, \quad m+1\leqslant i\leqslant n. \tag{61}$$

We proceed as with the solution of (7), (8), to find that the solution of (57), (59) is

$$\sum_{j=1}^{n} M_{ij}{}^*Y_j^\mu=-u_iY_i^\mu, \tag{62}$$

$$\mathrm{Re}u_i=0, \quad i=1,\cdots,m, \tag{63}$$

$$\geqslant 0, \quad i=m+1,\cdots,n. \tag{64}$$

When N_{ij} and X_i^μ are reintroduced, (62) and (58) become[15]

$$\sum_{j=1}^{n-1} N_{ij}{}^*X_j^\mu=-\sum_{k=1}^{i} u_kY_k^\mu, \quad i=1,\cdots,n-1, \tag{65}$$

$$\sum_{k=1}^{n} u_kY_k^\mu=0. \tag{66}$$

The existence of a solution of (62), and therefore of (65), (66), is guaranteed by (53). Our problem is therefore reduced to writing down the conditions on Y_i^μ which are necessary and sufficient for $N_{ij}(Z)$ to be determined up to a real positive factor by (63)–(66). This approach to the problem is closely analogous to that suggested following (13) for solving (8).

Because of the formal covariance of Eqs. (63)–(66) under complex Lorentz transformations, the uniqueness of N_{ij} is simply the requirement that

(i) No complex vector y^μ exist such that $Y_i^\mu+y^\mu\in D$, i.e., such that $Y_i^\mu\to Y_i^\mu+y^\mu$ preserve (52), (53).

(ii) The u_i be determined up to a real positive common factor by (63), (64), (66).

(iii) The X_i^μ be linearly independent. The first condition insures that the X_i^μ determine the $Y_i^\mu\in D$, the second that Y_i^μ determine the u_i. Together these conditions ensure that the right-hand sides of (65) be expressible in terms of X_i^μ. The third requirements says that (65) be solvable, giving N_{ij} as unique functions of X_i^μ, and hence because of invariance, as unique functions of X_{ij} (up to a real positive factor).

When condition (iii) is satisfied, (53) is preserved only if y^μ lies in the $(n-1)$-dimensional complex space spanned by the n Y_i^μ. In the following we may therefore treat Y_i^μ and y^μ as complex $(n-1)$-vectors.

Condition (i) requires the nonexistence of a complex

[12] What remains to be done is to continue through some of the "corners" formed by the F_{ij}' curves to the $\mathfrak{F}$ curve of reference 4. Our inability to do this by means of integral representations is due to the difficulty of incorporating the Jacobi identity into the Streater representation for the double commutator.

[13] As explained in the introduction we treat only that part of the perturbation theory problem which is essential for our applications. The reader is referred to reference 3 for an introduction and what is, in some respects, a broader treatment.

[14] The Y_i^μ and X_i^μ of this section are introduced as a set of parameters, in terms of which the singularity domain can be defined. They should not be interpreted as spacetime coordinates, although it will be shown later that the X_i^μ may be so interpreted without error.

[15] It is interesting to notice that (66) implies (53). Thus the same boundary would have been obtained without imposing (53) in the first place.

vector y^μ that satisfy

$$(Y_i+y)^2=r_i\geqslant 0, \quad r_i \text{ real.} \tag{67}$$

This condition, being nonlinear in y^μ, is very difficult to handle, and we shall be content to explore the consequences of the weaker condition:

(i') No complex vector y^μ exist such that

$$\mathrm{Im}(Y_i y)=0 \tag{68}$$

$$\mathrm{Re}(Y_i y)\geqslant 0, \quad i>m, \tag{69}$$

$$y^2=\text{real and positive.} \tag{70}$$

Another way to say this is that we consider only solutions of (67) such that ρy^μ is a solution when y^μ is, ρ being an arbitrary positive number. In this way the problem is, essentially, reduced to a linear one.

It is now necessary to go back to (64), and to prove that $\mathrm{Re}u_i$, for $i>m$, can vanish only on a non-measurable part of *BMD*. First note that the manifold on which $|(Y_iY_j)|=0$, and all $Y_i^2=0$, that is the case $m=0$, is mapped on a manifold in Z space of real dimension n^2-n-2. Since the real dimension of *BMD* is n^2-n-1, the case $m=0$ can henceforth be neglected. For $m>0$, the real dimension of the mapping in Z space of the manifold defined by (53) and (60)–(64) is exactly $(n^2-n)+m-2-(m-1)=n^2-n-1$. The submanifold on which one of $\mathrm{Re}u_i$, $i>m$, vanishes, is of one lower dimension, and does not, therefore, make up a measurable part of *BMD*.

Now, if we multiply (68) or (69) by u_i, sum over i, and compare with (63), (64), and (66), we find that (68), (69) can have no solution, except possibly with all equalities in (69). Actually the existence of u_i is even more closely connected with that of y^μ. In fact, if $m>0$, the existence of u_i is equivalent to (a) the nonexistence of any solution of (68), (69) with inequalities and (b) the existence of a $m-1$ parameter family of solutions with equalities. We next show that if $m\geqslant 3$, then a member of this family can be found that satisfies (70), which means that this case does not give any part of the boundary.

Every member of the $m-1$ parameter family of solutions just referred to may be represented by a point in the y^2 plane. The aggregate of points in the y^2 plane which represents solutions will be referred to as the allowed part of the y^2 plane. What we have to prove is that this allowed part includes part of the real positive axis, when $m\geqslant 3$. First we note that the allowed part is a domain. In fact, if $m\geqslant 3$, the family of solutions contains, in general, $m-1\geqslant 2$ independent parameters. The mapping onto the y^2 plane therefore gives a two-dimensional manifold, except in very special degenerate cases. Indeed, the condition that the allowed part of the y^2 plane be of dimension less than 2 is equivalent to imposing $m-2$ new conditions on the Z_{ij}, which cannot be satisfied on a measurable part of *BMD*.

Next we show that the allowed part, in the cases in which it is a domain, is the whole y^2 plane. Indeed, if it were not, it would have a boundary, on which a complex number $N(y^2)$ could be defined, with the significance of a normal, which would be determined up to a real positive factor by

$$\mathrm{Re}\{N^*(ydy)\}\geqslant 0, \tag{71}$$

for any dy^μ such that $y^\mu+dy^\mu$ belongs to the family of solutions. The condition on dy^μ is

$$(dyY_i)=\text{real}, \quad \text{and} \quad \text{zero for } i>m. \tag{72}$$

Whatever $y^\mu\neq 0$, the sign of dy^μ is not restricted. Hence the sign of N is not defined, contrary to the assumption. Therefore, no boundary exists, and the allowed part of the y^2 plane includes the real axis. Therefore the cases $m\geqslant 3$ do not contribute to the boundary. This result was known[4] for $n=3$, and conjectured as well as made plausible[5] for $n=4$.

We have thus seen that only the cases $m=1$ and $m=2$ are interesting. As we have replaced condition (i) by the weaker condition (i') we do not know what part, if any, of these two manifolds make up *BMD*. We also do not know what part of *BMD* constitutes the boundary of the singularity domain.[16] Nevertheless, we proceed to write down the analytical equations for the "one-mass" ($m=1$) and "two-mass" ($m=2$) manifolds, hoping that their form might give useful hints in the problem of finding the holomorphy envelope of the field theory domain.

The one-mass manifold is trivially written down. All we have to do is to write down (53) in terms of Z_{ij}, using (54), (55), (60), and (61). Then we have to calculate u_i, in order to impose condition (64). To obtain the two-mass manifold we impose the reality of $u_1/u_2=\rho$, and eliminate the other u_i from (66). Writing the result in terms of Z_{ij}, using (54), (55), (60), and (61), we end up with two equations involving the three real numbers a_1, a_2, and ρ. In this case too, the u_i must be explicitly calculated, in order to impose condition (64). Carrying out this program, we specialize to the case $n=3$.

It is convenient to introduce, rather than the Z_{ij} defined by (55), the Z_i, $i=1, 2, 3$, defined by (51). Then the solution of (66) for the u_i is, with $Y_1^2=a_1=0$,

$$\frac{u_1}{u_3}=\frac{-2a_2(Z_3-a_3)-(Z_1-a_2)(Z_2-a_2-a_3)}{(Z_1-a_2)^2}, \tag{73}$$

$$\frac{u_2}{u_3}=-\frac{Z_3-a_3}{Z_1-a_2}. \tag{74}$$

Inserting this into (66) we get

$$a_3(Z_1-a_2)^2+(Z_2-a_2-a_3)(Z_1-a_2)(Z_3-a_3)+a_2(Z_3-a_3)^2=0. \tag{75}$$

[16] What we do know, from reference 3, is that the singularity domain is bounded by cuts and by part of *BMD*.

We obtain a one-mass manifold simply by putting $a_3=0$, which gives either $Z_3=0$ or

$$Z_1Z_2+a_2(Z_3-Z_1-Z_2)+a_2^2=0, \tag{76}$$

which is just the F_{12}' curve. Of course the other F_{ij}' curves are obtained by puting either $a_1=a_2=0$ or $a_2=a_3=0$. The "relevance criterion" (64) reduces to, using (76),

$$\mathrm{Im}Z_1\,\mathrm{Im}Z_2>0, \tag{77}$$

which is just the relevance criterion (34) for the F_{12}' curve.

The two-mass manifold, on the other hand, is obtained by leaving a_2 and a_3 nonzero in (73)–(75), and imposing $u_2/u_3=\rho=$ real:

$$(Z_3-a_3)+\rho(Z_1-a_2)=0. \tag{78}$$

Inserting this result into (75) we get

$$Z_2=(a_3+\rho a_2)(1+\rho^{-1}). \tag{79}$$

Since only two real parameters, namely ρ and $(a_3+\rho a_2)$ are involved in (78), (79), we have the surprising result that the two-mass manifold is four-dimensional, and therefore it does not contribute to the boundary $(n^2-n-1=5)$.

The over-all results of this section are (a) only two kinds of manifolds, the $m=1$ and the $m=2$ manifolds, contribute to the perturbation theory boundary, and (b) in the special case $n=3$, we find that the $m=2$ manifold is degenerate.

D. Comparison

In the particular case $n=3$ we were lucky to discover immediately that the two-mass manifolds are the same as those encountered in the field theoretical problem. For higher values of n we are less fortunate. On the other hand it is very difficult, from the parametrized equation for two surfaces, to prove that they are distinct. In this section we therefore compare the surfaces obtained in Secs. (A) and (B) with those of Sec. (C), by trying to identify the respective normals N_{ij}.

Suppose that a point in Z_{ij} space is on the perturbation theory boundary. We want to determine under what circumstances this boundary is tangent to the field theoretical boundary at that point. If that is the case, we have two alternative definitions of the normal, namely[17]

$$\sum_{j=1}^{n-1} N_{ij}{}^*X_j{}^\mu=-\sum_{k=1}^{i} u_kY_k{}^\mu=-i\lambda_i\,\mathrm{Im}X_i{}^\mu, \tag{80}$$

where

$$\sum_{k=1}^{n} u_kY_k{}^\mu=0, \quad \mathrm{Re}u_i>0 \quad \text{if} \quad Y_i{}^2=0, \tag{81}$$

$$Y_i{}^2\geqslant 0, \quad \mathrm{Re}u_i=0 \quad \text{if} \quad Y_i{}^2>0,$$

$$\lambda_i>0, \quad (\mathrm{Im}X_i)^2=0. \tag{82}$$

The most obvious consequences of (80)–(82) are the following; assuming all $u_i\neq 0$:

$$Y_1{}^2=Y_n{}^2=0, \tag{83}$$

$$\mathrm{Im}Y_i{}^\mu=0 \quad \text{if} \quad Y_i{}^2>0. \tag{84}$$

When $n=3$, (83) means that we are on the one-mass manifold, while (84) can always be made to hold by applying a complex Lorentz transformation. We next find

$$u_2{}^2Y_2{}^2=2\lambda_1\lambda_2(\mathrm{Im}X_1,\,\mathrm{Im}X_2). \tag{85}$$

On the S curve, the right-hand side vanishes, and we would have $(Y_iY_j)=0$. On the F curve, the right-hand side is positive, while the left-hand side is negative. But on the F' curve, both sides of (85) are negative. Using (84), and[17] $X_1{}^\mu=Y_i{}^\mu-Y_{i+1}{}^\mu$, (80) reduces to

$$\lambda_1\,\mathrm{Re}u_1{}^{-1}=\lambda_3\,\mathrm{Re}u_3{}^{-1}=1. \tag{86}$$

Hence we have the expected result, that the boundary of perturbation theory agrees with the boundary of the secondary domain when they are both relevant,

II. THE FOUR-POINT FUNCTION

In the following we rely heavily on Part I to apply our methods to the four-point function. In order to minimize the inconvenience caused by the necessity of frequent references, this part is organized along the same lines as the foregoing: (A) the primitive domain, (B) analytic completion, (C) perturbation theory, and (D) comparison.

A. The Primitive Domain

The discussion of Part I needs no modification up to Eq. (23). We note that it was proved that only $n-1=3$ among the 6 vectors $a_i{}^\mu=\mathrm{Re}X_i{}^\mu$, $b_i{}^\mu=\mathrm{Im}X_i{}^\mu$ can be linearly independent at a point that is mapped on the boundary $BM\mathcal{T}$ in Z space.

We must distinguish three cases, according to whether the $b_i{}^\mu$ are colinear, coplanar, or neither. In the following, we discuss the last possibility only. A parallel treatment of the other two cases leads quickly to the conclusion that manifolds of too low dimensions are obtained.

There exists a Lorentz frame such that, on the boundary,

$$\begin{aligned} X_1{}^\mu&=\{\tfrac{1}{2}(r_1+w_1),\,\tfrac{1}{2}(r_1-w_1),\,s_1,\,t_1\},\\ X_2{}^\mu&=\{\tfrac{1}{2}(r_2+w_2),\,t_2,\,\tfrac{1}{2}(r_2-w_2),\,s_2\},\\ X_3{}^\mu&=\{\tfrac{1}{2}(r_3+w_3),\,s_3,\,t_3,\,\tfrac{1}{2}(r_3-w_3)\}. \end{aligned} \tag{87}$$

[17] Here we made a seemingly unwarranted indentification of the X_i of Secs. A and B with the X_i of Sec. C.[14] That this can be justified is shown in Part II, Sec. D.

The condition that only 3 among the 6 vectors $\mathrm{Re}X_i^\mu$, $\mathrm{Im}X_i^\mu$ be linearly independent is that the vector $\{1, -1, -1, -1\}$, which is normal to the imaginary parts, be normal to the real parts as well, that is

$$r_i+s_i+t_i=0. \tag{88}$$

In analogy with the 3-point case we see that (11) reduces[18] to 6 homogeneous and 3 inhomogeneous equations for the six N_{ij}^*. A novel feature, however, is that the six homogeneous equations are not automatically solvable. We therefore get new conditions on the parameters by setting the secular determinant equal to zero. Notice that, because of (88), (87) contains only 12 real parameters. Since the dimension of $BM\mathcal{T}$ is 11, only one more real condition can be imposed. Therefore, the vanishing of the determinant must be an identity in the w_i. If we calculate, e.g., the highest power of w_i, we find that it is

$$w_1w_2w_3(s_1s_2s_3-t_1t_2t_3). \tag{89}$$

We impose the condition

$$s_1s_2s_3=t_1t_2t_3, \tag{90}$$

and find that a solution of the 6 homogeneous equations then exists, giving all the N_{ij}^* up to a common complex factor ρ. Inserting this result into the three inhomogeneous equations, we find that λ_1/λ_2 and λ_2/λ_3 are real, as they must be, and that they are positive iff (assuming $\mathrm{Im}w_i>0$)

$$s_it_j>0, \quad (ij)=(12),\ (23),\ (31). \tag{91}$$

Using (87) we may write down expressions for $Z_{ij}=(X_iX_j)$. The diagonal elements may be solved for w_i [Eq. (22)], and used to eliminate w_i from the three off-diagonal elements. In this way we obtain the following three complex equations, involving five independent real parameters [$(ij)=(12),\ (23),\ (31)$]:

$$4(s_i+t_i)(s_j+t_j)[Z_{ij}-t_it_j-s_is_j-s_it_j+t_is_j] -(Z_{ii}+2t_i^2)(Z_{jj}+2s_j^2)=0, \tag{92}$$

$$s_1s_2s_3=t_1t_2t_3. \tag{93}$$

The "relevance conditions" (23) become

$$(s_i+t_i)\,\mathrm{Im}Z_{ii}<0, \quad i=1,\ 2,\ 3. \tag{94}$$

An interesting form of (92) is

$$2s_j^2[Z_{ii}+2k_{ij}Z_{ij}+k_{ij}^2Z_{jj}] +[Z_{ii}Z_{jj}-2\gamma_{ij}Z_{ij}+\gamma_{ij}^2]=0, \tag{95}$$

where

$$k_{ij}=-t_i/s_j, \quad \gamma_{ij}=2(s_is_j+t_it_j+s_it_j), \quad (ij)=(12),\ (23),\ (31). \tag{96}$$

Results equivalent to these were first obtained by Jost.[19]

[18] In this case it would be simpler to solve (13).

[19] R. Jost, to be published in the Proceedings of the Naples Conference, 1959.

The particular form (95) was first written down by Kleitman.[20]

In addition to the cuts and the four-point boundary that has just been determined, we obtain the three-point boundary by setting any one of the λ_i equal to zero. The result is the following set of S curves and F_{ij} curves:

$$Z_{ii}+2k_{ij}Z_{ij}+k_{ij}^2Z_{jj}=0, \quad k_{ij}>0, \tag{97}$$

$$Z_{ii}Z_{jj}-2\gamma_{ij}Z_{ij}+\gamma_{ij}^2=0, \quad \gamma_{ij}>0. \tag{98}$$

Here i, j is 1, 2 or 2, 3 or 3, 1.

B. Analytic Completion

There are 24 Wightman four-point functions, but only 12 distinct primitive domins. These may be divided into three groups of four domains each, such that Streater's theorem is applicable to four pairs of domains within each group, but not to any pair of domains from different groups. One such group consists of the following four domains

$$\begin{aligned} &ABCD: \ Y_1-Y_2, \ Y_2-Y_3, \ Y_3-Y_4\in\mathcal{T}',\\ &BACD: \ Y_2-Y_1, \ Y_1-Y_3, \ Y_3-Y_4\in\mathcal{T}',\\ &ABDC: \ Y_1-Y_2, \ Y_2-Y_4, \ Y_4-Y_3\in\mathcal{T}',\\ &BADC: \ Y_2-Y_1, \ Y_1-Y_4, \ Y_4-Y_3\in\mathcal{T}'. \end{aligned} \tag{99}$$

Streater's theorem may be applied to, e.g., the first two and to the last two, to give

$$\text{E.H.}\{ABCD\cup BACD\}=\{Y_1-Y_3,\ Y_2-Y_3, Y_3-Y_4\in\mathcal{T}'\}\cap\{Y_1-Y_2\in\mathcal{T}'\}, \tag{100}$$

$$\text{E.H.}\{ABDC\cup BADC\}=\{Y_1-Y_4,\ Y_2-Y_4, Y_4-Y_3\in\mathcal{T}'\}\cap\{Y_1-Y_2\in\mathcal{T}'\}. \tag{101}$$

Each of these domains will be called a secondary domain. As in the three-point case, a secondary domain is the envelope of holomorphy of two primitive domains.

The boundary of (100) may easily be determined. Except for the $(Y_1-Y_2)^2$ cut it may be obtained from the boundary of the primitive domain by means of the substitution $X_1^\mu\to X_1^\mu+X_2^\mu$. If we introduce the notation

$$Z_i=Z_{ii}=X_i^2, \quad i=1,\ 2,\ 3,$$

$$X_4=(X_1+X_2)^2, \quad Z_5=(X_2+X_3)^2, \quad Z_6=(X_1+X_2+X_3)^2, \tag{102}$$

the curves (98) become

$$Z_4Z_2+\gamma_{12}(Z_1-Z_2-Z_4)+\gamma_{12}^2=0, \quad \gamma_{12}>0, \tag{103}$$

$$Z_2Z_3-\gamma_{23}(Z_5-Z_2-Z_3)+\gamma_{23}^2=0, \quad \gamma_{23}>0, \tag{104}$$

$$Z_4Z_3+\gamma_{31}(Z_6-Z_3-Z_4)+\gamma_{31}=0, \quad \gamma_{31}>0. \tag{105}$$

These are all, except for the minus sign in (104), part of the corresponding perturbation theory boundary manifolds. From (97) are obtained, in a like manner,

[20] G. Källen, Lecture notes of the summer school at Les Houches, 1960.

two S' curves and one S curve, and the modified form of (95) may immediately be written down.

Streater[3] has also succeeded in finding a characterization of the "quartic domain" which is the holomorphy envelope of the union of the four primitive domains (99).

Theorem B.[3] The envelope of holomorphy of the union of the four domains (99) is the domain

$$U_1(X_1,X_2,X_3)\bigcap\{X_1\in\mathcal{T}'\}\bigcap\{X_3\in\mathcal{T}'\}, \quad (106)$$

where U_1 is the common domain of analyticity of all functions

$$F(X_1,X_2,X_3)=\int^{i(p\cdot X_1+q\cdot X_2+r\cdot X_3)} f(p,q,r) \times d^4p\,d^4q\,d^4r, \quad (107)$$

whose Fourier transforms $f(p,q,r)$ vanishes unless

$$q>p>0, \quad q>r>0. \quad (108)$$

Here $p>0$ means $p\in V_+$, the forward lightcone.

In order to determine $U_1(X_1,X_2,X_3)$, write the exponential in (107) in terms of $(q-p)$, $(q-r)$, p, and r. The most general form is

$$p\cdot X_1+q\cdot X_2+r\cdot X_3=(q-p)\cdot X+(q-r)\cdot(X_2-X) +p(X_1+X)+r\cdot(X_2+X_3-X),$$
$$X^\mu=a_1X_1{}^\mu+a_2X_2{}^\mu+a_3X_3{}^\mu. \quad (109)$$

Writing $s^\mu=q^\mu-p^\mu$, $t^\mu=q^\mu-r^\mu$, (107) becomes

$$F(X_1,X_2,X_3)=\int \exp\{i[s\cdot X+t\cdot(X_2-X)+p\cdot(X_1+X) +r\cdot(X_2+X_3-X)]\}\cdot f'(p,r,s,t)d^4p\,d^4r\,d^4s\,d^4t, \quad (110)$$

$$f'(p,r,s,t)=f(p,r+t,r)\delta(s-t+p-r). \quad (111)$$

The function $F(X_1,X_2,X_3)$ is analytic if X^μ can be chosen such that the coefficients of s, t, p, r in (110) all lie in the future tube. We call this domain U_1':

$$X_i\in U_1'(X_1,X_2,X_3) \text{ iff there exists an } X, \text{ such that } X,\ X_2-X,\ X_1+X,\ X_2+X_3-X\in\mathcal{T}'. \quad (112)$$

Since $f'(p,r,s,t)$ is not the most general function of the four arguments, $F(X_1,X_2,X_3)$ may be regular in a larger domain, and we can only conclude that $U_1'\subset U_1$ and hence that the envelope of holomorphy of the four domains (99) is bigger than or equal to

$$U_1'(X_1,X_2,X_3)\bigcap\{X_1\in\mathcal{T}'\}\bigcap\{X_3\in\mathcal{T}'\}. \quad (113)$$

There are two domains that can be obtained from (113) by permutations, namely

$$U_2'(X_1,X_2,X_3)\bigcap\{X_1+X_2\in\mathcal{T}'\}\bigcap\{X_2+X_3\in\mathcal{T}'\}, \quad (114)$$

$$U_3'(X_1,X_2,X_3)\bigcap\{X_2\in\mathcal{T}'\}\bigcap\{X_1+X_2+X_3\in\mathcal{T}'\}, \quad (115)$$

where U_2' and U_3' may be defined in a manner similar to (112):

$$X_i\in U_2' \text{ iff there exists an } X, \text{ such that } X,\ -X_2-X,\ X_1+X_2+X,\ X_3-X\in\mathcal{T}', \quad (116)$$

$$X_i\in U_3' \text{ iff there exists an } X, \text{ such that } X,\ -X_2-X_3-X,\ X_1+X_2+X_3+X,\ -X_3-X\in\mathcal{T}'. \quad (117)$$

The problem of computing the envelope of holomorphy of the union of the 12 primitive domains is now reduced to finding the envelope of holomorphy of the union of (113)–(115). We shall content ourselves with showing how the boundary of (112), say, may be found.

The method of the normal is easily adaptable to finding the mapping of U_1' directly, without explicitly calculating U_1'. The procedure is exactly analogous to the treatment of (56). Writing x, x_i for the imaginary parts of X, X_i, we find at once

$$iN_{1i}X_i{}^\mu=f_1{}^\mu+\sigma_1(x_1+x)^\mu,$$
$$iN_{2i}X_i{}^\mu=f_1{}^\mu+f_2{}^\mu+\sigma_1(x_1+x)^\mu+\sigma_2x^\mu,$$
$$iN_{3i}X_i{}^\mu=f_1{}^\mu+f_2{}^\mu-f_3{}^\mu+\sigma_1(x_1+x)^\mu+\sigma_2x^\mu-\sigma_3(x_2-x)^\mu =f_4{}^\mu+\sigma_4(x_2+x_3-x)^\mu. \quad (118)$$

Here $f_i{}^\mu$ are real vectors in V_+, and $f_i{}^\mu=0$ unless the coefficient of σ_i vanishes. The σ_i vanish unless their respective coefficients are lightlike. The solution of (118) is straightforward. Explicit results will be published elsewhere.

Before ending this section we mention that Streater[21] has proposed that the following theorems might hold:

Theorem C (conjectured).[21] The envelope of holomorphy of the union of the domains $ABCD$, $ABDC$, $ACDB$, and $ADCB$ is

$$\{X_1,\ X_1+X_2,\ X_1+X_2+X_3\in\mathcal{T}'\} \bigcap\{X_2,\ X_2+X_3\in\mathcal{T}'\}\bigcap\{X_3\in\mathcal{T}'\}. \quad (119)$$

Theorem D (conjectured).[21] The envelope of holomorphy of the union of the domains $ABCD$ and $BDCA$ is

$$\{X_1,\ X_1+X_2,\ X_3,\ X_1+X_2+X_3\in\mathcal{T}'\} \bigcap\{X_2,\ X_2+X_3\in\mathcal{T}'\}. \quad (120)$$

The boundaries of (119) and (120) are easily determined by our methods. Thus to find the boundary of (120) we would consider

$$\sum_i N_{ij}{}^*X_j{}^\mu=-i\lambda_1b_1{}^\mu-i\lambda_2(b_1{}^\mu+b_2{}^\mu)-i\lambda_3(b_1{}^\mu+b_2{}^\mu+b_3{}^\mu),$$
$$\sum_i N_{2j}{}^*X_j{}^\mu=-i\lambda_2(b_1{}^\mu+b_2{}^\mu)-i\lambda_3(b_1{}^\mu+b_2{}^\mu+b_3{}^\mu), \quad (121)$$
$$\sum_i N_{3j}{}^*X_j{}^\mu=-i\lambda_3(b_1{}^\mu+b_2{}^\mu+b_3{}^\mu)-i\lambda_4b_3{}^\mu.$$

[21] R. F. Streater, Nuovo cimento **15**, 937 (1960).

C. Perturbation Theory

First a word about the three-point parts of the four-point boundary. These are obtained by setting one of the u_i equal to zero. Suppose $u_4=0$; then the right-hand side of (65), and *a fortiori* the left-hand side, become independent of Y_4^μ. The problem then reduces to that of a three-point function depending on Y_1^μ, Y_2^μ, and Y_3^μ. The interesting point is that one obtains no F_{ij}' curve such that X_i^μ, X_j^μ are conjugate X variables, but all the others. This selectivity mirrors that of other sources of singularity of the four-point function in perturbation theory. In this particular respect the perturbation theory boundary agrees with the boundaries of the quartic domains (113)–(115).

Turning now to the four-point part of the boundary, we solve (66). Taking $a_1=a_4=0$, and writing Y_{ij} for (Y_iY_j), we find

$$u_1/u_3=-(Y_{34}+\rho Y_{24})/Y_{14}, \tag{122a}$$

$$u_4/u_3=-(Y_{13}+\rho Y_{12})/Y_{14}. \tag{122b}$$

$$\rho=\frac{u_2}{u_3}=\frac{Y_{23}Y_{14}-Y_{12}Y_{34}-Y_{13}Y_{24}}{2Y_{12}Y_{24}-a_2Y_{14}} \tag{123a}$$

$$=\frac{2Y_{13}Y_{34}-a_3Y_{14}}{Y_{23}Y_{14}-Y_{12}Y_{34}-Y_{13}Y_{24}}. \tag{123b}$$

We consider first the one-mass manifold, with $a_1=a_2=a_4=0$, $a_3=a>0$. Then the last equality in (123), which expresses the solubility of (66) [it is the expansion of (53)], is the only equation. To write it in terms of the Z-space variables we introduce the notation (102), and obtain[22]

$$[(Z_2-a)Z_6-(Z_3-a)Z_1-(Z_4-a)Z_5]^2 =4Z_1Z_5[Z_3Z_4+a(Z_6-Z_3-Z_4)+a_3^2]. \tag{124}$$

As we have stressed before, we do not know all the relevance criteria for this manifold,[16] but (63) gives

$$\mathrm{Im}(u_i/u_3)>0, \quad i=1, 2, 4. \tag{125}$$

This simply means that the relevance of the one-mass manifold changes at the intersection with one of the two-mass manifolds.

On the two-mass manifold, with $a_1=a_4=0$, a_2, $a_3\neq 0$, ρ is real, and (123) gives[22]

$$(Z_2-a_2-a_3)Z_6-(Z_1-a_2)(Z_3-a_3)-(Z_4-a_3)(Z_5-a_2) =2\rho[Z_1Z_5+a_2(Z_6-Z_1-Z_5)+a_2^2] \tag{126a}$$

$$=2\rho^{-1}[Z_3Z_4+a_3(Z_6-Z_3-Z_4)+a_3^2]. \tag{126b}$$

[22] The forms (124) and (126) agree with the results of reference 3. The relevance criteria (125) and (127) do not appear to have been given in that paper, however.

The relevance criteria (63) become

$$\mathrm{Im}\left(\frac{a_3+\rho a_2-Z_3-\rho Z_5}{Z_6}\right)>0,$$

$$\mathrm{Im}\left(\frac{a_3+\rho a_2-Z_4-\rho Z_1}{Z_6}\right)>0. \tag{127}$$

Perhaps a more interesting presentation of (126) is the following

$$(\rho Z_6)^{-1}[a_3+\rho a_2-Z_3-\rho Z_5][a_3+\rho a_2-Z_4-\rho Z_1] =Z_2-(a_3+\rho a_2)(1+\rho^{-1}), \tag{128}$$

$$\rho^2=\frac{Z_1Z_5+a_2(Z_6-Z_1-Z_5)+a_2^2}{Z_3Z_4+a_3(Z_6-Z_3-Z_4)+a_3^2}. \tag{129}$$

We see the analogy with the corresponding two-mass manifold in the three-point case; (128) corresponds to (79), and (129) to (74). Indeed, while the numerator and denominator of (74) vanish on cuts, those of (129) vanish on F_{ij}' curves. Notice that (128) involves ρ and $(a_3+\rho a_2)$ only. The same is true of (79), and "by accident" (74) also depends on these two parameters. This "accident" causes the degeneracy of the $m=2$ manifold for $n=3$. For $n=4$ there is no accidental degeneracy, however. Another aspect of (128) is that it represents a nearly successful attempt to simplify (127) (the right-hand side is "nearly" real).

We found above that the three-point parts of the four-point perturbation theory boundary bears significant resemblance to the extended field theoretical domains. No such agreement is apparent for the four-point parts. The most striking difference is that the perturbation theory manifolds are obtained in terms of one and three parameter representations, while field theory yields five parameter representations. We know of no way of deciding whether or not in a given case the number of parameters can be reduced without destroying the analytic form of the equations. The method of the following section is therefore indispensible in this case.

D. Comparison

A common feature of all the domains discussed in Sec. A, both the primitive domains and the proven or conjectured envelopes of holomorphy, is that they are all intersections of mappings of tubes. We shall now prove that no part of the one- or two-mass manifolds of the last section is the mapping of a tube.

In order to apply our method of comparing the two normals, it is necessary to justify the dual roles of both Y_i^μ and X_i^μ. Suppose our attention be fixed on a particular point ${}^0Z_{ij}$ in Z_{ij} space. In our discussion of

the axiomatic approach, and, in particular, in the definition of N_{ij}, there occur vectors X_i^μ. Let ${}^0Z_{ij}$ be on the boundary of one of the four point field-theoretical domains. Then values ${}^0X_i^\mu$ of X_i^μ can be found, such that $({}^0X_i, {}^0X_j)={}^0Z_{ij}$, and

$$\sum_j {}^0N_{ij}{}^* {}^0X_j^\mu = i\Lambda_i^\mu. \tag{130}$$

Here, Λ_i^μ is a set of real vectors. Thus, in the case of the primitive domain, $\Lambda_i^\mu = -\lambda_i \operatorname{Im}{}^0X_i^\mu$. The important point is that Λ_i^μ are always real for the mapping of tubes.

In our discussion of perturbation theory, complex vectors X_i^μ and Y_i^μ likewise occur. Suppose that the point ${}^0Z_{ij}$ is on the perturbation theory boundary. Then particular values of $X_i^\mu = Y_i^\mu - Y_{i+1}^\mu$ can be found which satisfy $(X_iX_j)={}^0Z_{ij}$, and

$$\sum_j N_{ij}{}^* X_i^\mu = -\sum_{k=1}^{i} u_k Y_k^\mu. \tag{131}$$

However, we have no assurance that X_j^μ equals ${}^0X_j^\mu$, so that the comparison of (130) and (131) is not immediately possible. We note, however, that (131), along with all the subsidiary conditions on u_i and Y_i^μ, are form-invariant under complex Lorentz transformations. We may therefore apply a complex Lorentz transformation which carries X_j^μ into ${}^0X_j^\mu$ without changing the form of (131). This is possible because $(X_i,X_j)={}^0Z_{ij}=({}^0X_i,{}^0X_j)$. The result is

$$\sum_j N_{ij}{}^* {}^0X_i^\mu = -\sum_{k=1}^{i} u_k {}^0Y_i^\mu, \quad {}^0X_i^\mu = {}^0Y_i^\mu - {}^0Y_{i+1}^\mu. \tag{132}$$

Now (130) may be compared with (131). If the two boundaries touch (rather than cross) at ${}^0Z_{ij}$, then ${}^0N_{ij}$ and N_{ij} must be equal up to a positive over-all factor. We therefore conclude that

$$iu_i {}^0Y_i^\mu = \text{real}. \tag{133}$$

But this is possible on a 10-dimensional subspace of Z_{ij} space only. Therefore, the perturbation theory boundary cannot be the boundary of the mapping of a tube. Q.E.D.

The meaning of this is that, if perturbation theory is to be of any help in finding the holomorphy envelope for the four-point function, then the partial results obtained up to now bear little similarity to the final result.

ACKNOWLEDGMENTS

It is a pleasure to thank V. Glaser, R. F. Streater, and A. Wightman for helpful conversations, and G. Källen for informative correspondence. I am grateful to CERN, the University of Bergen, the University of Pennsylvania, and especially to A. Klein for the hospitality extended to me. This work was supported in part by the Atomic Energy Commission.

Support of a Field in p Space

G. F. DELL'ANTONIO*,†,‡
Northwestern University, Evanston, Illinois
(Received May 31, 1961)

The notion of generalized free field is introduced, as an obvious extension of a discrete superposition of independent free fields with different masses. The following assumptions are also made: there is an underlying Hilbert space $\mathfrak{H}$ (positive-definite metric), the theory is Lorentz invariant, the vacuum belongs to $\mathfrak{H}$ and is there unique, the spectrum of the energy-momentum operator is—apart from the origin—completely contained within the region $p^2 \geq \epsilon^2$, $p_0 > 0$. It is then shown that a necessary condition for a cyclic field to have support in p^2 only on a finite interval of the positive real axis, is that $A(x)$ be a generalized free field. In the Appendix a similar result is derived under slightly weaker conditions.

1. INTRODUCTION AND DEFINITIONS

LET $\mathfrak{H}$ be a Hilbert space on which a unitary representation $U(a,\Lambda)$ of the inhomogeneous proper Lorentz group $L_+^+(a,\Lambda)$ is defined. Let π^μ be the infinitesimal generators of translations. We assume that there exists in $\mathfrak{H}$, and is there unique, the eigenstate of π^μ to the eigenvalue zero. This state will be called the "vacuum" and denoted by $|\Omega\rangle$. Apart from this state, the spectrum[1] of $\pi^2 \equiv \pi^\mu \pi_\mu$ is assumed to be bounded from below by a positive number, say $\epsilon^2 > 0$. It is also assumed that, apart from the vacuum, the spectrum of π^0 is positive definite.

We now introduce the definition of "generalized free fields"[2] restricting ourselves to real, scalar fields; the definition can be extended in an obvious way to the case in which the field is not real and/or transforms according to a finite-order representation of the homogeneous proper Lorentz group. The theorems to be proven below (Theorems I and II) remain valid also in this case, with obvious modifications in form. We restrict ourselves throughout this note to real scalar fields only to simplify the algebra involved in the proofs. We shall also assume that all quantities we introduce are tempered distributions; the proofs remain valid, however, also in more general cases.

Definition: Let $A(x)$ be a real field, $A(x) = A^+(x)$. Define

$$a(\varphi) \equiv \int a(p)\varphi(p)dp, \qquad a(p) = \int e^{ipx} A(x) d^4x, \tag{1.1}$$

$$\varphi^*(p) = \varphi(-p), \qquad a^\dagger(\varphi) = a(\varphi^*).$$

We shall say that $A(x)$ is a generalized free field if: $a(\varphi) = 0$ if $\varphi(p)$ has support only at $p^2 < 0$ or at $p = 0$ and there exists a measure $\mu(\sigma)$ such that the following identities hold:

$$[b^\dagger(\psi), b(\psi')] = -\int_{p_0, p^2 > 0} d^4p \psi^*(p)\psi'(p)$$
$$[b_i^\dagger(\varphi), b_j(\varphi')] = \delta_{m_i m_j} \int d^3p \varphi^*(\mathbf{p})\varphi'(\mathbf{p}) \tag{1.2}$$

all other commutators $= 0$

$$b(\psi) \equiv \int_{\substack{p_0 > 0 \\ p^2 > 0}} d^4p b(p)\psi(p); \quad b_i(\varphi) = \int d^3p b(\mathbf{p}, m_i)\varphi(\mathbf{p})$$

with

$$a(p) = \theta(p_0) \int_0^\infty d\mu(\sigma) b(\sigma, \mathbf{p}) \delta(p^2 - \sigma) + \theta(-p_0) \int_0^\infty d\mu(\sigma) b^\dagger(\sigma, -\mathbf{p}) \delta(p^2 - \sigma). \tag{1.3}$$

$\psi(p)$, $\varphi(\mathbf{p})$ are infinitely many times differentiable functions with compact support. $\{m_i\}$ is the set of points to which the measure $\mu(\sigma)$ attributes finite weight. Notice that, if

$$d\mu(\sigma) = \sum_1^N C_i \delta(\sigma - m_i^2),$$

$A(x)$ is a superposition of N independent real free fields. Equation (1.3) takes in fact the form

$$a(p) = \theta(p) \textstyle\sum_i C_i b_i(\mathbf{p}) p_0^{\frac{1}{2}} \delta(p^2 - m^2) + \theta(-p_0) \sum_i C_i b_i^\dagger(-\mathbf{p}) p_0^{\frac{1}{2}} \delta(p^2 - m^2) \tag{1.3'}$$

with $b_i(\mathbf{p}) \equiv p_0^{-\frac{1}{2}} b(\sigma, \mathbf{p})\,|_{\sigma = m_i^2}$ and (1.2) reads

$$[b_i(\mathbf{k}), b_j^\dagger(\mathbf{p})] = \delta_{ij}\delta(\mathbf{k} - \mathbf{p}); \quad [b_i(\mathbf{p}), b_j(\mathbf{k})] = [b_i^\dagger(\mathbf{p}), b_j^\dagger(\mathbf{k})] = 0. \tag{1.4}$$

2. CONSEQUENCES OF A RESTRICTION ON THE SUPPORT OF $a(p)$

The purpose of the present article is to give a proof of the following statement:

Theorem I. If a field $A(x)$ is cyclic in $\mathfrak{H}$ with respect to the vacuum and it satisfies the following conditions:

$$U(a,\Lambda)A(x)U^{-1}(a,\Lambda) = A(\Lambda x + a), \tag{1}$$

---

* On leave of absence from the Istituto Nazionale di Fisica Nucleare, Milano, Italy.

† Supported by a National Science Foundation grant.

‡ Now at Istituto di Fisica Teorica dell' Università, Napoli, Italy.

[1] Metric $+I, -I, -I, -I$.

[2] An analogous definition has been given by O. Greenberg (Proceedings of the 1961 Washington meeting of the American Physical Society).

where $U(a,\Lambda)$ is a representation of the inhomogeneous proper Lorentz group.

$$[A(x),A(y)]=0 \tag{2}$$

when $(x-y)^2<0$.

$$a(p)\neq 0 \tag{3}$$

only[3] for $0\leqslant p^2\leqslant M^2$ (M^2 is some positive number), then $A(x)$ is a generalized free field.[4]

The first step in the proof of the theorem will be to show (Lemma I) that, under conditions (1)–(3) (and with our assumption about the vacuum[5] and the spectrum of π^2), the commutator $[A(x),A(y)]$ is a c number. This result, together with Theorem II (to be proven in the next section) will then give the proof of Theorem I.

Lemma I. Under conditions (1)–(3) of Theorem I, the quantity $F(x,y)\equiv[A(x),A(y)]$ is a c number.

Proof: Consider $F_\phi(x,y)\equiv\langle\phi|[A(x),A(y)]|\Omega\rangle$ where $|\phi\rangle$ is an eigenstate[6] of π^μ. We shall first prove that

$$F_\phi(x,y)=0 \tag{2.1}$$

if

$$p_\phi{}^2>0. \tag{2.2}$$

Since the vacuum is, according to our assumptions, the only eigenstate of π^2 in $\mathfrak{H}$ for which (2.2) is not satisfied, (2.1) implies that

$$[A(x),A(y)]|\Omega\rangle=C(x,y)|\Omega\rangle, \tag{2.3}$$

where $C(x,y)$ is a certain function of (distribution in) x, y. To prove (2.1) let us consider separately the two functions

$$F_\phi'(x,y)\equiv\langle\phi|A(x)A(y)|\Omega\rangle \tag{2.4}$$

and

$$F_\phi{}^2(x,y)\equiv\langle\phi|A(y)A(x)|\Omega\rangle. \tag{2.5}$$

Inserting in (2.4) a complete set of eigenstates of π^μ we get

$$F_\phi'(x,y)=\sum_n\langle\phi|A(x)|n\rangle\langle n|A(y)|\Omega\rangle=$$

(the symbol $\sum_n$ stands for summation over discrete indices and integration over a continuum)

$$=\sum_n e^{-ip_\phi\cdot x+ip_n\cdot(x-y)}\langle\phi|A(0)|n\rangle\langle n|A(0)|\Omega\rangle,$$

where p_n is defined through $\pi^\mu|n\rangle=p_n{}^\mu|n\rangle$. We now define a function $F_\phi'(z_1z_2)$ by

$$F_\phi'(z_1z_2)\equiv e^{-ip_\phi\cdot z_1}\sum_n e^{ip_n(z_1-z_2)}\times\langle\phi|A(0)|n\rangle\langle n|A(0)|\Omega\rangle. \tag{2.6}$$

This function is analytic in every finite region of the z_1, z_2 space provided only

$$\mathrm{Im}(z_1-z_2)\in V^+ \ (V^+ \text{ is the forward lightcone}). \tag{2.6'}$$

We also have

$$\lim_{\substack{\mathrm{Im}z_i\to 0\\ \mathrm{Im}(z_1-z_2)\in V_+}} F_\phi'(z_1z_2)=F_\phi'(x,y).$$

We consider now

$$F_\phi{}^2(x,y)=\sum_n e^{-ip_\phi\cdot y+ip_n\cdot(y-x)}\langle\phi|A(0)|n\rangle\langle n|A(0)|\Omega\rangle$$

and introduce a new function

$$F_\phi{}^2(z_1z_2)=\sum_n e^{-ip_\phi\cdot z_2+ip_n(z_2-z_1)}\langle\phi|A(0)|n\rangle\langle n|A(0)|\Omega\rangle.$$

It will be convenient for what follows to consider as independent variables z_1 and $z_3=z_1-z_2$. $F_\phi'(z_1,z_3)$ is analytic for $\mathrm{Im}z_3\in V_+$ and every finite value of z_1. $F_\phi{}^2(z_1z_3)$ will be rewritten

$$F_\phi{}^2(z_1z_2)=e^{-ip_\phi\cdot z_1+ip_\phi\cdot z_3}\int dp e^{-ip\cdot z_3}G(p) \tag{2.7}$$

with

$$G(p)=\sum_\alpha\langle\phi|A(0)|p,\alpha\rangle\langle p,\alpha|A(0)|\Omega\rangle. \tag{2.8}$$

We want now to show that $G(p)$ has compact support. The restrictions on the spectra of $A(p)$ and π^2 imply

$$\langle\phi|A(0)|n\rangle\langle n|A(0)|\Omega\rangle=0 \tag{2.9}$$

unless

$$\epsilon^2\leqslant p_n{}^2\leqslant M^2, \quad p_n\in V_+ \tag{2.10}$$

and

$$0\leqslant(p_n-p_\phi)^2\leqslant M^2. \tag{2.11}$$

Let us choose the reference frame in which $\mathbf{p}_\phi=0$. Then conditions (2.10), (2.11) read

$$\epsilon^2\leqslant(p_n{}^0{}^2-\mathbf{p}_n{}^2)\leqslant M^2, \tag{2.12}$$

$$0\leqslant(p_n{}^0-p_\phi{}^0)^2-\mathbf{p}_n{}^2\leqslant M^2, \tag{2.13}$$

which can be solved to give

$$\max\{\epsilon,(p_\phi{}^0{}^2-M^2+\epsilon^2)/2p_\phi{}^0\}\leqslant p_n{}^0\leqslant(p_\phi{}^0{}^2+M^2)/2p_\phi{}^0. \tag{2.14}$$

Together with (2.10), (2.14) shows that the support of $G(p)$ is compact. It follows[7] that $\int dp e^{-ipz_3}G(p)$ is continuous for $\mathrm{Im}z_3=0$ and can be extended to be an entire function for $\mathrm{Im}z_3\neq 0$. The factor $e^{ip_\phi\cdot(z_3-z_1)}$ is also entire and therefore such is $F_\phi{}^2(z_1z_2)$. In addition

$$\lim_{\mathrm{Im}z_1,z_2\to 0} F_\phi{}^2(z_1z_2)=F_\phi{}^2(x_1x_2).$$

We notice then that both $F_\phi'(z_1z_2)$ and $F_\phi{}^2(z_1z_2)$ are analytic in the domain[8] D:

$$D: \ \mathrm{Im}z_3\in V_+, \ \mathrm{Im}z_1 \text{ finite} \quad (z_3\equiv z_1-z_2).$$

It is important for what follows that all real points belong to the boundary of D.

[3] The theorem is also true under the weaker assumption $a(p)=0$ for $p^2<0$, $a(p)|\Omega\rangle=0$ for $p^2>M^2$.

[4] Or a superposition of derivatives thereof, if $A(x)$ is not a scalar.

[5] A result analogous to Theorem I can be obtained relaxing the condition that there is only one eigenstate of π^μ to the eigenvalue zero. A discussion of this case is given in the Appendix.

[6] One should, strictly speaking, introduce "wave packets." This is irrelevant for the proof and amounts to a certain complication in the algebra involved; we shall therefore make free use of "eigenstates of π^μ."

[7] L. Schwartz, *Theorie des Distributions* (Hermann & Cie, Paris, France, 1951), Tome II, p. 128, Théorème XVI.

[8] As a matter of fact, both are entire functions.

According to (2.2) if $(x-y)^2<0$, $F_\phi'(x,y)=F_\phi^2(x,y)$. The function $F_\phi(z_1z_2)\equiv F_\phi'(z_1z_2)-F_\phi^2(z_1z_2)$, analytic in D, has therefore zero as the boundary value on (a two-dimensional) part of the boundary of D. The "edge-of-the-wedge" theorem[9] assures then that $F_\phi(z_1z_2)\equiv 0$ and therefore that also its boundary values on the remaining part of the boundary of D are zero, i.e.,

$$F_\phi(x,y)\equiv F_\phi'(x,y)-F_\phi^2(x,y) \quad \text{all } x,y. \tag{2.15}$$

Equation (2.1) is thereby proven. One can see that the arguments which led to (2.15) fail if $p_\phi=0$ or if $p_\phi^2=0$ (in this case one cannot chose a reference frame in which $\mathbf{p}_\phi=0$; choosing $p_\phi^0=p_\phi^1=p_\phi$; $p_\phi^2=p_\phi^3=0$ the inequalities (2.12) and 2.13) give $p_n^0-p_n^1\leqslant M^2/2p_\phi$, $p_n^0\geqslant|\mathbf{p}_n|$ and the region characterized by these restrictions is not finite).

So far, we have proven that

$$[A(x),A(y)]|\Omega\rangle=C(x,y)|\Omega\rangle, \tag{2.3}$$

where $C(x,y)$ is a c-number function of x, y. Consider now

$$\langle\phi|\{[A(x),A(y)]-C(x,y)\}|\psi\rangle, \tag{2.16}$$

where $|\phi\rangle$, $|\psi\rangle\in\mathfrak{H}$. Since $A(x)$ is, by assumption, cyclic in $\mathfrak{H}$, we will have proven that

$$[A(x),A(y)]=C(x,y) \tag{2.17}$$

if we can show that

$$0=\langle\Omega|A(x_1)\cdots A(x_n)\{[A(x),A(y)] -C(x,y)\}A(y_1)\cdots A(y_m)|\Omega\rangle \tag{2.18}$$

for all configurations $(x_1\cdots x_n)$ and $(y_1\cdots y_m)$ and all values of the indices m, n. Consider the functions[10]

$$G_1(x_1\cdots y_m,x,y) \equiv\langle\Omega|A(x_1)\cdots A(x_m)A(x)A(y)A(y_1)\cdots A(y_m)|\Omega\rangle,$$

$$G_2(x_1\cdots y_m,x,y) \equiv\langle\Omega|A(x_1)\cdots A(x_n)A(y)A(x)A(y_1)\cdots A(y_m)|\Omega\rangle,$$

$$G_3(x_1\cdots y_m,x,y) \equiv C(x,y)\langle\Omega|A(x_1)\cdots A(x_n)A(y_1)\cdots A(y_m)|\Omega\rangle.$$

Since we have assumed that the energy-momentum vector of all states in $\mathfrak{H}$ lies in the forward lightcone, one can find three functions $G_1(z_1\cdots Z_n\zeta_1\cdots\zeta_m x,y)$, $G_2(z_1\cdots\zeta_m,x,y)$, $G_3(z_1\cdots\zeta_m,x,y)$, analytic in the tube S:

$$S:\ \begin{matrix}\mathrm{Im}(z_i-z_{i+1})\in V_- & \mathrm{Im}z_n\in V_+\\ \mathrm{Im}(\zeta_i-\zeta_{i+1})\in V_- & \mathrm{Im}\zeta_1\in V_-\end{matrix}\ \text{real parts arbitrary}$$

and such that $G_i(x_1\cdots y_m,x,y)$ is the boundary value of $G_1(z_1\cdots\zeta_m,x,y)$ when $\mathrm{Im}z_i\to 0$, $\mathrm{Im}\zeta_i\to 0$, $(z_i,\zeta_j)\in S$. Let us now choose the y_j's in such a way that

$$(x-y_j)^2<0, \quad (y-y_j)^2<0 \text{ all } j\text{'s}.$$

Using (2.2), we see that, on this part of the boundary of S,

$$\begin{aligned}&G_1(x_1\cdots y_m,x,y)-G_2(x_1\cdots y_m,x,y)-G_3(x_1\cdots y_m,x,y)\\ &\quad=\langle\Omega|A(x_1)\cdots A(x_n)\{[A(x),A(y)]\\ &\qquad\qquad -C(x,y)\}A(y_1)\cdots A(y_m)|\Omega\rangle\\ &\quad=\langle\Omega|A(x_1)\cdots A(x_n)A(y_1)\cdots A(y_m)\\ &\qquad\times\{[A(x),A(y)]-C(x,y)\}|\Omega\rangle=0, \text{ by } (2.3).\end{aligned}$$

Therefore

$$G(z_i,\zeta_j,x,y)\equiv G_1(z_i,\zeta_j,x,y)-G_2(z_i,\zeta_j,x,y)-G_3(z_i,\zeta_j,x,y)$$

is analytic in S and goes to zero on (a n-dimensional) part of the boundary of S.

According to the edge-of-the-wedge theorem, therefore, $G(z_i,\zeta_j,x,y)\equiv 0$ in S and, since all real points are on the boundary of S, $G(x_1\cdots y_m,x,y)=0$ for all x_i, y_j. Equation (2.18) is thereby proven and (2.17) follows. (Notice that $\mathfrak{H}$ is the closure of the set of all vectors which can be given the form

$$|\phi\rangle=\sum_n\int\phi_n(x_1\cdots x_n)A(x_1)\cdots A(x_n)|\Omega\rangle dx_1\cdots dx_n,$$

where $\phi_n(x_1\cdots x_n)$ is a function with compact support.)

3. CONDITIONS UNDER WHICH A FIELD IS A GENERALIZED FREE FIELD

Our next task will be to prove:

Theorem II. The following three are equivalent statements:

$$A(x) \text{ is a generalized free field,} \tag{3.1}$$

$$[A(x),A(y)]=G(x-y) \quad [G(x-y) \text{ is, in general, an operator}], \tag{3.2}$$

$$[A(x),A(y)]=C(x,y) \quad [C(x,y) \text{ is a } c \text{ number}], \tag{3.3}$$

provided

(a) the transformation properties of $A(x)$ under the inhomogeneous Lorentz group are

$$U(a,\Lambda)A(x)U^{-1}(a,\Lambda)=A(\Lambda x+a),$$

(b) $A(x)$ is cyclic,
(c) $A(x)$ is local,
(d) $A(x)$ is a tempered distribution in x and an operator in a Hilbert space $\mathfrak{H}$,
(e) a state with lowest energy exists in $\mathfrak{H}$ and is there unique. We shall call this state the vacuum and its symbol will be $|\Omega\rangle$.

We prove first the equivalence of (3.2) and (3.3).

Proposition 1. (3.2) implies (3.3).

Proof: Let π^μ be on $\mathfrak{H}$ the infinitesimal generator of translations; let $|n\rangle$ be an eigenstate of π^μ to the eigenvalue[11] p_n^μ.

[9] H. Bremerman, R. Oehme, and J. G. Taylor, Phys. Rev. **109**, 2178 (1958); F. T. Dyson, *ibid.* **110**, 579 (1958); H. Epstein, J. Math. Phys. **1**, 525 (1960).

[10] From here on, the proof follows R. Jost, Lectures at the International School of Physics, Spring 1959, Naples.

[11] See reference 6.

By assumption the expression

$$\langle n|G(x-y)|\Omega\rangle\equiv\langle n|[A(x),A(y)]|\Omega\rangle \tag{3.4}$$

depends on x and y only through the difference $x-y$; therefore

$$\left(\frac{\partial}{\partial x^\mu}+\frac{\partial}{\partial y^\mu}\right)\langle n|G(x-y)|\Omega\rangle=0. \tag{3.4'}$$

On the other end, if π^μ is the infinitesimal generator of translations, we have

$$\langle n|[A(x),A(y)]|\Omega\rangle$$

$$=e^{i\frac{1}{2}p_n\cdot(x+y)}\left\langle n\left|\left[A\left(\frac{x-y}{2}\right),A\left(\frac{y-x}{2}\right)\right]\right|\Omega\right\rangle. \tag{3.5}$$

Therefore,

$$\left[\left(\frac{\partial}{\partial x^\mu}+\frac{\partial}{\partial y^\mu}\right)e^{i\frac{1}{2}p_n(x+y)}\right]$$

$$\times\left\langle n\left|\left[A\left(\frac{x-y}{2}\right),A\left(\frac{y-x}{2}\right)\right]\right|\Omega\right\rangle=0$$

for all $|n\rangle$ and for all values of $x+y$. This implies

$$\left[A\left(\frac{x-y}{2}\right),A\left(\frac{y-x}{2}\right)\right]|\Omega\rangle=\bar{G}(x-y)|\Omega\rangle$$

$$=[A(x),A(y)]|\Omega\rangle$$

where

$$\bar{G}(x-y)\equiv\langle\Omega|G(x-y)|\Omega\rangle. \tag{3.6}$$

By standard methods, using the locality and cyclicity of $A(x)$ and the positive-definiteness of the energy, one concludes

$$G(x-y)\equiv\bar{G}(x-y)\quad\text{Q. E. D.} \tag{3.7}$$

From (3.7) we also see that $G(x-y)$ is an invariant function of its arguments.

Proposition 2. (3.3) implies (3.2).

Proof. By assumption,

$$C(x,y)=\langle\Omega|[A(x),A(y)]|\Omega\rangle$$

$$=\left\langle\Omega\left|\left[A\left(\frac{x-y}{2}\right),A\left(\frac{y-x}{2}\right)\right]\right|\Omega\right\rangle=C(x-y)$$

Q. E. D.

We shall now prove the equivalence of (3.2) and (3.3).

Proposition 3. (3.1) implies (3.3).

Proof. This is a part of the definition of generalized free field.

Proposition 4. (3.3) implies (3.1).

Proof. Let us introduce the Fourier transform of $A(x)$

$$a(p)=\int e^{ipx}A(x)dx. \tag{3.8}$$

By assumption,

$$[\pi^\mu,A(x)]=i\frac{\partial A(x)}{\partial x^\mu}. \tag{3.9}$$

Equation (3.9) can be rewritten as

$$[\pi^\mu,a(p)]=-p^\mu a(p), \tag{3.9'}$$

from which follows:

Lemma 2. $[a(p),a(k)]=0$ for all k if $p^2<0$.

Proof. Since there is no state in $\mathfrak{H}$ with a space-like energy-momentum four vector, we have

$$a(p)|\Omega\rangle=0\quad\text{if}\quad p^2<0,$$

and also

$$\langle\Omega|a(p)=0\quad\text{if}\quad p^2<0,$$

therefore

$$\langle\Omega|[a(p),a(k)]|\Omega\rangle\quad\text{if}\quad p^2<0$$

and Lemma 2 follows since the commutator is a c number. We write now $A(x)\equiv B(x)+C(x)$ where

$$B(x)=\int_{p^2\geq 0}e^{ipx}a(p)d^4p,$$

$$C(x)=\int_{p^2<0}e^{ipx}a(p)d^4p.$$

It follows from Lemma 2:

$$[B(x),C(y)]=0=[C(x),C(y)], \tag{3.10}$$

and moreover

$$C(x)|\Omega\rangle=0. \tag{3.11}$$

We shall now prove Lemma 3.

Lemma 3. If $A(x)$ is cyclic, $B(x)$ is also cyclic.

Proof. Cyclicity of $A(x)$ means the vectors

$$|\phi\rangle\equiv\int\phi(x_1\cdots x_n)A(x_1)\cdots A(x_n)|\Omega\rangle dx_1\cdots dx_n$$

form a complete basis in $\mathfrak{H}$. But

$$|\phi\rangle=\int\phi(x_1\cdots x_n)A(x_1)\cdots A(x_n)|\Omega\rangle dx_1\cdots dx_n$$

$$=\int\phi(x_1\cdots x_n)B(x_1)\cdots B(x_n)|\Omega\rangle dx_1\cdots dx_n$$

due to (3.10), (3.11). (3.12)

Therefore, also the vectors

$$\int\phi(x_1\cdots x_n)B(x_1)\cdots B(x_n)|\Omega\rangle dx_1\cdots dx_n$$

form a complete basis in $\mathfrak{H}$ and this completes the proof of the lemma.

Lemma 4. $C(x)=0$.

Proof.

$$\langle\chi|C(x)|\psi\rangle=\sum_n\int dy_1\cdots dy_n\psi(y_1\cdots y_n)$$
$$\times\langle\chi|C(x)A(y_1)\cdots A(y_n)|\Omega\rangle=0$$

by (3.10) and (3.11), for every pair of states $|\chi\rangle$, $|\psi\rangle$. We also have

$$[[a(p),a(k)],\pi^\mu]=(p^\mu+k^\mu)[a(p),a(k)]. \quad (3.13)$$

By assumption, the commutator is a c number. Equation (3.13) therefore implies

$$(p^\mu+k^\mu)[a(p),a(k)]=0. \quad (3.14)$$

We conclude that $[a(p),a(k)]$ is an invariant function with support, in the variable $(p+k)$, at the origin [the invariance follows from that of $G(x-y)$].

The support property for $a(p)$ as a function of p^2, expressed in Lemma 4, allows an invariant decomposition of $a(p)$ into positive- and negative-frequency parts:

$$a(p)=\theta(p_0)\mathbf{b}(p)+\theta(-p_0)\mathbf{b}^\dagger(-p). \quad (3.15)$$

Let μ be a positive measure on the positive real axis. We introduce a new field $b(\mathbf{p},\sigma)$ on $\mathfrak{H}$ defined through

$$\int\mathbf{b}(\mathbf{p},\sigma)\varphi(\sigma)d\sigma=\int d\mu(\sigma)b(\mathbf{p},\sigma)\varphi(\sigma) \quad (3.16)$$

for all $\varphi(\sigma)\in S$. The measure μ has to be such that $b(\mathbf{p},\sigma)$ has no δ-like singularities. Our aim will be to show that the measure μ can be so chosen that the field $b(\mathbf{p},\sigma)$ satisfies the "generalized canonical commutation relations" (3.2). We shall rewrite (3.16) in the form

$$\mathbf{b}(\mathbf{p},\sigma)=\int d\mu(\rho)\delta(\sigma-\rho)b(\mathbf{p},\rho) \quad (3.16')$$

with the understanding that all the expressions in what follows have to be taken in the sense of the theory of distributions. The measure μ consists of one part which gives finite weight only to a finite (or denumerably infinite) number of points and a second part μ' which assigns zero weight to isolated points. We therefore have

$$\mathbf{b}(\mathbf{p},p^2)=\sum_i C_i(2p^0)^{\frac{1}{2}}b_i(\mathbf{p})\delta(p^2-m_i^2)$$
$$+\int d\mu'(\sigma)\delta(\sigma-p^2)b(\mathbf{p},\sigma), \quad (3.17)$$

where the C_i's are positive numbers. (The factor $(2p^0)^{\frac{1}{2}}$ is added for reasons of normalization.)

On the other end, $[\mathbf{b}(p),\mathbf{b}^\dagger(k)]$ is a c number, and therefore equal to its vacuum-expectation value. The most general form of such expectation value is known[12] and we can write therefore

$$\int d\mu(\sigma)\int d\mu(\lambda)\delta(\sigma-p^2)\delta(\lambda-k^2)[b(\mathbf{p},\sigma),b^\dagger(\mathbf{k},\lambda)]$$
$$=\delta^4(p-k)\Big[\sum_i d_i\delta(p^2-\mu_i^2)+\int d\beta(\theta)\delta(p^2-\theta)\Big], \quad (3.18)$$

where $\{d_i\}$ is a finite (or denumerably infinite) set of positive numbers and β is a Lorentz-invariant, positive measure which assigns zero weight to single points. Substituting (3.17) into (3.18) one has

$$\sum_{ij}C_iC_j(4p^0k^0)^{\frac{1}{2}}[b_i(\mathbf{p},m_i^2),b_j^\dagger(\mathbf{k},m_j^2)]\delta(p^2-m_i^2)\delta(k^2-m_j^2)$$
$$+\sum_i C_i(2p^0)^{\frac{1}{2}}\int d\mu'(\sigma)\delta(\sigma-k^2)[b_i(\mathbf{p},m_i^2),b^\dagger(\mathbf{k},\sigma)]\delta(p^2-m_i^2)$$
$$+\sum_i C_i(2k^0)^{\frac{1}{2}}\int d\mu'(\sigma)\delta(\sigma-p^2)[b(\mathbf{p},\sigma),b_i^\dagger(\mathbf{k},m_i^2)]\delta(k^2-m_i^2)$$
$$+\int d\mu'(\sigma)\int d\mu'(\lambda)\delta(\sigma-p^2)\delta(\lambda-k^2)[b(\mathbf{p},\sigma),b^\dagger(k,\lambda)]$$
$$=\sum_i d_i(2p^0)\delta(p^2-\mu_i^2)\delta(k^2-\mu_i^2)\delta^3(\mathbf{p}-\mathbf{k})+\delta^4(p-k)\int d\beta(\theta)\delta(p^2-\theta). \quad (3.19)$$

We now recall that $[\mathbf{b}(\mathbf{p},p^2),\mathbf{b}^\dagger(\mathbf{k},k^2)]=0$ if $\mathbf{k}\neq\mathbf{p}$. Computing $[\mathbf{b}(\mathbf{p},f),\mathbf{b}^\dagger(\mathbf{k},g)]$, $\mathbf{b}(\mathbf{p},f)=\int\mathbf{b}(\mathbf{p},p^2)f(p^2)dp_2$ and choosing properly the support of f, g one then deduces

$$[b_i(\mathbf{p},m_i^2),b_j^\dagger(\mathbf{k},m_j^2)]=0 \text{ if } \mathbf{k}\neq\mathbf{p} \text{ or } m_i^2\neq m_j^2, \quad (3.20)$$
$$[b(\mathbf{p},\sigma),b^\dagger(\mathbf{k},\lambda)]=0 \text{ if } \mathbf{k}\neq\mathbf{p} \text{ or } \sigma\neq\lambda, \quad (3.21)$$
$$[b(\mathbf{p},\sigma),b_i^\dagger(\mathbf{k},m_i^2)]=0 \text{ if } \mathbf{k}\neq\mathbf{p} \text{ or } \sigma\neq m_i^2. \quad (3.22)$$

Equation (3.21) implies

$$\delta(p^2-m_i^2)\delta(k^2-m_j^2)[b(\mathbf{p},m_i^2),b^\dagger(k,m_j^2)]$$
$$=\alpha\delta(p^2-m_i^2)\delta(k^2-m_i^2)\delta(\mathbf{p}-\mathbf{k})\delta_{m_i^2m_j^2},$$

where α is some constant, δ_{rs} is the Kronecker symbol.

[12] H. Lehmann, Nuovo cimento **11**, 342 (1959); G. Kallen, Helv. Phys. Acta **25**, 417 (1952).

If we choose, in (3.17),

$$C_i = d_i^{\frac{1}{2}}, \tag{3.23}$$

we have therefore, from (3.19),

$$[b(\mathbf{p},m_i^2), b^\dagger(\mathbf{k},m_j^2)] = \delta_{m_j m_i}\delta(\mathbf{k}-\mathbf{p}) \tag{3.24}$$

and also $m_i = \mu_i$. From (3.22) we have moreover

$$[b(\mathbf{p},\sigma), b_i^\dagger(\mathbf{k},m_i^2)] = \beta_1\delta(m_i^2-\sigma)\delta(\mathbf{p}-\mathbf{k}) + \beta_2\delta_{m_i^2,\sigma}\delta(\mathbf{p}-\mathbf{k}).$$

Using the fact that the measure μ' attributes weight zero to isolated points, one can show that $\beta_1 = 0$. The second and third term in (3.19) are then proportional to $\beta_2 \int_{\sigma=m_i^2} d\mu'(\sigma) = 0$. From (3.21) we have finally, using invariance considerations[13]

$$[b(\mathbf{p},\sigma), b^\dagger(\mathbf{k},\lambda)] = g(\sigma)\delta(\mathbf{p}-\mathbf{k})\delta((\sigma-\mathbf{p}^2)^{\frac{1}{2}} - (\lambda-\mathbf{k}^2)^{\frac{1}{2}}).$$

Equation (3.19) implies therefore

$$\int d\mu'(\sigma) \int d\mu'(\lambda)\delta(\sigma-k^2)\delta(\sigma-\lambda)g(\sigma)\delta(\mathbf{p}-\mathbf{k})\delta(p^0-k^0) = \delta^4(p-k)\int d\beta(\theta)\delta(p^2-\theta). \tag{3.25}$$

Let ρ be the distribution defined by $(\rho,f) \equiv \int f d\beta$, where β is the measure which appears in (3.18). By assumption, ρ is a tempered distribution. We shall assume that "the square root of ρ" exists in the following sense: There exists a positive measure ν on E such that[14]

$$(\rho,f) = \nu(\sigma f) \quad \text{where} \quad \sigma_z = \nu_y[\delta(x-y)].$$

In particular, if ρ is a function integrable on every finite subset of E and if $(\rho(x))^{\frac{1}{2}}$ exists, then the measure ν is given formally by $d\nu = [\rho(x)]^{\frac{1}{2}}dx$. In fact

$$\sigma(x) = \int [\rho(y)]^{\frac{1}{2}}\delta(x-y)dy = [\rho(x)]^{\frac{1}{2}}$$

and

$$\nu(\sigma f) = \int [\rho(x)]^{\frac{1}{2}}[\rho(x)]^{\frac{1}{2}} f(x)dx = \int \rho(x)\delta(x)dx \equiv (\rho,f).$$

If one chooses[15]

$$\mu = \nu, \tag{3.26}$$

one has, for all β-measurable functions,

$$\beta(f) = \int d\mu'(\sigma)\int d\mu'(\lambda) f(\sigma)\delta(\sigma-\lambda). \tag{3.27}$$

$$[b(\mathbf{p},p_0), b^\dagger(\mathbf{k},k_0)] = \delta(\mathbf{p}-\mathbf{k})\delta(p_0-k_0). \tag{3.28}$$

With the choices indicated in (3.23), (3.26) (we assume that the two-point function $\langle\Omega|[A(x),A(y)]|\Omega\rangle$ of the field A is given), we have then relations (3.24), (3.28) which, as we said, stand for the more precise forms:

$$[b_i(\varphi), b_j^\dagger(\varphi')] = \delta_{ij}(\varphi,\varphi') \quad \varphi = \varphi(\mathbf{p}), \tag{3.24'}$$

$$[b(\psi), b^\dagger(\psi')] = (\psi,\psi') \quad \psi = \psi(\mathbf{p},p_0). \tag{3.28'}$$

All other commutators are zero. From Eqs. (3.15), (3.19), (3.24'), and (3.28'), Proposition 4 follows. Propositions 1 to 4 complete the proof of Theorem II. To Theorem II we may add the following corollary.

Corollary. From the arguments of L. S. Z.[16] one can see that only $b_i(\mathbf{p})$ and not $b(\mathbf{p},\sigma)$ [as defined in (3.21)] can satisfy an asymptotic condition in their sense. If such condition is postulated and the resulting free fields are taken to form a cyclic algebra with the vacuum as cyclic vector, then $b(\mathbf{p},\sigma) = 0$.

This follows immediately from the identity

$$[b(\psi), b_i^\dagger(\varphi)] = [b(\psi), b_i(\varphi)] = 0 \tag{3.29}$$

for all i, ψ, φ and the equivalence, for free fields, of cyclicity with respect to the vacuum and irreducibility. Equation (3.29), in fact, implies $b(\psi) = c$ number; on the other hand, from $\langle\Omega|A(x)|\Omega\rangle = 0$ and $\langle\Omega|b_i(\varphi)|\Omega\rangle = 0$ it follows $0 = \langle\Omega|b(\psi)|\Omega\rangle = b(\psi)$ Q. E. D. Therefore, if an asymptotic condition of the L. S. Z. type is introduced, the first statement in Theorem II should be strengthened to read

(1') $A(x)$ is a discrete superposition of independent free fields.

ACKNOWLEDGMENTS

The author would like to thank Dr. W. D. McGlinn for useful discussions and Professor A. J. F. Siegert for his hospitality at Northwestern University. The support of the National Science Foundation is also gratefully acknowledged.

APPENDIX[17]

Theorem I'. Under the assumptions of Theorem I, but allowing for the existence in $\mathfrak{H}$ of N orthonormal eigenstates $|\Omega\rangle_i$ $i = 1 \cdots N$, of π^μ to the eigenvalue zero, $A(x)$ takes the following form:

$$A(x) = \sum_i A_i(x)P_i. \tag{A.1}$$

In (A.1), the P_i's are projection operators into orthogonal subspaces $\mathfrak{H}_i$ of $\mathfrak{H}$ ($\sum_{1i}^N \oplus \mathfrak{H}_i = \mathfrak{H}$) and $A_i(x)$ [the restriction of $A(x)$ to $\mathfrak{H}_i$] is in $\mathfrak{H}_i$ a generalized free field cyclic with respect to $|\omega\rangle_i$, the unique (in $\mathfrak{H}_i$) eigenstate of π^μ to the eigenvalue zero.

[13] P. Methee, Comm. Math. Helv. 28, 225 (1954).

[14] E is the positive real axis.

[15] We shall not give here the detailed proof that this choice defines indeed $b(\mathbf{p},\sigma)$ as a bona-fide operator on $\mathfrak{H}$; let us however remark that, if $d\beta(\sigma) = 0$ [and therefore $d\mu(\sigma) = 0$] for $\sigma_1 < \sigma < \sigma_2$, then $a(\mathbf{p},\sigma) = 0$ on the same interval. In fact, $d\beta(\sigma) = 0$ for $\sigma_1 < \sigma < \sigma_2$ implies $\|a(\mathbf{p},\sigma)|\Omega\rangle\| = 0$ on the same interval, and this, together with $[a(\mathbf{p},\sigma), a(\mathbf{k},\lambda)] = 0$ $\sigma \neq \lambda$ and the cyclicity of $A(x)$ with respect to $|\Omega\rangle$, leads to $a(\mathbf{p},\sigma) = 0$, $\sigma_1 < \sigma < \sigma_2$.

[16] H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo cimento 1, 205 (1955); also O. W. Greenberg and A. S. Wightman, "The asymptotic conditions in quantum field theory" (unpublished).

[17] We would like to thank Professor E. C. G. Sudarshan for an interesting discussion which originated this Appendix.

To prove this theorem we notice that the equality (2.1) is still valid, since nowhere in its proof the uniqueness of the vacuum was used. Now, however, (2.3) does not follow from (2.1). One has rather

$$[A(x),A(y)]|\Omega\rangle_i = i\sum_j C_{ij}(x-y)|\Omega\rangle_j. \quad (A.2)$$

We now notice that the matrix $\|C\|$:

$$C_{ij} = -i_j\langle\Omega|[A(x),A(y)]|\Omega\rangle_i$$

is Hermitian (strictly speaking, the matrices $\|C_f\|$: $C_f{}^{ij} = \int C_{ij}(x)f(x)dx$ are Hermitian for all real $f(x)$ with compact support).

In fact

$$C_{ji}{}^*(x-y) = i(_i\langle\Omega|[A(x),A(y)]|\Omega\rangle_j)^* = -i_j\langle\Omega|[A(x),A(y)]|\Omega\rangle_i = C_{ij}(x-y). \quad (A.3)$$

The matrix $\|C\|$ can therefore be reduced to diagonal form by a unitary matrix $U_{ij}(x-y)$ which will in general depend on $x-y$.

Let us consider the N orthonormal vectors $|\omega\rangle_i \in \mathfrak{H}$ defined by

$$|\omega\rangle_i = \sum_j [U^{-1}(x-y)]_{ij}|\Omega\rangle_j. \quad (A.4)$$

The $|\omega\rangle_i$'s are evidently eigenstates of π^μ to the eigenvalue zero. From (A.2) and (A.4) one also derives

$$[A(x),A(y)]|\omega\rangle_i = ib_i(x-y)|\omega\rangle_i, \quad (A.5)$$

where $b_i(x-y)$ is the ith eigenvalue of the matrix $\|C\|$. Let $\mathfrak{H}_i$ be the subspace of $\mathfrak{H}$ defined as follows: $\mathfrak{H}_i$ is the closure (with the topology of $\mathfrak{H}$) of the set of vectors that are obtained from $|\omega\rangle_i$ by the action of finite-order polynomials in $A(x)$. In other words, $\mathfrak{H}_i$ is the subspace of $\mathfrak{H}$ in which $A(x)$ is cyclic with respect to the state $|\omega\rangle_i$.

We want now to show that, if $b_i(x) \neq b_j(x)$, then

$$\mathfrak{H}_i \cap \mathfrak{H}_j = 0.$$

Let $|\psi\rangle \in \mathfrak{H}_i \cap \mathfrak{H}_j$. Consider $[A(x),A(y)]|\psi\rangle$. We shall have proven that

$$[A(x),A(y)]|\psi\rangle = ib_i(x-y)|\psi\rangle \quad (A.6)$$

if we can show that

$$\langle\chi|[A(x),A(y)]A(x_1)\cdots A(x_n)|\omega\rangle_i = ib_i(x-y)\langle\chi|A(x_1)\cdots A(x_n)|\omega\rangle_i \quad \text{all } |\chi\rangle \in \mathfrak{H}. \quad (A.7)$$

In fact, by assumption, $|\psi\rangle \in \mathfrak{H}_i$ and $A(x)$ is cyclic in $\mathfrak{H}_i$ with respect to $|\omega\rangle_i$.

The proof that (A.4) implies (A.7) is completely analogous to the proof that (2.3) implies (2.15) and will not be repeated here. In the same way, considering now $|\psi\rangle$ as a vector in $\mathfrak{H}_j$, we can deduce

$$[A(x),A(y)]|\psi\rangle = ib_j(x-y)|\psi\rangle. \quad (A.8)$$

But (A.6) and (A.8) are contradictory unless $|\psi\rangle$ is the null vector [we have assumed $b_i(x-y) \neq b_j(x-y)$]. Therefore,

$$\mathfrak{H}_i \cap \mathfrak{H}_j = 0 \quad \text{if} \quad b_i(\xi) \neq b_j(\xi). \quad (A.9)$$

Let M be the number of distinct $b_j(\xi)$'s. Then

$$\sum_1^M{}_i \oplus \mathfrak{H}_i' \subset \mathfrak{H}, \quad (A.10)$$

where $\mathfrak{H}_i'$ is the subspace on which $[A(x),A(y)]$ takes on the value $b_i(x-y)$. We assume that $A(x)$ is cyclic with respect to the set $\{|\omega\rangle_i\}$. In other words, we assume that

$$U_i\mathfrak{H}_i' = \mathfrak{H}. \quad (A.11)$$

From (A.10) and (A.11) it follows

$$\sum_1^M{}_i \oplus \mathfrak{H}_i' = \mathfrak{H}. \quad (A.12)$$

This and (A.13),

$$[A(x),A(y)]|\psi\rangle = ib_i(x-y)|\psi\rangle \quad \text{if} \quad |\psi\rangle \in \mathfrak{H}_i', \quad (A.13)$$

lead then to

$$[A(x),A(y)] = i\sum_1^M{}_j b_j(x-y)P_j', \quad P_j'\mathfrak{H} = \mathfrak{H}_j. \quad (A.14)$$

From (A.12) and the identity

$$A(x)\mathfrak{H}_j' \subset \mathfrak{H}_j', \quad (A.15)$$

it follows

$$[A(x),P_i'] = 0 \text{ all } i\text{'s and } x\text{'s}. \quad (A.16)$$

From (A.16) it is now possible to show that also

$$[\pi^\mu,P_i'] = 0, \quad (A.17)$$

where the π^μ are the infinitesimal generators of translations. In fact, one has

$$[\pi^\mu,A(x)] = i\frac{\partial A(x)}{\partial x^\mu}, \quad (A.18)$$

and therefore

$$[P_j',[\pi^\mu,A(x)]] = 0 \text{ all } j\text{'s}, \quad (A.19)$$

and, due to (A.16), also that

$$[A(x),[P_j',\pi^\mu]] = 0. \quad (A.20)$$

On the other hand,

$$[P_j',\pi^\mu]|\omega\rangle_i = 0 \text{ all } i, j, \mu.$$

Therefore, $[P_j',\pi^\mu] = 0$ on $\mathfrak{H}_i'$, all i, j, and, since

$$\mathfrak{H} = \sum_1^M{}_i \oplus \mathfrak{H}_i',$$

$[P_j',\pi^\mu] = 0$ on $\mathfrak{H}$. With similar arguments one can prove that

$$[U(a,\Lambda),P_j'] = 0 \text{ on } \mathfrak{H}. \quad (A.21)$$

We shall now show that, if $\mathfrak{H}_i'$ contains s_i vacua $|\omega\rangle_{ik}$, $k = 1\cdots s$, then

$$\mathfrak{H}_i' = \sum_1^{s_i}{}_k \oplus \mathfrak{H}_{ik}$$

where $|\omega\rangle_{ik}\in\mathcal{H}_{ik}$ and $A(x)$ is cyclic in $\mathcal{H}_{ik}$ with respect to $|\omega\rangle_{ik}$. Equation (A.17) has the consequence that the restriction of π^μ to $\mathcal{H}_i'$ provides a representation of the infinitesimal generators of translation and this, together with the assumption that no state with negative energy exists in $\mathcal{H}$, implies that the energy-momentum p^μ of the states in $\mathcal{H}_i'$ (apart from $|\omega\rangle_{ik}$, $k=1\cdots s$) lies in the forward lightcone (and in fact satisfies also the requirement $p^2>\epsilon^2$).

This allows an invariant decomposition of the restriction of $A(x)$ to $\mathcal{H}_i'$ into positive and negative frequencies (Lemmas 2–4 in the text; the proof does not make essential use of the uniqueness of the vacuum).

We shall now prove that

$$\mathcal{H}_{ik}\bigcap\mathcal{H}_{ij}=0 \quad \text{for} \quad k\neq j \tag{A.22}$$

by showing that the two sets $\{P(A_i)|\omega\rangle_{ik}\}$ and $\{P'(A_i)|\omega\rangle_{ij}\}$ are mutually orthogonal if $k\neq j$. $P(A_i)$ and $P'(A_i)$ are (smeared) polynomials of any finite order in the $A_i(x)$'s. The two sets are dense in $\mathcal{H}_{ik}$, $\mathcal{H}_{ij}$, respectively; if $|\psi\rangle\in\mathcal{H}_{ik}\bigcap\mathcal{H}_{ij}$, $\||\psi\rangle\|=1$, there exist two vectors $|\psi_0\rangle\in\{P(A_i)|\omega\rangle_{ik}\}$ and $|\phi_0\rangle\in\{P'(A_i)|\omega\rangle_{ij}\}$, $\||\psi_0\rangle\|=\||\phi_0\rangle\|=1$, such that

$$\||\psi\rangle-|\psi_0\rangle\|<\epsilon, \tag{A.23}$$

$$\||\phi\rangle-|\phi_0\rangle\|<\epsilon, \tag{A.24}$$

with ϵ arbitrarily small. Conditions (A.23) and (A.24) imply

$$|\langle\psi|\psi_0\rangle|<\epsilon^{\frac{1}{2}}, \quad |\langle\psi|\phi_0\rangle|<\epsilon^{\frac{1}{2}}, \tag{A.25}$$

and also

$$\||\psi\rangle\|+|\langle\psi|\psi_0\rangle|+|\langle\psi|\phi_0\rangle|+|\langle\psi_0|\phi_0\rangle|<\epsilon. \tag{A.26}$$

If $\langle\psi_0|\phi_0\rangle=0$, then (A.25), (A.25′), and (A.26) imply

$$\||\psi\rangle\|<\epsilon. \tag{A.27}$$

Since ϵ is arbitrary, (A.27) implies $|\psi\rangle=0$. It remains to be proven that the two sets $\{P(A_i)|\omega\rangle_{ik}\}$ and $\{P'(A_i)|\omega\rangle_{ij}\}$ are mutually orthogonal. This is true if

$$_{ik}\langle\omega|A_i(x_1)\cdots A_i(x_n)|\omega\rangle_{ij}=0, \quad k\neq j \tag{A.28}$$

for all configurations $\{x_1\cdots x_n\}$. To prove (A.28) we make use of the decomposition of $A_i(x)$ into positive (A_i^+) and negative (A_i^-) frequencies and remember that $A_i^+(x)|\omega\rangle_{ik}=0$ for all k, since such state would have negative energy. We also know that $[A_i(x),A_i(y)]$ and therefore also $[A_i^+(x),A_i^-(y)]\equiv ib_i(x-y)$ is a c number. We can, therefore, follow precisely the same steps as in the reduction of a polynomial in a free field to its ordered form, to obtain

$$_{ik}\langle\omega|A_i(x_1)\cdots A_i(x_n)|\omega\rangle_{ij}=C(x_1\cdots x_n)_{ik}\langle\omega|\omega\rangle_{ij}=0, \quad \text{if} \quad i\neq j,$$

where $C(x_1\cdots x_n)$ is a properly symmetrized product of $b_i(x-y)$'s. This completes the proof that $\mathcal{H}_{ik}\bigcap\mathcal{H}_{ij}=0$ if $k\neq j$. Introducing projection operators P_{ik} ($P_{ik}\mathcal{H}_i=\mathcal{H}_{ik}$), one can easily see that

$$[P_{ik},A_i(x)]=0, \tag{A.29}$$

$$[P_{ik},\pi^\mu]=0. \tag{A.30}$$

From (A.12), (A.16), (A.17), (A.22), (A.29), and (A.30), we conclude

$$\mathcal{H}=\sum_1^N{}_i\oplus\mathcal{H}_i, \quad [A(x),P_i]=[\pi^\mu,P_i], \tag{A.31}$$

where $A_i(x)$ is cyclic with respect to $|\omega\rangle_i$. One also has $[U(a,\Lambda),P_i]=0$ and

$$[A(x),A(y)]=i\sum_1^N{}_j C_j(x-y)P_j.$$

Therefore, $[A_i(x),A_i(y)]$ is a c number and $A_i(x)$ is cyclic in $\mathcal{H}_i$ with respect to $|\omega\rangle_i$. In $\mathcal{H}_i$, $|\omega\rangle_i$ is the unique vacuum (i.e., eigenstate of π_i^μ to the eigenvalue zero). We can use Theorem I and conclude that $A_i(x)$ is a generalized free field. We have, finally, the result: under the conditions of Theorem I′,

$$A(x)=\textstyle\sum_i A_i(x)P_i,$$

where $A_i(x)$ are generalized free fields; they are characterized by the "eigenvalues" $iC_k(x-y)$ of the matrix $\|b\|$: $b_{ij}=i_j\langle\Omega|[A(x),A(y)]|\Omega\rangle_i$. P_i are projection operators into orthogonal subspaces $\mathcal{H}_i$ ($\sum_1^N{}_i\oplus\mathcal{H}_i=\mathcal{H}$).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Nature of the Axioms of Relativistic Quantum Field Theory. I*

E. C. G. SUDARSHAN AND K. BARDAKCI
Department of Physics and Astronomy, University of Rochester, Rochester, New York

(Received April 25, 1961)

The formulation of field theories by means of Wightman functions is studied. It is shown that, given two field theories that satisfy all the axioms, one can construct a family of Wightman fields with the same properties by a process of superposition of Wightman functions. The condition of unitarity is formulated without reference to asymptotic conditions, and it is proved that the Wightman fields constructed by the above superposition process (starting with "unitary" fields) fail to preserve unitarity, and *a fortiori*, the standard asymptotic condition.

1. INTRODUCTION

IN the search for a dynamical scheme for describing elementary particle phenomena consistent with relativistic invariance and quantum mechanical principles, the theory of quantized fields has been favored with more study and has provided more insight than any other scheme. The use of manifestly covariant local Lagrangians as a starting point and the use of perturbation expansions lead to questionable mathematical operations with infinite quantities. In view of this, during the last few years the study of general field theories without starting with any specific Lagrangian has received much attention.[1] The more fundamental part of such a program concerns the study of an abstract axiom system more or less suggested by earlier Lagrangian theories. In such a study it is worthwhile to know if the axioms are independent and whether they are compatible; while the axioms are "related" to general physical requirements their truth is neither "self-evident" nor can one trust intuitive "physical" justifications for the compatibility of these axioms.

Among the set of axioms usually taken as characterizing quantized fields, these comments apply particularly to the so-called "asymptotic condition"[2] which enables one to relate the field operators to particle scattering amplitudes. The somewhat provisional nature of this axiom has been noted before; and perhaps not unconnected with this is the fact that the other "field axioms" have been the subject of a structure analysis by Wightman.[3] Making use of the tools developed in this brilliant study we show in this paper that the "asymptotic condition" is an independent axiom and that one can construct systems satisfying all other axioms but not this axiom provided that at least one quantum field theory yielding a nontrivial scattering matrix exists. In the course of this study we have been able to construct several examples of fields with a trivial scattering matrix.

In Sec. 2 we review Wightman's theory and construct certain elementary families of Wightman fields using the technique of vacuum expectation values. Section 3 discusses the weak axiom of asymptotic particle interpretation and the normalization of the field. The main result of the present paper is to show that almost all members of the families of fields constructed in Sec. 2 do not satisfy the (weak) axiom of asymptotic particle interpretation; this result is stated and proved in Sec. 4. Certain related comments are made in the concluding section.

2. FAMILIES OF WIGHTMAN FIELDS

According to Wightman,[3] a quantum field theory is defined in terms of a Hilbert space $\mathfrak{H}$ and a set of hermitian linear operators (more specifically, operator-valued distributions) $\phi(x)$ labeled by a four-vector x provided the following conditions are satisfied:

(I) Manifest Lorentz invariance. There must exist unitary operators $U(a,\Lambda)$ such that

$$\phi(\Lambda x+a)=U(a,\Lambda)\phi(x)U^{-1}(a,\Lambda)$$

for every proper orthochronous inhomogeneous Lorentz transformation.

(II) Absence of negative energy states. The spectrum of the Hamiltonian operator must be nonnegative, the Hamiltonian being defined as the hermitian generator of time translations.

(III) Local commutativity. The commutator of two field operators at space-like points must vanish,

$$[\phi(x),\phi(y)]=0 \quad \text{for} \quad (x-y)^2<0.$$

(IV) The existence of the "vacuum" state. There exists a unique state $|0\rangle$ invariant under all $U(a,\Lambda)$.

We now form the vacuum expectation values of products of n field operators labeled by the points $x_1, x_2, \cdots, x_n$:

$$W^{(n)}(x_1,x_2,\cdots,x_n)\equiv W^{(n)}(\{x\}) =\langle 0|\phi(x_1)\cdots\phi(x_n)|0\rangle. \quad (1)$$

It can then be shown that, as a consequence of the conditions imposed on the Hilbert space $\mathfrak{H}$ and the

* Supported by the U. S. Atomic Energy Commission.

[1] See, for example, the Proceedings of the "Colloque sur les Problèmes Mathématiques de la Théorie Quantique des Champs" (Lille, 1957); see also, "Problemi Matematici della Teoria Quantistica delle Particelle e dei Campi" Suppl. Nuovo cimento **14** (1959) and references given there.

[2] R. Haag: Dan. Mat. Fys. Medd. **29**, No. 1 (1955); H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo cimento **1**, 205 (1955); O. W. Greenberg, Ph.D. thesis, Princeton University 1956 (unpublished).

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

linear operators $\phi(x)$, this set of functions labeled by the four-vector variables (hereafter called Wightman functions) has the following properties:

(i) $W^{(n)}(\{\Lambda x+a\})=W^{(n)}(\{x\})$, (Lorentz invariance).

(ii) $W^{(n)}(\{x\})$ is the boundary value of a complex function $W^{(n)}(\{z\})$ analytic for $\mathrm{Im}\{z\}$ in the backward light cone (absence of negative energies).

(iii) $W^{(n)}(\{x\})=W^{(n)}(\{x'\})$, where $\{x'\}$ is any permutation of the n variables $\{x\}$, provided the permuted variables have space-like separations (local commutativity).

(iv)

$$\sum_{r,s}\int\cdots\int f_r^*(x_1,\cdots,x_r)W^{(r+s)}(x_1,\cdots,x_r,y_1,\cdots y_s)\times f_s(y_1,\cdots,y_s)d^4x_1\cdots d^4x_r d^4y_1\cdots d^4y_s\geqslant 0, \quad (2)$$

where f_r are suitable arbitrary functions (positive definite metric). Wightman has also shown[3] that these conditions are sufficient, that is, given a set of functions $W^{(n)}(\{x\})$ satisfying these conditions, one can construct a theory of a (neutral scalar) field satisfying the four conditions stated at the beginning of this section which has these functions for its vacuum expectation values.

Before the field theory so defined can be used to describe a model of relativistic quantum theory of particles, one must introduce some particle concepts. The structure satisfying only the conditions introduced in this section is a more general system; we shall refer to this structure as a "Wightman field."

We now state two obvious properties of a Wightman field in terms of its Wightman functions in the form of two theorems.

Theorem I (*scale change*). If $W^{(n)}(\{x\})$ are a set of Wightman functions, the set of functions $k^nW^{(n)}(\{x\})$ defines a Wightman field for every real number k.

This statement is immediately verified by noting that if $\phi(x)$ is the Wightman field which corresponds to $W^{(n)}(\{x\})$, then $k\phi(x)$ corresponds to $k^nW^{(n)}(\{x\})$.

Theorem II (*convexity*). If $W_1^{(n)}(\{x\})$ and $W_2^{(n)}(\{x\})$ are two sets of Wightman functions, the convex set

$$W^{(n)}(\{x\})=\lambda W_1^{(n)}(\{x\})+(1-\lambda)W_2^{(n)}(\{x\}) \quad (3)$$

defines a Wightman field provided the real number λ lies between 0 and 1.

The theorem is proved by noting that the functions $W^{(n)}(\{x\})$ satisfy all the conditions imposed on Wightman functions: Lorentz invariance, analyticity in the future tube, permutation symmetry for space-like separated arguments, and finally the condition specified by Eq. (2). Hence they define a Wightman field. Note that, in this case, it is not easy to construct the field operator in a simple manner but these functions satisfy all the conditions imposed on Wightman functions; hence they define a Wightman field. If λ is real but not necessarily in the interval $0\leqslant\lambda\leqslant 1$ then all conditions are satisfied except positive definiteness; even this last condition may be satisfied in special cases [as is seen by considering $W^{(n)}(\{x\})$ and $W_1^{(n)}(\{x\})$].

Thus, given one Wightman field we can construct an infinite number of distinct Wightman fields using Theorem I; however, out of this infinite set, a specific choice can be made by stating a normalization condition. We shall state such a condition in the next section. Theorem II allows us to construct an infinite set of Wightman fields (normalized, if so required) from two (or more) distinct Wightman fields. Let us call the set of all Wightman fields $W^{(n)}(\{x\})$ generated by $W_1^{(n)}(\{x\})$ and $W_2^{(n)}(\{x\})$ the "family"; every point in this family is labeled by a parameter λ. We have remarked above that while $0\leqslant\lambda\leqslant 1$ is allowed in all cases, values of λ outside this interval are not necessarily forbidden. It is then interesting to state the following theorem regarding the boundedness of the allowed values of λ:

Theorem III (*semibounded families*). There exists either a lower limit λ_1 or an upper limit λ_2 (or both) such that for either $\lambda<\lambda_1$ or $\lambda_2<\lambda$ (or both) the combinations

$$V^{(n)}(\{x\})=\lambda W_1^{(n)}(\{x\})+(1-\lambda)W_2^{(n)}(\{x\})$$

cannot be a set of Wightman functions.

To prove the existence of such limits, we use the positive definiteness condition showing that these are violated for sufficiently large negative or positive values of λ. Consider in particular $W_1^{(2)}(\{x\})$, which is non-negative according to (2). It cannot be everywhere zero without making the field operator $\phi_1(x)$ trivial. Choose any suitable testing function $f(y)$ such that

$$\int f^*(y)W_1^{(2)}(x+\tfrac{1}{2}y,\, x-\tfrac{1}{2}y)f(y)d^4y=1,$$

and let

$$\int f^*(y)W_2^{(2)}(x+\tfrac{1}{2}y,\, x-\tfrac{1}{2}y)f(y)d^4y=a\geqslant 0.$$

Then,

$$\int f^*(y)V^{(2)}(x+\tfrac{1}{2}y,\, x-\tfrac{1}{2}y)f(y)d^4y=a+\lambda(1-a),$$

which becomes negative $\lambda<-a/|1-a|$ or for $a/|a-1|>\lambda$ according as a is less than or greater than unity. Hence, the statement made in the theorem is proved.

This demonstration however does not guarantee that provided $\lambda_1<\lambda<\lambda_2$ the set $V^{(n)}(\{x\})$ are Wightman functions since the positive definiteness condition in its complete form may still be violated; it may even be violated for other testing functions using $W^{(2)}(\{x\})$ only. However, from Theorem II we know that there exists the nontrivial family $0\leqslant\lambda\leqslant 1$ at least. In general the family is, of course, larger.

It is also remarkable that of the original fields obeyed canonical commutation relations the family of Wightman fields so generated also satisfy canonical commutation relations. This statement is consequent upon the identification of all matrix elements of the commutator of the field and its time derivative (at the same time) in terms of the Wightman functions.

3. ASYMPTOTIC PARTICLE INTERPRETATION AND THE SCATTERING MATRIX

If this *field* theory is to become a theory of interacting *particles*, one must introduce particle variables into the theory and identify at least some subspace of the Hilbert space $\mathfrak{H}$ as being associated with the particle states. Such a program[4] has so far not been carried out except for free fields. There is however another type of particle interpretation which is less ambitious in the sense that certain linear combinations of vacuum expectation values of the fields are identified with a scattering amplitude for "asymptotically free" particles.[5] Since there are certain properties to be satisfied by the scattering amplitude *this identification* in turn *imposes some restrictions on the Wightman* fields. However the scattering amplitudes themselves provide only an incomplete characterization of the field; and it appears that without the use of sufficiently strong additional postulates, the scattering amplitudes do *not* determine the Wightman field. In support of this, it is known that one can construct several distinct Wightman fields with a trivial associated scattering amplitude.[6]

It is conventional[2] to state the requirement of an asymptotic particle interpretation in terms of an appropriately stated "asymptotic condition" and then to "derive" the scattering amplitude in terms of certain linear combinations of vacuum expectation values. We shall follow the alternative method of stating the connection between the scattering amplitude and the vacuum expectation values as the additional axiom. This apparently arbitrary procedure has certain advantages: first of all, unlike the other axioms of quantum field theory, the asymptotic condition has so far been stated only in unsatisfactory forms and their plausibility is not immediately obvious. The best defense seems to be that it leads to a covariant expression for the scattering amplitude; but the expression itself could be obtained by other means, say for example, by a formal summation of the perturbation series.[7] Secondly the question of completeness of the particle scattering states which is generally a prerequisite to the axiomatization of the asymptotic condition seems too strong; it is conceivable that the field Hilbert space is considerably larger than the particle Hilbert space.

We shall hence take as an axiom the following condition[8]:

The scattering matrix element related to the transition to a state containing r particles with four-momenta $p_1, \cdots, p_r$ from a state containing s particles with four-momenta $q_1, \cdots, q_s$ (with $p_1^2=\cdots=q_s^2=\mu^2$) is given by the expression

$$S(p_1,\cdots,p_r;q_1,\cdots,q_s)$$

$$=\int d^4x_1\cdots d^4x_r d^4y_1\cdots d^4y_s$$

$$\times\Lambda(p_1,x_1)\cdots\Lambda(p_r,x_r)\Lambda(-q_1,y_1)\cdots\Lambda(-q_s,y_s)$$

$$\times\langle 0|T[\phi(x_1),\cdots,\phi(y_s)]|0\rangle, \quad (4a)$$

where

$$\Lambda(p,x)=\frac{-i}{(2\pi)^4}e^{ipx}(\Box_x^2-\mu^2) \quad (4b)$$

and μ is a "mass" parameter. Hence, the T product vacuum expectation value is defined in terms of the Wightman functions by the equations

$$\langle 0|T\{\phi(x_1),\cdots,\phi(x_n)\}|0\rangle=W^{(n)}(x_1,\cdots x_n) \quad (5a)$$

for $x_1^0>x_2^0>\cdots>x_n^0$,

$$\langle 0|T\{\phi(x_1),\cdots,\phi(x_n)\}|0\rangle$$
$$=\langle 0|T\{\phi(x_1'),\cdots,\phi(x_n')\}|0\rangle, \quad (5b)$$

where $x_1', \cdots, x_n'$ are any permutations of $x_1, \cdots, x_n$. (Asymptotic particle interpretation.)

At this point, we must restrict our further discussion to Wightman fields for which the T-product vacuum expectation values exist. Given any Wightman field we can now calculate the particle scattering matrix in terms of this identification; but there is no guarantee that the scattering matrix so defined satisfies the conditions imposed on a scattering matrix, in particular unitarity. It is considered further necessary that the one-particle states are "steady" so that the S-matrix elements connecting one-particle states to any other state vanish identically (and that the two-particle scattering is elastic below the three-particle threshold). This condition can be used to normalize the field operator:

$$\int d^4x\Lambda(p,x)\int d^4y\Lambda(-q,y)\langle 0|T\{\phi(x),\phi(y)\}|0\rangle$$
$$=(2\pi)^4\delta(p-q)\delta(p^2-\mu^2) \quad (6)$$

with $p^2=q^2\to\mu^2$. It then follows that if $W^{(n)}(\{x\})$ denotes the Wightman functions for this normalized field of mass μ then $k^nW^{(n)}(\{x\})$ defines a field which is not normalized except for the special case $k=\pm 1$. The

[4] A. S. Wightman and S. S. Schweber, Phys. Rev. **98**, 812 (1955).
[5] This point of view is somewhat more general than the classification of particle interpretations discussed by Wightman and Schweber (reference 4).
[6] H. J. Borchers, Nuovo cimento **15**, 784 (1960).
[7] See, for example, Y. Nambu, Phys. Rev. **98**, 803 (1955).
[8] This choice is very closely related to the work of K. Nishijima, Phys. Rev. **119**, 485 (1960).

fields defined in terms of two (or more) sets of normalized Wightman functions in the form $W^{(n)}=\lambda W_1^{(n)}+(1-\lambda)W_2^{(n)}$ is also normalized in the above manner if and only if the masses are identical.

The axiom of asymptotic particle interpretation introduced here is weaker than the usual asymptotic condition in the sense that we do not assume either the completeness of the many particle states nor the existence of asymptotic fields. But if the asymptotic condition is postulated as an axiom of the theory in addition to the axioms for a Wightman field, we can *derive* the expression for the particle scattering matrix yielding the so-called reduction formulas.[9] Thus the axiom of asymptotic particle interpretation for a Wightman field yields a more general system than the Wightman field with the stronger axiom of asymptotic condition. Needless to say everything we have proved in the following sections apply *a fortiori* to fields satisfying the usual system of axioms including the asymptotic condition. We now proceed to show that Wightman fields in general do *not* have an asymptotic particle interpretation.

4. WIGHTMAN FIELDS WITHOUT ASYMPTOTIC PARTICLE INTERPRETATION

In terms of the scattering matrix S one may define the scattering amplitude f in the standard manner; and then note that the scattering amplitude so defined is *linearly* related to the Wightman functions. The unitarity relation imposed on $f(p_1,\cdots,p_r;q_1,\cdots,q_s)$ is

$$f(p_1,\cdots,p_r;q_1,\cdots,q_s)-f^*(q_1,\cdots,q_s;p_1,\cdots,p_r)$$
$$=i\sum_{n=0}^{\infty}\int d^4k_1\delta(k_1^2-\mu^2)\theta(k_1^0)\cdots\int d^4k_n\delta(k_n^2-\mu^2)$$
$$\times\theta(k_n^0)f(p_1,\cdots,p_r;k_1,\cdots,k_n)$$
$$\times f^*(q_1,\cdots,q_s;k_1,\cdots,k_n) \quad (7)$$

or symbolically,

$$(f-f^+)=iff^+. \quad (7')$$

In the summation, most of the terms contribute nothing since energy and momentum must be conserved if the scattering amplitude is not to vanish. Let f_1 and f_2 be the scattering amplitudes for two Wightman fields with asymptotic particle interpretation defined by their Wightman functions $W_1^{(n)}$ and $W_2^{(n)}$. We shall further specialize than to correspond to the same "mass." If we now define a field in terms of the Wightman functions

$$W^{(n)}=\lambda W_1^{(n)}+(1-\lambda)W_2^{(n)}$$

in view of the linear relation between the Wightman function and the scattering amplitude, it follows that the scattering amplitude f for this Wightman field is simply given by

$$f=\lambda f_1+(1-\lambda)f_2.$$

Using the unitarity condition, Eq. (7), twice it is now possible to derive the relation

$$\{\lambda f_1+(1-\lambda)f_2\}\{\lambda f_1^+ +(1-\lambda)f_2^+\}$$
$$=\lambda f_1f_1^+ +(1-\lambda)f_2f_2^+, \quad (8')$$

which may be written

$$\lambda(1-\lambda)\sum_{n=0}^{\infty}\int d^4k_1\delta(k_1^2-\mu^2)\theta(k_1^0)\cdots\int d^4k_n\delta(k_n^2-\mu^2)$$
$$\times\theta(k_n^0)g(p_1,\cdots,p_r;k_1,\cdots,k_n)$$
$$\times g^*(q_1,\cdots q_s;k_1,\cdots,k_n)=0, \quad (8a)$$

with

$$g=f_1-f_2. \quad (8b)$$

If we now specialize to the case of elastic scattering, the integrand is nonnegative and the vanishing of the integral implies that either $g=0$ identically or $\lambda(1-\lambda)=0$. In the first case the two Wightman fields must have the same scattering matrix and all the Wightman fields in the allowed family $\lambda_1\leqslant\lambda\leqslant\lambda_2$ yield the *same* scattering matrix; the second case is trivial. We may now prove the following theorem.

Theorem IV (*equivalent scattering matrices*). A Wightman field defined in terms of the Wightman functions

$$W^{(n)}=\sum\lambda_\alpha W_\alpha^{(n)}, \quad \sum\lambda_\alpha=1, \quad \lambda_\alpha\geqslant 0,$$

the functions $W_\alpha^{(n)}$ admitting asymptotic particle interpretations with the same "mass," has an asymptotic particle interpretation if and only if all the Wightman fields have the same scattering matrix.

This more general statement is proved essentially the same way as used above; one derives in place of (8′) the equation

$$\sum_{\alpha>\beta}\lambda_\alpha\lambda_\beta(f_\alpha^+-f_\beta^+)(f_\alpha-f_\beta)=0,$$

from which it follows that $f_\alpha=f_\beta$ unless λ_α or λ_β vanishes provided all the λ_α are nonnegative. Note that, unlike the case of two fields only, here the condition $\lambda_\alpha\geqslant 0$ cannot be simply relaxed; in general, on grounds of continuity, one expects the domain of values of λ_α (with sum unity) for which the theorem holds is somewhat larger in view of the demonstration above regarding only two fields.

5. DISCUSSION

The results of the preceding section imply that the axiom of asymptotic particle interpretation is independent of the other axioms of field theory and is not derivable from them; a conclusion already indicated by the existence of several distinct fields with the same S matrix. We have actually used only a weaker axiom

[9] H. Lehmann, K. Symanzik, and W. Zimmerman, Nuovo cimento 1, 205 (1955).

in this connection in the sense that we have neither required detailed properties of the field mass spectrum nor the completeness of the many-particle states. Our systems are correspondingly more general and the "unitarity conditions" are imposed only on the Fourier transforms of the time-ordered combinations

$$\langle 0|T\{\phi(x_1),\cdots,\phi(x_n)\}|0\rangle$$

of the Wightman functions for momenta on the mass shell; without additional restrictions this is not sufficient to determine the field in any sense. Yet here we see that the unitarity requirement on the particle scattering matrix excludes most Wightman fields from having an asymptotic particle interpretation.

Perhaps the weakest point of the present investigation is that it has not provided any example of a field theory with asymptotic particle interpretation with a nontrivial scattering matrix; rather it asserts that if there exists at least one such theory there exists an infinity of Wightman fields not having an asymptotic particle interpretation belonging to the family generated by this one field together with the free field of the same mass.

We have worked here within the framework of the conventional axiomatization of quantum field theory. If the purpose of the field theory is only to provide a quantum theory of interacting particles invariant under the complex Lorentz group, the conventional axiomatization is too rigid in that it imposes "physical requirements" on the field. This is most easily seen in the case of the axiom of positive definiteness: in a theory where the physical particle states do not form a complete set of states in the generalized Hilbert space in which the field operators are defined, it is sufficient if the particle states constitute a subspace with positive definite metric. That these considerations are not devoid of physical interest is seen from the example of the quantized Maxwell field. One of the present authors has discussed[10] examples of quantum field theories formulated in terms of a generalized Hilbert space with an indefinite metric where again the physical particle states are not complete in the generalized space but constitute only a subspace with positive definite metric. In such theories the physical interpretation requires an interpretive postulate and the precise form of this postulate depends on the dynamics of the field.

Our investigations also provide several examples of Wightman fields with a trivial scattering matrix. In addition to a trivial scale change $W^{(n)} \to k^n W^{(n)}$, we also have more generally

$$W^{(n)}=\sum_\alpha \lambda_\alpha k_\alpha{}^n W_\alpha{}^{(n)}, \quad \sum_\alpha \lambda_\alpha=1, \quad \lambda_\alpha \geqslant 0, \qquad (9)$$

which provide Wightman fields, the functions $W_\alpha{}^{(n)}$ corresponding to known theories; say either free fields with arbitrary masses, or the Wick polynomials of free fields or terminating Haag expansions.[6] By a limiting procedure in forming such linear combinations one can produce any two-point function

$$\langle 0|\phi(x)\phi(y)|0\rangle=\int d\rho(m^2)\Delta^{(1)}(m;x-y), \qquad (10)$$

(where $\Delta^{(1)}(m;x-y)$ is the two-point Wightman function for a free field of mass m) by taking for the Wightman functions

$$W^{(n)}(\{x\})=\int d\rho(m^2)W^{(n)}(m;\{x\}), \qquad (11)$$

where $W^{(n)}(m;\{x\})$ are the Wightman functions for a free field of mass m, and $\rho(m^2)$ is a nonnegative measure. But all these fields have a trivial scattering matrix.

Finally the present study illustrates the validity of Wightman's statement[3] that the consequences of positive definiteness are distinct from the consequences of unitarity. The Wightman fields constructed above satisfy positive definiteness but do not yield unitary scattering matrices, while certain indefinite metric theories (including quantum electrodynamics)[10] provide examples of theories in which the field operators are defined in a generalized Hilbert space but the scattering matrices are unitary.

ACKNOWLEDGMENTS

We would like to thank Dr. A. S. Wightman, Dr. O. W. Greenberg, Dr. G. Feinberg, and T. F. Jordan for useful discussions; and Dr. Y. Nambu and Dr. I. Bialynicki-Birula for a critical reading of the manuscript.

[10] E. C. G. Sudarshan, Phys. Rev. **123**, 2183 (1961).

Dynamical Mappings of Density Operators in Quantum Mechanics*

THOMAS F. JORDAN AND E. C. G. SUDARSHAN
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received February 21, 1961)

The most general dynamical law for a quantum mechanical system is studied with particular reference to the necessary and sufficient conditions for such a law to represent Hamiltonian dynamics. The main results are stated in the form of three theorems.

I. INTRODUCTION

THE most general description of the state of a quantum mechanical system is afforded by the von Neumann density operator, and may be defined as a "real" linear functional which maps non-negative Hermitian operators on a Hilbert space to non-negative numbers and maps the unit operator to unity. It is well known that, in special cases, one can specify the states in terms of normalized vectors of the Hilbert space. The dynamical law is then usually given as a unitary transformation on these vectors. We refer to this law as Hamiltonian dynamics. But the most general dynamical law for a quantum mechanical system is to be formulated in terms of the density operator and may be described by a linear mapping of the set of density operators into itself. The question immediately arises as to the conditions under which a linear mapping of the space of all operators into itself maps the subset of density operators into itself, and then as to the conditions under which such a mapping represents Hamiltonian dynamics. This problem has been investigated[1] for the restricted case of a system described by a finite dimensional vector space. In this paper we will answer these questions for operators defined on any Hilbert space.

Since the density operators form a convex set, the possible dynamical mappings also form a convex set. It may then appear that all dynamical mappings can be formed as "probabalistic" combinations of some simple set of extremal mappings in the same fashion as all density operators can be formed as mixtures of pure state operators. That this set of extremal mappings cannot be limited to the Hamiltonian mappings is evident from the existence of a mapping of all pure state operators to a single pure state operator. Hence, such a limitation would be an additional physical postulate. If we limit ourselves to mappings of pure states to pure states as described by a linear mapping on the Hilbert space then, such a postulate is implicitly assumed.

In Sec. II we develop some properties of the convex set of density operators and the linear space to which they belong and in Sec. III we prove three theorems which contain the main results outlined above.

II. THE OPERATOR SPACE AND THE CONVEX SUBSET OF DENSITY OPERATORS

The quantum mechanical state of a physical system can be specified by a density operator ρ which satisfies the conditions[2]:

$$(\phi,\rho\psi)=(\rho\phi,\psi) \qquad \text{(Hermiticity)}, \qquad (1)$$

$$(\phi,\rho\phi)\geq 0 \qquad \text{(positive-definiteness)}, \qquad (2)$$

$$\mathrm{Tr}(\rho)=1 \qquad \text{(normalization)}, \qquad (3)$$

where ϕ and ψ are any vectors of the Hilbert space $\mathfrak{H}$ on which the operator ρ is defined. It is well known[3] that an operator which satisfies these conditions has a purely discrete spectrum with real non-negative eigenvalues. From this it follows that $\mathrm{Tr}(\rho^2)\leq 1$, the equality holding if and only if ρ has just one nonzero eigenvalue, or equivalently if and only if $\rho^2=\rho$ which is the condition that ρ be a projection operator. In the latter case we say that ρ represents a pure state. Any density operator can then be expanded in its spectral representation as

$$\rho=\sum_i a_i\rho^{(i)},$$

where $\rho^{(i)}$ are orthogonal projection operators and a_i are real positive coefficients satisfying $\sum_i a_i=1$.

If we consider the linear operators on $\mathfrak{H}$ as themselves forming a vector space, we can define an inner product in this space by

$$(\rho_1,\rho_2)=\mathrm{Tr}(\rho_1^+\rho_2).$$

Let $\mathfrak{L}$ be the linear space of all operators ρ for which $\|\rho\|^2=\mathrm{Tr}(\rho^+\rho)<\infty$. Then all density operators belong to $\mathfrak{L}$. If $\rho=a\rho^{(1)}+(1-a)\rho^{(2)}$, where $\rho^{(1)}$ and $\rho^{(2)}$ are density operators and $0\leq a\leq 1$, then by an application of Schwartz's inequality we get that

$$\begin{aligned}(\rho,\rho)\leq a^2(\rho^{(1)},\rho^{(1)})+(1-a)^2(\rho^{(2)},\rho^{(2)})\\ +2a(1-a)[(\rho^{(1)},\rho^{(1)})(\rho^{(2)},\rho^{(2)})]^{\frac{1}{2}}\\ \leq a^2+(1-a)^2+2a(1-a)=1,\end{aligned}$$

where the equality holds only if $\rho^{(1)}=\rho^{(2)}$ and $(\rho^{(1)},\rho^{(1)})=1$. Hence, if ρ represents a pure state, it cannot be

* Supported in part by the U. S. Atomic Energy Commission.

[1] E. C. G. Sudarshan, P. M. Mathews, and J. Rau, Phys. Rev. 121, 920 (1961). We refer the reader to this paper for a discussion of the physical motivation for the problem considered in the present paper and also for physical examples which illustrate the basic ideas and possible applications.

[2] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, New Jersey, 1955), Chap. IV. For a general discussion of density operators see, e.g., U. Fano, Revs. Modern Phys. 29, 74 (1957).

[3] von Neumann, reference 2, p. 189.

formed as a linear combination with positive coefficients of any two distinct density operators. For any $\rho^{(1)}$ and $\rho^{(2)}$ as above, ρ is also a density operator, so the density operators form a convex set. We will call the set of all ρ formed as above for all values of a, $0<a<1$, the line segment between $\rho^{(1)}$ and $\rho^{(2)}$. We then define the set of extremal elements of the convex set as the set of all elements which do not belong to a line segment between any two distinct elements. From the above remarks, it is clear that the extremal elements of the set of all density operators is the set of operators which represent pure states, and that all other elements can be formed by positive combinations of these.

We will denote by $\psi\phi^+$ a linear operator defined on $\mathfrak{K}$ by

$$(\phi^{(r)},\psi\phi^+\phi^{(s)})=(\phi^{(r)},\psi)(\phi,\phi^{(s)}),$$

where ψ and ϕ are vectors and the $\phi^{(r)}$ form an orthonormal basis in $\mathfrak{K}$. Then the operators $\phi^{(r)}\phi^{(s)+}$ form an orthonormal basis in $\mathfrak{L}$. For

$$\begin{aligned}(\phi^{(r)}\phi^{(s)+},\phi^{(r')}\phi^{(s')+})&=\mathrm{Tr}\{(\phi^{(r)}\phi^{(s)+})^+\phi^{(r')}\phi^{(s')+}\}\\&=\sum_{n,m}(\phi^{(m)},\phi^{(s)})(\phi^{(r)},\phi^{(n)})(\phi^{(n)},\phi^{(r')})(\phi^{(s')},\phi^{(m)})\\&=\delta_{rr'}\delta_{ss'}\end{aligned}$$

and any operator ρ in $\mathfrak{L}$ can be expanded as

$$\rho=\sum_{rs}\rho_{rs}\phi^{(r)}\phi^{(s)+}$$

where

$$\rho_{rs}=(\phi^{(r)}\phi^{(s)+},\rho)=(\phi^{(r)},\rho\phi^{(s)})$$

and

$$\sum_{rs}|\rho_{rs}|^2=\mathrm{Tr}(\rho^+\rho)<\infty.$$

Since we are interested mainly in density operators, it is of interest that the pure state operators $\phi^{(r)}\phi^{(r)+}$ for all r and

$$\begin{aligned}&(1/\sqrt{2})(\phi^{(r)}+\phi^{(s)})(1/\sqrt{2})(\phi^{(r)}+\phi^{(s)})^+\\&\quad=\tfrac{1}{2}\phi^{(r)}\phi^{(r)+}+\tfrac{1}{2}\phi^{(s)}\phi^{(s)+}+\tfrac{1}{2}(\phi^{(r)}\phi^{(s)+}+\phi^{(s)}\phi^{(r)+})\\&(1/\sqrt{2})(\phi^{(r)}+i\phi^{(s)})(1/\sqrt{2})(\phi^{(r)}+i\phi^{(s)})^+\\&\quad=\tfrac{1}{2}\phi^{(r)}\phi^{(r)+}+\tfrac{1}{2}\phi^{(s)}\phi^{(s)+}+\tfrac{1}{2}i(\phi^{(s)}\phi^{(r)+}-\phi^{(r)}\phi^{(s)+})\end{aligned}\qquad(4)$$

for all r, s, $r<s$, form a linearly independent set which spans $\mathfrak{L}$.[4]

III. DYNAMICAL MAPPINGS

The most general dynamical transformation on the system is represented by a linear mapping of the set of density operators into itself. But since there are sets of density operators spanning $\mathfrak{L}$, this uniquely defines a linear mapping of $\mathfrak{L}$ into itself,[5]

$$\rho\rightarrow\rho'=A\rho.\qquad(5)$$

We shall call a linear mapping on $\mathfrak{L}$ which maps the set of density operators into itself a dynamical mapping. The properties of such a mapping are described by the following.

Theorem 1. Necessary and sufficient conditions for a linear operator A on $\mathfrak{L}$ to give a dynamical mapping are:

For any set of basis vectors $\phi^{(r)}$ in $\mathfrak{K}$

(a) $(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=(\phi^{(s)}\phi^{(r)+},A\phi^{(s')}\phi^{(r')+})^*$,

(b) $\sum_r(\phi^{(r)}\phi^{(r)+},A\phi^{(r')}\phi^{(s')+})=\delta_{r's'}$,

(c) The operator ω defined on $\mathfrak{K}$ by $(\phi^{(s)},\omega\phi^{(r')})=(\phi^{(r)}\phi^{(r)+},A\phi^{(r')}\phi^{(s')+})$ is positive definite for each choice of r. In particular, this implies that

$$(\phi^{(r)}\phi^{(r)+},A\phi^{(r')}\phi^{(r')+})\geq 0.$$

When these conditions are satisfied, Hermitian operators are mapped to Hermitian operators, the trace is preserved, and positive definite operators are mapped to positive definite operators.

Proof. We write Eq. (5) as

$$(\phi^{(r)},\rho'\phi^{(s)})=\sum_{r's'}(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')})(\phi^{(r')},\rho\phi^{(s')}).\qquad(5a)$$

If A satisfies (a), we can deduce that ρ' is Hermitian when ρ is Hermitian. If A satisfies (b) we can deduce that $\mathrm{Tr}(\rho')=\mathrm{Tr}(\rho)$. If A satisfies (c), then $(\phi^{(r)},\rho'\phi^{(r)})=\sum_{r's'}(\phi^{(s')},\omega\phi^{(r')})(\phi^{(r')},\rho\phi^{(s')})=\mathrm{Tr}(\omega\rho)\geq 0$ if ρ is positive definite. Since this holds for each $\phi^{(r)}$ belonging to any set of basis vectors in $\mathfrak{K}$, we deduce that ρ' is positive definite if ρ is positive definite. The sufficiency of these conditions as well as the final statement of the theorem have thus been proved. The necessity of (a) is obtained by noting that if ρ is taken to be $\frac{1}{2}(\phi^{(n)}\phi^{(m)+}+\phi^{(m)}\phi^{(n)+})$ or $\frac{1}{2}i(\phi^{(n)}\phi^{(m)+}-\phi^{(m)}\phi^{(n)+})$, each of which is, according to (4), a real combination of pure state operators, then ρ' must be Hermitian, or from (5a),

$$\begin{aligned}&\tfrac{1}{2}\{(\phi^{(r)}\phi^{(s)+},A\phi^{(n)}\phi^{(m)+})+(\phi^{(r)}\phi^{(s)+},A\phi^{(m)}\phi^{(n)+})\}^*\\&\quad=\tfrac{1}{2}\{(\phi^{(s)}\phi^{(r)+},A\phi^{(n)}\phi^{(m)+})+(\phi^{(s)}\phi^{(r)+},A\phi^{(m)}\phi^{(n)+})\},\\&\tfrac{1}{2}\{(\phi^{(r)}\phi^{(s)+},A\phi^{(n)}\phi^{(m)+})+(\phi^{(r)}\phi^{(s)+},A\phi^{(m)}\phi^{(n)+})\}^*\\&\quad=-\tfrac{1}{2}\{(\phi^{(s)}\phi^{(r)+},A\phi^{(n)}\phi^{(m)+})-(\phi^{(s)}\phi^{(r)+},A\phi^{(m)}\phi^{(n)+})\},\end{aligned}$$

from which (a) follows. Similarly, using these same operators for ρ we see from (4) that we must have $\mathrm{Tr}(\rho')=0$ while if $\rho=\phi^{(n)}\phi^{(n)+}$ we must have $\mathrm{Tr}(\rho')=1$. Using (5a), these imply (b). To prove the necessity of (c) we note that for every pure state operator $\rho=\psi\psi^+$

[4] This space of operators with the inner product defined by the trace has been considered by J. Schwinger, Proc. Natl. Acad. Sci. U. S. 46, 257 (1960).

[5] To avoid confusion between the two types of operators we will use capital letters A for operators on $\mathfrak{L}$ and Greek letters ρ, ω, σ for operators on $\mathfrak{K}$ (elements of $\mathfrak{L}$). Greek letters ϕ, ψ, χ, ξ will denote vectors in $\mathfrak{K}$, while small letters a will denote scalars.

we must have

$$(\phi^{(r)},\rho'\phi^{(r)})=\sum_{r's'}(\phi^{(s')},\omega\phi^{(r')})(\phi^{(r')},\psi\psi^{+}\phi^{(s')})$$
$$=\sum_{r's'}(\phi^{(s')},\psi)^{*}(\phi^{(s')},\omega\phi^{(r')})(\phi^{(r')},\psi)$$
$$=(\psi,\omega\psi)\geq 0$$

for every $\phi^{(r)}$ belonging to any set of basic vectors in $\mathcal{H}$. This completes the proof of Theorem 1.

In order to characterize the Hamiltonian dynamical transformations we will need one more definition. If the dynamical mapping maps all pure state operators to pure state operators then for each normalized vector ϕ in $\mathcal{H}$ we have $\phi\phi^{+}=\rho\rightarrow\rho'=\phi'\phi'^{+}$. The mapping $a\phi\rightarrow a\phi'$ of $\mathcal{H}$ into $\mathcal{H}$ will be called the mapping induced on $\mathcal{H}$ by the dynamical mapping. Note that this induced mapping is not necessarily linear and is defined only to within a (unimodular) phase factor. When we can choose these phase factors so as to make the mapping linear, we will say that the dynamical mapping induces a linear mapping on $\mathcal{H}$.

Theorem 2. Equivalent necessary and sufficient conditions for a dynamical transformation to represent Hamiltonian dynamics are[6]:

(i) There exists a linear unitary operator ω on $\mathcal{H}$ such that $\rho\rightarrow\rho'=\omega\rho\omega^{+}$. That is the operator A has the form

$$(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=(\phi^{(r)},\omega\phi^{(r')})(\phi^{(s)},\omega\phi^{(s')})^{*}.$$

(ii) The dynamical mapping gives a mapping of the set of pure state operators into itself and induces a linear mapping on $\mathcal{H}$.

(iii) For each member $\chi^{(i)}$ of any set on basis vectors in $\mathcal{H}$, there exists a normalized vector $\xi^{(i)}$ such that the dynamical transformation maps

$$\chi^{(i)}\chi^{(j)+}\quad\text{to}\quad\xi^{(i)}\xi^{(j)+}.$$

(iv) The operator A of the dynamical mapping can be factored in the form

$$(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=(\phi^{(r)},\omega\phi^{(r')})(\phi^{(s)},\sigma\phi^{(s')})^{*},$$

where ω and σ are linear operators on $\mathcal{H}$.

Proof. (i) represents the usual form of Hamiltonian dynamics; we need only prove that (i) implies (ii) implies (iii) implies (iv) implies (i).

That (i) implies (ii) is obvious. For $\phi\phi^{+}\rightarrow\phi'\phi'^{+}=\omega\phi\phi^{+}\omega^{+}=\omega\phi(\omega\phi)^{+}$ induces the linear mapping

$$a\phi\rightarrow a\phi'=a\omega\phi\quad\text{on}\quad\mathcal{H}.$$

To prove that (ii) implies (iii) we note that for any $\chi^{(i)}$ belonging to a set of basis vectors in $\mathcal{H}$, the pure state operator $\chi^{(i)}\chi^{(i)+}$ is mapped to a pure state operator, say $\xi^{(i)}\xi^{(i)+}$. Also, according to (ii), all the $\xi^{(i)}$ must be normalized vectors in $\mathcal{H}$. Since the induced mapping on $\mathcal{H}$ must be linear it can be determined by $\xi^{(i)}=\omega\chi^{(i)}$. Using Eq. (4) and the fact that $(1/\sqrt{2})(\chi^{(i)}+\chi^{(j)})$ and $(1/\sqrt{2})(\chi^{(i)}+i\chi^{(j)})$ are mapped to $(1/\sqrt{2})(\xi^{(i)}+\xi^{(j)})$ and $(1/\sqrt{2})(\xi^{(i)}+i\xi^{(j)})$, respectively, we see that $\frac{1}{2}(\chi^{(i)}\chi^{(j)+}+\chi^{(j)}\chi^{(i)+})$ and $\frac{1}{2}i(\chi^{(i)}\chi^{(j)+}-\chi^{(j)}\chi^{(i)+})$ are mapped to $\frac{1}{2}(\xi^{(i)}\xi^{(j)+}+\xi^{(j)}\xi^{(i)+})$ and $\frac{1}{2}i(\xi^{(i)}\xi^{(j)+}-\xi^{(j)}\xi^{(i)+})$ from which it follows that $\chi^{(i)}\chi^{(j)+}$ is mapped to $\xi^{(i)}\xi^{(j)+}$ which establishes (iii).

To obtain (iv) we use Eq. (5a) with (iii) to write

$$(\phi^{(r)},\xi^{(i)}\xi^{(j)+}\phi^{(s)})$$
$$=\sum_{r's'}(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})(\phi^{(r')},\chi^{(i)}\chi^{(j)+}\phi^{(s')}).$$

Then,

$$\sum_{ij}(\phi^{(r)},\xi^{(i)})(\chi^{(i)},\phi^{(r'')})(\phi^{(s'')},\chi^{(j)})(\xi^{(j)},\phi^{(s)})$$
$$=\sum_{r's'}(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})\sum_{i}(\phi^{(r')},\chi^{(i)})(\chi^{(i)},\phi^{(r'')})$$
$$\times\sum_{j}(\phi^{(s'')},\chi^{(j)})(\chi^{(j)},\phi^{(s')})$$
$$=\sum_{r's'}(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})\delta_{r'r''}\delta_{s's''}$$

since both the $\phi^{(r)}$ and the $\chi^{(i)}$ were assumed to form sets of basis vectors in $\mathcal{H}$. Hence $(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=\sum_i(\phi^{(r)},\xi^{(i)}\chi^{(i)+}\phi^{(r')})\sum_j(\phi^{(s)},\xi^{(j)}\chi^{(j)+}\phi^{(s')})^{*}$ and setting $\omega=\sigma=\sum_i\xi^{(i)}\chi^{(i)+}$ gives (iv).

Finally to show that (iv) implies (i) we use condition (a) of Theorem 1 which in the factored form of (iii) is

$$(\phi^{(r)},\omega\phi^{(r')})(\phi^{(s)},\sigma\phi^{(s')})^{*}=(\phi^{(s)},\omega\phi^{(s')})^{*}(\phi^{(r)},\sigma\phi^{(r')}).$$

Then $\sigma=c\omega$ where c is a real number. For $(\phi^{(r)},\sigma\phi^{(r')})=0$ if and only if $(\phi^{(r)},\omega\phi^{(r')})=0$, and for all $(\phi^{(r)},\sigma\phi^{(r')})\neq 0$, $(\phi^{(s)},\sigma\phi^{(s')})\neq 0$ we have

$$\frac{(\phi^{(r)},\omega\phi^{(r')})}{(\phi^{(r)},\sigma\phi^{(r')})}=\frac{(\phi^{(s)},\omega\phi^{(s')})^{*}}{(\phi^{(s)},\sigma\phi^{(s')})^{*}}=\frac{1}{c}.$$

Then,

$$(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=(\phi^{(r)},c^{\frac{1}{2}}\omega\phi^{(r')})(\phi^{(s)},c^{\frac{1}{2}}\omega\phi^{(s)})^{*}.$$

Now if $\chi^{(i)}$ is any set of basis vectors in $\mathcal{H}$, let $\xi^{(i)}=c^{\frac{1}{2}}\omega\chi^{(i)}$. Then using condition (b) of Theorem 1 we have that

$$(\xi^{(i)},\xi^{(j)})=\sum_{rr's'}(\phi^{(r)},c^{\frac{1}{2}}\omega\phi^{(r')})(\phi^{(r)},c^{\frac{1}{2}}\omega\phi^{(s')})^{*}$$
$$\times(\phi^{(r')},\chi^{(j)})(\chi^{(i)},\phi^{(s')})$$
$$=\sum_{r's'}(\chi^{(i)},\phi^{(s')})\delta_{r's'}(\phi^{(r')},\chi^{(j)})=(\chi^{(i)},\chi^{(j)})=\delta_{ij}.$$

Then all the $\xi^{(i)}$ are non-null distinct vectors and form a basis in $\mathcal{H}$; $c^{\frac{1}{2}}\omega$ must be a unitary operator which gives (i) and completes the proof of Theorem 2.

[6] Another criteria for a Hamiltonian mapping, that the mapping preserve multiplication properties, has been given by J. Schwinger, Proc. Natl. Acad. Sci. U. S. 46, 570 (1960).

To show that the condition of the linear induced mapping in (ii) of the preceding theorem is actually needed we will include two examples (one a one-to-one mapping, the other not) of dynamical mappings which map the pure state operators to pure state operators, but are not Hamiltonian, in the proof of the following theorem which describes some of the possible mappings.

Theorem 3. The set of possible dynamical mappings, or the set of operators A giving these mappings, forms a convex set. The set of extremal elements of this set contains those which map all pure state operators to pure state operators. These include Hamiltonian mappings and (one-to-one and non-one-to-one) mappings which induce a nonlinear mapping on $\mathfrak{K}$.

Proof. If two operators $A^{(1)}$ and $A^{(2)}$ each give a dynamical mapping of $\mathfrak{L}$, then if $0\leq a\leq 1$, $A=aA^{(1)}+(1-a)A^{(2)}$ also gives a dynamical mapping. For if ρ is any density operator in $\mathfrak{L}$, it must be mapped to a density operator $\rho^{(1)\prime}$ by $A^{(1)}$ and to a density operator $\rho^{(2)\prime}$ by $A^{(2)}$. But then A maps ρ to $\rho'=a\rho^{(1)\prime}+(1-a)\rho^{(2)\prime}$ which is also a density operator. Hence the set of possible dynamical mappings or the set of operators A giving these mappings forms a convex set.

If a mapping takes all pure state operators to pure state operators it cannot be on the line segment between two distinct mappings. For this would mean that at least one pure state operator would be on the line segment between two distinct density operators.

We have seen examples of Hamiltonian mappings. As an example of a dynamical mapping which gives a mapping of the set of all pure state operators one-to-one onto itself but induces a nonlinear mapping on $\mathfrak{K}$,[7] we consider the following: Let $\phi^{(r)}\phi^{(s)+}$ be mapped to $\phi^{(s)}\phi^{(r)+}$ for all r, s. Consider any pure state operator $\psi\psi^+$, $\psi=\sum_r(\phi^{(r)},\psi)\phi^{(r)}$. Then

$$\psi\psi^+=\sum_{rs}(\phi^{(r)},\psi)(\psi,\phi^{(s)})\phi^{(r)}\phi^{(s)+}$$

is mapped to

$$\sum_{rs}(\phi^{(r)},\psi)(\psi,\phi^{(s)})\phi^{(s)}\phi^{(r)+}=\sum_{rs}(\phi^{(s)},\psi)(\psi,\phi^{(r)})\phi^{(r)}\phi^{(s)+}=\sum_{rs}(\phi^{(r)},\psi)^*(\psi,\phi^{(s)})^*\phi^{(r)}\phi^{(s)+}=\psi'\psi'^+,$$

where $\psi'=\sum_r(\phi^{(r)},\psi)^*\phi^{(r)}$. All pure state operators are clearly mapped one-to-one to pure state operators so we do have a dynamical mapping, but the induced mapping $\psi\rightarrow\psi'$ is clearly not linear.

An example of a non-one-to-one mapping is the mapping of all pure state operators to a single pure state operator, $\phi^{(r)}\phi^{(r)+}\rightarrow\psi\psi^+$, $\phi^{(r)}\phi^{(s)+}\rightarrow 0$ for $r\neq s$. Note that this induces a nonlinear mapping on $\mathfrak{K}$. For, according to Eq. (4), $(1/\sqrt{2})(\phi^{(r)}+\phi^{(s)})\rightarrow\psi\neq\sqrt{2}\psi$. These examples complete the proof of Theorem 3.

We have seen that the set of all dynamical mappings is larger than the convex subset having the Hamiltonian mappings as its boundary. To limit ourselves to this latter subset would require an additional physical postulate. The non-one-to-one mappings of pure state operators to pure state operators could be thought of as describing a kind of measurement process but this does not exhaust the non-Hamiltonian mappings of pure state operators to pure state operators. It is interesting to note that if one describes the mappings of pure states to pure states in terms of linear mappings of the vectors in $\mathfrak{K}$ the postulate limiting these to Hamiltonian mappings is implicitly contained in the linearity. From the density operator point of view this cannot readily be interpreted as resulting from the kinematical structure of the theory.

As a final note we observe that the operator A for a dynamical mapping can have the form

$$(\phi^{(r)}\phi^{(s)+},A\phi^{(r')}\phi^{(s')+})=(\phi^{(r)},\omega\phi^{(s)})(\phi^{(r')},\sigma\phi^{(s')})$$

only if $(\phi^{(r')},\sigma\phi^{(s')})=\delta_{r's'}$ and ω is a density operator. For condition (b) of Theorem 1 requires that $\sum_r(\phi^{(r)},\omega\phi^{(r)})(\phi^{(r')},\sigma\phi^{(s')})=\delta_{r's'}$ which implies that $(\phi^{(r')},\sigma\phi^{(s')})=\delta_{r's'}$ and $\mathrm{Tr}(\omega)=1$. Conditions (a) and (c) then require that ω be Hermitian and positive definite.

[7] Such nonlinear mappings are used for representing antilinear discrete operations in quantum mechanics. The most familiar example is time inversion; see E. P. Wigner, Gottinger Nachr., 546, (1932); J. Math. Phys. **1**, 409 (1960); R. G. Sachs, *Nuclear Theory* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1953), Appendix; another example is charge conjugation in one-particle theories, see e.g., L. L. Foldy, Phys. Rev. **102**, 568 (1956).

A Note on k-Commutative Matrices*

D. W. ROBINSON
Brigham Young University, Provo, Utah

(Received April 3, 1961)

Let A and B be square matrices over a field in which the minimum polynomial of A is completely reducible. It is shown that A is k commutative with respect to B for some non-negative integer k if and only if B commutes with every principal idempotent of A. The proof is brief, simplifying much of the previous study of k-commutative matrices. The result is also used to generalize some well-known theorems on finite matrix commutators that involve a complex matrix and its transposed complex conjugate.

INTRODUCTION

THE study of matrix commutators of higher order has received attention from several authors.[1] In particular, W. E. Roth[2] considered what he called k-commutative matrices. The main purpose of this note is to prove briefly a useful characterization of these matrices.

Let A and B be n by n matrices over a field F in which the minimum polynomial $\prod_\alpha(x-\alpha)^{s_\alpha}$ of A is completely reducible. If $[A,B]=AB-BA$ denotes the commutator of A and B, then $[_{(k)}A,B]$ is defined recursively by

$$[_{(0)}A,B]=B \quad \text{and} \quad [_{(k)}A,B]=[A,[_{(k-1)}A,B]]$$

for $k>0$.[3] A is said to be k commutative with respect to B if and only if $[_{(k)}A,B]=0$, and $[_{(j)}A,B]=0$ implies $j\geqq k$. Clearly, A is k commutative with respect to B for at most one non-negative integer k.

THEOREM

A is k commutative with respect to B for some non-negative integer k if and only if B commutes with every principal idempotent[4] of A.

Preliminary to the proof of this theorem, the following lemma is demonstrated.

LEMMA

If E_α is a principal idempotent of A and $[A,B]$ commutes with E_α, then B commutes with E_α.

To prove this, let $E_\alpha'=I-E_\alpha$. Since E_α commutes with A,

$$[A,E_\alpha BE_\alpha']=E_\alpha[A,B]E_\alpha'.$$

Thus, since $E_\alpha E_\alpha'=0$, under the hypothesis of the lemma, A commutes with $E_\alpha BE_\alpha'$. But since E_α is a polynomial in A, E_α also commutes with $E_\alpha BE_\alpha'$. Hence, $E_\alpha BE_\alpha'=0$. By a similar argument, $E_\alpha' BE_\alpha=0$. Finally, $E_\alpha B=E_\alpha BE_\alpha=BE_\alpha$.

The proof of the theorem is now given.

NECESSITY

The result is obvious in case $k=0$. Thus, suppose that A is k commutative with respect to B for some $k>0$, and let E_α be any principal idempotent of A. Since A commutes with $[_{(k-1)}A,B]$, and E_α is a polynomial in A, then E_α commutes with $[_{(k-1)}A,B]$. Hence, by repeated use of the lemma above, E_α commutes with $[_{(j)}A,B]$, $j=k-1, k-2, \cdots, 0$. In particular, E_α commutes with B.[5]

SUFFICIENCY

Let $A_\alpha=(A-\alpha I)E_\alpha$ for α any characteristic value of A with associated principal idempotent E_α. Thus, if B commutes with every principal idempotent of A, it follows by induction on $k\geqq 1$ that

$$[_{(k)}A,B]=\sum_\alpha \sum_{j=0}^{k} (-1)^j \binom{k}{j} A_\alpha^{k-j} B A_\alpha^{j},$$

where the first sum is taken over all of the distinct characteristic values of A, and where the usual notation for the binomial coefficient is used. Since A_α is nilpotent, by choosing k sufficiently large, the sum on the right is zero, and the desired conclusion is obtained.

Moreover, since A_α is nilpotent of order equal to the index s_α of α, the following result due to Roth[6] is also a consequence of the preceding equation.

COROLLARY 1

Let m be the largest of the indices associated with the distinct characteristic values of A. If A is k commutative with respect to B, then $k<2m$.

Furthermore, for any scalar polynomial $\phi(x)$, since the index of the characteristic value $\phi(\alpha)$ of $\phi(A)$ is at most m, and the principal idempotent of $\phi(A)$ associated with $\phi(\alpha)$ is the sum $\sum E_\beta$ over all of the distinct characteristic values β of A such that $\phi(\beta)=\phi(\alpha)$, the following result is also immediate.

---

* Presented to the American Mathematical Society, April 22, 1961.

[1] For a survey of these results see O. Taussky, Am. Math. Monthly **64**, 229 (1957).

[2] W. E. Roth, Trans. Am. Math. Soc. **39**, 483 (1936).

[3] See also, M. Marcus and N. A. Khan, J. Research Natl. Bur. Standards **64B**, 51 (1960), and M. F. Smiley, Am. Math. Soc. Notices **7**, 927 (1960).

[4] For the definition and properties of the principal idempotents of a matrix, see, for example, N. Jacobson, *Lectures in Abstract Algebra* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1953), Vol. II, pp. 130–132.

[5] See also, W. E. Roth, Trans. Am. Math. Soc. **39**, 483 (1936), Theorem 9.

[6] Reference 5, Theorem 5.

COROLLARY 2

Let m be defined as in Corollary 1, and let $\phi(x)$ and $\theta(x)$ be polynomials over F. If A is k commutative with respect to B for some k, then $\phi(A)$ is j commutative with respect to $\theta(B)$ for some $j<2m$.

Roth[7] considered only the case $\theta(x)=x$, and showed under this condition that $j\leqq k$. However, this stronger inequality is not in general valid for every polynomial $\theta(x)$.

As an application of the preceding results, some remarks are now given concerning commutators that involve a complex matrix A and its transposed complex conjugate A^*.

First, as is well known, the principal idempotents of a normal matrix are Hermitian. More generally, the following is now demonstrated.

COROLLARY 3

Any complex matrix A is k commutative with respect to A^* for some non-negative integer k if and only if the principal idempotents of A are Hermitian.

To prove this, it is first observed that the principal idempotents of A^* are the transposed complex conjugates of the principal idempotents E_α of A. Thus, since $E_\alpha{}^*$ commutes with A^*, if $E_\alpha=E_\alpha{}^*$, then by the theorem above A is k commutative with respect to A^* for some non-negative integer k. Conversely, if E_α commutes with A^*, then it also commutes with $E_\alpha{}^*$. But any normal idempotent matrix is Hermitian.

Finally, as an application of Corollary 1 above, a well-known theorem is generalized.

COROLLARY 4[8]

The commutator $[A,A^*]$ is k commutative with respect to A for some non-negative integer k if and only if A is normal.

It is only necessary to prove that $C=[A,A^*]$ and $[_{(k)}C,A]=0$, for some positive integer k, implies $C=0$. But, since C is diagonable, applying Corollary 1 with $m=1$, $[C,A]=0$. Thus, by a theorem of Jacobson,[9] C is nilpotent. But any diagonable nilpotent matrix is necessarily zero.

[7] Reference 5, Theorem 3.

[8] For a proof of this corollary, in case either $k=1$ or $k=2$, see also T. Kato and O. Taussky, J. Wash. Acad. Sci. **46**, 38 (1956).

[9] N. Jacobson, Ann. Math. **36**, 877 (1935).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER–DECEMBER, 1961

New Possibilities for a Unified Field Theory

W. ISRAEL AND R. TROLLOPE
Mathematics Department, University of Alberta, Edmonton, Alberta, Canada

(Received December 8, 1960)

The introduction of nonsymmetric g_{ik} in unified field theories of the Einstein-Schrödinger type is open to the objection, on group-theoretical grounds, that the symmetric and antisymmetric parts transform independently. This objection does not apply to the use of nonsymmetric $\Gamma_{ik}{}^{\mu}$, since these quantities are irreducible under the "extended group," consisting of the point transformations and the Einstein λ transformations.

We consider a theory based on symmetric g_{ik} and nonsymmetric $\Gamma_{ik}{}^{\mu}$. The Lagrangian L is assumed to depend only on g_{ik} and the contracted curvature tensor R_{ik} (this insures the λ invariance and transposition invariance of the theory). For simplicity, we suppose further that L involves R_{ik} rationally and, at most, quadratically.

The resulting theory is able to account satisfactorily for the main feature of gravitation, electromagnetism, and their interaction. In particular, the theory yields the correct equations of motion for charged masses. The electromagnetic tensor is associated with the skew part of R_{ik}, and the λ transformations correspond roughly to the gauge transformations of electrodynamics.

1. INTRODUCTION

AN important feature of the unified field theory of Einstein and Schrödinger is the property of "λ invariance," i.e., invariance of the field equations under the group of transformations

$$'\Gamma_{ik}{}^{\mu}=\Gamma_{ik}{}^{\mu}+\delta_i{}^{\mu}\frac{\partial\lambda}{\partial x^k},\quad 'g_{ik}=g_{ik},\tag{1.1}$$

where the function $\lambda(x)$ is arbitrary. This invariance property depends essentially on the invariance of the curvature tensor

$$R_{.ijk}{}^{h}\equiv\Gamma_{ik,j}{}^{h}-\Gamma_{ij,k}{}^{h}+\Gamma_{\alpha j}{}^{h}\Gamma_{ik}{}^{\alpha}-\Gamma_{\alpha k}{}^{h}\Gamma_{ij}{}^{\alpha}$$

and its contractions under the transformations (1.1). Quite generally, let us consider theories whose field equations are derived from a variational principle

$$\delta\int\mathfrak{L}(g,\Gamma)d^4x=0\tag{1.2}$$

by independently varying the g_{ik} and the $\Gamma_{ik}{}^{\mu}$ (Palatini

COROLLARY 2

Let m be defined as in Corollary 1, and let $\phi(x)$ and $\theta(x)$ be polynomials over F. If A is k commutative with respect to B for some k, then $\phi(A)$ is j commutative with respect to $\theta(B)$ for some $j<2m$.

Roth[7] considered only the case $\theta(x)=x$, and showed under this condition that $j\leqq k$. However, this stronger inequality is not in general valid for every polynomial $\theta(x)$.

As an application of the preceding results, some remarks are now given concerning commutators that involve a complex matrix A and its transposed complex conjugate A^*.

First, as is well known, the principal idempotents of a normal matrix are Hermitian. More generally, the following is now demonstrated.

COROLLARY 3

Any complex matrix A is k commutative with respect to A^* for some non-negative integer k if and only if the principal idempotents of A are Hermitian.

To prove this, it is first observed that the principal idempotents of A^* are the transposed complex conjugates of the principal idempotents E_α of A. Thus, since $E_\alpha{}^*$ commutes with A^*, if $E_\alpha=E_\alpha{}^*$, then by the theorem above A is k commutative with respect to A^* for some non-negative integer k. Conversely, if E_α commutes with A^*, then it also commutes with $E_\alpha{}^*$. But any normal idempotent matrix is Hermitian.

Finally, as an application of Corollary 1 above, a well-known theorem is generalized.

COROLLARY 4[8]

The commutator $[A,A^*]$ is k commutative with respect to A for some non-negative integer k if and only if A is normal.

It is only necessary to prove that $C=[A,A^*]$ and $[_{(k)}C,A]=0$, for some positive integer k, implies $C=0$. But, since C is diagonable, applying Corollary 1 with $m=1$, $[C,A]=0$. Thus, by a theorem of Jacobson,[9] C is nilpotent. But any diagonable nilpotent matrix is necessarily zero.

[7] Reference 5, Theorem 3.

[8] For a proof of this corollary, in case either $k=1$ or $k=2$, see also T. Kato and O. Taussky, J. Wash. Acad. Sci. **46**, 38 (1956).

[9] N. Jacobson, Ann. Math. **36**, 877 (1935).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER–DECEMBER, 1961

New Possibilities for a Unified Field Theory

W. ISRAEL AND R. TROLLOPE
Mathematics Department, University of Alberta, Edmonton, Alberta, Canada

(Received December 8, 1960)

The introduction of nonsymmetric g_{ik} in unified field theories of the Einstein-Schrödinger type is open to the objection, on group-theoretical grounds, that the symmetric and antisymmetric parts transform independently. This objection does not apply to the use of nonsymmetric $\Gamma_{ik}{}^{\mu}$, since these quantities are irreducible under the "extended group," consisting of the point transformations and the Einstein λ transformations.

We consider a theory based on symmetric g_{ik} and nonsymmetric $\Gamma_{ik}{}^{\mu}$. The Lagrangian L is assumed to depend only on g_{ik} and the contracted curvature tensor R_{ik} (this insures the λ invariance and transposition invariance of the theory). For simplicity, we suppose further that L involves R_{ik} rationally and, at most, quadratically.

The resulting theory is able to account satisfactorily for the main feature of gravitation, electromagnetism, and their interaction. In particular, the theory yields the correct equations of motion for charged masses. The electromagnetic tensor is associated with the skew part of R_{ik}, and the λ transformations correspond roughly to the gauge transformations of electrodynamics.

1. INTRODUCTION

AN important feature of the unified field theory of Einstein and Schrödinger is the property of "λ invariance," i.e., invariance of the field equations under the group of transformations

$$'\Gamma_{ik}{}^{\mu}=\Gamma_{ik}{}^{\mu}+\delta_i{}^{\mu}\frac{\partial\lambda}{\partial x^k},\quad 'g_{ik}=g_{ik},\tag{1.1}$$

where the function $\lambda(x)$ is arbitrary. This invariance property depends essentially on the invariance of the curvature tensor

$$R_{.ijk}{}^{h}\equiv\Gamma_{ik,j}{}^{h}-\Gamma_{ij,k}{}^{h}+\Gamma_{\alpha j}{}^{h}\Gamma_{ik}{}^{\alpha}-\Gamma_{\alpha k}{}^{h}\Gamma_{ij}{}^{\alpha}$$

and its contractions under the transformations (1.1). Quite generally, let us consider theories whose field equations are derived from a variational principle

$$\delta\int\mathfrak{L}(g,\Gamma)d^4x=0\tag{1.2}$$

by independently varying the g_{ik} and the $\Gamma_{ik}{}^{\mu}$ (Palatini

method of variation). It is then evident that *all theories whose Lagrangian densities $\mathfrak{L}$ are built up solely from g_{ik} and $R_{.ijk}^{h}$ will be characterized by the property of* λ *invariance.* Physically, it seems plausible that λ invariance is in some way related to the gauge invariance of electrodynamics.

The basic assumption in the formulation of Einstein's theory is that the symmetric g_{ik} and Γ_{ik}^{μ} of general relativity are to be replaced by nonsymmetric quantities. What *a priori* justification can be given for this assumption? To deal first with the affinity, a convincing argument in favor of a nonsymmetric extension of Γ_{ik}^{μ} is the fact that this extension is accompanied, for a theory of the type mentioned above, by an enlargement of the invariance group through the inclusion of the λ transformations (1.1). (This situation has an analog in the transition from special to general relativity, where the replacement of the scalar gravitational potential ϕ by the ten potentials g_{ik} is motivated by the general covariance of the new theory, a scalar theory being essentially only Lorentz-covariant in the sense that Lorentz frames have a privileged status.)

On the other hand, no enlargement of the invariance group results when we subjoin an antisymmetric part to g_{ik}. In particular, the field equations admit no transformations which "mix" the symmetric and antisymmetric parts $\underset{\frown}{g}_{ik}$ and $\underset{\smile}{g}_{ik}$ [as the transformations (1.1) do with $\underset{\frown}{\Gamma}_{ik}^{\mu}$ and $\underset{\smile}{\Gamma}_{ik}^{\mu}$]. The two parts transform as separate and independent entities, and no real unification is achieved by combining them. *If, then, group-theoretical considerations are accepted as a basic guiding principle in the construction of a unified theory, it will be logically most economical and satisfactory to retain the symmetry of the fundamental tensor g_{ik}, while admitting nonsymmetric Γ_{ik}^{μ}.*

A further disadvantage of introducing a nonsymmetric g_{ik} into the Lagrangian lies in the excessive freedom (from the physical point of view) which is thereby afforded to the solutions of the resulting field equations. Thus, Einstein's theory admits nontrivial solutions even for "flat" space-time ($R_{.ijk}^{h}=0$), as has been shown in detail by Wyman and Zassenhaus.[1] This must be regarded as an unsatisfactory trait of a theory which purports to give a geometrical description of physical fields. It is probable that almost any field theory which operates with a non-symmetric g_{ik} would be open to this objection. The difficulty is connected with the fact that the flatness condition does not necessarily imply $\underset{\smile}{g}_{ik}=0$. We shall revert to this point later on (Sec. 7).

In the present paper we shall investigate the possibility of constructing a unified field theory of gravitation and electromagnetism based on the following postulates:

(1) *The fundamental quantities are a symmetric tensor g_{ik} and a nonsymmetric affine connection Γ_{ik}^{μ}.*

(2) *The field equations are derivable from a variational principle*

$$\delta\int\sqrt{-g}\,L d^4x=0, \tag{1.2}$$

where (to insure λ invariance) L is assumed to depend only on g_{ik} and $R_{.ijk}^{h}$.

The analysis is simplified if we assume further that $R_{.ijk}^{h}$ enters rationally and at most quadratically into L, and only in the form of its *contraction* $R_{ij}\equiv R_{.ija}^{\alpha}$. The most general form for L is then

$$L=aR+\frac{b}{2}R^2+\frac{\alpha}{2}\underset{\smile}{R}_{ik}\underset{\smile}{R}^{ik}+\frac{\beta}{2}\underset{\frown}{R}_{ik}\underset{\frown}{R}^{ik}. \tag{1.3}$$

Here, a, b, α, β are constants,

$$R\equiv g^{\alpha\beta}R_{\alpha\beta},$$

and g_{ik} and its inverse g^{ik} are used to lower and raise indices in the familiar way, e.g.,

$$R^{ik}=g^{i\alpha}g^{k\beta}R_{\alpha\beta}.$$

In writing down (1.3), we have been greatly assisted by the postulated symmetry of g_{ik}, which restricts enormously the number of invariants which can be formed out of g_{ik} and R_{ik}.

It will be shown in the following sections that the present theory is able to give a satisfactory account of Maxwell-Einstein fields. In Sec. 2 we derive the field equations and Bianchi identities from the variational principle (1.2). The weak-field solutions of the theory are then compared (Sec. 4) with the equations of Maxwell and the Einstein gravitational equations. (Concerning the physical interpretation of the theory, we may mention here in anticipation that the electromagnetic tensor is taken to be proportional to $\underset{\smile}{R}_{ik}$. The tensor $\underset{\smile}{\Gamma}_{i\alpha}^{\alpha}$ then corresponds roughly to the 4-potential, and the λ transformations to the gauge transformations.)

Section 5 is devoted to a brief discussion of the equations of motion of test particles. Finally, in Sec. 6 we obtain rigorous, spherically symmetric solutions for two particular cases ($b=\beta=0$ and $a=\beta=0$, respectively). These solutions recall in some respects the nonlinear electrodynamics of Born and Infeld.[2] Some concluding remarks of a general nature will be found in Sec. 8.

We close this introduction with a comment on the principle of *transposition invariance*, which has played an important role in the development of Einstein's unified field theory. In his later presentations of the theory, Einstein[3] formulated this principle as follows. Introduce the quantities U_{ik}^{μ}, defined by

$$U_{ik}^{\mu}=\Gamma_{ik}^{\mu}-\Gamma_{i\alpha}^{\alpha}\delta_k^{\mu}.$$

[1] M. Wyman and H. Zazzenhaus, Phys. Rev. **110**, 228 (1958).

[2] M. Born, Proc. Roy. Soc. (London) **A143**, 410 (1934); M. Born and L. Infeld, *ibid.* **A144**, 425 (1934).

[3] A. Einstein and B. Kaufman, Ann. Math. Princeton, **62**, 128 (1955). See also B. Kaufman in *Fünfzig Jahre Relativitätstheorie* (Birkhäuser Verlag, Basle, 1956), p. 227.

A tensor $A_{ik}(U)$ is called "transposition symmetric" if it goes over into its transpose under the transformation

$$U_{ik}{}^{\mu} \rightarrow \tilde{U}_{ik}{}^{\mu} \equiv U_{ki}{}^{\mu}. \tag{1.4}$$

The principle of transposition invariance requires that the field equations stay valid under the transformation (1.4). The field equations of the present theory are transposition invariant in this sense, since the Lagrangian (1.3) is built entirely out of transposition symmetric quantities. One sees this immediately on examining the expression for R_{ik} in terms of U:

$$R_{ik} = U_{ik,\alpha}{}^{\alpha} - U_{i\beta}{}^{\alpha} U_{\alpha k}{}^{\beta} + \tfrac{1}{3} U_{i\alpha}{}^{\alpha} U_{\beta k}{}^{\beta}.$$

If we had permitted our Lagrangian to contain the curvature tensor in forms other than its contraction R_{ik}, the transposition invariance of the theory would have been forfeited. For instance,

$$S_{jk} = R_{.\alpha jk}{}^{\alpha} = (U_{\alpha k,j}{}^{\alpha} - U_{\alpha j,k}{}^{\alpha}) - \tfrac{1}{3}(U_{k\alpha,j}{}^{\alpha} - U_{j\alpha,k}{}^{\alpha})$$

is not transposition symmetric.

2. THE FIELD EQUATIONS AND BIANCHI IDENTITIES

Our notation will, in general, follow that of Einstein,[4] but with one simplification, affecting the notation for covariant derivatives. In this paper, we shall be concerned almost entirely with covariant derivatives of tensors and tensor-densities of rank *two*, and in all these cases it will be only the "mixed, $+\ -$" covariant derivative which turns up. Hence, no confusion should result if we suppress the $+,-$ subscripts. If A_{ik} is a covariant tensor, $\mathfrak{A}^{ik}$ a contravariant tensor-density, we shall write

$$A_{ik;\mu} \equiv A_{\underset{+}{i}\ \underset{-}{k};\mu} = A_{ik,\mu} - A_{\alpha k}\Gamma_{i\mu}{}^{\alpha} - A_{i\alpha}\Gamma_{\mu k}{}^{\alpha}$$

$$\mathfrak{A}_{;\mu}{}^{ik} \equiv \mathfrak{A}_{;\mu}{}^{\underset{+}{i}\ \underset{-}{k}} = \mathfrak{A}_{,\mu}{}^{ik} + \mathfrak{A}^{\alpha k}\Gamma_{\alpha\mu}{}^{i} + \mathfrak{A}^{i\alpha}\Gamma_{\mu\alpha}{}^{k} - \mathfrak{A}^{ik}\Gamma_{\underset{\sim}{\mu\alpha}}{}^{\alpha}$$

the comma denoting partial differentiation. It will also be convenient to define $M_{ik}(A)$, the "Maxwellian" of the tensor $A_{\mu\nu}$, by

$$M_{ik}(A) = \tfrac{1}{4} g_{ik} A_{\alpha\beta} A^{\alpha\beta} - \tfrac{1}{2} g^{\alpha\beta}(A_{i\alpha}A_{k\beta} + A_{\alpha i}A_{\beta k}). \tag{2.1}$$

Observe that M_{ik} is a symmetric tensor with vanishing trace.

Our variational principle is

$$\delta \int \sqrt{-g}\, L d^4x = 0 \tag{2.2}$$

for arbitrary, independent variations δg^{ik}, $\delta\Gamma_{ik}{}^{\mu}$ which vanish on the boundary of integration. From the expression (1.3) for L, it is readily verified that

$$\delta(\sqrt{-g}\, L) = \mathfrak{F}^{ik}\delta R_{ik} + \sqrt{-g}\, W_{ik}\delta g^{ik}, \tag{2.3}$$

where

$$\mathfrak{F}^{ik} = \sqrt{-g}\,\{(a+bR)g^{ik} + \alpha \underset{\sim}{R}{}^{ik} + \beta \underline{R}{}^{ik}\}, \tag{2.4}$$

$$W_{ik} = a(\underline{R}_{ik} - \tfrac{1}{2}Rg_{ik}) + bR(\underline{R}_{ik} - \tfrac{1}{4}Rg_{ik}) - \alpha M_{ik}(R_{\smile}) - \beta M_{ik}(R_{\frown}), \tag{2.5}$$

and $M_{ik}(R_{\smile})$, $M_{ik}(R_{\frown})$ denote the Maxwellians of $R_{\underset{\sim}{\mu\nu}}$, $R_{\underline{\mu\nu}}$, respectively. The tensor W_{ik} has been symmetrized, since δg^{ik} is symmetric.

Noting the Palatini relation

$$\delta R_{ik} = (\delta\Gamma_{\underset{+}{i}\underset{\sim}{\mu}}{}^{\mu})_{;k} - (\delta\Gamma_{\underset{+}{i}\ \underset{-}{k}}{}^{\mu})_{;\mu},$$

we find, after some manipulation, that the first term of (2.3) can be written

$$\mathfrak{F}^{ik}\delta R_{ik} = \mathfrak{N}_{\mu}{}^{ik}\delta\Gamma_{ik}{}^{\mu} + (\mathfrak{F}^{\alpha\mu}\delta\Gamma_{\alpha\beta}{}^{\beta} - \mathfrak{F}^{\alpha\beta}\delta\Gamma_{\alpha\beta}{}^{\mu})_{,\mu}, \tag{2.6}$$

where

$$\mathfrak{N}_{\mu}{}^{ik} = \mathfrak{F}_{;\mu}{}^{ik} - \mathfrak{F}_{;\alpha}{}^{i\alpha}\delta_{\mu}{}^{k} - \mathfrak{F}^{ik}\Gamma_{\mu} - \mathfrak{F}^{i\alpha}\Gamma_{\alpha}\delta_{\mu}{}^{k}, \tag{2.7}$$

$$\Gamma_i \equiv \Gamma_{\underset{\sim}{i\alpha}}{}^{\alpha}.$$

The term in parentheses on the right of (2.6) is a divergence, whose integral will vanish if $\delta\Gamma_{ik}{}^{\mu} = 0$ on the boundary. Thus, (2.2), (2.3), and (2.6) lead to the *field equations*

$$\mathfrak{N}_{\mu}{}^{ik} = 0, \tag{2.8}$$

$$W_{ik} = 0. \tag{2.9}$$

The customary way of obtaining the *Bianchi identities* is to make use of the invariance of the action integral with respect to infinitesimal coordinate transformations. We shall find it slightly more convenient to consider the invariance of

$$I \equiv \int \mathfrak{F}^{ik} R_{ik} d^4x = \int \sqrt{-g}\,(2L - aR) d^4x.$$

Under the infinitesimal transformation

$$x^{i'} = x^i + \epsilon\xi^i(x),$$

the "substantial variation" $\delta\mathfrak{F}^{ik} \equiv \mathfrak{F}^{ik'}(x) - \mathfrak{F}^{ik}(x)$ is given by

$$\delta\mathfrak{F}^{ik} = \epsilon(\mathfrak{F}^{i\alpha}\xi_{,\alpha}{}^{k} + \mathfrak{F}^{\alpha k}\xi_{,\alpha}{}^{i} - \mathfrak{F}^{ik}\xi_{,\alpha}{}^{\alpha} - \mathfrak{F}_{,\alpha}{}^{ik}\xi^{\alpha}). \tag{2.10}$$

For the variation of I, we obtain, assuming $\xi^i = 0$ on the boundary of integration,

$$I' - I = 0 = \int (\mathfrak{F}^{ik}\delta R_{ik} + R_{ik}\delta\mathfrak{F}^{ik}) d^4x$$

$$= \int (\mathfrak{N}_{\mu}{}^{ik}\delta\Gamma_{ik}{}^{\mu} + R_{ik}\delta\mathfrak{F}^{ik}) d^4x \tag{2.11}$$

by (2.6). *Let us now assume that one half of the field*

[4] A. Einstein, *The Meaning of Relativity* (Princeton University Press, Princeton, New Jersey, 1953), Appendix II, 4th ed.

equations, viz. (2.8), is satisfied. Then the first term on the right of (2.11) disappears, and the second term yields, using (2.10) and after integration by parts,

$$0=\int\{-(\mathfrak{g}^{\alpha\beta}R_{\alpha k})_{,\beta}-(\mathfrak{g}^{\alpha\beta}R_{k\beta})_{,\alpha}+(\mathfrak{g}^{\alpha\beta}R_{\alpha\beta})_{,k}-\mathfrak{g}_{,k}{}^{\alpha\beta}R_{\alpha\beta}\}\xi^k d^4x. \quad (2.12)$$

It can be verified from (2.7) that

$$\mathfrak{N}_k{}^{ki}=-2\mathfrak{g}_{,k}{}^{\underset{\smile}{ik}},$$

so that

$$\mathfrak{g}_{,k}{}^{\underset{\smile}{ik}}=0 \quad (2.13)$$

is a consequence of (2.8). Using this result, we find, after some rearrangement, that (2.12) leads to

$$(\mathfrak{g}^{\underline{\alpha\beta}}R_{\underline{\alpha k}})_{,\beta}-\tfrac{1}{2}\mathfrak{g}^{\underline{\alpha\beta}}R_{\underline{\alpha\beta},k}=\tfrac{1}{2}\mathfrak{g}^{\underset{\smile}{\alpha\beta}}R_{[\underset{\smile}{\alpha\beta},k]}. \quad (2.14)$$

These are the Bianchi identities of our theory modulo the field equations (2.8). The square brackets indicate the cyclic divergence

$$F_{[\underline{ik},l]}\equiv F_{\underline{ik},l}+F_{\underline{li},k}+F_{\underline{kl},i}.$$

Note that (2.13) can be written in the equivalent form

$$(\sqrt{-g}\,R^{\underset{\smile}{ik}})_{,k}=0, \quad (2.15)$$

which we shall later identify with one half of Maxwell's equations (Sec. 4).

3. DISCUSSION OF THE FIELD EQUATIONS

The field equations (2.8) may be formally simplified, following Schrödinger, by the introduction of a new affine connection $^*\Gamma_{ik}{}^{\mu}$, defined by

$$^*\Gamma_{ik}{}^{\mu}=\Gamma_{ik}{}^{\mu}+\tfrac{2}{3}\delta_i{}^{\mu}\Gamma_k, \quad (3.1)$$

we then have

$$^*\Gamma_i=0, \quad (3.2)$$

$$\mathfrak{g}_{;\mu^*}{}^{ik}=\mathfrak{g}_{;\mu}{}^{ik}-\mathfrak{g}^{ik}\Gamma_\mu+\tfrac{2}{3}\mathfrak{g}^{i\alpha}\Gamma_\alpha\delta_\mu{}^k, \quad (3.3)$$

where $;\cdots^*$ indicates covariant differentiation with respect to $^*\Gamma_{ik}{}^{\mu}$. With the aid of (3.2), (3.3) and (2.7), it is easy to show that

$$\mathfrak{N}_\mu{}^{ik}-\tfrac{1}{3}\mathfrak{N}_\alpha{}^{i\alpha}\delta_\mu{}^k=\mathfrak{g}_{;\mu^*}{}^{ik}.$$

Therefore the equations

$$\mathfrak{g}_{;\mu^*}{}^{ik}=0, \quad ^*\Gamma_i=0 \quad (3.4)$$

are equivalent to (2.8). We shall make use of this equivalence in the following sections.

Turning now to Eqs. (2.9), we shall consider two cases, corresponding to $a\neq 0$ and $a=0$.

Case I: $a\neq 0$. From (2.5) we obtain $g^{ik}W_{ik}=-aR$. Hence (2.9) yields

$$R=0 \quad (a\neq 0). \quad (3.5)$$

There is no loss of generality in taking $a=1$. Equation (2.9) now simplifies to

$$R_{\underline{ik}}=\alpha M_{ik}(R_{\smile})+\beta M_{ik}(R_{\frown}). \quad (3.6)$$

If we regard $R_{\underset{\smile}{\mu\nu}}$ as proportional to the electromagnetic field, then the first term on the right-hand side represents the electromagnetic energy tensor.[5] At first sight, it is disconcerting to find also the Maxwellian of a *symmetric* tensor on the right-hand side. However, as we shall now see, there is always one solution of (3.6) for which this additional term vanishes identically.

According to a remarkable algebraic theorem due to Rainich,[6] *a set of necessary and sufficient conditions that a symmetric tensor T_{ik} be expressible as the Maxwellian of an anti-symmetric tensor is*

$$g^{\alpha\beta}T_{\alpha\beta}=0, \quad M_{ik}(T)=0. \quad (3.7)$$

It follows that *all solutions of the equation*

$$R_{\underline{ik}}=\alpha M_{ik}(R_{\smile}) \quad (3.8)$$

satisfy (3.6). The condition

$$M_{ik}(R_{\frown})=0$$

which makes (3.6) deducible from (3.8), is satisfied as a consequence of (3.5) and (3.8).

Equations (3.6) may be thought of as a set of ten quadratic equations for the ten quantities $R_{\underline{ik}}$. There are twenty solutions for $R_{\underline{ik}}$, one of which is (3.8). Of the remaining solutions, some may be complex, and others will lead to physically unacceptable results, such as negative energy densities, improper behavior at spatial infinity, etc. Pending a detailed examination of the additional solutions, it appears best to exclude them from the present considerations by setting $\beta=0$ in the field equations.

Collecting our results, we have, as our field equations for the case $a\neq 0$, $\beta=0$:

$$\left.\begin{aligned}&\mathfrak{g}_{;\mu^*}{}^{ik}=0, \quad ^*\Gamma_i=0, && (3.9)\\ &\mathfrak{g}_{,k}{}^{\underset{\smile}{ik}}=0, && (3.10)\\ &R_{\underline{ik}}=\alpha M_{ik}(R_{\smile}), && (3.11)\\ &R=0. && (3.12)\end{aligned}\right\}\text{I}$$

Here,

$$\mathfrak{g}^{ik}=\sqrt{-g}\,(g^{ik}+\alpha R^{\underset{\smile}{ik}}), \quad (3.13)$$

and $^*\Gamma_{ik}{}^{\mu}$ is defined by (3.1).

Case II: $a=0$. In this case we can no longer deduce $R=0$ from (2.9). In fact the field equations leave R

[5] In order that the interaction between gravitation and electromagnetism should have the correct sign, it is necessary that the constant α in (3.8) be *negative.*

[6] G. Y. Rainich: Trans. Am. Math. Soc. **27**, 106 (1925). See also C. Misner and J. A. Wheeler: Ann. Phys. **2**, 525 (1957).

completely undetermined. By inspection of (2.4), (2.5) it is seen immediately that if $a=0$, the field equations (2.8), (2.9) are invariant under the transformation

$$g_{ik} \rightarrow \mu g_{ik}, \quad \Gamma_{ik}{}^{s} \rightarrow \Gamma_{ik}{}^{s},$$

implying

$$R \rightarrow (1/\mu)R,$$

with μ an arbitrary function. Hence the field equations fix only the combination Rg_{ik}, not R and g_{ik} separately.

The equations (2.9) may be written

$$bR(R_{\underline{ik}}-\tfrac{1}{4}Rg_{ik})=\alpha M_{ik}(R_{\smile})+\beta M_{ik}(R_{\frown}). \quad (3.14)$$

Let

$$H_{ik}\equiv R_{\underline{ik}}-\tfrac{1}{4}Rg_{ik}. \quad (3.15)$$

Then

$$g^{ik}H_{ik}=0, \quad (3.16)$$

and

$$M_{ik}(H)=M_{ik}(R_{\frown})+\tfrac{1}{2}RH_{ik}. \quad (3.17)$$

All solutions of

$$\left(b+\frac{\beta}{2}\right)RH_{ik}=\alpha M_{ik}(R_{\smile}) \quad (3.18)$$

are solutions of (3.14). For from (3.16) and (3.18), we have, according to Rainich's theorem,

$$M_{ik}(H)=0.$$

Using (3.17), we then find that (3.14) is a consequence of (3.18). This proves our statement.

As before, (3.14) admits other solutions besides (3.18), which we can exclude by setting $\beta=0$. Doing this, we obtain as our field equations for $a=0$,

$$\left.\begin{aligned} &\mathfrak{g}_{;\mu^*}{}^{ik}=0, \quad {}^*\Gamma_i=0, && (3.19)\\ &\mathfrak{g}_{,k}{}^{\underset{\smile}{ik}}=0, && (3.20)\\ &bR(R_{\underline{ik}}-\tfrac{1}{4}Rg_{ik})=\alpha M_{ik}(R_{\smile}), && (3.21)\\ &\mathfrak{g}^{ik}=\sqrt{-g}\,(bRg^{ik}+\alpha R^{\underset{\smile}{ik}}). && (3.22) \end{aligned}\right\} \text{II}$$

This system of equations has to be supplemented by an additional equation determining R. The simplest condition to impose is that R is a (nonvanishing) constant, and we may as well write, with no further loss of generality,

$$bR=1. \quad (3.23)$$

Then (3.21) and (3.22) may be written

$$R_{\underline{ik}}-(1/4b)g_{ik}=\alpha M_{ik}(R_{\smile}) \quad (3.21a)$$

$$\mathfrak{g}^{ik}=\sqrt{-g}\,(g^{ik}+\alpha R^{\underset{\smile}{ik}}). \quad (3.22a)$$

These differ from the corresponding equations I only by the presence of the "cosmological term" $(1/4b)g_{ik}$.

4. APPROXIMATE SOLUTIONS FOR WEAK FIELDS AND PHYSICAL INTERPRETATION OF THE THEORY

It is convenient to define ${}^{\dagger}s_{ik}$ as the covariant tensor "inverse" to the contravariant tensor density $\mathfrak{g}^{ik}$, so that we have the relations

$${}^{\dagger}s_{i\alpha}\mathfrak{g}^{k\alpha}=\delta_i{}^{k}(-\det\mathfrak{g}^{\mu\nu})^{\frac{1}{2}}, \quad \det{}^{\dagger}s_{ik}=\det\mathfrak{g}^{ik}. \quad (4.1)$$

The equations $\mathfrak{g}_{;\mu^*}{}^{ik}=0$ are then equivalent to ${}^{\dagger}s_{ik;\mu^*}=0$.

Let us consider the field equations I [Eqs. (3.9) to (3.12)]. We shall examine the linearized form of these equations for weak gravitational and electromagnetic fields. Specifically, we shall assume

$${}^{\dagger}s_{ik}=\eta_{ik}+\gamma_{ik}, \quad (4.2)$$

where η_{ik} is the Minkowski tensor:

$$\eta_{11}=\eta_{22}=\eta_{33}=-1, \quad \eta_{44}=1, \quad \eta_{ik}=0 \quad (i\neq k),$$

and the γ_{ik} are small quantities whose squares and products will be neglected.

By (4.1), we then obtain

$$\mathfrak{g}^{ik}=\eta^{ik}+\gamma_{\mu\nu}(\tfrac{1}{2}\eta^{\mu\nu}\eta^{ik}-\eta^{i\nu}\eta^{\mu k}).$$

Equation (3.13) yields

$$\alpha\sqrt{-g}\,R^{\underset{\smile}{ik}}=\gamma_{\underset{\smile}{\mu\nu}}\eta^{i\mu}\eta^{k\nu},$$

or

$$\alpha R_{\underset{\smile}{ik}}=\gamma_{\underset{\smile}{ik}} \quad (4.3)$$

to the first order in $\gamma_{\alpha\beta}$. Also,

$$g_{ik}=\eta_{ik}+\gamma_{\underline{ik}}. \quad (4.4)$$

Equations I are, in the present approximation,

$${}^{\dagger}s_{ik;\mu^*}=0, \quad (4.5)$$

$$\gamma_{,k}{}^{\underset{\smile}{ik}}=0, \quad (4.6)$$

$$R_{\underline{ik}}=0. \quad (4.7)$$

From (4.5), which reads in expanded form

$$\gamma_{ik,\mu}-\eta_{\alpha k}\,{}^*\Gamma_{i\mu}{}^{\alpha}-\eta_{i\alpha}\,{}^*\Gamma_{\mu k}{}^{\alpha}=0,$$

we obtain by the usual cyclic permutation and addition,

$${}^*\Gamma_{ik}{}^{\mu}=\tfrac{1}{2}\eta^{\mu\alpha}(\gamma_{\alpha k,i}+\gamma_{i\alpha,k}-\gamma_{ki,\alpha}).$$

Hence by virtue of (3.1),

$$\Gamma_{ik}{}^{\mu}=-\tfrac{2}{3}\delta_i{}^{\mu}\Gamma_k+\tfrac{1}{2}\eta^{\mu\alpha}(\gamma_{\alpha k,i}+\gamma_{i\alpha,k}-\gamma_{ki,\alpha}).$$

The vector Γ_k is assumed to be of the same order as γ_{ik}. Computing the Ricci tensor from the linearized formula

$$R_{ik}=\Gamma_{i\alpha,k}{}^{\alpha}-\Gamma_{ik,\alpha}{}^{\alpha},$$

we find

$$R_{ik}=\tfrac{2}{3}(\Gamma_{k,i}-\Gamma_{i,k})+\tfrac{1}{2}\eta^{\alpha\beta}(\gamma_{ki,\alpha\beta}+\gamma_{\alpha\beta,ik}-\gamma_{\alpha i,\beta k}-\gamma_{\beta k,i\alpha}). \quad (4.8)$$

We are still at liberty to impose four coordinate conditions which will not affect the quasi-Minkowskian character of our coordinates. Let us choose these (as in general relativity) to be

$$\eta^{\alpha\beta}(\gamma_{\alpha i,\beta}-\tfrac{1}{2}\gamma_{\alpha\beta,i})=0. \tag{4.9}$$

From (4.9) we obtain by differentiation,

$$\eta^{\alpha\beta}(\gamma_{\alpha i,\beta k}+\gamma_{\alpha k,i\beta}-\gamma_{\alpha\beta,ik})=0.$$

This enables us to simplify (4.8) down to

$$R_{ik}=\tfrac{2}{3}(\Gamma_{k,i}-\Gamma_{i,k})+\tfrac{1}{2}\Box\gamma_{ki},\quad \Box\equiv\eta^{\alpha\beta}\partial_\alpha\partial_\beta. \tag{4.10}$$

The Eqs. (4.7) now lead to

$$\Box\underset{\sim}{\gamma_{ik}}=0$$

which are the linearized field equations of gravitation, if we interpret ${}^{\dagger}\underset{\sim}{s_{ik}}$ (or perhaps g_{ik}) as the metric tensor. (In the following section it will be shown that the choice of ${}^{\dagger}\underset{\sim}{s_{ik}}$ for the metric tensor leads to the correct equations of motion.)

If $\underset{\sim}{R_{ik}}$ is assumed proportional to the electromagnetic field, then (4.6) [or, in its rigorous form (2.15)] is most naturally interpreted as the *first* Maxwell tetrad (the one which asserts the vanishing of the magnetic current).

Further, we have from (4.3) and (4.10),

$$\Box\underset{\sim}{\gamma_{ik}}=\tfrac{2}{3}(\Gamma_{k,i}-\Gamma_{i,k})+\frac{1}{\alpha}\underset{\sim}{\gamma_{ki}}.$$

This equation shows that Γ_i, is, in a very rough sense, the vector potential of electromagnetism. Taking the cyclic divergence, we get

$$\Box\underset{\sim}{\gamma_{[ik,l]}}=-\frac{1}{\alpha}\underset{\sim}{\gamma_{[ik,l]}}. \tag{4.11}$$

This is compatible with, although not equivalent to, the second tetrad of Maxwell's theory.[7]

Define (rigorously)

$$J^\mu\equiv(-\det{}^{\dagger}\underset{\sim}{s_{ik}})^{-\frac{1}{2}}\frac{1}{3!}\epsilon^{\mu ikl}\underset{\sim}{R_{[ik,l]}}$$

so that J^μ is proportional to the 4-current. The equation of conservation of charge is then (rigorously) satisfied:

$$\nabla_\mu J^\mu=0.$$

Here, ∇_μ indicates the covariant derivative with respect to the Christoffel symbols of ${}^{\dagger}\underset{\sim}{s_{ik}}$.

We have already pointed out that the constant α must be negative.[5] Accordingly, (4.11) may be written

$$\Box J^\mu=(1/\lambda^2)J^\mu, \tag{4.12}$$

where $\lambda=\sqrt{-\alpha}$ is a fundamental length, which we shall assume to be of subatomic order. For a static spherically symmetric charge distribution with $J^4=\rho(r)$, (4.12) reduces to

$$\Delta\rho=(1/\lambda^2)\rho$$

with the solution

$$\rho=Ae^{-r/\lambda}/r. \tag{4.13}$$

This form of solution is of course only valid for weak fields (i.e., large r). In particular, the singularity which (4.13) displays for $r=0$ may be illusory. According to (4.13), the charge density falls off very rapidly with distance. We have in effect, a particle with a radius of order λ. In general, at distances from "sources" large compared with λ, the 4-current vanishes to a high approximation, i.e., the second Maxwell tetrad

$$\underset{\sim}{\gamma_{[ik,l]}}=0$$

is satisfied.

So far, we have considered only the field equations I which constitute a particular case $\alpha\neq0$, $\beta=0$ of our general field equations (2.8), (2.9). If $\alpha=0$, we obtain in place of I, the "cosmological" field equations II. In this case, the general character of our solutions may be expected to be the same as before, but silhouetted now against the background of a de Sitter universe, instead of flat space-time.

5. THE EQUATIONS OF MOTION

Our discussion up to this point already indicates with fair certainty that the present theory correctly describes the interaction between gravitation and electricity. Indeed, we found in the previous section that the linearized equations for $\underset{\sim}{R_{\mu\nu}}$ are in agreement with Maxwell's equations. And in the next higher approximation the quantity $M_{\mu\nu}(\underset{\sim}{R})$, proportional to the electromagnetic stress tensor, appears on the right-hand side of the equations for $\underset{\sim}{R_{\mu\nu}}$. It seems plausible to infer that the predictions of the theory regarding the gravitational effects of weak electromagnetic fields will agree with the results of general relativity, and also that the present field equations will lead to the correct equations of motion for charged masses.

We shall now examine this latter point more closely by giving a brief sketch, following Infeld and Callaway,[8] of how the equations of motion can be derived by the EIH method. Our discussion will be restricted to the field equations I.

We assume that the field is "quasi-static," i.e., that the derivatives of the field quantities with respect to x^4 are of order ϵ times the space derivatives, where ϵ is small. The tensor ${}^{\dagger}s_{\alpha\beta}$ is then expanded in powers of ϵ

[7] Equations similar in structure to (4.11) appear in many theories based on modifications of Einstein's unified field theory. See for instance, E. Schrödinger, *Space-Time Structure* (Cambridge University Press, New York, 1950); B. Kursunoglu, Phys. Rev. **88**, 1369 (1952); W. B. Bonnor, Proc. Roy. Soc. (London) **A226**, 366 (1954).

[8] L. Infeld, Acta Phys. Polon. **10**, 284 (1950); J. Callaway, Phys. Rev. **92**, 1567 (1953).

as follows:

$$^{\dagger}s_{ik}=-\delta_{ik}+\epsilon^2 \underset{2}{h_{ik}}+\epsilon^4 \underset{4}{h_{ik}}+\cdots,$$

$$^{\dagger}s_{i4}=\epsilon^3 \underset{3}{h_{i4}}+\cdots,$$

$$^{\dagger}s_{44}=1+\epsilon^2 \underset{2}{h_{44}}+\epsilon^4 \underset{4}{h_{44}}+\cdots.$$

We assume that the field is "quasistatic" i.e., that the derivatives of the field quantities with respect to x^4

To the second and third order in ϵ, it is found, as in Sec. 4, that $\underset{\sim}{h_{\alpha\beta}}$ satisfies Maxwell's equations at distances from the particles large compared with the subatomic length λ:

$$\underset{2}{\breve{h}}_{ik,k}=0, \quad \underset{2}{\breve{h}}_{[ik,l]}=0, \tag{5.1}$$

$$\underset{3}{\breve{h}}_{4k,k}=0, \quad \underset{3}{\breve{h}}_{4i,k}+\underset{3}{\breve{h}}_{k4,i}+\underset{2}{\breve{h}}_{ik,4}=0. \tag{5.2}$$

Note also

$$\alpha \underset{2}{\breve{R}}_{ik}=\underset{2}{\breve{h}}_{ik}. \tag{5.3}$$

A solution of (5.1) appropriate for N slowly moving point charges is

$$\underset{2}{\breve{h}}_{ik}=\epsilon_{ikl}\phi_{,l}, \tag{5.4}$$

where

$$\phi=\sum_{i=1}^{N}\phi(i), \quad \phi(k)=e(k)/r(k),$$

and $r(k)$ is the distance from the kth charge to the field point. The third-order field equations are satisfied if we write

$$\underset{3}{\breve{h}}_{4k}=-\epsilon_{4krs}\{\sum_{i=1}^{N}\phi(i)\dot{\xi}_r(i)\}_{,s},$$

where $\xi_m(i)$, the coordinates of the ith particle, are slowly varying functions of time, and the dot indicates the time-derivative.

Let $\mathrm{P}_{\mu\nu}$ be the Ricci tensor formed from the Christoffel symbols of $^{\dagger}\underline{s_{\alpha\beta}}$, which we shall identify with the metric tensor. Write

$$\tilde{\mathrm{P}}_{\mu\nu}=\mathrm{P}_{\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}\eta^{\alpha\beta}\mathrm{P}_{\alpha\beta},$$

$$\tilde{R}_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}\eta^{\alpha\beta}R_{\alpha\beta},$$

$$\Lambda_{\mu\nu}=\tilde{R}_{\mu\nu}-{}_{\mu}\tilde{\mathrm{P}}_{\nu}.$$

To the second and third orders, $\tilde{R}_{\mu\nu}$ coincides with $\tilde{\mathrm{P}}_{\mu\nu}$, i.e., $\underset{2}{\Lambda_{\mu\nu}}=\underset{3}{\Lambda_{\mu\nu}}=0$. Hence the field equations

$$\underset{2}{R_{ik}}=0, \quad \underset{2}{R_{44}}=0, \quad \underset{3}{R_{4k}}=0$$

can be solved exactly as in the theory of gravitational motion.[9] We need not enter into details here.

[9] A. Einstein and L. Infeld, Can. J. Math. 1, 209 (1949).

Consider now the fourth-order field equations

$$\underset{4}{R_{ik}}=\alpha M_{ik}(\underset{2}{R})$$

which may be put into the form

$$\underset{4}{\tilde{\mathrm{P}}_{ik}}+\underset{4}{\Lambda_{ik}}+\underset{4}{T_{ik}}=0 \tag{5.5}$$

with

$$\underset{4}{T_{ik}}=-\alpha \underset{4}{M_{ik}}(\underset{\sim}{R})$$

$$=-\frac{1}{\alpha}\{\tfrac{1}{4}\underset{2}{\breve{h}}_{mn}\underset{2}{\breve{h}}^{mn}\eta_{ik}-\eta^{rs}\underset{2}{\breve{h}}_{ir}\underset{2}{\breve{h}}_{ks}\}.$$

Since $\underset{2}{\breve{h}}_{ik}$ is given by (5.4), it is clear that $\underset{4}{T_{ik}}$ is proportional to the electromagnetic stress tensor for purely Coulomb forces.

The quantities $\underset{4}{\Lambda_{ik}}$ can be expressed in terms of $\underset{2}{h_{\alpha\beta}}$, $\underset{3}{h_{\alpha\beta}}$ by solving the equations

$$^{\dagger}s_{ik;\mu^*}=0 \tag{5.6}$$

for the affinity to the fourth order, from which $\underset{4}{\underline{R_{ik}}}$ and hence $\underset{4}{\Lambda_{ik}}$ can be computed. Now, the relations (5.6) are the same as those which hold in Einstein's theory. Hence, noting (6.8) below, we may take over the results of Infeld and Callaway,[8] who found that

$$-\underset{4}{\Lambda_{ik}}=(\underset{2}{\breve{h}}_{is}\underset{2}{\breve{h}}_{kp,s}+\eta_{ip}\phi_{,rk}\phi_{,r}-\eta_{ik}\phi_{,rp}\phi_{,r})_{,p}.$$

We now integrate equations (5.5) over a closed surface surrounding (only) the lth particle:

$$\int^{(l)}(\underset{4}{\tilde{\mathrm{P}}_{ik}}+\underset{4}{\Lambda_{ik}}+\underset{4}{T_{ik}})n^i dS=0. \tag{5.7}$$

The integral of each of the three terms in (5.7) is independent of the size and shape of the surface, since to the fourth order,

$$\underset{4}{\tilde{\mathrm{P}}}_{ik,k}=0, \quad \underset{4}{\Lambda_{ik,k}}=0, \quad \underset{4}{T_{ik,k}}=0.$$

The integral of $\underset{4}{\tilde{\mathrm{P}}_{ik}}$ yields, as is well known[9] a term proportional to the inertial resistance $-m(l)\ddot{\xi}(l)$ of the lth particle, plus the gravitational force acting on it. The integral of $\underset{4}{\Lambda^{ik}}$ may be shown to vanish.[8] Hence the last term of (5.7) must account for the contribution of the electrical forces; and, indeed, it is clear from the physical meaning of the Maxwell stress tensor, that the integral of $\underset{4}{T_{ik}}$ will yield the Coulomb force acting on the lth particle. The influence of the magnetic field does not appear in this approximation, since we have assumed the particles to be moving slowly.

6. STATIC SPHERICALLY SYMMETRIC SOLUTIONS

It can be shown[10] that the most general spherically symmetric, static tensor field ${}^{\dagger}s_{ik}$ is expressible in terms of polar coordinates in the form

$$ {}^{\dagger}s_{ik}=\begin{bmatrix} -e^{\mu} & 0 & 0 & w \\ 0 & -r^2 & r^2 v\sin\theta & 0 \\ 0 & -r^2 v\sin\theta & -r^2\sin^2\theta & 0 \\ -w & 0 & 0 & e^{\nu} \end{bmatrix}, \tag{6.1} $$

where μ, ν, v, and w are functions of r only.

Our object is to obtain solutions to field equations I and II for the special case $v=0$, $w\neq 0$. The more interesting case of $w=0$, $v\neq 0$ leads to serious mathematical difficulties, (cf. the treatment by Wyman[11] of the corresponding case in Einstein's theory). In view of the arguments of Sec. 4 it would seem that the solution considered here corresponds to the field of a magnetic pole rather than that of an electric charge.

We first consider solutions to the set of field equations I. With $v=0$ in (6.1) we use Eqs. (4.1) to obtain $\mathfrak{s}^{ik}$. Substitution of $\mathfrak{s}^{ik}$ in (3.13) enables us to find $\sqrt{-g}$, g^{ik}, and $R.\overset{ik}{\smile}$ The results are:

$$ g_{ik}=\operatorname{diag}\left[-e^{\mu}(1-A)^{\frac{1}{2}}, \frac{-r^2}{(1-A)^{\frac{1}{2}}}, \frac{-r^2\sin^2\theta}{(1-A)^{\frac{1}{2}}}, e^{\nu}(1-A)^{\frac{1}{2}}\right], \tag{6.2} $$

$$ \mathfrak{s}\overset{14}{\smile}=-\mathfrak{s}\overset{41}{\smile}=-[A/(1-A)]^{\frac{1}{2}}r^2\sin\theta, \tag{6.3} $$

$$ R_{\underset{\smile}{14}}=-R_{\underset{\smile}{41}}=\frac{w}{a}(1-A)^{\frac{1}{2}}, \tag{6.4} $$

where

$$ A\equiv w^2 e^{-(\mu+\nu)}. \tag{6.5} $$

The remaining $\mathfrak{s}^{\underset{\smile}{ik}}$ and $R_{\underset{\smile}{ik}}$ are identically zero.

Considering (3.10), we note that it is satisfied identically except for the case $i=4$, which yields

$$ \frac{\partial}{\partial r}\{[A/(1-A)]^{\frac{1}{2}}r^2\sin\theta\}=0. \tag{6.6} $$

Integration of (6.6) leads to

$$ A=l^4/(l^4+r^4), \tag{6.7} $$

where the length l is a constant of integration. If we denote by ${}^*R_{ik}$ the Ricci tensor formed from the ${}^*\Gamma_{ik}{}^{\mu}$, then it follows from (3.1) that

$$ {}^*R_{ik}=R_{ik}+\frac{2}{3}\left(\frac{\partial\Gamma_i}{\partial x_k}-\frac{\partial\Gamma_k}{\partial x_i}\right), \quad {}^*R_{\underset{\smile}{ik}}=R_{\underset{\smile}{ik}}. \tag{6.8} $$

Hence (3.11) may be replaced by

$$ {}^*R_{\underset{\smile}{ik}}=\alpha M_{ik}(R_{\smile}). \tag{6.9} $$

Using (6.2) and (6.3) we find that (6.9) reduces to the following set of equations:

$$ \begin{aligned} {}^*R_{\underset{\smile}{ik}}&=0, \quad i\neq k, \\ {}^*R_{11}+\tfrac{1}{2}\alpha g^{44}(R_{\underset{\smile}{14}})^2&=0, \\ {}^*R_{22}-\tfrac{1}{2}\alpha g_{22}g^{11}g^{44}(R_{\underset{\smile}{14}})^2&=0, \\ {}^*R_{33}-\tfrac{1}{2}\alpha g_{33}g^{11}g^{44}(R_{\underset{\smile}{14}})^2&=0, \\ {}^*R_{44}+\tfrac{1}{2}\alpha g^{11}(R_{\underset{\smile}{14}})^2&=0. \end{aligned} \tag{6.10} $$

In order to deal with (6.10) we need expressions for the ${}^*\Gamma_{ik}{}^{\mu}$ which can be obtained by solving the algebraic equations (3.9) for the 64 ${}^*\Gamma_{ik}{}^{\mu}$, from which ${}^*R_{ik}$ may be computed. Papapetrou[10] has solved a set of equations of the type (3.9), with a fundamental tensor of the form (6.1), and has computed a Ricci tensor which is identical with our ${}^*R_{ik}$. He has shown that the nonvanishing components of ${}^*R_{ik}$ are the following:

$$ \begin{aligned} {}^*R_{11}&=-\frac{1}{2}\nu''+\frac{\nu'}{4}(\nu'-\mu')-\frac{\mu'}{r}+2\left(\frac{A}{r}\right)' \\ &\quad +\frac{2A}{r}\left(\nu'-\frac{1}{2}\mu'+\frac{2A}{r}\right), \\ {}^*R_{22}&=\frac{{}^*R_{33}}{\sin^2\theta}=\tfrac{1}{2}re^{-\mu}(\nu'-\mu')+e^{-\mu}-1+2Ae^{-\mu}, \\ {}^*R_{44}&=-\tfrac{1}{2}(\nu'e^{\nu-\mu})'+\frac{\nu'e^{\nu-\mu}}{4}(\nu'-\mu'-4/r) \\ &\quad -\left(\frac{A}{4}e^{\nu-\mu}\right)'+\frac{A}{r}e^{\nu-\mu}\left(3\nu'-2\mu'-\frac{14}{r}+\frac{8A}{r}\right), \\ {}^*R_{\underset{\smile}{14}}&=-{}^*R_{\underset{\smile}{41}}=-2\left(\frac{w}{r}e^{-\mu}\right)'-\frac{4w}{r^2}e^{-\mu}. \end{aligned} \tag{6.11} $$

It follows from (6.10) that

$$ g^{11}\,{}^*R_{11}-g^{44}\,{}^*R_{44}=0 \tag{6.12} $$

and when the explicit expressions for the g^{ik} and ${}^*R_{ik}$ are substituted in (6.12) we obtain [noting (6.3)]

$$ \left(\frac{\nu'+\mu'}{r}+\frac{4A}{r^2}\right)(1-A)=0. $$

But according to (6.7)

$$ 1-A=\frac{r^4}{r^4+l^4}>0, $$

so it follows that

$$ \frac{\nu'+\mu'}{r}+\frac{4A}{r^2}=0. \tag{6.13} $$

Since (6.13) may be written

$$ \frac{d}{dr}\left[\mu+\nu+\log\left(\frac{r^4}{r^4+l^4}\right)\right]=0, $$

[10] A. Papapetrou, Proc. Roy. Irish Acad. **A52**, 69 (1948).

[11] M. Wyman, Can. J. Math. **2**, 427 (1950). For the general case, see W. B. Bonnor, Proc. Roy. Soc. (London) **A210**, 427 (1952).

we conclude that

$$e^{\mu+\nu}\left(\frac{r^4}{r^4+l^4}\right)=\zeta^2, \tag{6.14}$$

ζ being a constant of integration.

It is reasonable to assume that for large r the ${}^{\dagger}s_{ik}$ should approach the metric of flat space-time, which implies that we must have $\zeta^2=1$. Hence, using (6.5),

$$e^{\nu}=\left(1+\frac{l^4}{r^4}\right)e^{-\mu}, \quad w=\pm\frac{l^2}{r^2}. \tag{6.15}$$

Returning now to Eqs. (6.10) we find, using (6.2), (6.4), and (6.15), that the explicit expression of the third equation is

$$\nu'-\mu'+\frac{2}{r}(1-e^{\mu})+\frac{4A}{r}=\frac{l^4e^{\mu}}{\alpha r(l^4+r^4)^{\frac{1}{2}}}. \tag{6.16}$$

Equation (6.13) enables us to eliminate ν' from (6.16) whereupon μ is determined by

$$\mu'-\frac{1}{r}(1-e^{\mu})+\frac{l^4}{2\alpha r}\cdot\frac{e^{\mu}}{(l^4+r^4)^{\frac{1}{2}}}=0. \tag{6.17}$$

Let $e^{-\mu}=1-2m(r)/r$, then (6.17) becomes

$$m'=-\frac{l^4}{4\alpha}(l^4+r^4)^{-\frac{1}{2}}, \tag{6.18}$$

so that

$$m=m_0+\frac{l^4}{4\alpha}\int_r^{\infty}\frac{d\xi}{(l^4+\xi^4)^{\frac{1}{2}}}, \tag{6.19}$$

where m_0 is a constant of integration. Hence the equation

$$e^{-\mu}=1-\frac{2m_0}{r}+\frac{e^2}{r}\int_r^{\infty}\frac{d\xi}{(l^4+\xi^4)^{\frac{1}{2}}}, \tag{6.20}$$

together with (6.15), determines all the components of ${}^{\dagger}s_{ik}$. In (6.20), we have defined the constant e by

$$e^2=-l^4/2\alpha, \tag{6.21}$$

recalling that α is negative.[5] It may be verified that the full set of Eqs. (6.10) is satisfied by our solution.

The asymptotic form of (6.15), (6.20) for large r $(r\gg l)$,

$$e^{-\mu}=e^{\nu}=1-(2m_0/r)+(e^2/r^2),$$

is in agreement with the general relativistic line-element for the field of a particle with mass proportional to m_0 and (magnetic) charge proportional to e.

Equations (6.4), (6.8), and the last of (6.11) give a partial determination of the vector Γ_i. (This vector is of course arbitrary to a certain extent because of the λ invariance of our theory.) It is found that Γ_1 is arbitrary, $\Gamma_2=\Gamma_3=0$ (spherical symmetry), and Γ_4 is given by a somewhat complicated expression, whose asymptotic form for large r is

$$\Gamma_4\sim\text{const}\pm\tfrac{3}{2}(l^2/\alpha)(1/r).$$

Our expression (6.20) for $e^{-\mu}$ will be everywhere nonsingular if the magnetic pole-strength and mass are related by

$$m_0=\tfrac{1}{2}e^2\int_0^{\infty}\frac{d\xi}{(l^4+\xi^4)^{\frac{1}{2}}}.$$

However, e^{ν} still becomes infinite when $r\to 0$. It is not possible in the present theory to obtain a completely nonsingular metric tensor ${}^{\dagger}s_{ik}$ representing a magnetic monopole.

So much for the field equations I. We turn now to the alternative field equations II, Eqs. (3.19) to (3.22). As the method of solution is much the same as before, it will be sufficient to record the main results. If ${}^{\dagger}s_{ik}$ is assumed to be given by (6.1) with $v=0$, it is found that

$$g_{ik}=\frac{1}{bR}\,\text{diag}[-e^{\mu}(1-A)^{\frac{1}{2}}, -r^2/(1-A)^{\frac{1}{2}}, -r^2\sin^2\theta/(1-A)^{\frac{1}{2}}, e^{\nu}(1-A)^{\frac{1}{2}}],$$

$$e^{\nu}=\left(1+\frac{l^4}{r^4}\right)e^{-\mu},$$

$$e^{-\mu}=1-\frac{2m_0}{r}+\frac{e^2}{r}\int_r^{\infty}\frac{d\xi}{(\xi^4+l^4)^{\frac{1}{2}}}-\frac{1}{4br}\int_0^r(l^4+\xi^4)^{\frac{1}{2}}d\xi,$$

$$w=\pm l^2/r^2.$$

A is given by (6.7) and e by (6.21). The asymptotic form of this solution is

$$e^{\nu}=e^{-\mu}=1-\frac{2m_0}{r}+\frac{e^2}{r^2}-\frac{1}{12b}r^2, \quad (r\gg l)$$

and represents, according to general relativity, the field of a magnetic particle in a de Sitter universe.

7. SOLUTIONS FOR FLAT SPACE-TIME

The existence of non-trivial solutions of Einstein's unified field equations for "flat" space-time ($R_{.ijk}{}^h=0$) has already been alluded to as being a defect of the Einstein theory (Sec. 1). In the case of the field equations I of the present theory, it will now be shown that the assumption

$$R_{.ijk}{}^h=0 \tag{7.1}$$

necessarily implies that the electromagnetic field R_{ik} vanishes (this is obvious) and that the line-element is reducible to the Minkowskian form.

If (7.1) holds, the field equations I simplify to

$${}^{\dagger}s_{ik,\mu}-{}^{\dagger}s_{\alpha k}\,{}^*\Gamma_{i\mu}{}^{\alpha}-{}^{\dagger}s_{i\alpha}\,{}^*\Gamma_{\mu k}{}^{\alpha}=0, \quad {}^*\Gamma_i=0, \tag{7.2}$$

$$\mathfrak{g}^{ik}=\sqrt{-g}\,g^{ik}. \tag{7.3}$$

From (7.3) and (4.1) it follows that ${}^{\dagger}s_{ik}=g_{ik}$, a symmetric tensor. Then (7.2) may be solved in the conventional way for the ${}^{*}\Gamma_{ik}{}^{\mu}$, which turn out to be the Christoffel symbols $\left\{ {\mu \atop ik} \right\}$ of g_{ik}. Hence, by (3.1) we have

$$\Gamma_{ik}{}^{\mu}=\left\{ {\mu \atop ik} \right\}-\tfrac{2}{3}\delta_i{}^{\mu}\Gamma_k. \tag{7.4}$$

Substituting (7.4) into (7.1), we find

$$R_{ijk}{}^{h}=\mathrm{P}_{ijk}{}^{h}+\tfrac{2}{3}\delta_i{}^{h}(\Gamma_{j,k}-\Gamma_{k,j}),=0, \tag{7.5}$$

where $\mathrm{P}_{.ijk}{}^{h}$ is the curvature tensor formed from $\left\{ {\mu \atop ik} \right\}$. Contracting with respect to h,i and observing that $\mathrm{P}_{.\alpha jk}{}^{\alpha}\equiv 0$, we obtain from (7.5),

$$\Gamma_{j,k}-\Gamma_{k,j}=0 \tag{7.6}$$

so that

$$\mathrm{P}_{.ijk}{}^{h}=0. \tag{7.7}$$

Equation (7.7) expresses the flatness of the Riemannian space having the metric tensor g_{ik}. Consequently, g_{ik} can be reduced to Minkowskian form. According to (7.6), Γ_k is equal to an arbitrary gradient: this is merely a reflection of the λ invariance of the theory.

The question of flatness does not arise in connection with our second set of field equations II, since here the existence of a nonvanishing curvature invariant R is presupposed.

8. CONCLUDING REMARKS

As Einstein[12] has pointed out, there are two distinct points of view from which a field theory may be regarded as "unified."

(1) The field quantities should appear as unified, covariant entities which are irreducible under the invariance group of the theory. (For instance, the electric and magnetic fields experience a unification in this sense under the Lorentz group of special relativity.)

(2) The Lagrangian of the theory should not be expressible as the sum of several invariant parts, but should be a formally unified entity.

Einstein's theory, whose field equations are derived from the variational principle

$$\delta\int(-g)^{\frac{1}{2}}Rd^4x=0, \tag{8.1}$$

and which operates with nonsymmetric g_{ik}, $\Gamma_{ik}{}^{\mu}$ appears to be satisfactorily unified from the second point of view, but fails to satisfy (1), since the g_{ik} are not irreducible.

Precisely the reverse appears true of the theory developed in this paper. Our field quantities are irreducible under the "extended group," comprising the group of coordinate transformation and the λ transformations (1.1) but our Lagrangian (1.3) is not unified according to criterion (2). However, the latter is a somewhat loose requirement which can be met by a variety of formal devices. Let us write $\mathbf{g}$, $\mathbf{R}$ for the matrices $g^{\alpha\beta}$, $R_{\alpha\beta}$, and regard their matrix product as expressed dimensionlessly in terms of some suitable microscopic standard of length (e.g., the length λ of Sec. 4). Then we may, for instance, consider the action principle

$$\delta\int(-g)^{\frac{1}{2}}Ld^4x=0, \qquad L=\mathrm{Tr}\,f(\mathbf{gR}), \tag{8.2}$$

where f is an arbitrary analytic function such that $f(0)=0$. Expanding f in powers of the matrix $\mathbf{gR}$, we obtain

$$L=f'(0)g^{\mu\nu}R_{\mu\nu}+\tfrac{1}{2}f''(0)R_{\mu\nu}R^{\nu\mu}+\cdots, \tag{8.3}$$

which agrees with (1.3) up to the quadratic terms, if $b=0$, $\beta=-\alpha$. From this point of view, the present paper may be regarded as a preliminary exploration of the consequences of (8.2) in the case of weak fields, for which the higher terms in (8.3) can presumably be neglected.

More generally, a Lagrangian of the form

$$L=\mathrm{Tr}\,f(\mathbf{gR},\,\mathbf{gR}^{\mathrm{T}})$$

might be considered, where $\mathbf{R}^{\mathrm{T}}$ denotes the transposed matrix. This Lagrangian is λ invariant, transposition symmetric, formally unified, and reduces to (1.3) (with $b=0$) in the quadratic approximation. The difficulty is to find some way of discriminating among the innumerable possibilities which present themselves once Einstein's simple choice $L=\mathrm{Tr}\ (gR)$ is abandoned as inadequate.

We remark finally that an approach to the problem of unification resembling ours in one or two respects has been given in a series of profound studies by Lanczos.[13] Postulating a Lagrangian which is homogeneous quadratic in the curvature, Lanczos is led to field equations of cosmological type, the "cosmological constant" being now interpreted as a microscopic constant. Lanczos' treatment differs basically from ours in not straying outside the framework of Riemannian geometry, and in not making use of the Palatini method of variation.[14]

ACKNOWLEDGMENT

We should like to express our sincere thanks to Professor Max Wyman for many helpful discussions and suggestions.

[12] A. Einstein, Ann. Math. Princeton **46**, 578 (1945).

[13] See the comprehensive review: C. Lanczos, Revs. Modern Phys. **29**, 337 (1957).

[14] Quadratic action principles using the Palatini method of variation have been considered for the case of symmetric g_{ik} and $\Gamma_{ik}{}^{\mu}$ by G. Stephenson, Nuovo cimento **9**, 263 (1958), and P. W. Higgs, Nuovo cimento **11**, 816 (1959). The nonsymmetric case has recently been discussed by G. Stephenson, Proc. Cambridge Phil. Soc. **56**, 247 (1960). No attempt is made in these papers to develop a full-scale theory.

Approximate Stress Energy Tensor for Gravitational Fields*

A. H. TAUB
University of Illinois, Digital Computer Laboratory, Urbana, Illinois
(Received April 28, 1961)

An invariant formulation in Minkowski space-time of an approximation to the Einstein theory of gravitation is given. In this formulation a tensor is introduced which may be interpreted as the approximate stress energy tensor of the gravitational field. Conservation laws involving this tensor and the material stress energy tensor are formulated. The behavior of these tensors under "gauge transformations" of the weak gravitational fields is discussed. The classical limit of the conservation of energy equation is studied and the results are compared to some observations of Bondi on a possible analog of the Poynting vector for a gravitational field.

1. INTRODUCTION

IT is the main purpose of this paper to formulate and discuss conservation laws in invariant form in Minkowski space-time for an approximate version of the Einstein theory of gravitation. These laws will involve the approximate energy and momentum of the material and gravitational fields. The discussion will be mainly concerned with a first approximation to the Einstein theory but may be extended to higher approximations. We shall relate the results obtained to some observations of Bondi[1] concerning an analog to the Poynting vector for classical time-dependent gravitational fields.

The Minkowski space-time will be used as the underlying space in which the discussion will take place. In principle, any fixed Riemannian space-time may be used. There are, however, two reasons for choosing the Minkowski one: (a) with this choice the Newtonian approximation is readily obtained from the first approximation given below by neglecting terms of the order of $1/c^2$ and (b) the underlying space-time admits a ten parameter group of motions, the inhomogeneous Lorentz group. Use is made of the latter fact in formulating conserved quantities.

The approximate theory mentioned above is obtained by considering the metric tensor $g_{\mu\nu}$ of space-time as defined over the Minkowski space as a convergent power series expansion in

$$k=8\pi G/c^2=1.864\times10^{-27}\text{ cm g}^{-1}, \tag{1.1}$$

where G is Newton's constant of gravitation and c is the velocity of light of the special theory of relativity.

We assume that

$$g_{\mu\nu}=\eta_{\mu\nu}+kh_{\mu\nu}+\tfrac{1}{2}k^2h_{(2)\mu\nu}+\cdots, \tag{1.2}$$

where $\eta_{\mu\nu}$ is the metric tensor of the Minkowski space-time. The coordinate system in which Eq. (1.2) holds may be an arbitrary one. The equations satisfied by the $g_{\mu\nu}$, that is $h_{\mu\nu}$, $h_{(2)\mu\nu}\cdots$ will be derived from the Einstein field equations

$$G_{\mu\nu}=-kc^2T_{\mu\nu}, \tag{1.3}$$

where

$$G_{\mu\nu}=R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R. \tag{1.4}$$

$R_{\mu\nu}$ is the Ricci tensor and R the scalar curvature tensor formed from the $g_{\mu\nu}$. The tensor $T_{\mu\nu}$ is the stress energy tensor of the matter "creating" the gravitational field.

Both the tensors $G_{\mu\nu}$ and $T_{\mu\nu}$ may be considered as functions of k and written as

$$G_{\mu\nu}=G_{(0)\mu\nu}+kG_{(1)\mu\nu}+\tfrac{1}{2}k^2G_{(2)\mu\nu}+\cdots, \tag{1.5}$$

$$T_{\mu\nu}=T_{(0)\mu\nu}+kT_{(1)\mu\nu}+\tfrac{1}{2}k^2T_{(2)\mu\nu}+\cdots. \tag{1.6}$$

It is evident from Eq. (1.2) that

$$G_{(0)\mu\nu}\equiv0. \tag{1.7}$$

The following discussion will center about the discussion of the equations

$$[(G_{\mu\nu}+kc^2T_{\mu\nu})g^{\nu\rho}]_{;\rho}=0, \tag{1.8}$$

which are consequences of Eqs. (1.3). In Eq. (1.8), the semicolon denotes the covariant derivative with respect to the metric tensor $g_{\mu\nu}$. Because of the Bianchi identities, we have

$$(G_{\mu\nu}g^{\nu\rho})_{;\rho}\equiv0. \tag{1.9}$$

If Eqs. (1.5) and (1.6) are substituted into Eqs. (1.3), and the resulting equations are regarded as identities in k, we obtain

$$G_{(n)\mu\nu}=-nc^2T_{(n-1)\mu\nu}. \tag{1.10}$$

These equations may be regarded as differential equations for the determination of the $h_{(n)\mu\nu}$ in terms of $h_{(m)\mu\nu}$ and $T_{(m)\mu\nu}(m=1,2,\cdots n-1)$. The $T_{(m)\mu\nu}$ must be such that

$$\left[\frac{d^m}{dk^m}(kc^2T_{\mu\nu}g^{\nu\rho})_{;\rho}\right]_{k=0}=0, \tag{1.11}$$

where $T_{\mu\nu}$ is given by Eq. (1.6), and the $g^{\nu\rho}$ are functions of k which satisfy

$$\frac{dg^{\nu\rho}}{dk}=-g^{\nu\sigma}\frac{dg_{\sigma\tau}}{dk}g^{\tau\rho}. \tag{1.12}$$

* This work was supported in part by the National Science Foundation.

[1] H. Bondi, "Les proprietes physiques des ondes gravitationelles," Colloque International Sur Les Theories Relativistes de la Gravitation, Royaumont, 21–27 June 1959 (to be published).

We also require that

$$(T_{(0)\mu\nu}\eta^{\mu\rho})_{,\rho}=0, \tag{1.13}$$

where the comma denotes the covariant derivative with respect to the tensor $\eta_{\mu\nu}$.

Equations (1.12) and (1.13) are equations for the determination of the $T_{(n)\mu\nu}$ in terms of $h_{(r)\mu\nu}$ and $T_{(s)\mu\nu}$ with $r=1, 2, \cdots n$ and $s=0, 2, \cdots n-1$.

2. CALCULATION OF $G_{(1)\nu}{}^{\mu}$

We begin our discussion by considering the expansion of the Christoffel symbols as power series in k. Thus,

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}=\tfrac{1}{2}g^{\mu\rho}(g_{\nu\rho|\sigma}+g_{\rho\sigma|\nu}-g_{\nu\sigma|\rho}), \tag{2.1}$$

where we have used the notation

$$\partial g_{\nu\rho}/\partial x^{\sigma}=g_{\nu\rho|\sigma}. \tag{2.2}$$

We may write

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}=\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(0)}+k\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(1)}+\frac{k^2}{2!}\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(2)}+\cdots,$$

where

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(n)}=\left(\frac{d^n}{dk^n}\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}\right)_{k=0}. \tag{2.3}$$

Thus

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(0)}=\tfrac{1}{2}\eta^{\mu\rho}(\eta_{\nu\rho|\sigma}+\eta_{\rho\sigma|\nu}-\eta_{\sigma\nu|\rho}), \tag{2.4}$$

the Christoffel symbol calculated from the η's. In a Galilean coordinate system,

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(0)}=0. \tag{2.5}$$

It may be verified that

$$A_{\nu\sigma}{}^{\mu}=\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(1)}=\tfrac{1}{2}\eta^{\mu\rho}(h_{\nu\rho,\sigma}+h_{\rho\sigma,\nu}-h_{\sigma\nu,\rho}), \tag{2.6}$$

where, as above the comma denotes, the covariant derivative with respect to the tensor $\eta_{\mu\nu}$. It follows from Eq. (2.6) that

$$\left\{\begin{matrix}\mu\\ \nu\mu\end{matrix}\right\}_{(1)}=\tfrac{1}{2}\eta^{\mu\rho}h_{\mu\rho,\nu}=\tfrac{1}{2}h_{,\nu}, \tag{2.7}$$

where

$$h=\eta^{\mu\rho}h_{\mu\rho}. \tag{2.8}$$

The Ricci tensor is defined by the equation

$$R_{\mu\nu}=-\left\{\begin{matrix}\sigma\\ \mu\nu\end{matrix}\right\}_{|\sigma}+\left\{\begin{matrix}\sigma\\ \mu\sigma\end{matrix}\right\}_{|\nu}-\left\{\begin{matrix}\rho\\ \mu\nu\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \rho\sigma\end{matrix}\right\}+\left\{\begin{matrix}\tau\\ \mu\sigma\end{matrix}\right\}\left\{\begin{matrix}\sigma\\ \tau\nu\end{matrix}\right\}. \tag{2.9}$$

Hence,

$$R_{(0)\mu\nu}=0$$

and

$$R_{(1)\mu\nu}=-\left\{\begin{matrix}\sigma\\ \mu\nu\end{matrix}\right\}_{(1),\sigma}+\left\{\begin{matrix}\sigma\\ \mu\sigma\end{matrix}\right\}_{(1),\nu}. \tag{2.10}$$

On substituting from Eqs. (2.6) and (2.7) into Eq. (2.10), we obtain

$$R_{(1)\mu\nu}=-\tfrac{1}{2}\eta^{\rho\sigma}(h_{\rho\nu,\mu}+h_{\rho\mu,\nu}-h_{\mu\nu,\rho}-\eta_{\rho\nu}h_{,\mu})_{,\sigma}. \tag{2.11}$$

Since $R_{(0)\mu\nu}=0$, we have

$$R_{(1)}=\eta^{\mu\nu}R_{(1)\mu\nu}=-\eta^{\rho\sigma}(h_{\rho\alpha,\beta}\eta^{\alpha\beta}-h_{,\rho})_{,\sigma}. \tag{2.12}$$

Thus we may write

$$\begin{aligned}G_{(1)\mu\nu}&=R_{(1)\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}R_{(1)}\\&=-\tfrac{1}{2}\eta^{\sigma\rho}(k_{\rho\nu,\mu}+k_{\mu\rho,\nu}-k_{\mu\nu,\rho}-\eta_{\mu\nu}k_{\rho\alpha,\beta}\eta^{\alpha\beta})_{,\sigma}\\&=-\tfrac{1}{2}\eta^{\rho\sigma}(k_{\rho\nu,\mu}-k_{\mu\nu,\rho}-\eta_{\mu\nu}k_{\rho\alpha,\beta}\eta^{\alpha\beta}+\eta_{\rho\nu}k_{\mu\alpha,\beta}\eta^{\alpha\beta})_{,\sigma},\end{aligned} \tag{2.13}$$

where

$$k_{\mu\nu}=h_{\mu\nu}-\tfrac{1}{2}\eta_{\mu\nu}h. \tag{2.14}$$

Because the Minkowski space is a flat space it follows that

$$G_{(1)}{}^{\mu}{}_{\nu}=\eta^{\tau\mu}G_{(1)\tau\nu}, \tag{2.15}$$

and the order of covariant differentiation is immaterial that is,

$$t\cdots{}_{,\alpha\beta}=t\cdots{}_{,\beta\alpha}.$$

It may be verified by using Eqs. (2.14) and (2.15) that

$$G_{(1)}{}^{\mu}{}_{\nu,\mu}=0. \tag{2.16}$$

3. CALCULATION OF $G_{(2)}{}^{\mu\nu}$

In this section we shall evaluate the above tensor in terms of $h_{\mu\nu}$, $h_{(2)\mu\nu}$ and their first and second derivatives. We shall show that it may be written as a sum of two tensors. One of these contains the second derivatives of $h_{\mu\nu}$ and $h_{(2)\mu\nu}$ and has a vanishing divergence. The other is a function of the $h_{\mu\nu}$, its first derivatives, and $G_{(1)\mu\nu}$.

It follows from Eqs. (2.1) that

$$\left\{\begin{matrix}\mu\\ \nu\sigma\end{matrix}\right\}_{(2)}=B_{\nu}{}^{\mu}{}_{\sigma}-2h_{\rho}{}^{\mu}A_{\nu\sigma}{}^{\rho}, \tag{3.1}$$

where

$$B_{\nu\sigma}{}^{\mu}=\eta^{\mu\rho}\tfrac{1}{2}(h_{(2)\nu\rho,\sigma}+h_{(2)\sigma\rho,\nu}-h_{(2)\nu\sigma,\rho}) \tag{3.2}$$

and $A_{\nu\sigma}{}^{\rho}$ is defined by Eqs. (2.6).

By differentiating Eq. (2.9) twice with respect to k and setting $k=0$ we obtain

$$\begin{aligned}R_{(2)\sigma\tau}=&-\left\{\begin{matrix}\rho\\ \sigma\tau\end{matrix}\right\}_{(2),\rho}+\left\{\begin{matrix}\rho\\ \sigma\rho\end{matrix}\right\}_{(2),\tau}\\&-2A_{\sigma\tau}{}^{\rho}A_{\rho\lambda}{}^{\lambda}+2A_{\sigma\lambda}{}^{\rho}A_{\rho\tau}{}^{\lambda}.\end{aligned} \tag{3.3}$$

Since

$$G^{\mu\nu}=(g^{\mu\sigma}g^{\nu\tau}-\tfrac{1}{2}g^{\mu\nu}g^{\sigma\tau})R_{\sigma\tau},$$

it follows that

$$G_{(2)}{}^{\mu\nu}=(\eta^{\mu\sigma}\eta^{\nu\tau}-\tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\tau})R_{(2)\sigma\tau} -2(h^{\mu\sigma}\eta^{\nu\tau}+h^{\nu\tau}\eta^{\mu\sigma}-\tfrac{1}{2}h^{\mu\nu}\eta^{\sigma\tau}-\tfrac{1}{2}h^{\sigma\tau}\eta^{\mu\nu})R_{(1)\sigma\tau}. \quad (3.4)$$

In view of Eqs. (2.10) and (3.3), we may write the above equation as

$$G_{(2)}{}^{\mu\nu}=\Big[(\eta^{\mu\sigma}\eta^{\nu\tau}-\tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\tau})\Big(-\left\{\begin{matrix}\alpha\\ \sigma\tau\end{matrix}\right\}_{(2)}+\left\{\begin{matrix}\rho\\ \sigma\rho\end{matrix}\right\}_{(2)}\delta_\tau{}^\alpha\Big) +2(h^{\mu\sigma}\eta^{\nu\tau}+h^{\nu\tau}\eta^{\mu\sigma}-\tfrac{1}{2}h^{\mu\nu}\eta^{\sigma\tau}-\tfrac{1}{2}h^{\sigma\tau}\eta^{\mu\nu}) \times(A_{\sigma\tau}{}^\alpha-A_{\sigma\rho}{}^\rho\delta_\tau{}^\alpha)\Big]_{,\alpha} +2(\eta^{\mu\sigma}\eta^{\nu\tau}-\tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\tau}) \times(A_{\sigma\lambda}{}^\rho A_{\rho\tau}{}^\lambda-A_{\sigma\tau}{}^\rho A_{\rho\lambda}{}^\lambda)-2(h^{\mu\sigma}\eta^{\nu\tau}+h^{\nu\tau}\eta^{\mu\sigma} -\tfrac{1}{2}h^{\mu\nu}\eta^{\sigma\tau}-\tfrac{1}{2}h^{\sigma\tau}\eta^{\mu\nu})_{,\alpha}(A_{\sigma\tau}{}^\alpha-A_{\sigma\rho}{}^\rho\delta_\tau{}^\alpha).$$

It may be verified that

$$G_{(2)}{}^{\mu\nu}=H^{\mu\nu}+2c^2k^{-1}E^{\mu\nu}, \quad (3.5)$$

where

$$H^{\mu}{}_{,\nu}{}^{\nu}=0 \quad (3.6)$$

and

$$H^{\mu\nu}=\big[(\eta^{\mu\sigma}\eta^{\nu\tau}-\tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\tau})(-B_{\sigma\tau}{}^\alpha+\delta_\tau{}^\alpha B_{\sigma\rho}{}^\rho) +(l^{\mu\alpha}-\tfrac{1}{4}l\eta^{\mu\alpha})_{,\tau}\eta^{\tau\nu}+(l^{\nu\alpha}-\tfrac{1}{4}l\eta^{\nu\alpha})_{,\tau}\eta^{\mu\tau} -(l^{\mu\nu}-\tfrac{1}{4}l\eta^{\mu\nu})_{,\tau}\eta^{\tau\alpha}-\eta^{\mu\nu}(l^{\alpha\tau}-\tfrac{1}{4}l\eta^{\alpha\tau})_{,\tau}+h_{,\tau}{}^{\nu\alpha}h^{\tau\mu} -h^{\mu\nu}h_{,\tau}{}^{\alpha\tau}+h_{,\tau}{}^{\mu\alpha}h^{\tau\nu}-h_{,\tau}{}^{\mu\nu}h^{\alpha\tau}\big]_{,\alpha}-h_{,\alpha}{}^{\mu\tau}h_{,\tau}{}^{\alpha\nu} +h_{,\alpha}{}^{\mu\alpha}h_{,\tau}{}^{\nu\tau}-\big[h(h^{\rho\mu}\eta^{\alpha\nu}+h^{\alpha\nu}\eta^{\rho\mu}-h^{\mu\nu}\eta^{\alpha\rho}-\eta^{\mu\nu}h^{\rho\alpha})\big]_{,\rho\alpha} +\big[hh_{,\rho}(\eta^{\rho\mu}\eta^{\alpha\nu}-\eta^{\mu\nu}\eta^{\rho\alpha})\big]_{,\alpha}, \quad (3.7)$$

with

$$l^{\mu\nu}=h^{\mu}\rho h^{\rho\nu}. \quad (3.8)$$

$$2c^2k^{-1}E^{\mu\nu}=2(\eta^{\mu\sigma}\eta^{\nu\tau}-\tfrac{1}{2}\eta^{\mu\nu}\eta^{\sigma\tau})(A_{\sigma\lambda}{}^\rho A_{\rho\tau}{}^\lambda-A_{\sigma\tau}{}^\rho A_{\tau\lambda}{}^\lambda) -2(h^{\mu\sigma}\eta^{\nu\tau}+h^{\nu\tau}\eta^{\mu\sigma}-\tfrac{1}{2}h^{\mu\nu}\eta^{\sigma\tau}-\tfrac{1}{2}h^{\sigma\tau}\eta^{\mu\nu})_{,\alpha} \times(A_{\sigma\tau}{}^\alpha-A_{\sigma\rho}{}^\rho\delta_\tau{}^\alpha) +\big[h(k^{\rho\mu}\eta^{\alpha\nu}+k^{\alpha\nu}\eta^{\rho\mu}-k^{\mu\nu}\eta^{\alpha\rho}-\eta^{\mu\nu}k^{\rho\alpha})_{,\rho}\big]_{,\alpha} +h_{,\alpha}{}^{\mu\tau}h_{,\tau}{}^{\alpha\nu}-h_{,\alpha}{}^{\mu\alpha}h_{,\tau}{}^{\nu\tau}, \quad (3.9)$$

where, as before,

$$k^{\mu\nu}=h^{\mu\nu}-\tfrac{1}{2}\eta^{\mu\nu}h.$$

Equation (3.9) may be shown to be equivalent to

$$2c^2k^{-1}E^{\mu\nu}=-k_{\tau,\rho}{}^\mu k_{,\sigma}{}^{\tau\rho}\eta^{\sigma\nu}-k_{\tau,\rho}{}^\nu k_{,\sigma}{}^{\tau\rho}\eta^{\sigma\mu}+k_{\rho,\tau}{}^\mu k_{,\lambda}{}^{\nu\rho}\eta^{\tau\lambda} +k_{,\rho}{}^{\rho\tau}k_{,\tau}{}^{\mu\nu}-k_{,\rho}{}^{\mu\rho}k_{,\lambda}{}^{\nu\lambda}+\tfrac{1}{2}k_{,\sigma}{}^{\tau\rho}k_{\tau\rho,\lambda}\eta^{\sigma\mu}\eta^{\lambda\nu} -\tfrac{1}{4}h_{,\rho}h_{,\sigma}\eta^{\rho\mu}\eta^{\sigma\nu}-2hG_{(1)}{}^{\mu\nu} +\tfrac{1}{2}\eta^{\mu\nu}\big[k_{\alpha\sigma,\tau}k_{,\rho}{}^{\alpha\tau}\eta^{\rho\sigma}-\tfrac{1}{2}k_{\alpha\sigma,\tau}k_{,\rho}{}^{\alpha\sigma}\eta^{\rho\tau} +\tfrac{1}{4}h_{,\rho}h_{,\sigma}\eta^{\rho\sigma}\big]. \quad (3.10)$$

It follows from Eqs. (3.5) and (3.6) that

$$G_{(2),\nu}{}^{\mu\nu}=2c^2k^{-1}E_{,\nu}{}^{\mu\nu}. \quad (3.11)$$

In view of the field equations, that is Eqs. (1.10) and Eq. (2.16), this equation may be written as

$$(T_0{}^{\mu\nu}+kT_{(1)}{}^{\mu\nu}+E^{\mu\nu})_{,\nu}=0. \quad (3.12)$$

Equation (3.11) holds for arbitrary $h_{\mu\nu}$ and $h_{(2)\mu\nu}$, that is, it is an identity in these quantities.

Equations (3.12) are the approximate equations of motion of the matter represented by the tensor $T^{\mu\nu}$ [cf. Eq. (1.6)], correct to terms involving k^2. They are written in invariant form in the Minkowski space. If $h_{\mu\nu}$ is interpreted as a tensor in this space (approximately) representing the gravitational field created by the matter, then we may regard the last term in Eqs. (3.12) as the "stress-energy" tensor of the gravitational field.

In a later section, we shall evaluate the right-hand side of Eqs. (3.9) for a particular choice of the tensor $T_{(0)\mu\nu}$, that is, for a particular choice of $h_{\mu\nu}$. We first make some remarks concerning some consequences of these equations.

4. CONSERVATION EQUATIONS

We may write Eqs. (3.12) as

$$(M^{\mu\nu}+E^{\mu\nu})_{,\mu}=0, \quad (4.1)$$

where

$$M^{\mu\sigma}=T_{(0)}{}^{\mu\sigma}+kT_{(1)}{}^{\mu\sigma}. \quad (4.2)$$

Multiplying Eq. (4.1) by an arbitrary vector field of Minkowski space-time λ_ν and summing, we obtain

$$(\lambda_\nu(M^{\mu\nu}+E^{\mu\nu}))_{,\mu}-\tfrac{1}{2}(M^{\mu\nu}+E^{\mu\nu})(\lambda_{\mu,\nu}+\lambda_{\nu,\mu})=0,$$

since both $M^{\mu\nu}$ and $E^{\mu\nu}$ are symmetric.

When λ_μ is a Killing vector of Minkowski space-time, that is when

$$\lambda_{\mu,\nu}+\lambda_{\nu,\mu}=0, \quad (4.3)$$

we obtain the conservation equation

$$(\lambda_\nu(M^{\mu\nu}+E^{\mu\nu}))_{,\mu}=0. \quad (4.4)$$

As is well known, the general solution of Eqs. (4.3) is given by

$$\lambda_\mu=F_{\mu\nu}x^\nu+a_\mu \quad (4.5)$$

in a Galilean coordinate system in Minkowski space-time where

$$F_{\mu\nu}=-F_{\nu\mu}$$

and are independent of x^ρ and the a_μ are constants. There are thus ten linearly independent λ_μ and associated with each of these there is a conservation theorem of the form of Eq. (4.4).

The four-vectors which in a galilean coordinate system have the coordinate

$$\lambda_\mu{}^{(\alpha)}=\eta_{\alpha\mu}, \quad \alpha=1, 2, 3, 4 \quad (4.6)$$

will be said to be associated with the conservation of energy and momentum.

Equation (4.4) implies that

$$\int \lambda_\nu (M^{\mu\nu}+E^{\mu\nu}) n_\mu d^3v = 0, \tag{4.7}$$

where the integral is taken over a closed three-dimensional hypersurface in Minkowski space-time and

$$n_\mu d^3v = (-\eta)^{\frac{1}{2}} \epsilon_{\mu\nu\sigma\tau} \frac{\partial x^\nu}{\partial u}\frac{\partial x^\sigma}{\partial v}\frac{\partial x^\tau}{\partial w} du dv dw, \tag{4.8}$$

if u, v, and w are variables giving a parameterization of the hypersurface.

For use in later sections, we derive an equation based on the Bianchi identity

$$G_{;\nu}{}^{\mu\nu} = G_{|\nu}{}^{\mu\nu} + G^{\rho\nu}\left\{ {\mu \atop \rho\mu} \right\} + G^{\mu\rho}\left\{ {\nu \atop \mu\rho} \right\} = 0.$$

If this equation is differentiated twice with respect to k and k is set equal to zero, we obtain

$$G_{(2),\nu}{}^{\mu\nu} = -2G_{(1)}{}^{\rho\mu}\left\{ {\nu \atop \rho\nu} \right\}_{(1)} - 2G_{(1)}{}^{\nu\rho}\left\{ {\mu \atop \nu\rho} \right\}_{(1)},$$

that is,

$$G_{(2)\mu}{}^{\nu},_\nu = G_{(1)}{}^{\nu\rho} h_{\nu\rho,\mu} - h_{,\nu} G_{(1)\mu}{}^{\nu} - 2G_{(1)}{}^{\nu\rho} h_{\nu\mu,\rho}.$$

In view of Eqs. (1.10), (2.16), and (4.2), this equation may be written as

$$M_{\mu,\nu}{}^{\nu} = \tfrac{1}{2}k[T_{(0)}{}^{\nu\rho} h_{\nu\rho,\mu} - h_{,\nu} T_{(0)\mu}{}^{\nu} - 2T_{(0)}{}^{\nu\rho} h_{\rho\mu,\nu}]. \tag{4.9}$$

5. GAUGE INVARIANCE

In this section we shall discuss the effects of a transformation of coordinates in the Reimannian space-time on the tensors $h_{\mu\nu}$ and $h_{(2)\mu\nu}$ in the Minkowski space-time. We recall that, in any coordinate system,

$$h_{(n)\mu\nu} = (d^n g_{\mu\nu}/dk^n)_{k=0}. \tag{5.1}$$

Under the transformation of coordinates in the Reimannian space-time defined by the equations

$$y^\mu = y^\mu(x), \tag{5.2}$$

the tensor $g_{\mu\nu}$ transforms as

$$g_{\mu\nu}(x) = g_{\sigma\tau}{}^*(y(x)) y_{|\mu}{}^{\sigma} y_{|\nu}{}^{\tau}. \tag{5.3}$$

It then follows that, if the functions y^μ are independent of k, the quantities $h_{(n)\mu\nu}$ transform as tensors in the Minkowski space-time under the transformation given by Eqs. (5.2) where these are now interpreted as a transformation of coordinates in the Minkowski space-time.

If the functions y^μ depend on k, it follows from Eqs. (5.2) and (5.3) that

$$h_{(n)\mu\nu}{}^* \equiv (d^n g_{\mu\nu}{}^*/dk^n)_{k=0}$$

is not the tensor transform in the Minkowski space-time of $h_{(n)\mu\nu}$. In this case, Eqs. (5.2), which may be written as

$$y^\mu = y^\mu(x;k), \tag{5.4}$$

may be interpreted either as a transformation of coordinates for fixed k or as a congruence of curves for variable k.

It is sufficient to discuss the case where Eqs. (5.4) are such that

$$(y_{|\nu}{}^{\mu})_{k=0} = \delta_\nu{}^\mu. \tag{5.5}$$

That is,

$$y^\mu = x^\mu + k f^\mu(x;k). \tag{5.6}$$

For a general transformation of the form of (5.4) is obtained from (5.6) by following it by a transformation independent of k. Let us write

$$(d^n y^\mu/dk^n)_{k=0} = a_{(n)}{}^\mu \tag{5.7}$$

and set

$$a_{(1)}{}^\mu = a^\mu. \tag{5.8}$$

The functions a^μ are the components of a contravariant vector field, the vector field tangent to the congruence of curves (5.4) at the point x^μ. In fact, under the transformation of coordinates

$$x'^\sigma = g^\sigma(x),$$

with the definition

$$y'^\sigma = g^\sigma(y),$$

we have

$$a'^\sigma = \left(\frac{dy'^\sigma}{dk}\right)_{k=0} = \left(\frac{\partial g^\sigma}{\partial y^\tau}\right)_{k=0} a^\tau = \frac{\partial x'^\sigma}{\partial x^\tau} a^\tau.$$

However, the functions $a_{(2)}{}^\mu(x)$ do not have a vector transformation law. Indeed we have

$$a_{(2)}{}'^\sigma = \left(\frac{d^2y'^\sigma}{dk^2}\right)_{k=0} = \frac{\partial x'^\sigma}{\partial x^\tau} a_{(2)}{}^\tau + \frac{\partial^2 x'^\sigma}{\partial x^\rho \partial x^\tau} a^\rho a^\tau.$$

Note that the quantity

$$b^\sigma = a_{(2)}{}^\sigma + \left\{ {\sigma \atop \tau\rho} \right\} a^\rho a^\tau \tag{5.9}$$

does obey the transformation law of a vector.

It follows from Eqs. (5.3) and the definition (5.1) that under the transformation (5.4) subject to Eq. (5.5)

$$h_{\mu\nu}{}^* = h_{\mu\nu} - a_{\mu,\nu} - a_{\nu,\mu}, \tag{5.10}$$

$$\begin{aligned} h_{(2)\mu\nu}{}^* &= h_{(2)\mu\nu} - b_{\mu,\nu} - b_{\nu,\mu} - 2\eta_{\sigma\tau} a_{,\mu}{}^\sigma a_{,\nu}{}^\tau \\ &\quad - 2(h_{\mu\nu,\rho}{}^* a^\rho + h_{\mu\sigma}{}^* a_{,\nu}{}^\sigma + h_{\nu\sigma}{}^* a_{,\mu}{}^\sigma) \\ &= h_{(2)\mu\nu} - 2(h_{\mu\nu,\rho} a^\rho + h_{\mu\sigma} a_{,\nu}{}^\sigma + h_{\nu\sigma} a_{,\mu}{}^\sigma) \\ &\quad - (b_\mu - 2a_{\mu,\sigma} a^\sigma),_\nu - (b_\nu - 2a_{\nu,\sigma} a^\sigma),_\mu, \end{aligned} \tag{5.11}$$

where the vector b^μ is given by Eqs. (5.9) and we have made use of the fact that $\eta_{\sigma\tau}$ is the metric tensor of a flat space.

In case

$$h_{\mu\nu}=h_{(2)\mu\nu}=0, \qquad h_{\mu\nu}{}^{*}=-(a_{\mu,\nu}+a_{\nu,\mu}), \tag{5.12}$$

$$h_{(2)\mu\nu}{}^{*}=-(b_{\mu}-2a_{\mu,\sigma}a^{\sigma})_{,\nu}-(b_{\nu}-2a_{\nu,\sigma}a^{\sigma})_{,\mu}. \tag{5.13}$$

That is, even when the Reimannian space-time is flat but a non-Galilean coordinate system is used which arises from a Galilean one by a transformation of the type given by Eqs. (5.4), the quantities $h_{\mu\nu}$ and $h_{(2)\mu\nu}$ need not vanish. However, they are of the form given by Eqs. (5.12) and (5.13).

We shall call the transformation

$$h_{\mu\nu}\rightarrow h_{\mu\nu}{}^{*}, \qquad h_{(2)\mu\nu}\rightarrow h_{(2)\mu\nu}{}^{*},$$

where the $h_{\mu\nu}{}^{*}$ and $h_{(2)\mu\nu}{}^{*}$ are given by Eqs. (5.10) and (5.11) a gauge transformation. It is the transformation induced on these tensors by the coordinate transformation (5.4). When $h_{\mu\nu}{}^{*}$ and $h_{(2)\mu\nu}{}^{*}$ are substituted into Eqs. (2.13) and (3.4), we will obtain quantities we shall denote as $G_{(1)}{}^{\mu\nu*}$ and $G_{(2)}{}^{\mu\nu*}$. These are the coefficients of the first and second powers of k in the expansion of the tensor

$$R^{*\mu\nu}-\tfrac{1}{2}g^{\mu\nu*}R^{*}=G^{*\mu\nu},$$

which may be obtained from the tensor $G^{\mu\nu}$ by using the fact that $G^{*\mu\nu}$ arises from $G^{\mu\nu}$ by means of the transformation (5.4) and the transformation law

$$G^{*\mu\nu}(x)=G^{\sigma\tau}(x)y_{|\sigma}{}^{\mu}y_{|\tau}{}^{\nu}. \tag{5.14}$$

It follows from this equation by setting $k=0$, by differentiating with respect to k and setting $k=0$, and by differentiating twice with respect to k and setting $k=0$ that

$$G_{(0)}{}^{*\mu\nu}=G_{(0)}{}^{\mu\nu}=0, \qquad G_{(1)}{}^{*\mu\nu}=G_{(1)}{}^{\mu\nu}, \tag{5.15}$$

and

$$G_{(2)}{}^{*\mu\nu}=G_{(2)}{}^{\mu\nu}-2[G_{(1),\rho}{}^{\mu\nu}a^{\rho}-G_{(1)}{}^{\mu\tau}a_{,\tau}{}^{\nu}-G_{(1)}{}^{\tau\nu}a_{,\tau}{}^{\mu}]. \tag{5.16}$$

These equations may also be derived by substituting $h_{\mu\nu}{}^{*}$ and $h_{(2)\mu\nu}{}^{*}$ into the equations defining $G_{(1)}{}^{*\mu\nu}$ and $G_{(2)}{}^{*\mu\nu}$ as functions of these quantities. Since Eqs. (3.10) hold identically in $h_{\mu\nu}$ and $h_{(2)\mu\nu}$, we have

$$G_{(2),\tau}{}^{*\sigma\tau}=2c^{2}k^{-1}E_{,\tau}{}^{*\sigma\tau}=2c^{2}k^{-1}E_{,\tau}{}^{\sigma\tau}+2G_{(1)}{}^{\sigma\rho}a_{,\tau\rho}{}^{\tau}+2G_{(1)}{}^{\rho\tau}a_{,\tau\rho}{}^{\sigma} =2(c^{2}k^{-1}E^{\sigma\tau}+G^{\sigma\tau}a_{,\rho}{}^{\rho}+G^{\tau\rho}a_{,\rho}{}^{\sigma})_{,\tau} \tag{5.17}$$

as follows from Eqs. (5.16) and (2.16).

It is a consequence of Eqs. (5.15), (5.16), and (1.10) that

$$T_{(0)}{}^{*\mu\nu}+kT_{(1)}{}^{*\mu\nu}=T_{(0)}{}^{\mu\nu}+kT_{(1)}{}^{\mu\nu}-k[T_{(0),\rho}{}^{\mu\nu}a^{\rho}-T_{(0)}{}^{\mu\tau}a_{,\tau}{}^{\nu}-T_{(0)}{}^{\tau\nu}a_{,\tau}{}^{\mu}]. \tag{5.18}$$

We now define the vector

$$\lambda_{\mu}{}^{*}=\lambda_{\mu}-k(a^{\rho}\lambda_{\mu,\rho}+\lambda_{\rho}a_{,\mu}{}^{\rho}), \tag{5.19}$$

where λ_{μ} is one of the Killing vectors of Minkowski space-time, that is, λ_{μ} satisfies Eqs. (4.3). It may be verified as a consequence of Eqs. (5.18) and (5.19) that

$$\lambda_{\mu}{}^{*}M^{*\mu\nu}=\lambda_{\mu}M^{\mu\nu}-k(a^{\rho}(\lambda_{\mu}T_{(0)}{}^{\mu\nu})_{,\rho}-\lambda_{\mu}T_{(0)}{}^{\mu\rho}a_{,\rho}{}^{\nu}), \tag{5.20}$$

where terms in k^{2} have been neglected, and $M^{\mu\nu}$ and $M^{*\mu\nu}$ are defined by means of Eqs. (4.2) and the corresponding equations for the starred quantities.

If we multiply Eqs. (5.20) by $(1-ka_{,\sigma}{}^{\sigma})$ we then obtain to the same accuracy

$$(1-ka_{,\sigma}{}^{\sigma})\lambda_{\mu}{}^{*}M^{\mu\nu*}=\lambda_{\mu}M^{\mu\nu}-k[(a^{\rho}\lambda_{\mu}T_{(0)}{}^{\mu\nu})_{,\rho}-\lambda_{\mu}T_{(0)}{}^{\mu\rho}a_{,\rho}{}^{\nu}],$$

and hence

$$[(1-ka_{,\sigma}{}^{\sigma})\lambda_{\mu}{}^{*}M^{\mu\nu*}]_{,\nu}=(\lambda_{\mu}M^{\mu\nu})_{,\nu}-k[a^{\nu}(\lambda_{\mu}T_{(0)}{}^{\mu\rho})_{,\rho}]_{,\nu}=(\lambda_{\mu}M^{\mu\nu})_{,\nu}. \tag{5.21}$$

The first form of Eq. (5.21) holds for an arbitrary vector λ_{μ}. The second form of this equation follows from the first form by virtue of the fact that λ_{μ} is a Killing vector and $T_{(0)}{}^{\mu\nu}$ has a vanishing divergence.

It follows from Eq. (5.21), by integration over a region of Minkowski space-time bounded by a closed three-dimensional hypersurface, that

$$\int(1-ka_{,\sigma}{}^{\sigma})\lambda_{\mu}{}^{*}M^{\mu\nu*}n_{\nu}d^{3}v=\int\lambda_{\mu}M^{\mu\nu}n_{\nu}d^{3}v =-\int\lambda_{\mu}E^{\mu\nu}n_{\nu}d^{3}v. \tag{5.22}$$

It is of course a consequence of the first of Eqs. (5.17) that

$$\int\lambda_{\mu}M^{\mu\nu*}n_{\nu}d^{3}v=-\int\lambda_{\mu}E^{\mu\nu*}n_{\nu}d^{3}v. \tag{5.23}$$

The difference between the surface integral of $E^{\mu\nu*}$ and that of $E^{\mu\nu}$ is due to the fact that the hypersurface in Minkowski space-time into which the hypersurface defined by the equations

$$x^{\mu}=x^{\mu}(u,v,w)$$

transforms under the transformation defined by Eqs. (5.4) differs from the former one. Thus we see that, although the gravitational energy tensor $E^{\mu\nu}$ is not gauge invariant, the conserved quantities computed from it are related by Eqs. (5.22) which take into account the fact that the gauge transformations arise from coordinate transformations in the Riemannian space-time.

6. BONDI'S RELATION

In this section we shall compare Eq. (4.4) with

$$\lambda_\nu=\eta_{4\nu},$$

where the $\eta_{\mu\nu}$ are evaluated in a Galilean coordinate system, to a set of equations first derived by Bondi[1] from classical arguements. We shall derive his relations by studying the classical limit of the Einstein field equations for weak fields for the case where $T_{(0)}{}^{\mu\nu}$ is the stress energy tensor of a perfect fluid. In forming the classical limit we shall neglect terms involving $1/c^2$. The tensor $E^{\mu\nu}$ will also be evaluated for this special case in the classical limit.

McVittie[2] has given the $h_{\mu\nu}$ associated via Eqs. (1.10) with such a $T_{(0)}{}^{\mu\nu}$. In the notation used above, McVittie's results may be written as follows:

In a Galilean coordinate system, let

$$h_{\mu\nu}=-2\varphi\delta_\mu{}^4\delta_\nu{}^4+(\varphi+2g_{(\mu)}/c^2)\eta_{\mu\nu}, \tag{6.1}$$

then

$$h=2(\varphi+\psi/c^2), \tag{6.2}$$

where

$$\psi=\sum_{\mu=1}^{4} g_{(\mu)} \tag{6.3}$$

and

$$k_{\mu\nu}=-2\varphi\delta_\mu{}^4\delta_\nu{}^4-2c^{-2}\psi_{(\mu)}\eta_{\mu\nu} \tag{6.4}$$

with

$$2\psi_{(\mu)}=\psi-2g_{(\mu)}, \tag{6.5}$$

hence

$$\sum_{\mu=1}^{4}\psi_{(\mu)}=\psi. \tag{6.6}$$

It may be verified from Eqs. (2.13) that

$$G_{(1)}{}^{44}=c^2\delta^{ij}\varphi_{,ij}+\psi_{(4),ij}\delta^{ij}+\sum_{(l)}\psi_{(l),ll},$$

$$G_{(1)}{}^{4i}=-c^2\varphi_{,i4}-\psi_{(4),i4}-\psi_{(i),i4},$$

$$G_{(1)}{}^{ii}=c^2\varphi_{,44}+\psi_{(4),44}-c^2\psi_{(i),\rho\sigma}\eta^{\rho\sigma}+2\psi_{(i),ii}-c^2\Big(\sum_l\psi_{(l),ll}\Big),$$

$$G_{(1)}{}^{ij}=c^2(\psi_{(i)}+\psi_{(j)})_{,ij}\quad i\neq j.$$

The tensor $T_{(0)}{}^{\mu\nu}$ may now be calculated from Eq. (1.10) where the above quantities are used for $G_{(1)}{}^{\mu\nu}$. If in the resulting equations we neglect the terms in $1/c^2$, we obtain for the classical limit

$$\begin{aligned}
&T_{(0)}{}^{44}=-\delta^{ij}\varphi_{,ij}=\rho,\\
&T_{(0)}{}^{4i}=\varphi_{,i4}=\rho U_i,\\
&T_{(0)}{}^{ii}=-\varphi_{,44}-2\psi_{(i),ii}+\psi_{(i),kl}\delta^{kl}+\sum_{(l)}\psi_{(l),ll}=\rho U_i{}^2+p,\\
&T_{(0)}{}^{ij}=-(\psi_{(i)}+\psi_{(j)})_{,ij}=\rho U_iU_j.
\end{aligned} \tag{6.7}$$

[2] G. C. McVittie, *General Relativity and Cosmology* (Chapman and Hall, Ltd., London, England, 1956), Secs. 6.1 and 6.2.

The extreme right-hand sides of Eqs. (6.7) are obtained from the classical limit of the relativistic stress energy tensor of a fluid

$$T^{\mu\nu}=\rho\left(1+\frac{\epsilon}{c^2}+\frac{p}{\rho c^2}\right)u^\mu u^\nu-\frac{p}{c^2}\eta^{\mu\nu},$$

with

$$u^4=\frac{1}{(1-v^2/c^2)^{\frac{1}{2}}},\quad u^i=\frac{U_i}{(1-v^2/c^2)^{\frac{1}{2}}},\quad v^2=\sum_i U_i{}^2.$$

The quantities $\psi_{(i)}$ are not arbitrary but must be chosen so that a set of equations called consistency equations by McVittie must be satisfied. These equations are determined from the requirement that the ten equations (5.7) determine the five quantities ρ, p, and U_i. When these are satisfied we find

$$\begin{aligned}
&\rho=-\varphi_{,ij}\delta^{ij},\\
&\rho U_i=+\varphi_{,i4},\quad i,j=1,2,3,\\
&p=-\varphi_{,44}+\chi,
\end{aligned} \tag{6.8}$$

where χ is determined by the integrable equations

$$\chi_{,l}=\sum_{i=1}^{3}\left(\frac{\varphi_{,4l}\varphi_{,4j}}{\varphi_{,kl}\delta^{kl}}\right)_{,j},\quad l=1,2,3. \tag{6.9}$$

We next evaluate the right-hand side of Eq. (4.9) in the classical limit, that is, by substituting from Eq. (6.7) for $T_{(0)}{}^{\mu\nu}$ and from Eqs. (6.1) and (6.2) with the terms in $1/c^2$ omitted for $h_{\mu\nu}$ and h.

Equation (4.9) then becomes

$$M_{\alpha,}{}^{\mu}=\tfrac{1}{2}k[\varphi_{,\alpha}T_{(0)}{}^{\mu\rho}(-2\delta_\mu{}^4\delta_\rho{}^4+\eta_{\mu\rho})-4\varphi_{,\mu}(T_{(0)\alpha}{}^{\mu}-T_{(0)}{}^{4\mu}\delta_\alpha{}^4)]. \tag{6.10}$$

If we set $\alpha=4$ in this equation we obtain, on neglecting terms in $1/c^2$ inside the parentheses,

$$M_{4,}{}^{\mu}=-\tfrac{1}{2}k\varphi_{,4}T_{(0)}{}^{44}=\tfrac{1}{2}k\varphi_{,4}\delta^{ij}\varphi_{,ij}. \tag{6.11}$$

The first of these equations may be written as

$$\begin{aligned}
M_{4,}{}^{\mu}&=-\tfrac{1}{2}k[(\varphi T_{(0)}{}^{44})_{,4}-\varphi T_{(0),}{}^{44}]\\
&=-\tfrac{1}{2}k[(\varphi T_{(0)}{}^{44})_{,4}+\varphi T_{(0),i}{}^{4i}]\\
&=-\tfrac{1}{2}k[(\varphi T_{(0)}{}^{44})_{,4}+(\varphi T_{(0)}{}^{4i})_{,i}-\varphi_{,i}T_{(0)}{}^{4i}],
\end{aligned} \tag{6.12}$$

since

$$T_{(0),\nu}{}^{4\nu}=T_{(0),4}{}^{44}+T_{(0),i}{}^{4i}=0.$$

The second of Eqs. (6.11) may be written as

$$\begin{aligned}
M_{4,}{}^{\mu}&=\tfrac{1}{2}k[(\varphi_{,4}\varphi_{,j}-\varphi\varphi_{,4j})_{,i}\delta^{ij}-\varphi_{,4i}\varphi_{,j}\delta^{ij}+(\varphi\varphi_{,4j})_{,i}\delta^{ij}]\\
&=\tfrac{1}{2}k[(\varphi_{,4}\varphi_{,j}-\varphi\varphi_{,4j})_{,i}\delta^{ij}-\varphi_{,i}T_{(0)}{}^{4i}+(\varphi T_{(0)}{}^{4i})_{,i}],
\end{aligned} \tag{6.13}$$

where we have used Eqs. (6.7). Subtracting Eq. (6.13) from (6.12), we then obtain

$$\varphi_{,i}T_{(0)}{}^{4i}=\tfrac{1}{2}(\varphi_{,4}\varphi_{,j}-\varphi\varphi_{,4j})_{,i}\delta^{ij}+(\varphi T_{(0)}{}^{4i})_{,i}+\tfrac{1}{2}(\varphi T_{(0)}{}^{44})_{,4}.$$

The scalar φ is related to the Newtonian potential V by the equation

$$V=4\pi G\varphi,$$

as is evident from the first of Eqs. (6.7). Hence, the above equation may be written as

$$\rho U_i V_{,j}\delta^{ij}=(8\pi G)^{-1}(V_{,4}V_{,i}-VV_{,4i})_{,j}\delta^{ij} +(\rho V U_i)_{,j}\delta^{ij}+\tfrac{1}{2}(\rho V)_{,4}. \quad (6.14)$$

Equation (6.14) has been derived by Bondi from purely classical arguments and has led him to suggest that the vector

$$p_i=(8\pi)^{-1}(V_{,4}V_{,i}-VV_{,4i})$$

is the gravitational analogue of the Poynting vector. That is, it represents the momentum of the gravitational radiation through unit area of a surface exterior to the moving matter.

When Eqs. (6.12) and (6.13) are added we obtain

$$M_{4,\mu}{}^{\mu}=\tfrac{1}{4}k[(\varphi\varphi_{,ij}\delta^{ij})_{,4}+(\varphi_{,4}\varphi_{,j}-\varphi\varphi_{,4j})_{,i}\delta^{ij}]$$

or

$$M_{4,\mu}{}^{\mu}=c^{-2}[-(\tfrac{1}{2}\rho V)_{,4} +(8\pi G)^{-1}(V_{,4}V_{,j}-VV_{,4j})_{,i}\delta^{ij}]. \quad (6.15)$$

This equation relates the four-dimensional divergence of the energy-momentum of the material field with the time rate of change of the potential energy of the mass distribution and the divergence of the vector p_i.

We shall compare Eq. (6.15) to the equation resulting from Eq. (4.4) for the case of a perfect fluid by choosing λ_ν as mentioned above. To do this we substitute for $k_{\mu\nu}$ from Eqs. (6.4), with the $1/c^2$ terms omitted, into Eqs. (3.10). We then obtain, on neglecting $1/c^2$ terms

$$E^{44}=-\tfrac{1}{2}k[+\tfrac{7}{2}\varphi_{,i}\varphi_{,j}\delta^{ij}+4\varphi\varphi_{,ij}\delta^{ij}],$$

or

$$E^{44}=-c^{-2}[-4V\rho+\tfrac{7}{8}(\pi G)^{-1}V_{,i}V_{,j}\delta^{ij}] =-c^{-2}[-\tfrac{1}{2}\rho V+\tfrac{7}{8}(\pi G)^{-1}(VV_{,i})_{,j}\delta^{ij}], \quad (6.16)$$

$$E^{4i}=\tfrac{1}{2}k[3\varphi_{,4}\varphi_{,i}+4\varphi\varphi_{,i4}],$$

or

$$E^{4i}=c^{-2}(4\pi G)^{-1}[3V_{,4}V_{,i}+4VV_{,i4}] =-c^{-2}[(8\pi G)^{-1}(V_{,4}V_{,i}-VV_{,4i}) -\tfrac{7}{8}(\pi G)^{-1}(VV_{,i})_{,4}]. \quad (6.17)$$

On substituting Eqs. (6.16) and (6.17) into Eq. (4.4), we obtain Eq. (6.15), since the time derivative of the second term in Eq. (6.16) cancels the space divergence of the second term of Eq. (6.17).

Thus we see that, in the classical limit, the gravitational stress energy tensor introduced above leads to a conservation of energy equation identical to that proposed by Bondi. However, the energy density and the analogue of the Poynting vector, the quantities E^{44} and E^{4i} differ somewhat from those he proposed. The differences are: The former quantity contains a term proportional to a three dimensional divergence (the three dimensional Laplacean operating on V^2) and the latter contains a term proportional to the time derivative of a gradient, (the gradient of V^2).

The derivation given above enables one to obtain the stresses E^{ij} as well as the quantities E^{44} and E^{4i}. These may be calculated by substituting from Eq. (6.4) into Eqs. (3.10) and neglecting terms involving $1/c^2$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Principles of Limiting Absorption and Limiting Amplitude in Scattering Theory. I. Schrödinger's Equation*

FAROUK M. ODEH
International Business Machines Corporation, Research Center, Yorktown Heights, New York
(Received July 11, 1961)

The outgoing solution of the time-independent Schrödinger equation, with a suitably restricted real potential, is shown to be the uniform limit of the square-integrable solutions of the same equation with complex energy as the imaginary part of the energy tends to zero. Under further restrictions on the potential, it is also shown that the solution to the initial-value problem for the time-dependent Schrödinger equation tends to the outgoing solution as time increases indefinitely.

1. INTRODUCTION

THE scattering problems which arise in classical quantum mechanics may be described in terms of an unperturbed wave incident upon a potential which is switched on at a certain instant of time $t=0$. Therefore, the scattered wave, which exists for $t\geq 0$, satisfies an initial value problem for the time-dependent Schrödinger equation. If the incident wave is time-harmonic, it is natural to assume that the scattered wave becomes time-harmonic with the same frequency as time becomes infinite. Then the steady-state solution, which is the limiting form of the scattered wave, will satisfy the time-independent Schrödinger equation. The question then arises as to what conditions should be imposed at infinity to ensure the uniqueness of the solution of this equation.

Three principles seem to be available for determining a unique solution of a scattering problem. These are:

(1) The radiation principle, which is that the solution should satisfy Sommerfeld's radiation condition.

(2) The principle of "limiting absorption" (Ignatowsky[1]) which states, in the case of Schrödinger's equation with energy $E>0$, that the solution is to be obtained as the limit of the bounded solutions of the same equation with energy $E+i\epsilon$ as the parameter ϵ tends to zero through positive values (assuming the time dependence is of the form $e^{-i\omega t}$). The principle has its roots in the theory of scattering of electromagnetic waves where it is known that the media through which such waves propagate are usually "absorptive." This means that in actual scattering phenomena, which are characterized by the fact that the wave number k does have a small positive imaginary part, energy is always dissipated. Consequently, it is expected that boundedness or vanishing conditions at infinity would insure the uniqueness of the solutions to such problems. The principle of limiting absorption states, roughly, that in order to solve the idealized scattering problem where no absorption occurs we must take the limit of the "physically attainable" solutions as the absorption approaches zero.

(3) The principle of "limiting amplitude." This consists of solving the initial value problem and then studying the limit of such a solution as time increases indefinitely. It is supposed that any part of the solution whose time dependence is different from that of the incident time-harmonic wave has a transient character, i.e., it will die out as time increases. Consequently, a stationary state with the same time dependence as the incident wave will emerge as the desired correct solution. This is physically the most reasonable of all three principles, though its applicability is rather limited.

In a previous paper[2] we considered the Dirichlet-Neumann problem associated with the three-dimensional Schrödinger wave equation with a suitably restricted real potential $V(\mathbf{r})$

$$L\psi\equiv[\Delta+k^2-V(\mathbf{r})]\psi(\mathbf{r})=0, \quad k>0, \qquad (1.1)$$

and proved that Sommerfeld's radiation condition does specify the solution uniquely. The primary aim of this paper is to apply the latter two principles to the same equation. Our main results are:

(i) The radiating solution of $L\psi=f$, where ψ is defined in the whole space or in the exterior of a regular closed bounded surface Σ and f is a suitably restricted function, exists and is the uniform limit of the L_2 solutions of the Schrödinger equation with complex energy k^2 as $\mathrm{Im}(k^2)\to 0$ through positive values.

(ii) Under rather severe restrictions on the operator $(-\Delta+V)$, we show that the radiating solution is the limit, as $t\to\infty$, of the solution to the time-dependent problem. Thus, we have shown the equivalence of the three methods described above for characterizing the solution.

The above results extend some of the results of Povzner[3] and Ladyzhenskaia[4] although the methods used are similar. Ladyzhenskaia proved that the time-dependent problem (for a wave-like equation) leads as time increases to the radiating solution under the assumptions that $(-\Delta+V)$ has no eigenvalues and that V is continuously differentiable ($C^{(1)}$) and is of com-

* This work is based on a part of the writer's dissertation, submitted to the Mathematics Department, University of California, Berkeley, and was supported in part by the Office of Naval Research.

[1] W. Ignatowsky, Ann. Phys. **18**, 495 (1905).

[2] C. Zemach and F. Odeh, Arch. Rational Mech. and Analysis **5**, 226 (1960).

[3] A. Y. Povzner, Mat. Sbornik **32** (74), 109 (1953).

[4] O. A. Ladyzhenskaia, Uspekhi Mat. Nauk **12** (3), 161 (1957).

pact support.[5] Povzner assumed that $V\in C^{(1)}$ and is $O(r^{-\mu})$, $\mu>3.5$ and proved the existence of the solution to the integral equation derived from $LG(\mathbf{r},\mathbf{s})=\delta(\mathbf{r}-\mathbf{s})$ (the so-called second resolvent equation) and its continuous dependence on k if k^2 does not belong to the point spectrum of $(-\Delta+V)$.[5] Assuming V to be also radial, Povzner proved that the point spectrum is contained in the nonpositive real axis. Recently Ikebe[6] proved, independently, the existence of G for all $k^2>0$ assuming V to be square integrable, Hoelder continuous (except for a finite number of singularities), and $O(r^{-\mu})$, $\mu>2$ at infinity. The conditions we impose on V are stated in Sec. 2 below and are less restrictive than any of the above mentioned.

The main restriction referred to in (ii) is that the operator $(-\Delta+V)$ have no point eigenvalues (i.e., bound states). If the potential were switched on at $t=-\infty$, as is usually done in scattering theory, one would expect that the exclusion of bound states would be unnecessary. A similar remark applies to the problem treated by Ladyzhenskaia.[4]

We outline the plan of the paper. In Sec. 2 it is shown, following Povzner,[3] that the question of the existence of G reduces to studying an inhomogeneous integral equation with a Fredholm alternative property and we study the homogeneous equation. In Sec. 3 the existence of the solution to $L\psi=f$ and its continuous dependence on k are proved. In Sec. 4, the time-dependent problem is reduced to a time-independent one by means of a Laplace transform, and the validity of the principle of limiting amplitude is shown. In Part II we prove that similar results to (i) and (ii) above hold also for the wave equation in an inhomogeneous medium.[6a]

2. THE RESOLVENT OPERATOR $[-\Delta+V-k^2]^{-1}$

Assumptions and Preliminaries

We shall use the same notation which was employed in Zemach and Odeh,[2] hence we define

$$I(\mathbf{r})=\int_D \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s},$$

where D is the whole space E_3 (or the exterior of a bounded closed regular surface Σ) and assume:

$$\text{(i)}\quad I(\mathbf{r})\to 0 \quad \text{as} \quad r\to\infty. \tag{2.1a}$$

$$\text{(ii)}\quad V(\mathbf{r}) \text{ is locally Hoelder-continuous}^7 \text{ and is bounded near infinity.} \tag{2.1b}$$

It is clear that (i), (ii) imply

$$\text{(iii)}\quad \max_{\mathbf{r}\in D} I(\mathbf{r})<\infty. \tag{2.1c}$$

Under these conditions the operator $H\equiv-\Delta+V$ has a unique self-adjoint extension defined on L_2 (Kato[8]) which we denote again by H. In scattering problems we usually ask for the solution ψ of $(H-E)\psi=f$, where E is the energy, which may be real or complex, and f is a known smooth function. If E is not real, then the inverse $(H-E)^{-1}$ exists in L_2 and depends analytically on E. In fact, if $\mathrm{Im}E>0$ and $k=\sqrt{E}$ is so defined that $\mathrm{Im}k>0$ then it is known (Povzner[3]) that $(H-E)^{-1}$ is an integral operator with a kernel $h(\mathbf{r},\mathbf{s};E)\equiv G(\mathbf{r},\mathbf{s};k)$ which is symmetric and, for a.e. $\mathbf{s}$,[9] belongs to $L_2(\mathbf{r})$ and satisfies

$$\left.\begin{aligned} &G(\mathbf{r},\mathbf{s},k)=G_0(\mathbf{r},\mathbf{s},k)-\int G_0(\mathbf{r},\mathbf{t})V(\mathbf{t})G(\mathbf{t},\mathbf{s})d\mathbf{t},\\ &\text{where}\\ &G_0(\mathbf{r},\mathbf{s})=\frac{1}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}. \end{aligned}\right\} \tag{2.2}$$

Noticing that (2.2) is the same equation which G would satisfy if $k^2>0$ and G is outgoing at infinity, we see that the problem of the existence of radiating solutions reduces to investigating the solvability of (2.2) when k^2 is real. Moreover, since both L_2 and radiating solutions are $O(r^{-1})$ at infinity by Theorem 1 in reference 2[10] we may restrict our solutions to the Banach space B of all functions ψ, which are continuous on D and which tend to zero uniformly at infinity. A natural norm on B is

$$\|\psi\|_B=\max_{\mathbf{r}\in D}|\psi(\mathbf{r})|. \tag{2.3}$$

Let the operator T_k be defined on B by

$$T_k\psi(\mathbf{r})=-\frac{1}{4\pi}\int\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}. \tag{2.4}$$

Since we have assumed that $\mathrm{Im}E\geqq 0$ we see that k lies in the first quadrant Q of the complex k plane. Equation (2.2) may now be written, for a.e. $\mathbf{s}$, in the form

$$\psi(\mathbf{r})=\phi(\mathbf{r})+T_k\psi(\mathbf{r}), \quad k\in Q \tag{2.2a}$$

where $\phi(\mathbf{r})$ is a known function.

The following lemmas demonstrate that if $k\in Q$ and $k^2\neq 0$ is not a negative eigenvalue of H then the inverse $(1-T_k)^{-1}$ exists as a bounded operator on B.

[5] Additional hypotheses should have been assumed in Ladyzhenskaia's paper, since Povzner's proof fails in the case $k=0$. A counter example is given in Sec. 2, and the correct hypotheses are stated in Lemma 4.1.

[6] T. Ikebe, Arch. Rational Mech. and Analysis **5**, 1–34 (1960).

[6a] F. M. Odeh, J. Math. Phys. **2**, 800 (1961).

[7] This may be easily extended to a potential having a finite number of square-integrable singularities. We then need to modify slightly the proofs of Lemma 2.1 and Lemma 2 in reference 2.

[8] T. Kato, Trans. Am. Math. Soc. **70**, 195 (1951).

[9] a.e. means almost every or almost everywhere.

[10] Theorems 1 and 2 in reference 2 have been proved for radiating wave functions only but they are valid for L_2 eigenfunctions also; see Odeh (reference 12).

Lemma 2.1: The operator T_k is a bounded linear operator on B to B.

Proof: Since

$$|T_k\psi(\mathbf{r})| \leqq (4\pi)^{-1}\|\psi\| \int \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s} \leqq C\|\psi\|,$$

it follows that $\|T_k\|$ is bounded. It is clear that $T_k\psi(r) \to 0$ uniformly as $r \to \infty$ by (2.1a). To prove the continuity in $\mathbf{r}$ of $T_k\psi$, consider

$$\int\left[\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}-\frac{e^{ik|\mathbf{r}'-\mathbf{s}|}}{|\mathbf{r}'-s|}\right]V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}$$

$$=\int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}-e^{ik|\mathbf{r}'-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}$$

$$+\int e^{ik|\mathbf{r}'-\mathbf{s}|}\left[\frac{1}{|\mathbf{r}-s|}-\frac{1}{|\mathbf{r}'-\mathbf{s}|}\right]V\psi d\mathbf{s}=I_1+I_2.$$

Now

$$|I_1| \leqq k|\mathbf{r}-\mathbf{r}'|\cdot\|\psi\|\int \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s}=O[\|\psi\|\cdot|\mathbf{r}-\mathbf{r}'|].$$

Let $B(\mathbf{r},\delta)$ denote the sphere $|\mathbf{s}-\mathbf{r}| \leqq \delta$, and let $\delta=|\mathbf{r}-\mathbf{r}'|^{\frac{1}{2}}$, then

$$|I_2| \leqq \|\psi\|\int_{D-B(\mathbf{r},\delta)} |V(\mathbf{s})|\left[\frac{1}{|\mathbf{r}-\mathbf{s}|}-\frac{1}{|\mathbf{r}'-\mathbf{s}|}\right]d\mathbf{s}$$

$$+\|\psi\|\int_{B(\mathbf{r},\delta)} |V(s)|\left[\frac{1}{|\mathbf{r}-\mathbf{s}|}+\frac{1}{|\mathbf{r}'-\mathbf{s}|}\right]d\mathbf{s}$$

$$\leqq \|\psi\||\mathbf{r}-\mathbf{r}'|^{\frac{1}{2}}\int \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s}+O[\|\psi\|\cdot|\mathbf{r}-\mathbf{r}'|]$$

$$=O[\|\psi\|\cdot|\mathbf{r}-\mathbf{r}'|^{\frac{1}{2}}] \quad \text{as} \quad \mathbf{r}\to\mathbf{r}'.$$

Therefore, $T_k\psi(r)$ is Hoelder-continuous.

Lemma 2.2: The operator T_k is compact on B.

Proof: Let $\{\psi_n\}$ be a sequence in B such that $\|\psi_n\| \leqq 1$. The proof of Lemma 2.1 shows that $\{T_k\psi_n\}$ is uniformly bounded, equicontinuous, and tends to zero uniformly with respect to n as $r \to \infty$. We can therefore select a subsequence of $\{T_k\psi_n\}$ which converges uniformly in $\mathbf{r}$ to a function in B.

We discuss now the homogeneous equation $\psi=T_k\psi$.

Lemma 2.3: Let $\psi \in B$ be a solution of

$$\psi=T_k\psi=-(4\pi)^{-1}\int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}, \quad (2.5)$$

where $\mathrm{Im}k=b \geqq 0$. Then $\psi(\mathbf{r})$ is $O(r^{-1}e^{-br})$ at infinity.

Proof: If ψ satisfies (2.5) then it is Hoelder-continuous by Lemma 2.1, hence $V\psi$ is Hoelder-continuous also. In this case (2.5) is equivalent to (Titchmarsh[11]):

$$[\Delta+k^2-V]\psi=0.$$

Using the notation of Theorem 1 in reference 2 we then have

$$\psi(\mathbf{r})=\psi_0(\mathbf{r})+\int_{D_0} G_0V\psi d\mathbf{s} \quad \mathbf{r}\in D_0, \quad (2.6)$$

where ψ_0 is a surface integral over Σ_{R_1} and is $O(r^{-1}e^{-br})$. The method of the proof of the theorem mentioned above shows that ψ itself is of the same order.

Remarks: Since the representations (2.5) and (2.6) are valid for any L_2 eigenfunctions of $-\Delta+V$ (Povzner[3], Lemma 7, Chap. II) we deduce from the above lemma that an eigenfunction of $[\Delta+E-V]\psi=0$ is $O(r^{-1}e^{-br})$ at infinity, where $E=-b^2$. If $E>0$ then there exist no eigenfunctions.[12]

Lemma 2.4: If ψ is a solution of (2.5), then it satisfies the radiation condition

$$r\left|\frac{\partial\psi}{\partial r}-ik\psi\right| \leqq M\epsilon(r), \quad (2.7)$$

where $\epsilon(r) \to 0$ as $r \to \infty$ and does not depend on ψ, and M is an upper bound for $|r\psi(\mathbf{r})|$.

Moreover, if $k^2>0$, then ψ vanishes identically.

Proof: Because of the corollary to Theorem 3 in reference 2 it is enough to prove (2.7). Consider

$$-4\pi\left[\frac{\partial\psi}{\partial r}-ik\psi\right]=ik\int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V\psi\left[\frac{\partial|\mathbf{r}-\mathbf{s}|}{\partial r}-1\right]d\mathbf{s}$$

$$-ik\int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|^2}V\psi\frac{\partial|\mathbf{r}-\mathbf{s}|}{\partial r}d\mathbf{s}=ikI_1+I_2.$$

Let $I_1=J_1+J_2$, where J_1 denotes the integral over the sphere $s<r^{\frac{1}{2}}$. To estimate J_1 we notice that

$$\frac{\partial|\mathbf{r}-\mathbf{s}|}{\partial r}-1=\left(1-\frac{s}{r}\cos\theta\right)\left[1-\frac{2s}{r}\cos\theta+\frac{s^2}{r^2}\right]^{-\frac{1}{2}}-1, \quad (2.8)$$

where θ is the angle between $\mathbf{r}$, $\mathbf{s}$.

The right side of (2.8) is $O(s^2r^{-2})$, hence there exists a constant C_1 such that

$$|J_1| \leqq C_1r^{-\frac{1}{2}}\int \frac{|V(\mathbf{s})\psi(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s} \leqq C_2r^{-\frac{1}{2}}. \quad (2.9)$$

[11] E. C. Titchmarsh, *Eigenfunction Expansions* (Oxford University Press, New York, 1958), Pt. II, Chap. 22 and pp. 10–11.

[12] F. Odeh, Technical Report No. 8, Department of Mathematics, University of California, August, 1960 (Chap. I).

Also

$$|J_2| \leqq \int_{s>r^{\frac{1}{2}}} \frac{|V(\mathbf{s})\psi(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|} d\mathbf{s} \leqq 2M \int \frac{|V(\mathbf{s})|}{s|\mathbf{r}-\mathbf{s}|} d\mathbf{s} \leqq \frac{2M}{r}\epsilon_1(r), \quad (2.10)$$

where $\epsilon_1(r) \to 0$ as $r \to \infty$.

A similar estimate holds for I_2 and the radiation condition (2.7) follows immediately.

Lemma 2.5: If $k^2 \neq 0$ is not a negative eigenvalue of H then the inverse $(I-T_k)^{-1}$ exists in B.

Proof: Since T_k is compact, the Riesz-Schauder theory (Dunford and Schwartz[13]) shows that $(I-T_k)^{-1}$ exists for all k such that the equation $\psi = T_k\psi$ implies that $\psi \equiv 0$. In view of Lemma 2.4 we may assume $\mathrm{Im}k > 0$ and hence Lemma 2.3 shows that ψ would be an L_2 solution of $[\Delta + k^2 - V]\psi = 0$. Since H is self-adjoint, ψ vanishes identically unless k^2 is a negative eigenvalue.

A counter-example: The case $k=0$ must be excluded in the preceding lemma (and in Povzner,[3] Theorem 2, Chap. II).

Consider

$$\psi(\mathbf{r}) = \begin{cases} 3-3r+r^2 & r \leqq 1 \\ r^{-1} & r \geqq 1 \end{cases}$$

$$V(\mathbf{r}) = \begin{cases} \dfrac{6(1-r^{-1})}{3-3r+r^2} & r \leqq 1 \\ 0 & r \geqq 1. \end{cases}$$

Then ψ is a regular solution of the homogeneous equation $\psi = T_0\psi$ and yet is not square integrable.

3. PRINCIPLE OF LIMITING ABSORPTION

In order to apply the results of the last section to the equation (2.2) we rewrite it in the form

$$G = G_0 + T_k G \quad (3.1)$$

and iterate (3.1); therefore we define

$$G_1 = G - G_0, \quad A_1 = T_k G_0.$$

Then G_1 satisfies

$$G_1 = A_1 + T_k G_1. \quad (3.2)$$

Lemma 3.1: $A_1(\mathbf{r},\mathbf{s}) \in B(\mathbf{r})$ for all $\mathbf{s}$.

The proof is similar to that of Lemma 2.1 and will be omitted.

We are now in a position to prove the existence of G and exhibit some of its properties.

Lemma 3.2: Suppose $k^2 \neq 0$ is not a negative eigenvalue of H. Then

(1) The kernel equation (3.1) has a unique solution $G(\mathbf{r},\mathbf{s})$ which is continuous in $\mathbf{r}$ except if $\mathbf{r}=\mathbf{s}$. G is the Green's function for the operator $[\Delta + k^2 - V]$ and behaves asymptotically like $O[r^{-1}e^{-br}]$, where $b = \mathrm{Im}k$.

(2) If $k^2 > 0$ and s is bounded, then G satisfies the radiation condition.

$$r[(\partial G/\partial r) - ikG] \to 0$$

uniformly with respect to $\mathbf{s}$, as $r \to \infty$. (3.3)

Proof: (1) Let G_1 be the unique solution of (3.2). Then set $G = G_0 + G_1$. G is again Hoelder continuous in $\mathbf{r}$, for for $\mathbf{r} \neq \mathbf{s}$, and Eq. (3.1) is then equivalent to $[\Delta + k^2 - V] \times G(\mathbf{r},\mathbf{s}) = -\delta(\mathbf{r}-\mathbf{s})$. The asymptotic behavior is proved exactly as in Lemma 2.3 (and Theorem 1 in reference 2). We notice here that if we assume that s is bounded (and smaller than R_1 of Lemma 2 in reference 2) then the order is uniform with respect to $\mathbf{s}$.

(2) Lemma (2.4) proves the radiation condition for each fixed $\mathbf{s}$. The uniformity follows from the uniform boundedness of $rG(\mathbf{r})$ for all $s < R_1$.

Remark: If we wish to prove the existence of a Green's function K which vanishes on a closed surface Σ and is outgoing at infinity, then we have to use, instead of G_0, the Green's function $K_0(\mathbf{r},\mathbf{s})$ of $(\Delta + k^2)$ which vanishes on Σ. Saunders[14] proved the existence of K_0 and in fact that $K_0 = G_0 +$ a certain dipole distribution over Σ.[15] In such a case K satisfies

$$K(\mathbf{r},\mathbf{s}) = K_0 + \int_D K_0(\mathbf{r},\mathbf{t})V(\mathbf{t})K(\mathbf{t},\mathbf{s})d\mathbf{t},$$

where D is the domain outside of Σ.

The same procedure that we used to prove the existence of G may be used to prove the existence of K. Thus it is possible to prove the existence of an outgoing solution of $[\Delta + k^2 - V]\psi = 0$ which assumes given boundary values on Σ.

The following two lemmas pave the way for Theorem 1 which formulates the principle of limiting absorption.

Lemma 3.3: The operator T_k is continuous in k in the uniform operator topology.

Proof: Let

$$I(\mathbf{r},k) = \int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|} V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}, \quad \text{where} \quad \|\psi\| = 1.$$

If $\epsilon > 0$ is given, then Lemma 2 in reference 2 shows that there exists a number R such that for all $\mathbf{r}$

$$\int_{s>R} \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|} d\mathbf{s} < \epsilon.$$

[13] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, Inc., New York, 1958), Pt. I, pp. 577–585.

[14] W. K. Saunders, Proc. Natl. Acad. Sciences **38** (4), 342 (1952).

[15] Except in an exceptional case where it is taken over a neighboring surface.

Now consider the difference

$$|I(\mathbf{r},k_1)-I(\mathbf{r},k_2)|$$

$$\leqq \int_{s>R} + \int_{s\leqq R} \leqq 2\epsilon + |k_1-k_2| \int_{s\leqq R} |V(\mathbf{s})|\,d\mathbf{s}$$

$$\to 0 \text{ as } k_1 \to k_2 \text{ independently of } \mathbf{r}.$$

Hence $I(\mathbf{r},k)$ is uniformly continuous with respect to $\mathbf{r}$ and therefore T_k is continuous in the sense of the norm.

Lemma 3.4: If $k^2 \neq 0$ is not a negative eigenvalue of H then $G(\mathbf{r},\mathbf{s},k)$ is continuous in k (except at $\mathbf{r}=\mathbf{s}$).

Proof: We have

$$G=G_0+G_1$$
$$=G_0+(I-T_k)^{-1}A_1.$$

Since T_k is continuous and $(I-T_k)^{-1}$ exists by Lemma 2.5, we conclude (Dunford and Schwartz[13]) that $(I-T_k)^{-1}$ is continuous (in the norm). Using assumption (2.1c) we can prove that A_1 is continuous in k uniformly in $\mathbf{r}$, $\mathbf{s}$ and the lemma is proved.

Theorem 1 (principle of limiting absorption): Let $\psi(\mathbf{r},\epsilon)$ be the unique L_2 solution of

$$[-\Delta-(E+i\epsilon)+V]\psi=f(\mathbf{r}), \tag{3.4}$$

where E, $\epsilon>0$ and f is integrable and of compact support.[16]

Then $\psi(\mathbf{r},\epsilon)$ tends uniformly in $\mathbf{r}$ as $\epsilon \to 0$ to a function $\psi(\mathbf{r})$ which is the unique outgoing solution of $[-\Delta-E+V]\psi=f$.

Proof: If $k^2=E+i\epsilon$, then

$$\psi(\mathbf{r},\epsilon)=\int G(\mathbf{r},\mathbf{s},k)f(\mathbf{s})d\mathbf{s}$$

$$=\int \frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}f(\mathbf{s})d\mathbf{s}+\int G_1(\mathbf{r},\mathbf{s};k)f(\mathbf{s})d\mathbf{s}. \tag{3.5}$$

Lemma (3.4) shows that $G_1(\mathbf{r},\mathbf{s},k)$ tends uniformly in $\mathbf{r}$, $\mathbf{s}$ to $G_1(\mathbf{r},\mathbf{s},\sqrt{E})$ as $\epsilon \to 0$. Since f is integrable, the right side of (3.5) tends uniformly in $\mathbf{r}$ to

$$\psi(\mathbf{r})=\int G(\mathbf{r},\mathbf{s},\sqrt{E})f(\mathbf{s})d\mathbf{s}. \tag{3.6}$$

Since G is the Green's function for $(-\Delta-E+V)$ which satisfies the radiation condition uniformly in $\mathbf{s}$ the last assertion of the theorem is proved.

4. PRINCIPLE OF LIMITING AMPLITUDE

We consider the scattering problem for a time-harmonic wave $\psi_0(\mathbf{r},t)=p(\mathbf{r})e^{-iEt}$ which is a solution of the unperturbed Schrödinger equation. It is incident on a potential V which is switched on at time $t=0$.[17] Using the time-dependent Schrödinger equation we find that the generated scattered wave $\psi(\mathbf{r},t)$ satisfies an equation of the form

$$i\frac{\partial\psi(\mathbf{r},t)}{\partial t}+(\Delta-V)\psi(\mathbf{r},t)=f(\mathbf{r})e^{-iEt}A(t), \quad t\geqq 0$$
$$\psi(\mathbf{r},0)=0. \tag{4.1}$$

Here $f(\mathbf{r})$ is determined by ψ_0 and V and is assumed to be integrable and $O(r^{-3-\epsilon})$, $\epsilon>0$, at infinity and $A(t)=1$. The function $A(t)$ describes the switching on process and we take it to be

$$A(t)=\begin{cases} t & 0\leqq t\leqq 1 \\ 1 & t\geqq 1. \end{cases} \tag{4.2}$$

Let $\chi(\mathbf{r},z)$ be the Laplace transform of ψ with respect to t. Then we have from (4.1)

$$(H-\lambda)\phi(\mathbf{r},\lambda)=f(\mathbf{r})\frac{e^{-i(E-\lambda)}-1}{(E-\lambda)^2} \quad \mathrm{Im}\lambda>0, \tag{4.3}$$

where $\lambda=iz$ and $\phi(\mathbf{r},\lambda)\equiv\chi(\mathbf{r},z)$.

Our first step will be to define $\phi(\lambda)$, as the unique solution of (4.3) in the closed half-plane $\mathrm{Im}\lambda\geqq 0$. The following lemma accomplishes this by imposing severe restrictions on the potential.

Lemma 4.1: Suppose

(i) H has no negative eigenvalues (4.4)

(ii) The equation $\psi=T_0\psi$ has no solutions in B.[18] (4.5)

Then the solution $\phi(\mathbf{r},\lambda)$ of (4.3) is analytic in $\mathrm{Im}\lambda>0$ and continuous in $\mathrm{Im}\lambda\geqq 0$ except at $\lambda=E$ where it has a pole-like singularity.

Proof: The analyticity of ϕ is a consequence of the fact that H is self-adjoint. The hypotheses of the lemma prove, in view of Lemmas 3.2 and 3.4, the existence of $G(\mathbf{r},\mathbf{s};k)$, $k=\sqrt{\lambda}$, and its continuous dependence on k for all $\mathrm{Im}\lambda\geqq 0$. The existence and the continuity of ϕ follow immediately.

Example: If V satisfies either of the following conditions:

$$\text{(a)} \quad \sup_{\mathbf{r}} \int \frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s}<4\pi$$

$$\text{(b)} \quad V(\mathbf{s})\geqq 0$$

then the hypotheses of the preceding lemma are satisfied. The proof is immediate.

To investigate the asymptotic behavior of $\phi(\mathbf{r},\lambda)$ as $\lambda \to \infty$ we first prove

[16] The theorem remains true if $f=O(r^{-3-\epsilon})$ at infinity.

[17] We may, on the other hand, assume that the incident wave is switched on in the presence of a potential. A similar equation to (4.1) would still hold in this case.

[18] Conditions (i) and (ii) are equivalent to requiring that any bounded solution of $\psi=T_k\psi$, $k^2\leqq 0$ vanishes identically.

Lemma 4.2: The norm of the operator $(I-T_k)^{-1}$ remains uniformly bounded as $|k|\to\infty$.

Proof: The continuity of $(I-T_k)^{-1}$ insures its boundedness on every compact set. Since

$$(I-T_k)^{-1}=(I+T_k)(I-T_k^2)^{-1}$$
$$=(I+T_k)(I+T_k^2+T_k^4+\cdots) \quad (4.6)$$

it is sufficient to prove that T_k^2 becomes a "small perturbation" as $|k|\to\infty$. Now

$$\|T_k^2\|=\sup_{\|\phi\|=1}\left[\sup_{\mathbf{r}}\left|\int\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V(\mathbf{s})d\mathbf{s}\times\int\frac{e^{ik|\mathbf{s}-\mathbf{y}|}}{|\mathbf{s}-\mathbf{y}|}V(\mathbf{y})\phi(\mathbf{y})d\mathbf{y}\right|\right]. \quad (4.7)$$

Let $I(V;\mathbf{r})$ denote the integral on the right side of (4.7), then given $\epsilon>0$, choose R_0 such that

$$\int_{D_0}\frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}d\mathbf{s}<\epsilon, \quad \text{independent of } \mathbf{r}$$

where $D_0=\{\mathbf{s}\,|\,s\geqq R_0\}$. This is possible by Lemma 2 in reference 2. Let

$$U(\mathbf{s})=\begin{cases}V(\mathbf{s}) & s\leqq R_0\\ 0 & s>R_0.\end{cases}$$

It has been proved (Ladyzhenskaia,[4] or Zemach and Klein[19]) that $I(U,\mathbf{r})\to 0$ as $|k|\to\infty$ independently of $\mathbf{r}$, therefore it is sufficient to prove that $I(V;\mathbf{r})$ can be approximated uniformly by $I(U;\mathbf{r})$. Consider the difference

$$[I(V;\mathbf{r})-I(U;\mathbf{r})]$$
$$=\int_D V(\mathbf{y})\phi(\mathbf{y})d\mathbf{y}\int_{D_0}\frac{e^{ik[|\mathbf{r}-\mathbf{s}|+|\mathbf{s}-\mathbf{y}|]}}{|\mathbf{r}-\mathbf{s}|\,|\mathbf{s}-\mathbf{y}|}V(\mathbf{s})d\mathbf{s}$$
$$+\int_{D_0}\cdots d\mathbf{y}\int_{D-D_0}\cdots d\mathbf{s}=L_1+L_2.$$

But

$$|L_1|\leqq\int_D|V(\mathbf{y})|\,d\mathbf{y}\int_{D_0}\frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|\,|\mathbf{s}-\mathbf{y}|}d\mathbf{s}$$
$$\leqq\int_D\frac{|V(\mathbf{y})|}{|\mathbf{r}-\mathbf{y}|}d\mathbf{y}\int_{D_0}\left[\frac{|V(\mathbf{s})|}{|\mathbf{r}-\mathbf{s}|}+\frac{|V(\mathbf{s})|}{|\mathbf{s}-\mathbf{y}|}\right]d\mathbf{s}$$
$$\leqq C\cdot 2\epsilon, \text{ where } C \text{ is some constant.}$$

Similarly $|L_2|\leqq 2C\epsilon$. Therefore $\|T_k\|^2\to 0$ as $|k|\to\infty$ and the Neumann series (4.6) then converges, hence the lemma follows.

[19] C. Zemach and A. Klein, Nuovo cimento **10**, 1078 (1958).

Lemma 4.3: The function $\phi(\mathbf{r},\lambda)$ behaves asymptotically like $O[|\lambda|^{-2}]$ as $|\lambda|\to\infty$.

Proof: In view of (4.3) we may write

$$\phi(\mathbf{r},\lambda)=\int G(\mathbf{r},\mathbf{s};k)f(\mathbf{s})\cdot\frac{e^{-i(E-\lambda)}-1}{(E-\lambda)^2}d\mathbf{s}, \quad k^2=\lambda.$$

Hence it is sufficient to consider the behavior of G. But

$$G=G_0+(I-T_k)^{-1}A_1, \quad (3.1a)$$

where G_0, A_1 are both uniformly bounded in k. The preceding lemma completes the proof.

We formulate now the principle of limiting amplitude.

Theorem 2: Let $\psi(\mathbf{r},t)$ denote the solution of Eq. (4.1) and suppose that the hypotheses of Lemma (4.1) are fulfilled. Then $e^{iEt}\psi(\mathbf{r},t)$ tends as $t\to\infty$ to the unique outgoing solution of

$$[\Delta+E-V]\psi(\mathbf{r})=f(\mathbf{r}). \quad (4.8)$$

Proof: By the inverse Laplace transform

$$\psi(\mathbf{r},t)=\frac{1}{2\pi i}\int_{\alpha-i\infty}^{\alpha+i\infty}e^{zt}\chi(\mathbf{r},z)dz \quad \alpha>0. \quad (4.9)$$

But

$$\chi(\mathbf{r},z)\equiv\phi(\mathbf{r},\lambda) \text{ is analytic in } z \text{ for } \mathrm{Re}z>0,$$

continuous for $\mathrm{Re}z\geqq 0$ except at $z=-iE$ where it has a simple "pole." Deforming the contour in (4.9) to $\mathrm{Re}z=0$, and noticing that $|\chi(\mathbf{r},z)|=O[|z|^{-2}]$ as $|z|\to\infty$, we get

$$e^{iEt}\psi(\mathbf{r},t)=\frac{1}{2}\int G(\mathbf{r},\mathbf{s},\sqrt{E})f(\mathbf{s})d\mathbf{s}$$
$$+P\cdot V\cdot\frac{1}{2\pi}e^{iEt}\int_{-\infty}^{+\infty}e^{i\beta t}\chi(\mathbf{r},i\beta)d\beta. \quad (4.10)$$

The first term on the right side of (4.10) furnishes $\frac{1}{2}\psi(\mathbf{r})$, where ψ is the outgoing and therefore unique solution of (4.8). In the Appendix we prove that the principal value of the second term exists for each t and tends, as $t\to\infty$, to $\frac{1}{2}\psi(\mathbf{r})$. The method used is a refinement of the Riemann-Lesbegue lemma.

ACKNOWLEDGMENTS

The writer wishes to express his sincere thanks to Professor B. Friedman, University of California, Berkeley, for his encouragement and help and to Professor J. Keller, New York University, for discussions concerning the presentation of this paper.

APPENDIX

Consider

$$P\cdot V\cdot\frac{1}{2\pi}e^{iEt}\int_{-\infty}^{+\infty}e^{i\beta t}\chi(\mathbf{r},i\beta)d\beta, \quad (1)$$

where

$$\chi(\mathbf{r},i\beta)=\frac{e^{-i(E+\beta)}-1}{(E+\beta)^2}u(\beta)$$

$$u(\beta)=\int G(\mathbf{r},\mathbf{s};\,k)f(\mathbf{s})d\mathbf{s}=(I-T_k)^{-1}F(\mathbf{r},k),\quad k^2=-\beta$$

and

$$F(\mathbf{r},k)=\int G_0(\mathbf{r},\mathbf{s};\,k)f(\mathbf{s})d\mathbf{s}.$$

We prove first that u satisfies a Lipshitz condition for all $\beta<0$; consider

$$u(\beta_1)-u(\beta_2)=(I-T_{k_1})^{-1}(I-T_{k_2})^{-1}\times\{(T_{k_1}-T_{k_2})F(\mathbf{r};\,k)\}. \quad (2)$$

Using the facts that $(I-T_k)^{-1}$ exists and that $F(\mathbf{r})=O(r^{-1})$ and is Lipshitzian in k, an argument similar to the proof of Lemma 3.3 proves that the right side of (2) is Lipshitzian in k and hence in β. Now let us turn back to (1). The integral over the region outside any fixed neighborhood of $\beta=-E$ tends to zero as $t\to\infty$ by the Riemann-Lesbegue lemma. Let $\epsilon>0$ be fixed and consider

$$\frac{1}{2\pi}e^{iEt}\left[\int_{|E+\beta|\leqq\epsilon}e^{i\beta t}u(\beta)\frac{e^{-i(E+\beta)}-1}{(E+\beta)^2}d\beta\right]$$
$$=\frac{1}{2\pi i}e^{iEt}\int_{|E+\beta|\leqq\epsilon}(\cos\beta t+i\sin\beta t)\frac{u(\beta)}{(E+\beta)}d\beta$$
$$+\text{terms which tend to zero as } t\to\infty. \quad (3)$$

Let I_1, I_2 denote the cos and sin terms in (3). Then, writing $\beta t=t[(\beta+E)-E]$ and substituting in (3), we get

$$I_1=\frac{1}{2\pi i}e^{iEt}\left\{\cos tE\int_{|E+\beta|\leqq\epsilon}\frac{\cos t(\beta+E)}{(\beta+E)}u(\beta)d\beta\right.$$
$$\left.+\sin tE\int\frac{\sin t(\beta+E)}{(\beta+E)}u(\beta)d\beta\right\}. \quad (4)$$

The first term in (4) tends to zero as $t\to\infty$ since u is Lipshitzian. I_2 exhibits a similar behavior and we are left with

$$\lim_{t\to\infty}\frac{e^{iEt}}{2\pi i}(\sin tE+i\cos tE)\int\frac{\sin t(\beta+E)}{(\beta+E)}u(\beta)d\beta$$
$$=\tfrac{1}{2}u(\mathbf{r},-E) \text{ as in the theory of Fourier series}$$
$$=\tfrac{1}{2}\psi(\mathbf{r}).$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Principles of Limiting Absorption and Limiting Amplitude in Scattering Theory. II. The Wave Equation in an Inhomogeneous Medium

FAROUK M. ODEH
International Business Corporation Research Center, Yorktown Heights, New York
(Received July 11, 1961)

This is a continuation of Part I with special emphasis on wave equations in an inhomogeneous medium.

1. STATEMENT OF THE PROBLEM

THE principles which were employed in Part I to characterize the solutions of Schrödinger's equation can be applied, also, to the wave equation in an inhomogeneous medium.[1] It is assumed, however, that the properties of such a medium will approach those of a homogeneous one as the space variable $\mathbf{r}$ approaches infinity. This is expressed mathematically by assuming that the wave velocity is a function of $\mathbf{r}$ which tends to a certain constant as $r\to\infty$. The problem will be to characterize the solutions of the reduced wave equation in a three-dimensional space:

$$[\Delta+(k^2/c^2(\mathbf{r}))]\psi(\mathbf{r})=f(\mathbf{r})\quad k>0, \quad (1.1)$$

where $f(\mathbf{r})$ is an integrable function which is $O(r^{-\mu})$, $\mu>3$, at infinity. If we assume that $c^2(\mathbf{r})\to 1$ as $r\to\infty$, it becomes possible to define a "potential" $V(\mathbf{r})$ through the relationship

$$[c^2(\mathbf{r})]^{-1}=1-V(\mathbf{r}). \quad (1.2a)$$

Equation (1.1) may now be written in the form

$$[\Delta+k^2-k^2V(\mathbf{r})]\psi(\mathbf{r})=f(\mathbf{r})\quad k>0. \quad (1.3)$$

We assume that:

(i) $V(\mathbf{r})$ is integrable, bounded near infinity and locally Hoelder-continuous.[1a] (1.2b)

(ii) $[1-V(\mathbf{r})]\geq\epsilon>0$ except at a finite number of points where $1-V(\mathbf{r})$ may tend to zero. (1.2c)

[1] F. M. Odeh, J. Math. Phys. 2, 794 (1961).

[1a] Except at a finite number of square-integrable singularities.

where

$$\chi(\mathbf{r},i\beta)=\frac{e^{-i(E+\beta)}-1}{(E+\beta)^2}u(\beta)$$

$$u(\beta)=\int G(\mathbf{r},\mathbf{s};\,k)f(\mathbf{s})d\mathbf{s}=(I-T_k)^{-1}F(\mathbf{r},k),\quad k^2=-\beta$$

and

$$F(\mathbf{r},k)=\int G_0(\mathbf{r},\mathbf{s};\,k)f(\mathbf{s})d\mathbf{s}.$$

We prove first that u satisfies a Lipshitz condition for all $\beta<0$; consider

$$u(\beta_1)-u(\beta_2)=(I-T_{k_1})^{-1}(I-T_{k_2})^{-1}\times\{(T_{k_1}-T_{k_2})F(\mathbf{r};\,k)\}. \quad (2)$$

Using the facts that $(I-T_k)^{-1}$ exists and that $F(\mathbf{r})=O(r^{-1})$ and is Lipshitzian in k, an argument similar to the proof of Lemma 3.3 proves that the right side of (2) is Lipshitzian in k and hence in β. Now let us turn back to (1). The integral over the region outside any fixed neighborhood of $\beta=-E$ tends to zero as $t\to\infty$ by the Riemann-Lesbegue lemma. Let $\epsilon>0$ be fixed and consider

$$\frac{1}{2\pi}e^{iEt}\left[\int_{|E+\beta|\leqq\epsilon}e^{i\beta t}u(\beta)\frac{e^{-i(E+\beta)}-1}{(E+\beta)^2}d\beta\right]$$
$$=\frac{1}{2\pi i}e^{iEt}\int_{|E+\beta|\leqq\epsilon}(\cos\beta t+i\sin\beta t)\frac{u(\beta)}{(E+\beta)}d\beta$$
$$+\text{terms which tend to zero as } t\to\infty. \quad (3)$$

Let I_1, I_2 denote the cos and sin terms in (3). Then, writing $\beta t=t[(\beta+E)-E]$ and substituting in (3), we get

$$I_1=\frac{1}{2\pi i}e^{iEt}\left\{\cos tE\int_{|E+\beta|\leqq\epsilon}\frac{\cos t(\beta+E)}{(\beta+E)}u(\beta)d\beta\right.$$
$$\left.+\sin tE\int\frac{\sin t(\beta+E)}{(\beta+E)}u(\beta)d\beta\right\}. \quad (4)$$

The first term in (4) tends to zero as $t\to\infty$ since u is Lipshitzian. I_2 exhibits a similar behavior and we are left with

$$\lim_{t\to\infty}\frac{e^{iEt}}{2\pi i}(\sin tE+i\cos tE)\int\frac{\sin t(\beta+E)}{(\beta+E)}u(\beta)d\beta$$
$$=\tfrac{1}{2}u(\mathbf{r},-E) \text{ as in the theory of Fourier series}$$
$$=\tfrac{1}{2}\psi(\mathbf{r}).$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Principles of Limiting Absorption and Limiting Amplitude in Scattering Theory. II. The Wave Equation in an Inhomogeneous Medium

FAROUK M. ODEH
International Business Corporation Research Center, Yorktown Heights, New York
(Received July 11, 1961)

This is a continuation of Part I with special emphasis on wave equations in an inhomogeneous medium.

1. STATEMENT OF THE PROBLEM

THE principles which were employed in Part I to characterize the solutions of Schrödinger's equation can be applied, also, to the wave equation in an inhomogeneous medium.[1] It is assumed, however, that the properties of such a medium will approach those of a homogeneous one as the space variable $\mathbf{r}$ approaches infinity. This is expressed mathematically by assuming that the wave velocity is a function of $\mathbf{r}$ which tends to a certain constant as $r\to\infty$. The problem will be to characterize the solutions of the reduced wave equation in a three-dimensional space:

$$[\Delta+(k^2/c^2(\mathbf{r}))]\psi(\mathbf{r})=f(\mathbf{r})\quad k>0, \quad (1.1)$$

where $f(\mathbf{r})$ is an integrable function which is $O(r^{-\mu})$, $\mu>3$, at infinity. If we assume that $c^2(\mathbf{r})\to 1$ as $r\to\infty$, it becomes possible to define a "potential" $V(\mathbf{r})$ through the relationship

$$[c^2(\mathbf{r})]^{-1}=1-V(\mathbf{r}). \quad (1.2a)$$

Equation (1.1) may now be written in the form

$$[\Delta+k^2-k^2V(\mathbf{r})]\psi(\mathbf{r})=f(\mathbf{r})\quad k>0. \quad (1.3)$$

We assume that:

(i) $V(\mathbf{r})$ is integrable, bounded near infinity and locally Hoelder-continuous.[1a] (1.2b)

(ii) $[1-V(\mathbf{r})]\geq\epsilon>0$ except at a finite number of points where $1-V(\mathbf{r})$ may tend to zero. (1.2c)

[1] F. M. Odeh, J. Math. Phys. 2, 794 (1961).

[1a] Except at a finite number of square-integrable singularities.

The conditions (1.2b) imposed above on the potential $V(\mathbf{r})$, and hence on the velocity $c(\mathbf{r})$, insure that the radiating solution of equation (1.3) is unique (Zemach and Odeh[2]). In order to prove the existence of such a solution and to apply principles (2) and (3) of Part I to this equation, one has to consider again the resolvent operator $R(k)=[\Delta+k^2-k^2V]^{-1}$. If $\mathrm{Im}k^2\neq 0$ (and k is defined so that $\mathrm{Im}k>0$) then this resolvent, on $L_2(\mathbf{r})$, is an integral operator with a symmetric kernel $G(\mathbf{r},\mathbf{s},k)$ which belongs to $L_2(\mathbf{r})$, for a.e. $\mathbf{s}$, and satisfies[3]

$$G(\mathbf{r},\mathbf{s},k)=G_0(\mathbf{r},\mathbf{s},k)-k^2\int G_0(\mathbf{r},\mathbf{t})V(\mathbf{t})G(\mathbf{t},\mathbf{s},k)d\mathbf{t},$$

where

$$G_0(\mathbf{r},\mathbf{s},k)=\frac{1}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}. \tag{1.4}$$

Since the solution $\psi(\mathbf{r})$ of (1.3) may be written in the form $\psi(\mathbf{r})=R(k)\cdot f(\mathbf{r})$, it is necessary to discuss the kernel equation (1.4) in the case of real k. This equation becomes exactly the same as the kernel equation (2.2) of Part I when the potential V in the latter one is replaced by k^2V. We shall, therefore, follow the same procedures previously employed and adopt similar notation. The same Banach space $B(\mathbf{r})$ is introduced as before and the operator T_k is defined on B by

$$T_k\psi(\mathbf{r})=\frac{-k^2}{4\pi}\int\frac{e^{ik|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|}V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s},\quad \mathrm{Im}k\geq 0.$$

Throughout the whole discussion, proofs will be given only if they differ materially from the corresponding ones in Part I.

2. THE RESOLVENT OPERATOR $[\Delta+k^2-k^2V]^{-1}$

In this section we show that the kernel equation (1.4) has a solution which is continuous in $\mathbf{r}$ a.e. whenever the condition $\mathrm{Im}k\geq 0$ is satisfied.

Lemma 2.1: The operator T_k is a linear bounded compact operator on B to B.

Lemma 2.2: Let $\psi\in B$ be a solution of

$$\psi(\mathbf{r})=T_k\psi=-k^2\int G_0(\mathbf{r},\mathbf{s},k)V(\mathbf{s})\psi(\mathbf{s})d\mathbf{s}, \tag{2.1}$$

where $\mathrm{Im}k=b\geq 0$. Then

(i) $\psi(\mathbf{r})=O(r^{-1}e^{-br})$ at infinity

(ii) $$\lim_{r\to\infty} r\left[\frac{\partial\psi}{\partial r}-ik\psi\right]=0. \tag{2.2}$$

Lemma 2.3: If $\psi=T_k\psi$ and $\mathrm{Im}k\geq 0$, then ψ vanishes identically.

Proof: The lemma is obvious if $k=0$. We distinguish three other cases:

(1) $k^2>0$: In this case ψ satisfies

$$[\Delta+k^2-k^2V(\mathbf{r})]\psi=0, \tag{2.3}$$

where we have used the Hoelder-continuity of both V and the solution ψ of (2.1) in deriving (2.3). Since Lemma 2.2 shows that ψ satisfies the radiation condition, it follows from the corollary to Theorem 3 in reference 2 that ψ vanishes identically.

(2) $k^2<0$, $\mathrm{Im}k=b>0$: Let S_R, Σ_R denote a sphere of radius R around the origin and its surface, respectively. If Green's first formula is applied to ψ and its conjugate ψ^*, we obtain

$$\int_{\Sigma_R}\psi\frac{\partial\psi^*}{\partial s}dA=\int_{S_R}\psi\Delta\psi^*d\mathbf{s}+\int_{S_R}|\nabla\psi|^2d\mathbf{s}, \tag{2.4}$$

where $dA,d\mathbf{s}$ denote surface and volume elements, respectively.

Let $R\to\infty$ in (2.4), then the left side of (2.4) tends to zero by Lemma 2.2. Substituting for $\Delta\psi^*$ in the right side, we get

$$\lim_{R\to\infty}\int_{S_R}|\psi(\mathbf{s})|^2[1-V(\mathbf{s})]d\mathbf{s}=0. \tag{2.5}$$

Equation (2.5), together with condition (1.2c), prove that ψ vanishes identically.

(3) $\mathrm{Im}k=b>0$, $\mathrm{Im}k^2\neq 0$: Use of Green's second formula in conjunction with ψ,ψ^* leads to

$$\lim_{R\to\infty}(\mathrm{Im}k^2)\int_{S_R}|\psi|^2[1-V(\mathbf{s})]d\mathbf{s}=0.$$

Therefore, ψ vanishes identically in this case also, and the proof is concluded.

Lemma 2.4: The inverse $(I-T_k)^{-1}$ exists as a bounded operator on B for all k such that $\mathrm{Im}k\geq 0$.

Proof: Since T_k is compact and $\psi=T_k\psi$ holds true only if $\psi\equiv 0$, the statement of the lemma follows from the Fredholm alternative property for compact operators.

An iteration procedure similar to the one in Sec. 3 of Part I now yields

Lemma 2.5: Suppose $\mathrm{Im}k\geq 0$. Then

(1) The kernel equation (1.4) has a unique solution $G(\mathbf{r},\mathbf{s},k)$ which is continuous in $\mathbf{r}$ except if $\mathbf{r}=\mathbf{s}$. G is the Green's function for the operator $[\Delta+k^2-k^2V]$.

(2) The Green's function G satisfies the radiation condition (2.2).

[2] C. Zemach and F. Odeh, Arch. Rational Mech. and Analysis 5, 226 (1960).

[3] For proof see A. Y. Povzner, Mat. Sbornik 32 (74), 109 (1953); or T. Ikebe, Arch. Rational Mech. and Analysis 5, 1–34 (1960). A proof of the existence and the uniqueness of G also follows from Sec. 2.

3. PRINCIPLES OF LIMITING ABSORPTION AND LIMITING AMPLITUDE

It was shown in the last section that, if $\mathrm{Im}k\geq 0$, there exists a unique solution $\psi(\mathbf{r},k)$ to the equation

$$[\Delta-k^2-k^2V(\mathbf{r})]\psi(\mathbf{r})=f(\mathbf{r}).$$

In fact $\psi(\mathbf{r})$ is then given by

$$\psi(\mathbf{r},k)=\int G(\mathbf{r},\mathbf{s};k)f(\mathbf{s})d\mathbf{s}. \tag{3.1}$$

The dependence of this solution on k is obviously governed by the behavior of G as a function of k. But, since $G=G_0+(I-T_k)^{-1}T_kG_0$ and $G_0(\mathbf{r},\mathbf{s},k)$ is a continuous function of k (except at $\mathbf{r}=\mathbf{s}$) it would be sufficient to consider the operator $(I-T_k)^{-1}$. The following lemma exhibits the smooth dependence of the operator on the parameter k.

Lemma 3.1: The operator $(I-T_k)^{-1}$ is analytic in k for $\mathrm{Im}k>0$ and is continuous in k for $\mathrm{Im}k\geq 0$ (in the uniform operator topology).

Proof: Since $(I-T_k)^{-1}$ exists, it suffices to consider the behavior of T_k itself (Dunford and Schwartz[4]). Let $\mathrm{Im}k_0=b>0$ and consider the difference

$$D(\mathbf{r},k)=\int\left[\frac{e^{ik|\mathbf{r}-\mathbf{s}|}-e^{ik_0|\mathbf{r}-\mathbf{s}|}}{|\mathbf{r}-\mathbf{s}|\,|k-k_0|}-ie^{ik_0|\mathbf{r}-\mathbf{s}|}\right]V(\mathbf{s})d\mathbf{s}.$$

Given $\epsilon>0$, choose R large enough that

$$\int_{s\geq R}|V(\mathbf{s})|\,d\mathbf{s}<\epsilon.$$

Then

$$D(\mathbf{r},k)=\int_{s\geq R}+\int_{s\leq R}\leq 2\epsilon+C|k-k_0|\times\int_{s\leq R}|\mathbf{r}-\mathbf{s}|e^{ik_0|\mathbf{r}-\mathbf{s}|}|V(\mathbf{s})|\,d\mathbf{s},$$

where C depends on k_0 only.

Therefore, $D(\mathbf{r},k)$ tends to zero as $k\to k_0$ independently of $\mathbf{r}$ and the analyticity of T_k for $\mathrm{Im}k>0$ is proved. The continuity follows in a manner similar to Lemma 3.3 in Part I.

We state now the principle of limiting absorption.

Theorem 1: Let $\psi(\mathbf{r},\epsilon)$ be the unique L_2 solution of

$$\{\Delta+(k^2+i\epsilon)[1-V(\mathbf{r})]\}\psi(\mathbf{r},\epsilon)=f(\mathbf{r}),$$

where $k^2,\epsilon>0$ and f is integrable and $O(r^{-\mu})$, $\mu>3$, at infinity. Then $\psi(\mathbf{r},\epsilon)$ tends uniformly in $\mathbf{r}$ as $\epsilon\to 0$ to a function $\psi(\mathbf{r})$ which is the unique outgoing solution of $[\Delta+k^2-k^2V]\psi=f$.

The proof invokes the continuity of G as a function of k and proceeds along the same lines of Theorem 1 in Part I.

We turn now to the task of characterizing the solution of (1.1) by means of employing the principle of limiting amplitude. Let us consider the time-dependent problem.

$$\left.\begin{aligned}\Delta\psi(\mathbf{r},t)-\frac{1}{c^2(\mathbf{r})}\psi_{tt}=f(\mathbf{r})e^{-ikt}A(t)\\ \psi(\mathbf{r},0)=\psi_t(\mathbf{r},0)=0\end{aligned}\right\}, \tag{3.2}$$

where

$$A(t)=\begin{cases}t/\epsilon & 0\leq t\leq\epsilon\\ 1 & t\geq\epsilon.\end{cases} \tag{3.3}$$

Let $\chi(\mathbf{r},z)$, $\mathrm{Re}z>0$, denote the Laplace transform of $\psi(\mathbf{r},t)$ with respect to t and $\phi(\mathbf{r},\lambda)\equiv\chi(\mathbf{r},z)$ where $\lambda=iz$. Then we deduce from (3.2), (3.3) that $\phi(\mathbf{r},\lambda)$ satisfies

$$[\Delta+\lambda^2-\lambda^2V(\mathbf{r})]\phi(\mathbf{r},\lambda)=f(\mathbf{r})\frac{1-e^{i\epsilon(\lambda-k)}}{(\lambda-k)^2},\qquad \mathrm{Im}\lambda>0, \tag{3.4}$$

where $V(\mathbf{r})$ is defined by (1.2a).

The properties of the function $\phi(\mathbf{r},\lambda)$ are summed up in the next lemma which can be proved by using the methods of Sec. 4 in Part I in conjunction with Lemmas 2.5, 3.1.

Lemma 3.2: Let $\mathrm{Im}\lambda\geq 0$. Then

(1) The equation (3.4) has a unique solution $\phi(\mathbf{r},\lambda)$ which belongs to $B(\mathbf{r})$ for every fixed λ.

(2) The function $\phi(\lambda)$ is analytic in λ for $\mathrm{Im}\lambda>0$ and is continuous in λ for $\mathrm{Im}\,\lambda\geq 0$ except at $\lambda=k$ where it has a pole-like singularity.

It is possible now to proceed in a fashion similar to the one employed in the last section of the first part and prove the analog of Theorem 2 there, namely:

Theorem 2: Let $\psi(\mathbf{r},t)$ denote the unique bounded solution of the time-dependent problem (3.2). Then $e^{ikt}\psi(\mathbf{r},t)$ tends as $t\to\infty$ to the unique outgoing solution of

$$\left[\Delta+\frac{k^2}{c^2(\mathbf{r})}\right]\psi(\mathbf{r})=f(\mathbf{r}).$$

ACKNOWLEDGMENT

The writer wishes to thank Professor M. Kline and Professor J. Keller of New York University, for suggesting the problem.

[4] N. Dunford and J. T. Schwartz, *Linear Operators* (Interscience Publishers, Inc., New York, 1958), Pt. I, pp. 577–585.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Foundations for a Treatment of the Scattering of Light by the Hydrodynamical and Statistical Atom Model*

MASAMI WAKANO
Palmer Physical Laboratory, Princeton University, Princeton, New Jeresey
(Received June 12, 1961)

A study has been made on the scattering of light by the hydrodynamical and statistical atom model. Bloch's treatment of the hydrodynamical equations of motion for this model is supplemented here by inclusion of the interaction with the electromagnetic field. We limited attention to oscillations of small amplitude. By correspondence principle arguments, general expressions were derived for the cross sections for absorption, coherent and incoherent scattering.

The energy can be expressed—following Bloch—as the energy of a Thomas-Fermi atom plus a Hamiltonian which is associated with departures from the Thomas-Fermi distribution. Using Bloch's quantization of this Hamiltonian and applying the method of quantum field theory, we rederived the correspondence principle results for elementary cross sections.

Then applying the correspondence theoretical argument to the matrix element, we rederived Heisenberg's result for the total intensity of the Compton scattering. We also apply the method of stationary phase to the hydrodynamical treatment and show that this method gives the same result as does the linear term in the momentum transfer in Heisenberg's expression, except for a numerical factor $3^{\frac{1}{2}}/2$—a point that was discussed by Bloch many years ago.

Application of the general formulas given here for angular distribution of Rayleigh and Compton scattering will require electronic machine calculations of higher modes of oscillation of the gas model of the atom analogous to those made by Wheeler and Fireman for $l=1$.

I. INTRODUCTION AND SUMMARY

WE discuss several processes which involve the interaction between light and atomic electrons. This subject has been so much discussed in the past and at the present we might think that there is no essential problem left except to secure accurate knowledge of atomic wave functions.

Certainly for the hydrogen atom and other elements of small Z, there is no problem to be studied. However, when the value of Z gets larger and larger, the number of degrees of freedom of the dynamical motion becomes so big that we need some simple and suitable approximation method which can be applied universally for all Z.

As such a method, we have the simple model of Thomas[1] and Fermi.[2] This model gives a smoothed-out representation of the charge distribution in the ground state. When an atom absorbs a photon, typically one electron jumps to an excited state—or to the continum. This is one of the processes upon which attention will be focussed here.

First we treat the interaction between light and atom in classical theory. The energy is taken up from electromagnetic field into vibration of the electron gas. This energy is the classical equivalent of photon absorption. Also, the electromagnetic moments induced in the electrified gas atom by the primary electromagnetic disturbance—and by any supplemental electromagnetic disturbance also from outside—generate secondary waves which are the classical equivalents of Rayleigh and Compton scattering. Therefore one has in the gas atom a far reaching model for treating the interaction of photons with atoms.

The Thomas-Fermi gas model has the disadvantage of exaggerating the number of low momentum electrons. This effect shows up in the circumstance that the density of electrons in the Thomas-Fermi atom falls off more slowly at large distances than do the experimental and Hartree-Fock values. Attempts have been made to remedy this shortcoming of the simple model. Dirac[3] derived a revised equation of state by statistical methods from Fock's equation for a many electron system. This Dirac equation includes the exchange effect. The correlation energy between electrons with antiparallel spins was calculated by Wigner[4] and this correction also has the same effect as Dirac's correction in cutting off the radius of the neutral atom to a finite value.

Recently, much work[5] has been done to include quantum corrections in a systematic way. The Thomas-Fermi model is treated as the lower limit of the approximation in $\hbar$. But all these corrections are obtained at the expense of the scaling law which is so important in the Thomas-Fermi equation. For this reason, no account is taken of such corrections in the following work.

When the energy of an incoming photon and the momentum transfer to an electron become large, sufficient to eject a K electron from a heavy atom ($Z\sim100$) to the continum, we have to use the relativistic wave functions for the electron or the relativistic equation of state developed by Chandrasekhar and we cannot use the Thomas-Fermi equation of state. Therefore, to test our theory we have to limit the transfer in energy and momentum below the K-absorption edge (Fig. 1).

* Based on the thesis submitted to Princeton University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, May, 1961.

[1] L. H. Thomas, Proc. Cambridge Phil. Soc. **23**, 542 (1926).

[2] E. Fermi, Z. Physik **48**, 73 (1928).

[3] P. A. M. Dirac, Proc. Cambridge Phil. Soc. **26**, 376 (1930).

[4] P. Gombas, Z. Physik **121**, 523 (1943).

[5] W. D. Theis, Z. Physik **142**, 503 (1955); W. Macke, Phys. Rev. **100**, 992 (1955); Am. J. Phys. **17**, 1 (1955); E. S. Fradkin, Soviet Phys.—JETP **64**, 5 (1957); S. Golden, Phys. Rev. **105**, 604 (1957); **107**, 1283 (1957); D. A. Kirzhnits, Soviet Phys.—JETP **5**, 64 (1957); L. C. R. Alfred, Phys. Rev. **121**, 1275 (1961); G. A. Baraff and S. Borowitz, Phys. Rev. **121**, 1704 (1961).

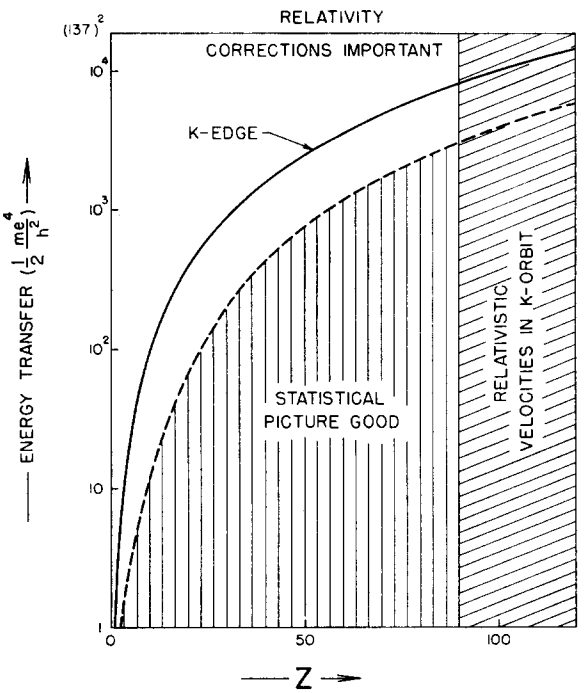

FIG. 1. Diagram showing the domain of atomic number Z and energy transfer from photon to atom where one can safely use the statistical atom model.

To treat the dynamical motion of the electron gas as simply and universally as Thomas and Fermi treated the hydrostatics of the gas model, Bloch[6] introduced the hydrodynamical proper modes in his famous work on the stopping power of a dilute gas for charged particles. He regards the electrons as a degenerate Fermi gas endowed with a characteristic pressure—density relation. The ground state has the property familiar from the work of Thomas and Fermi. This gas is capable of normal modes of oscillation about this equilibrium state. Bloch's model underlies the present analysis of the scattering of light by atoms.

In Bloch's theory of stopping power, the interaction between a passing particle and atomic electrons has been assumed to take place directly through action at the distance. To treat the scattering of light by atoms, we derive (Sec. II) the hydrodynamical equation of motion of a coupled system of electron gas and radiation field. Based on these equations and the Poisson equation, we derive the absorption cross section. The absorption cross section is obtained as the energy uptake by the atom when the atom jumps to excited states from the ground state. Then we give a general definition for the oscillator strength. This work differs from that of Wheeler and Fireman[7] only in this respect, that they used the dipole approximation whereas we use the general retardation factor.

[6] F. Bloch, Z. Physik 81, 363 (1933).

[7] J. A. Wheeler and E. L. Fireman, "A universal atomic photo-absorption curve," Aeronutronic Systems, Inc., a subsidiary of Ford Motor Company (1957).

Following Maxwell's theory of classical electrodynamics we derive (Sec. III) the differential cross section for coherent scattering. For a long wavelength of the incoming wave, we find the well-known standard Rayleigh formula which connects the differential cross section to oscillator strength.

The amplitude for coherent scattering turns out to be the sum of a "direct part" and a "dispersive part." The classical equivalent of a matrix element for the direct part makes no reference to any excitation of hydrodynamical oscillations, and when analyzed into partial waves contains contributions only of the type l (incident wave)$=l$ (outgoing wave). The dispersive part is a product of factors akin to matrix elements one for the incoming wave, one for the outgoing wave, each satisfying the relation l (hydrodynamical oscillation) $=l$ (electromagnetic wave)± 1.

We apply Einstein's argument of statistical mechanics to derive (Sec. IV) the differential cross section for incoherent scattering. We consider an atom illuminated with two classical electromagnetic waves. Under these two perturbations we solve the hydrodynamical equations to the second order and get the absorption cross section. From the absorption cross section we obtain the differential cross section for inelastic scattering. The formula has three terms, (1) direct scattering, (2) dispersive scattering, (3) internal scattering (tentative name). The internal scattering is connected with the complexity of the original equations of hydrodynamics. It says that if two modes are excited, then these two modes in combination excite another mode through the internal mechanism.

So far we discussed Bloch's hydrodynamical electron gas in classical terms as being capable of free vibration. We give (Sec. V) proper quantum field theoretical treatment to this free vibration. We divide the total Hamiltonian into the sum of a Hamiltonian of the ground state and a Hamiltonian which is associated with deviations from Thomas-Fermi's distribution. This Hamiltonian is quadratic in deviations in our approximation and has the form of an assembly of classical harmonic oscillators. We quantize this system and the radiation field. Following the perturbation method of quantum electrodynamics, we derive differential cross sections for coherent and incoherent scattering. Compared with the results obtained by classical methods, the new results have the same form in direct scattering and more or less the same form in dispersive scattering. Corresponding to the internal scattering we have a new term which is related with the simultaneous excitation or annihilation of two modes. However, the states which have two or more modes excited are not stationary but decay exponentially with time. Therefore, our method of derivation is not correct in its treatment of these higher order processes.

Finally, (Sec. VI) we discuss the Compton scattering. We take the direct scattering term. Using JWKB orbitals and the method of stationary phase, we rederive

Heisenberg's formula[8] for total intensity for the Compton scattering. Also in the hydrodynamical model we employ JWKB expressions for the characteristic modes of oscillation and use the method of stationary phase to calculate the scattering cross section. The result agrees up to the Bloch factor[9] of $3^{\frac{1}{2}}/2$ with Heisenberg-Bewilogua's[10] result for values of the scattering cross section for small momentum transfer. In this derivation we ascribe momentum[11] $\hbar\omega/(dp/mdn_0)^{\frac{1}{2}}$, and inertial mass[12] $\hbar\omega/(dp/mdn_0)$ to a phonon of energy $\hbar\omega$. We find that only at the distance where the momentum $\hbar\omega/(dp/mdn_0)^{\frac{1}{2}}$ of the phonon agrees with the value of the momentum transfer from the photon do we get a contribution to the total intensity.

II. HYDROELECTRODYNAMICS OF ELECTRON GAS

Equation of Motion

In this section we develop basic ideas and formulas which we use in this work, following Bloch and Wheeler-Fireman. We start with Euler's hydrodynamical equation of motion for the electron gas which interacts with the electromagnetic field

$$mn\frac{D\mathbf{v}}{Dt}=-\nabla p-n\nabla V+en\left(\mathbf{E}+\frac{\mathbf{v}}{c}\times\mathbf{H}\right). \quad (2.1)$$

Here e and m are the charge and mass of the electron, $n(x,y,z,t)$ is the number density of electrons, $\mathbf{v}(x,y,z,t)$ is the velocity distribution, and $p(x,y,z,t)$ is the hydrodynamical pressure. Pressure p arises from the zero-point energy of the degenerate quantum Fermi gas. We neglect exchange energy and temperature correction. We simply assume that the equation of state

$$P=\frac{h^2}{5m}\left(\frac{3}{8\pi}\right)^{2/3}n^{5/3} \quad (2.2)$$

is valid for hydrodynamically excited states as well as for the ground state.

V is the electrostatic potential, $\mathbf{E}$ and $\mathbf{H}$ are the transversal electric and magnetic field intensity. These quantities represent the total electrostatic and electromagnetic field inside the atom. Therefore electrons move not in the given applied (external) field but in the final (total) field including the effect of the induced charge-current distribution inside the atom. Inside the atom, we can neglect the electromagnetic interaction between electron and electron and take the electrostatic interaction alone because the electromagnetic interaction is proportional to v/c ($\ll 1$), where v is the velocity of an electron.

[8] W. Heisenberg, Physik Z. 32, 737 (1931).
[9] F. Bloch, Helv. Phys. Acta 7, 385 (1934).
[10] L. Bewilogua, Physik Z. 32, 740 (1931).
[11] L. Landau, J. Phys. (U.S.S.R.) 5, 71 (1941).
[12] H. Kramers, Physica 18, 653 (1952).

We rewrite Eq. (2.1) into the following form:

$$\frac{\partial}{\partial t}\left(\mathbf{v}+\frac{e}{mc}\mathbf{A}\right)=-\frac{1}{m}\nabla\int_0^n\frac{dP(n')}{n'}-\frac{1}{m}\nabla V-\nabla\frac{v^2}{2}+\left[\mathbf{v}\times\left[\nabla\times\left(\mathbf{v}+\frac{e}{mc}\mathbf{A}\right)\right]\right]. \quad (2.3)$$

Then we can introduce the velocity potential u defined by

$$\mathbf{v}+e\mathbf{A}/(mc)=-\nabla u. \quad (2.4)$$

We substitute (2.4) into (2.3) and get

$$\frac{\partial u}{\partial t}=\frac{1}{m}\int_0^n\frac{dP(n')}{n'}+\frac{V}{m}+\frac{1}{2}\left(\nabla u+\frac{e}{mc}\mathbf{A}\right)^2. \quad (2.5)$$

The field equation inside the atom is the Poisson equation

$$\nabla^2V=-4\pi e^2n. \quad (2.6)$$

The equation of continuity is

$$\partial n/\partial t=\mathrm{div}[n(\nabla u+e\mathbf{A}/(mc)]. \quad (2.7)$$

Three equations (2.5), (2.6), and (2.7) and the equation of state (2.2) supply the necessary and sufficient basis for the present analysis.

Small Vibrations, Equilibrium State, Orthogonality Relations

Bloch's hydrodynamical electron gas is capable of oscillations around the ground state. The deviation from the ground state can be treated, as usual, by the theory of small vibration. We expand hydrodynamical quantities in terms of deviations:

$$\begin{aligned}&\text{number density:} && n=n_0+n_1+n_2+\cdots,\\ &\text{electron pressure:} && p=p_0+p_1+p_2+\cdots,\\ &\text{velocity potential:} && u=0+u_1+u_2+\cdots,\\ &\text{electrostatic potential:} && V=V_0+V_1+V_2+\cdots.\end{aligned} \quad (2.8)$$

Succeeding terms are specified by the order of interaction with the external field. Suppose the interaction is switched off at a certain time. From that moment the atom remains stationary in the various excited states. *Exactly speaking*, if two or more modes are excited, this state is not stationary because these modes in combination excite another mode as discussed in Sec. IV. However, this process takes place through the 2nd-order perturbation. Therefore, as far as we are concerned with the lowest order, the states where two or more modes are excited are stationary and are characterized by harmonic oscillations. These oscillations we call free vibrations. Then equations which are linear in n_1, u_1, and V_1 are

$$\nabla^2V_1=-4\pi e^2n_1, \quad (2.9)$$

$$\partial n_1/\partial t=\nabla(n_0\nabla u_1), \quad (2.10)$$

$$m\partial u_1/\partial t=n_1(dP/n_0dn_0)+V_1. \quad (2.11)$$

We characterize the free vibration by proper frequency ω and by the indices of a spherical harmonic. The general solution is the superposition of the elementary solutions of the form

$$n_1^{(k)} = -\omega_k N_k(r,\theta,\phi)\sin(\omega_k t+\delta_k), \tag{2.12}$$

$$u_1^{(k)} = U_k(r,\theta,\phi)\cos(\omega_k t+\delta_k), \tag{2.13}$$

$$V_1^{(k)} = m\omega_k V_k(r,\theta,\phi)\sin(\omega_k t+\delta_k). \tag{2.14}$$

Inserting (2.12), (2.13), and (2.14) into (2.9), (2.10), and (2.11) we get

$$\nabla^2 V_k = (4\pi e^2/m)N_k, \tag{2.15}$$

$$-\omega_k^2 N_k = \nabla(n_0 \nabla U_k), \tag{2.16}$$

$$-U_k = -N_k(dp/mn_0 dn_0)+V_k. \tag{2.17}$$

Next we recall the derivation of the contribution of an individual mode of proper vibration to the energy uptake of the statistical atom. As usual we proceed with the variation principle. The total energy has the following form:

$$E=\int n\frac{mv^2}{2}d\tau+\frac{e^2}{2}\int\frac{n^{(a)}n^{(b)}}{r_{a,b}}d\tau_a d\tau_b - Ze^2\int\frac{n}{r}d\tau+\int d\tau\cdot n\int_0^n\frac{P(n')}{n'^2}dn'. \tag{2.18}$$

Let this energy be minimized with respect to n, keeping the total number of electrons $\int n d\tau = Z$ fixed. We get the unperturbed static solution n_0. Then let all free vibrations be excited, so that

$$\begin{aligned} n &= n_0+n_1 = n_0+\textstyle\sum_k C_k n_1^{(k)}, \\ u &= 0+u_1 = \textstyle\sum_k C_k u_1^{(k)}, \\ V &= V_0+V_1 = V_0+\textstyle\sum_k C_k V_1^{(k)}. \end{aligned} \tag{2.19}$$

To secure real n_1, u_1, and V_1 we claim that

$$C_k = C_{-k}^*,$$

where

$$C_k = C_{\omega k,l,m} \tag{2.20}$$

and

$$C_{-k} = C_{\omega k,l,-m}$$

since we define the spherical harmonic such that

$$Y_{l,m}(\theta,\phi) = Y_{l-m}^*(\theta,\phi). \tag{2.21}$$

(*Not* in accord with Condon's choice of phase!)

The energy increase ΔE is given by

$$\Delta E=\frac{1}{2}\int d\tau\left\{mn_0(\nabla u_1)^2+n_1 V_1 + n_1^2\left[\frac{d^2}{dn^2}\left(n\int_0^n\frac{P(n')}{n'^2}dn'\right)_{n_0}\right]\right\} \tag{2.22}$$

because terms of the first order in deviations from equilibrium vanish. This ΔE must be a constant of time *after* the perturbation is turned off. We can explicitly show this constancy by taking the time derivative of ΔE:

$$\frac{d}{dt}(\Delta E)=\int d\mathbf{S}\left\{\left(n_0 V_1+n_1\frac{dP}{dn_0}\right)\nabla u_1 + \frac{1}{8\pi e^2}\left[V_1\frac{\partial V_1}{\partial t}-\frac{\partial V_1}{\partial t}\nabla V_1\right]\right\}. \tag{2.23}$$

Then we impose the boundary conditions:

(1) Normal component of gas velocity vanishes on the boundary surface.

(2) Variations in potential vanish on the boundary surface.

We eventually extend the boundary surface to the infinity. Due to the boundary conditions $d(\Delta E)/dt=0$.

On the other hand, the energy increase due to free vibrations has the following form:

$$\Delta E=\sum_k |C_k|^2\cdot\frac{1}{2}m\omega_k^2\int N_k U_k^* d\tau \tag{2.24}$$

provided that we demand the orthogonality condition

$$\int N_j U_k^* d\tau = 0. \tag{2.25}$$

This condition annuls all the mixed terms which would otherwise contribute to (2.24) and which would vary periodically in time with a circular frequency of the form $(\omega_j-\omega_k)$. The absence of these terms ensures that ΔE remains constant in time.

Absorption Process

Now we are ready to apply the hydrodynamical theory to the absorption of light. Following Wheeler-Fireman, we sketch the general procedure. At time $t=0$, let the electromagnetic interaction be applied. As time goes on, various modes are excited. The amplitude of the individual mode varies with time. After a sufficiently long time, let the interaction be stopped. The atom will be left in a superposition of free vibrations. During the supply of the electromagnetic interaction, the energy for the individual mode increases linearly with time or oscillates with time. We pick up those terms which increase linearly with time, and get the absorption cross section.

We expand the deviations n_1, u_1, and V_1 in terms of an orthonormal complete set N_k, U_k, and V_k.

$$\begin{aligned} n_1 &= \sum_k g_k(t)N_k(r,\theta,\phi), \\ u_1 &= \sum_k h_k(t)U_k(r,\theta,\phi), \\ V_1 &= \sum_k d_k(t)U_k(r,\theta,\phi), \end{aligned} \tag{2.26}$$

where we have included the time dependence of stationary states into the expansion coefficients g_k, h_k, and d_k. The equation of motion, field equation, and the equation of continuity in first order in the external field are as follows:

$$\frac{\partial u_1}{\partial t}=n_1\frac{1}{mn_0}\frac{dp}{dn_0}+\frac{V_1}{m},$$

$$\nabla^2 V_1=-4\pi e^2 n_1, \tag{2.27}$$

$$\frac{\partial n_1}{\partial t}=\operatorname{div}\left[n_0\left(\nabla u_1+\frac{e}{mc}\mathbf{A}\right)\right].$$

Inserting Eq. (2.26) into Eq. (2.27) we get the following relations:

$$d_k(t)=-mg_k(t), \tag{2.28}$$

$$\dot{h}_k(t)=g_k(t), \tag{2.29}$$

$$\ddot{h}_k(t)+\omega_k^2 h_k(t)$$

$$=\left(\frac{e}{m\omega}\right)\operatorname{Re}\left[\frac{i\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_k^* N_k d\tau}e^{-i\omega t}\right]. \tag{2.30}$$

Here, the external classical field is given by

$$\mathbf{E}=\operatorname{Re}[\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}]$$

or

$$\mathbf{A}=\operatorname{Re}\left[\frac{c}{i\omega}\mathbf{E}_0 e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\right], \tag{2.31}$$

where Re means the real part.

The general solution of (2.30) is

$$h_k(t)=\operatorname{Re}\left\{\frac{A_k}{-i\omega_k}e^{-i\omega_k t}\right.$$

$$\left.+\frac{e/m}{-i\omega}P\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_k^* N_k d\tau(\omega_k^2-\omega^2)}e^{-i\omega t}\right\}. \tag{2.32}$$

P means the principal part and A_k is an arbitrary constant. We fix this A_k from the initial condition. We demand that at $t=0$, the number density deviation n_1 and the velocity potential u_1 vanish. Namely, we ask

$$\begin{aligned} h_k(t=0)&=0,\\ g_k(t=0)&=0. \end{aligned} \tag{2.33}$$

Then A_k takes the following form:

$$A_k=-\frac{e}{m}\left\{\operatorname{Re}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_k^* N_k d\tau}\right.$$

$$\left.+i\frac{\omega_k}{\omega}\operatorname{Im}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_k^* N_k d\tau}\right\}P\frac{1}{\omega_k^2-\omega^2}. \tag{2.34}$$

At time t, the interaction is stopped and the atom remains in a state of free vibration. Therefore we have the following relations between $g_k(t)$, $h_k(t)$, and $C_k(t)$:

$$\begin{aligned} -C_k\omega_k\sin(\omega_k t+\delta_k)&=g_k(t),\\ C_k\cos(\omega_k t+\delta_k)&=h_k(t). \end{aligned} \tag{2.35}$$

Hence

$$\Delta E(t)=\sum_k\left(\tfrac{1}{2}m\int U_k^* N_k d\tau\right)(|g_k(t)|^2+\omega_k^2|h_k(t)|^2). \tag{2.36}$$

By elementary calculations we get

$$|g_k(t)|^2+\omega_k^2|h_k(t)|^2=\left(\frac{e}{m}\right)^2 P\cdot\frac{\omega_k^2}{(\omega_k^2-\omega^2)^2}\left\{\left[\operatorname{Re}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_k\int U_k^* N_k d\tau}\right]^2 M_k^{(1)}(t)+\left[\operatorname{Im}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_k\int U_k^* N_k d\tau}\right]^2\right.$$

$$\left.\times M_k^{(2)}(t)+2\left[\operatorname{Re}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_k\int U_k^* N_k d\tau}\right]\left[\operatorname{Im}\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_k\int U_k^* N_k d\tau}\right]M_k^{(3)}(t)\right\}, \tag{2.37}$$

where

$$M_k^{(1)}(t)=\omega^{-2}(\omega_k^2\sin^2\omega t+\omega^2\cos^2\omega t-2\omega\omega_k\sin\omega t\sin\omega_k t-2\omega^2\cos\omega t\cos\omega_k t+\omega^2),$$
$$M_k^{(2)}(t)=\omega^{-2}(\omega^2\sin^2\omega t+\omega_k^2-2\omega\omega_k\sin\omega t\sin\omega_k t+\omega_k^2\cos^2\omega t-2\omega_k^2\cos\omega_k t\cos\omega t), \tag{2.38}$$
$$M_k^{(3)}(t)=\omega^{-2}(\omega_k^2-\omega^2)\sin\omega t(\cos\omega_k t-\cos\omega t).$$

We sum over states k:

$$\Delta E(t)=\int d\omega_k\frac{dN}{d\omega_k}\left(\frac{m}{2}\int U_k^*N_k d\tau\right)(|g_k(t)|^2+\omega_k^2|h_k(t)|^2), \tag{2.39}$$

where $dN/d\omega_k$ is the level density at frequency ω_k. It turns out that $M_k^{(3)}(t)$ contributes only to oscillating terms so that we neglect it. $M_k^{(1)}(t)$ and $M_k^{(2)}(t)$ contribute to terms which increase linearly with time. Therefore we keep those terms. The final result is

$$\frac{d}{dt}(\Delta E(t))=\frac{\pi e^2}{4m}\left[\frac{dN}{d\omega_k}\left|\frac{\int n_0(\mathbf{E}_0\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_k\left(\int U_k^*N_k d\tau\right)^{\frac{1}{2}}}\right|^2\right]_{\omega_k=\omega}. \tag{2.40}$$

Since the flux $\mathbf{S}$ of the incoming electromagnetic wave is

$$\mathbf{S}=\mathbf{n}\frac{c}{4\pi}|\mathbf{E}|^2=\mathbf{n}\frac{c}{8\pi}|\mathbf{E}_0|^2, \tag{2.41}$$

the absorption cross section becomes

$$\sigma(\omega)=\frac{2\pi^2e^2}{mc}\left[\frac{dN}{d\omega_k}\frac{\left|\int n_0(\mathbf{e}\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau\right|^2}{\omega_k^2\int U_k^*N_k d\tau}\right]_{\omega_k=\omega}. \tag{2.42}$$

This agrees with the result of Wheeler-Fireman when the general retardation factor $e^{i(\mathbf{k}\cdot\mathbf{r})}$ is replaced by the first-order term $i(\mathbf{k}\cdot\mathbf{r})$ (dipole approximation).

Oscillator Strength

If we define the oscillator strength f by

$$\sigma(\omega)=\frac{2\pi^2e^2}{mc}f(\omega), \tag{2.43}$$

then

$$f(\omega)=\left[\frac{dN}{d\omega_k}\frac{\left|\int n_0(\mathbf{e}\cdot\nabla U_k^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau\right|^2}{\omega_k^2\int U_k^*N_k d\tau}\right]_{\omega_k=\omega}. \tag{2.44}$$

The eigenvalue of Eqs. (2.15), (2.16), and (2.17) were studied by Wheeler and Fireman. They have shown that the spectrum is not discrete, as Bloch had assumed, but continuous.

III. COHERENT SCATTERING

Scattering Cross Section

In this section we apply Bloch's hydrodynamics to elastic scattering and analyze the differential cross section by the method of partial waves.

The external electromagnetic field produces a current inside the atom. This current is given, to the first order, by

$$\mathbf{j}_1=-\frac{e}{c}n_0\left(\nabla u_1+\frac{e}{mc}\mathbf{A}\right), \tag{3.1}$$

where u_1 is already obtained in (2.26) and (2.32). However, we have to replace the principal part P in (2.32) by the outgoing boundary condition.

From the induced current (3.1) we get the magnetic field $\mathbf{H}(\mathbf{r},t)$ outside the atom.

$$\mathbf{H}(\mathbf{r},t)=\frac{e^2}{2mc^2r}\left\{e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\mathbf{E}_0\int n_0e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau-\sum_\alpha\frac{\int n_0\nabla U_\alpha e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau\int n_0(\mathbf{E}_0\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_\alpha^*N_\alpha d\tau(\omega_\alpha^2-\omega^2-i\epsilon)}+\text{c.c.}\right\}\times\mathbf{n}. \tag{3.2}$$

From this expression we get the differential cross section for coherent scattering

$$\frac{d\sigma}{d\omega}=\left(\frac{e^2}{mc^2}\right)^2\left|\left\{\mathbf{e}_{\text{in}}\int n_0 e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau-\sum_\alpha\frac{\int n_0\nabla U_\alpha e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_\alpha^* N_\alpha d\tau(\omega_\alpha^2-\omega^2-i\epsilon)}\right\}\times\mathbf{n}\right|^2. \tag{3.3}$$

Rewriting $\mathbf{n}$ by the product of two orthogonal polarization vectors, $\mathbf{e}_{\text{out}}$ of the scattered wave, we get

$$\frac{d\sigma}{d\omega}=\left(\frac{e^2}{mc^2}\right)^2\sum_{\substack{\text{polarization of}\\ \text{scattered wave}}}\left|(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})\int n_0 e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau-\sum_\alpha\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_\alpha)e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_\alpha^* N_\alpha d\tau(\omega_\alpha^2-\omega^2-i\epsilon)}\right|^2. \tag{3.4}$$

Forward Dispersion Relation for Light with a Long Wavelength

Based on this equation, we discuss the dispersion relation. We take the scattering of light with a long wavelength. Let the polarization vector $\mathbf{e}_{\text{in}}$ point in the x direction. Since N_α forms a complete set of functions satisfying

$$\int N_\alpha U_\beta^* d\tau=0 \quad \text{for} \quad \alpha\neq\beta,$$

we expand the vector $\nabla(n_0e^{-i(\mathbf{k}\cdot\mathbf{r})})$ in terms of the N_α with coefficients which are vectors $\mathbf{C}_\alpha$ independent of position:

$$\nabla(n_0e^{-i(\mathbf{k}'\cdot\mathbf{r})})=\sum_\alpha\mathbf{C}_\alpha N_\alpha^*=-\sum_\alpha\frac{\int n_0e^{-i(\mathbf{k}'\cdot\mathbf{r})}\nabla U_\alpha d\tau}{\int U_\alpha N_\alpha^* d\tau}N_\alpha^*. \tag{3.5}$$

We now multiply both sides with $xe^{-i(\mathbf{k}\cdot\mathbf{r})}$ and integrate. Then we use Eq. (2.6) and get

$$-\int x\nabla(n_0e^{-i(\mathbf{k}'\cdot\mathbf{r})})e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau=\sum_\alpha\frac{\int n_0e^{-i(\mathbf{k}'\cdot\mathbf{r})}\nabla U_\alpha d\tau}{\int U_\alpha N_\alpha^* d\tau}\cdot\frac{1}{\omega_\alpha^2}\int\nabla\cdot(n_0\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}(-x)d\tau. \tag{3.6}$$

We assume

$$kr_1\ll 1, \tag{3.7}$$

where r_1 is the distance inside which half of all the electrons are included. Then, noting $\nabla\mathbf{x}=\mathbf{e}_{\text{in}}$, we can reduce Eq. (3.6) to

$$\mathbf{e}_{\text{in}}\int n_0e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau=\sum_\alpha\frac{1}{\omega_\alpha^2}\frac{\int n_0e^{-i(\mathbf{k}'\cdot\mathbf{r})}\nabla U_\alpha d\tau\cdot\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\int U_\alpha^* N_\alpha d\tau}. \tag{3.8}$$

Inserting this relation into (3.4), we get

$$\frac{d\sigma}{d\Omega}=\left(\frac{e^2}{mc^2}\right)^2\sum_{\substack{\text{polarization of}\\ \text{scattered wave}}}\left|\omega^2\sum_\alpha\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_\alpha)e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_\alpha^2(\omega_\alpha^2-\omega^2-i\epsilon)\int U_\alpha^* N_\alpha d\tau}\right|^2. \tag{3.9}$$

For the forward scattering, the calculations given at the end of this chapter show that if $\mathbf{e}_{\text{out}}$ points in y direction, then the scattering amplitude of (3.9) vanishes. Hence we have

$$\mathbf{e}_{\text{out}}=\mathbf{e}_{\text{in}}. \tag{3.10}$$

Thus we have

$$\frac{d\sigma}{d\Omega}=\left(\frac{e^2}{mc^2}\right)^2\cdot\left|\omega^2\sum_\alpha\frac{1}{\omega_\alpha{}^2-\omega^2-i\epsilon}\frac{\left|\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha{}^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau\right|^2}{\omega_\alpha{}^2\int U_\alpha{}^*N_\alpha d\tau}\right|^2. \tag{3.11}$$

Now we generalize the definition of the oscillator strength given previously by Eq. (2.43) to

$$f_{l,m}(\omega_\alpha)=\frac{\left|\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha{}^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau\right|^2}{\omega_\alpha{}^2\int U_\alpha{}^*N_\alpha d\tau}. \tag{3.12}$$

Inserting this relation into (3.11) we get

$$\frac{d\sigma}{d\Omega}=\left(\frac{e^2}{mc^2}\right)^2\cdot\left|\omega^2\sum_{l,m}\frac{df_{l,m}(\omega_\alpha)}{\omega_\alpha{}^2-\omega^2-i\epsilon}\right|^2, \tag{3.13}$$

where

$$df_{l,m}(\omega_\alpha)=f_{l,m}(\omega_\alpha)\frac{dN}{d\omega_\alpha}d\omega_\alpha. \tag{3.14}$$

This is the well-known standard classical Rayleigh formula which connects the forward differential cross section to the oscillator strength.

Deriving Eq. (3.13) by simple argument, Putnam[13] takes the photoelectric cross section of the hydrodynamical gas model from the preliminary calculations of Wheeler and Fireman. He fits their results by a simple analytic formula for $df/d\omega$ as a function of ω. He uses this representation of the oscillator strength in the Rayleigh formula and integrates to get the elastic scattering cross section as a function of energy for wavelengths great in comparison with atomic dimensions. Eq. (3.4) of course is not limited to the case of long wavelengths and also forward scattering.

We looked for a dispersion relation for nonforward elastic scattering, but we have not yet succeeded in establishing it.

Analysis by the Method of Partial Waves

The scattering amplitude given by Eq. (3.4) can be written in the following way:

$$M(\mathbf{e}_{\text{out}},\mathbf{k}';\mathbf{e}_{\text{in}},\mathbf{k})=(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})\int n_0e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau-\sum_{l'',m''}\int_0^\infty d\omega_\alpha''\frac{dN(\omega_\alpha'')}{d\omega_\alpha''}$$
$$\times\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha''})e^{-i(\mathbf{k}'\cdot\mathbf{x})}d^3x\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha{}^*)e^{i(\mathbf{k}\cdot\mathbf{y})}d^3y}{2\omega_\alpha''\int U_{\alpha''}{}^*N_{\alpha''}d\tau(\omega_\alpha''-\omega-i\epsilon)}, \tag{3.15}$$

where we have defined

$$\begin{aligned}U_{-\omega_\alpha,l,m}&=U_{\omega_\alpha,l,m},\\ N_{-\omega_\alpha,l,m}&=N_{\omega_\alpha,l,m}.\end{aligned} \tag{3.16}$$

[13] P. Putnam, "Photon scattering by statistical atom," Ph.D. thesis, Princeton (1960).

We analyze the scattering amplitude $M(\mathbf{e}_{\text{out}}, \mathbf{k}'; \mathbf{e}_{\text{in}}, \mathbf{k})$ in Eq. (3.15) by the methods of partial waves. We choose the coordinate system as in Fig. 2. Here,

$$\begin{aligned}\mathbf{e}_{\text{in}} &= (\cos\phi_{\text{in}}, \sin\phi_{\text{in}}, 0),\\ \mathbf{e}_{\text{out}} &= (\cos\phi_{\text{out}}\cos\Theta, \sin\phi_{\text{out}}, -\cos\phi_{\text{out}}\sin\Theta),\\ \mathbf{k}' &= k(\sin\Theta, 0, \cos\Theta).\end{aligned} \tag{3.17}$$

The direct scattering term is well known and given by

$$(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})4\pi\int n_0(r)j_0(|\mathbf{k}'-\mathbf{k}|r)r^2dr. \tag{3.18}$$

To evaluate the dispersive scattering term we employ expansion of the familiar type:

$$e^{-i(\mathbf{k}'\cdot\mathbf{r})} = 4\pi\sum_{l'=0}^{\infty}\sum_{m'=-l'}^{l'}(-i)^{l'}j_{l'}(kr)Y_{l',m'} \times(\Theta,0)Y_{l',m'}{}^*(\theta,\phi). \tag{3.19}$$

We assume for $U_{\alpha''}(r,\theta,\phi)$ the following form:

$$U_{\alpha''}(r,\theta,\phi) = f_{\omega'',l''}(r)Y_{l'',m''}(\theta,\phi). \tag{3.20}$$

For simplicity we write $f_{\alpha''}(r)$ instead of $f_{\omega'',l''}(r)$ in the following formulas. Then two factors, $\int n_0(r)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha}{}^*) \times e^{i(\mathbf{k}\cdot\mathbf{y})}d^3y$ and $\int n_0(r)(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha''})e^{-i(\mathbf{k}'\cdot\mathbf{x})}d^3x$ in the dispersive scattering term can be reduced further as below:

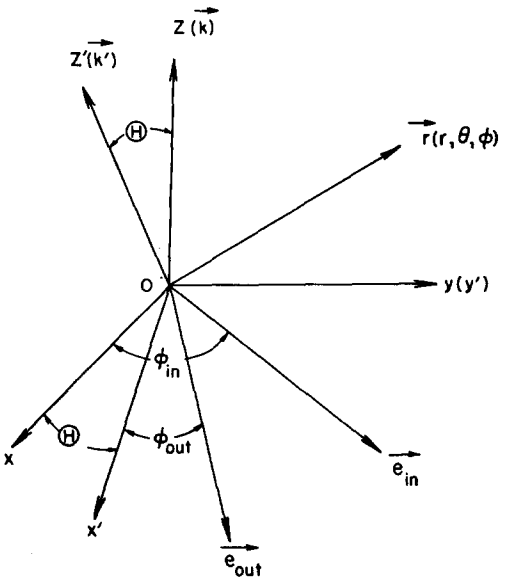

FIG. 2. The coordinate system that we use for the calculation of the matrix element, $M(\mathbf{e}_{\text{out}}, \mathbf{k}', \cdot\mathbf{e}_{\text{in}}, \mathbf{k})$ for the coherent scattering in Eq. (3.15). $\mathbf{k}$ and $\mathbf{k}'$ are the wave vectors of the incoming wave and the scattered wave. $\mathbf{e}_{\text{in}}$ and $\mathbf{e}_{\text{out}}$ also are the polarization vectors of the incoming wave and the scattered wave. We choose the direction of $\mathbf{k}$ as the z axis so that $\mathbf{e}_{\text{in}}$ lies in the (x,y) plane. Also, we take x and y axes such that $\mathbf{k}'$—taken as the z' axis—lies in the (z,x) plane.

$$\int n_0(r)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha''}{}^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau = e^{-im''\phi_{\text{in}}}\left(\frac{(l''-1)!}{(l''+1)!}\frac{(2l''+1)\pi}{4}\right)^{\frac{1}{2}}\sum_{l=0}^{\infty}(2l+1)i^l \times\left[K_1(l,l'')\int_0^{\infty}dr\cdot r^2n_0\frac{df_{\alpha''}}{dr}j_l(kr)+K_2(l,l'')\int_0^{\infty}dr\cdot rn_0f_{\alpha''}j_l(k\cdot r)\right], \tag{3.21}$$

where $m''=\pm1$ and

$$\begin{aligned}K_1(l,l'') &= -\frac{2(l-1)l}{(2l-1)(2l+1)}\delta_{l'',l-1}+\frac{2(l+1)(l+2)}{(2l+1)(2l+3)}\delta_{l'',l+1},\\ K_2(l,l'') &= \frac{2(l-1)^2l}{(2l-1)(2l+1)}\delta_{l'',l-1}+\frac{2(l+1)(l+2)^2}{(2l+1)(2l+3)}\delta_{l'',l+1}.\end{aligned} \tag{3.22}$$

From this result we see that the incoming partial wave of angular momentum l induces hydrodynamical oscillations of proper modes of angular momentum $(l-1)$ and $(l+1)$.

$$\begin{aligned}&\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha''})e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau\\ &= (\cos\phi_{\text{out}}\cos\Theta+im''\sin\phi_{\text{out}})\pi\left(\frac{(l''-1)!}{(l''+1)!}(2l''+1)\right)^{\frac{1}{2}}\sum_{l'=0}^{\infty}(-i)^{l'}(2l'+1)Y_{l',0}(\Theta,0)\\ &\times\left[K_1(l',l'')\int_0^{\infty}dr\,r^2n_0j_{l'}(kr)\frac{df_{\alpha''}}{dr}+K_2(l',l'')\int_0^{\infty}dr\,rn_0j_{l'}(kr)f_{\alpha''}\right]+\pi\left(\frac{(l''-1)!}{(l''+1)!}(2l''+1)\right)^{\frac{1}{2}}\\ &\times\sum_{l'\leqslant2}^{\infty}(-i)^{l'}\left(\frac{(l'-2)!}{(l'+2)!}(2l'+1)\right)^{\frac{1}{2}}Y_{l',2}(\Theta,0)\left[K_3(l',l'')\int_0^{\infty}dr\,r^2n_0j_{l'}(kr)\frac{df_{\alpha''}}{dr}+K_4(l',l'')\int_0^{\infty}dr\,rn_0j_{l'}(kr)f_{\alpha''}\right]\\ &-\cos\phi_{\text{out}}\sin\Theta2\pi\left(\frac{(l''-1)!}{(l''+1)!}(2l''+1)\right)^{\frac{1}{2}}\sum_{l'=0}^{\infty}(-i)^{l'}\left(\frac{(l'-1)!}{(l'+1)!}(2l'+1)\right)^{\frac{1}{2}}Y_{l',0}(\Theta,0)\\ &\times\left[K_5(l',l'')\int_0^{\infty}dr\,r^2n_0j_{l'}(kr)\frac{df_{\alpha''}}{dr}+K_6(l',l'')\int_0^{\infty}dr\,rn_0j_{l'}(kr)f_{\alpha''}\right],\end{aligned} \tag{3.23}$$

where $m''=\pm 1$ and

$$K_3(l',l'')=\frac{-2}{(2l''-1)(2l''+1)}\frac{(l''+1)!}{(l''-3)!}\delta_{l',l''-1}+\frac{2}{(2l''+1)(2l''+3)}\frac{(l''+3)!}{(l''-1)!}\delta_{l',l''+1},$$

$$K_4(l',l'')=-\frac{2(l''+1)(l''+1)!}{(2l''-1)(2l''+1)(l''-3)!}\delta_{l',l''-1}-\frac{2l''(l''+3)!}{(2l''+1)(2l''+3)(l''-1)!}\delta_{l',l''+1},$$

$$K_5(l',l'')=\frac{2(l''+1)l''!}{(2l''-1)(2l''+1)(l''-2)!}\delta_{l',l''-1}+\frac{2l''(l''+2)!}{(2l''+1)(2l''+3)l''!}\delta_{l',l''+1},\tag{3.24}$$

$$K_6(l',l'')=-\frac{2(l''+1)(l''+1)!}{(2l''-1)(2l''+1)(l''-2)!}\delta_{l',l''-1}+\frac{2l''(l''+2)!}{(2l''+1)(2l''+3)(l''-1)!}\delta_{l',l''+1}.$$

We conclude that, for the hydrodynamical oscillations of proper mode of angular momentum l'', the partial waves of angular momentum $(l''-1)$ and $(l''+1)$ of the scattered wave give the nonvanishing contribution. These expressions are well adapted to numerical evaluation.

IV. INCOHERENT SCATTERING

Use of Einstein's Statistical Argument to Derive Incoherent Scattering of Radiation

In this section we discuss the incoherent scattering of radiation in the framework of a classical picture, in which we treat both the oscillations of the electron gas and the perturbing effect of the electromagnetic radiation in classical terms.[14] We recall Einstein's derivation of the relation between coefficients $B_{0\to 1}$ and $B_{1\to 0}$ for absorption and stimulated emission (essentially classical concepts) with the spontaneous emission rate $A_{1\to 0}$.

The principle of detailed balance states that the number of transitions per second up has to equal, in equilibrium, the number of transitions down:

$$(B_{0\to 1}\rho_\omega)N_0=(B_{1\to 0}\rho_\omega+A_{1\to 0})N_1,\tag{4.1}$$

where the equilibrium numbers have the ratio

$$N_1/N_0=e^{\hbar\omega/kT}.\tag{4.2}$$

From this argument, Einstein derived, in a familiar way, the formula

$$A_{1\to 0}=(\hbar\omega^3/(2\pi)^3c^3)B_{1\to 0}.\tag{4.3}$$

In turn, $B_{1\to 0}(\omega)$ is related to the absorption cross section $\sigma(\omega')$, which has a sharp peak at $\omega'=\omega$, by the equation

$$B_{1\to 0}(\omega)=c/\hbar\omega\int\sigma(\omega')d\omega'.\tag{4.4}$$

Therefore, we have

$$A_{1\to 0}(\omega)=[\omega^2/(2\pi)^3c^2]\int\sigma(\omega')d\omega'.\tag{4.5}$$

Here, following Gregory Breit, we generalize this kind of reasoning to find the cross section for a slightly more complicated process where atoms are illuminated by the primary beam of frequency ω, of given direction and given polarization, and emit electromagnetic radiation of frequency ω', of given direction and given polarization.

We illuminate atoms with an additional beam which has the same frequency, direction, and polarization as does the secondary wave, and contains $\omega'^2/(2\pi c)^3$ photons per cm^2 and per second and per unit solid angle. Under these two perturbations, we solve the hydrodynamical equations to the second order and calculate the absorption cross section. We insert this expression into Eq. (4.5) and then we obtain the differential cross section for the incoherent scattering. We summarize the reasoning in the following formula:

$$\left(\frac{d\sigma}{d\Omega}\right)_{0\to k}=\frac{\left[\begin{array}{l}\text{Probability per second of transition from the ground state to the } k\text{th excited state when the}\\ \text{atom is illuminated by the primary radiation of circular frequency } \omega \textit{ and simultaneously} \text{ by}\\ \text{light of frequency } \omega' \text{ and of given polarization containing } \omega'^2/(2\pi c)^3 \text{ photons per cm}^2\text{, and}\\ \text{per second, and per unit solid angle.}\end{array}\right]}{[\text{Number of primary photons incident per cm}^2\text{ and per second}]},\tag{4.6}$$

where

$$\omega=\omega_k+\omega'.$$

This method was applied to the scattering of light by a hydrogen atom and it was shown (details of analysis not reported here) that the result for the differential cross section for incoherent scattering is identical with the formula obtained by the quantum field theory of radiation.

We rewrite this definition into a more convenient

[14] G. Breit (the author was told by Professor Wheeler that G. Breit had used this method).

form for calculations,

$$\left(\frac{d\sigma}{d\Omega}\right)_{0\to k}=\frac{\omega'^2}{8\pi^3c^2}\int_{\omega-\omega_k-\epsilon}^{\omega-\omega_k+\epsilon}\sum(\omega,\omega_1)d\omega_1 \tag{4.7}$$

and

$$\sum(\omega,\omega_1)=\frac{(d/dt)\Delta E_k^{(2)}(\omega,\omega_1)}{(\hbar\omega_k)}\Bigg/\frac{|\mathbf{E}_{\text{in}}{}^0|^2c}{8\pi\hbar\omega}\frac{|\mathbf{E}_{\text{out}}{}^0|^2c}{8\pi\hbar\omega_1}. \tag{4.8}$$

Here $\Delta E_k^{(2)}(\omega,\omega_1)$ is the energy uptake by the atom when the atom jumps to the kth excited state under two perturbations of frequency, ω and ω_1, respectively. To evaluate $\Delta E_k^{(2)}(\omega,\omega_1)$, we make a calculation of second order.

The equation of motion, field equation, and the equation of continuity in second order in the external field are as follows:

$$\frac{\partial u_2}{\partial t}=\frac{h^2}{3m^2}\left(\frac{3}{8\pi}\right)^{\frac{2}{3}}\left(\frac{n_2}{n_0^{\frac{1}{3}}}-\frac{1}{6}\frac{n_1^2}{n_0^{\frac{4}{3}}}\right)+\frac{V_2}{m}+\frac{1}{2}\left(\nabla u_1+\frac{e}{mc}\mathbf{A}\right)^2, \tag{4.9}$$

$$\nabla^2V_2=-4\pi e^2n_2,$$

$$\frac{\partial n_2}{\partial t}=\nabla\cdot\left[n_0\nabla u_2+n_1\left(\nabla u_1+\frac{e}{mc}\mathbf{A}\right)\right].$$

As before we expand n_2, u_2, and V_2 in terms of N_k, U_k, and V_k such that

$$n_2=\sum_k g_k^{(2)}(t)N_k(r,\theta,\phi),$$
$$u_2=\sum_k h_k^{(2)}(t)U_k(r,\theta,\phi), \tag{4.10}$$
$$V_2=\sum_k d_k^{(2)}(t)V_k(r,\theta,\phi).$$

Inserting Eqs. (4.10) into Eqs. (4.9) we get

$$d_k^{(2)}(t)=-mg_k^{(2)}(t),$$

$$\dot g_k^{(2)}(t)+\omega_k{}^2h_k^{(2)}(t)=\frac{1}{\int U_k{}^*N_kd\tau}\left\{\sum_{i,j}g_i^{(1)}(t)h_j^{(1)}(t)\int U_k{}^*\nabla\cdot(N_i\nabla U_j)d\tau+\frac{e}{mc}\sum_i g_i^{(1)}(t)\int U_k{}^*\nabla\cdot(N_i\mathbf{A})d\tau\right\}, \tag{4.11}$$

and

$$\dot h_k^{(2)}(t)=g_k^{(2)}(t)+\frac{1}{\int U_kN_k{}^*d\tau}\left\{-\frac{h^2}{18m^2}\left(\frac{3}{8\pi}\right)^{\frac{2}{3}}\sum_{i,j}g_i^{(1)}(t)g_j^{(1)}(t)\int N_k{}^*N_iN_jn_0^{-\frac{1}{3}}d\tau\right.$$
$$\left.+\frac{1}{2}\sum_{i,j}h_i^{(1)}(t)h_j^{(1)}(t)\int N_k{}^*(\nabla U_i\cdot\nabla U_j)d\tau+\frac{e}{mc}\sum_i h_i^{(1)}(t)\int N_k{}^*(\nabla U_i\cdot\mathbf{A})d\tau+\frac{e^2}{2m^2c^2}\int N_k{}^*\mathbf{A}^2d\tau\right\}.$$

We impose the same initial condition as before.

$$g_k^{(2)}(t=0)=0,$$
$$h_k^{(2)}(t=0)=0. \tag{4.12}$$

Let the interaction be stopped at t and the atom be left in a state of free vibration. Then the energy uptake $\Delta E_k^{(2)}(t)$ for mode k is given by the familiar formula

$$\nabla E_k^{(2)}(t)=\tfrac{1}{2}m\int U_kN_k{}^*d\tau(|g_k^{(2)}(t)|^2+\omega_k{}^2|h_k^{(2)}(t)|^2). \tag{4.13}$$

We insert these results into Eqs. (4.7) and (4.8) and get the final result

$$\left(\frac{d\sigma}{d\Omega}\right)_{0\to k}=\left(\frac{e^2}{mc^2}\right)^2\frac{\omega'}{\omega\omega_k}\frac{\hbar}{2m}\frac{1}{\int U_kN_k{}^*d\tau}\left|-(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})\omega_k\int N_k{}^*e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d\tau+\sum_i\frac{1}{\int U_i{}^*N_id\tau}\right.$$
$$\times\left\{\frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_i{}^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau}{\omega_i{}^2-\omega^2}\int e^{-i(\mathbf{k}'\cdot\mathbf{r})}\mathbf{e}_{\text{out}}\cdot(\omega_kN_k{}^*\nabla U_i-\omega N_i\nabla U_k{}^*)d\tau+\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_i{}^*)e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau}{\omega_i{}^2-(\omega-\omega_k)^2}\right.$$

(Continued on next page)

$$\times\int e^{i(\mathbf{k}\cdot\mathbf{r})}\mathbf{e}_{\text{in}}\cdot(\omega_k N_k^*\nabla U_i+(\omega-\omega_k)N_i\nabla U_k^*)d\tau\Big\}+\sum_{i,j}\frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_i^*)e^{i(\mathbf{k}\cdot\mathbf{r})}d\tau\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_j^*)e^{-i(\mathbf{k}'\cdot\mathbf{r})}d\tau}{\int U_i^*N_i d\tau\int U_j^*N_j d\tau(\omega_i^2-\omega^2)[\omega_j^2-(\omega-\omega_k)^2]}$$

$$\times\Big[-\omega_k\int N_k^*(\nabla U_i\cdot\nabla U_j)d\tau+\omega\int N_i(\nabla U_k^*\cdot\nabla U_j)d\tau-(\omega-\omega_k)\int N_j(\nabla U_k^*\cdot\nabla U_i)d\tau$$

$$+\left(\frac{3}{8\pi}\right)^{\frac{1}{3}}\frac{h^2}{9m^2}\omega\omega_k(\omega-\omega_k)\int N_k^*N_iN_jn_0^{-\frac{1}{3}}d\tau\Big]\Big|^2. \quad (4.14)$$

Comparing this expression with Kramers-Heisenberg[15]–Waller's[16] formula of wave mechanics, we see that the first term corresponds to the direct scattering. This correspondence is pursued in Sec. VI. The second and third terms are also familiar and correspond to the dispersive scattering. The last term is less familiar. One might argue that the correspondence theoretical method we used is the cause of this term. However, we applied our method to the simple system of hydrogen atom plus radiation field where the hydrogen atom is composed of a fixed force center and a bound electron. And we used Schrödinger's equation for this electron. Then we got the formula of Kramers, Heisenberg, and Waller. One might argue in another way that we omitted the electromagnetic interaction inside the atom and retained the electrostatic interaction. This electrostatic interaction eliminates itself in determining basic functions N_k, U_k, and V_k. Therefore, this elimination of the electrostatic interaction is reflected as the direct interaction between three modes i, j, and k. To check this point, let us drop the Poisson equation and the deviation in electrostatic potential. Then we see that all formulas which we obtained, Eqs. (2.25), (3.3), and (4.14), remain the same. The electrostatic interaction changes only N_k and U_k themselves. Thus we come to the conclusion that the complexity of the original hydrodynamical equations produced this term. The term in Eq. (4.14) which includes $\int N_k^*N_iN_jn_0^{-\frac{1}{3}}d\tau$ arises from the term in the first equation of (4.9) which includes n_1^2. This term shows that two modes i and j, excited by the external field, excite mode k through the compressional force. Terms in Eq. (4.14) which include $\int N_k^*(\nabla U_i\cdot\nabla U_j)d\tau$ and $\int N_i(\nabla U_k^*\cdot\nabla U_j)d\tau$ arise from terms in the first and third equations of (4.9) which include $(\nabla u_1)^2$ and $\nabla\cdot(n_1\nabla u_1)$, respectively. This kind of term is reasonable. *Flow* of the Fermi gas excited in mode i_1, and *flow* of the Fermi gas excited in mode i_2, build up in combination certain excesses and deficits of density which excite mode i_3. Indeed, suppose that at $t=0$ the external field be stopped and the atom be left in the state where two modes i and j are excited. Then, from Eq. (4.11), we see that through these terms, $\int N_k^*N_iN_jn_0^{-\frac{1}{3}}d\tau$, $\int N_k^*(\nabla U_i\cdot\nabla U_j)d\tau$, and $\int N_i\times(\nabla U_k^*\cdot\nabla U_j)d\tau$, the third mode k is excited.

[15] H. A. Kramers and W. Heisenberg, Z. Physik **31**, 681 (1925).
[16] I. Waller, Z. Physik **51**, 213 (1928).

V. QUANTIZATION OF THE OSCILLATING ATOM

Hamiltonian of the Free Field Which Is Quadratic in Deviations of Charge Density and Velocity Potential

Bloch's hydrodynamical electron gas is capable of oscillations around the ground state. We termed this oscillation the free vibration. In this section—following work of Bloch,[9] Wentzel,[17] and Bohr[18]—we give a proper quantum field theoretical treatment to this free vibration.

Let us start with total Hamiltonian of the coupled system of electron gas and radiation field:

$$H=\frac{m}{2}\int n\left(\nabla u+\frac{e}{mc}\mathbf{A}\right)^2d\tau+\frac{e^2}{2}\int\frac{n^{(a)}n^{(b)}}{r_{ab}}d\tau_a d\tau_b$$
$$-\int\frac{Ze^2}{r}nd\tau+\int d\tau n\int_0^n\frac{P(n^*)}{n^{*2}}dn^*$$
$$\equiv H_0+H_I. \quad (5.1)$$

Further we divide H_0 into the Hamiltonian of the ground state plus the Hamiltonian which describes deviations from Thomas-Fermi's distribution.

$$H_0=\frac{m}{2}\int n(\nabla u)^2d\tau+\frac{e^2}{2}\int\frac{n^{(a)}n^{(b)}}{r_{ab}}d\tau_a d\tau_b$$
$$-\int\frac{Ze^2}{r}nd\tau+\int d\tau\, n\int_0^n\frac{P(n^*)}{n^{*2}}dn^*$$
$$\equiv H_{0,0}+H_{0,2}, \quad (5.2)$$

[17] G. Wentzel, *Quantum Theory of Fields* (Interscience Publishers, Inc., New York, 1949).
[18] A. Bohr, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **26**, No. 14 (1952).

where

$$H_{0,0}=\frac{e^2}{2}\int\frac{n_0^{(a)}n_0^{(b)}}{r_{ab}}d\tau_a d\tau_b-\int\frac{Ze^2}{r}n_0 d\tau+d\tau n_0\int_0^n\frac{P(n^*)}{n^{*2}}dn^*$$

and

$$H_{0,2}=\frac{m}{2}\int n_0(\nabla u_1)^2 d+\frac{e^2}{2}\int\frac{n_1^{(a)}n_1^{(b)}}{r_{ab}}d\tau_a d\tau_b+\frac{1}{2}\int d\tau n_1^2\frac{1}{n_0}\frac{dP}{dn_0}. \quad (5.3)$$

The interaction Hamiltonian is given by

$$H_I=\frac{m}{2}\int n\left\{\frac{e}{mc}(\nabla u\cdot\mathbf{A}+\mathbf{A}\cdot\nabla u)+\left(\frac{e}{mc}\right)^2\mathbf{A}^2\right\}d\tau. \quad (5.4)$$

Canonically Conjugate Coordinate and Momentum

Let us forget the interaction H_I for a while and discuss the free vibration. We regard the system described by $H_{0,0}$ as vacuum and n_1 and u_1 as dynamical variables. As before, we expand these operators by N_α and U_α.

$$\begin{aligned}n_1(t)&=-\sum_{\omega_\alpha\geqslant 0}\sum_{l,m}C_{\omega_\alpha,l,m}(t)\omega_\alpha N_{\omega_\alpha,l,m}(r,\theta,\phi),\\ n_1(t)&=\sum_{\omega_\alpha\geqslant 0}\sum_{l,m}h_{\omega_\alpha,l,m}(t)U_{\omega_\alpha,l,m}(r,\theta,\phi).\end{aligned} \quad (5.5)$$

To secure real n_1 and u_1, we claim

$$\begin{aligned}C_{\omega_\alpha,l,m}(t)&=C_{\omega_\alpha,l,-m}^*(t)\\ \text{and}\qquad h_{\omega_\alpha,l,m}(t)&=h_{\omega_\alpha,l,-m}^*(t).\end{aligned} \quad (5.6)$$

With these operators, $C_{\omega_\alpha,l,m}$ and $h_{\omega_\alpha,l,m}$, we rewrite $H_{0,2}$ into the following form:

$$H_{0,2}=\sum_{\substack{\omega_\alpha\geqslant 0\\ l,m}}(|h_{\omega_\alpha,l,m}|^2+|C_{\omega_\alpha,l,m}|^2)\times\frac{m}{2}\omega_\alpha^2\int U_\alpha^* N_\alpha d\tau \equiv T+V. \quad (5.7)$$

From the equation of continuity we have

$$h_{\omega_\alpha,l,m}=\omega_\alpha^{-1}\dot{C}_{\omega_\alpha,l,m}. \quad (5.8)$$

We define the momentum $\pi_{\omega_\alpha,l,m}$ conjugate to $C_{\omega_\alpha,l,m}$ by

$$\pi_{\omega_\alpha,l,m}=\partial T/\partial\dot{C}_{\omega_\alpha,l,m}=C_{\omega_\alpha,l,m}^*\times\left(m\int U_{\omega_\alpha,l,m}^* N_{\omega_\alpha,l,m}d\tau\right). \quad (5.9)$$

Inserting (5.8) and (5.9) into (5.7) we get

$$H_{0,2}=\sum_\alpha\left\{\frac{1}{\left(2m\int U_\alpha^* N_\alpha d\tau\right)}|\pi_\alpha|^2+\left(\frac{1}{2}m\omega_\alpha^2\int U_\alpha^* N_\alpha d\tau\right)|C_\alpha|^2\right\}. \quad (5.10)$$

Here, and from now on, we use the symbol α instead of ω_α, l, and m. From the Hamiltonian, (5.10) we see that the hydrodynamical system can be regarded as an assembly of harmonic oscillators with frequency ω_α. We introduce creation and annihilation operators b_α^* and b_α of the hydrodynamical oscillation of the proper mode α in the familiar way:

$$\begin{aligned}C_\alpha&=\left(\frac{\hbar}{2B_\alpha\omega_\alpha}\right)^{\frac{1}{2}}(b_\alpha+b_{-\alpha}^*),\\ \pi_\alpha&=i(\tfrac{1}{2}\hbar B_\alpha\omega_\alpha)^{\frac{1}{2}}(b_\alpha^*-b_{-\alpha}),\end{aligned} \quad (5.11)$$

where

$$B_\alpha=m\int U_\alpha^* N_\alpha d\tau. \quad (5.12)$$

We define the number density operator n_α by

$$\begin{aligned}b_\alpha b_\alpha^*&=n_\alpha+1\\ \text{and}\qquad b_\alpha^* b_\alpha&=n_\alpha.\end{aligned} \quad (5.13)$$

Then the Hamiltonian $H_{0,2}$ takes a familiar form

$$H_{0,2}=\sum_\alpha\hbar\omega_\alpha(n_\alpha+\tfrac{1}{2}). \quad (5.14)$$

S-Matrix Formalism[19]: Matrix Elements for Coherent and Incoherent Scattering

Now we deal with a coupled system of hydrodynamical electron gas and radiation field. We use the interaction representation. The interaction Hamiltonian density operator $H_I(x_1,x_2,x_3,t)$ is given by

$$H_I(x)=\frac{m}{2}(n_0(\mathbf{x})+n_1(x))\left\{\frac{e}{mc}(\mathbf{A}(x)\cdot\nabla u_1(x)+\nabla u_1(x)\cdot\mathbf{A}(x))+\frac{e^2}{m^2c^2}\mathbf{A}^2(x)\right\}. \quad (5.15)$$

[19] S. Tomonaga, Progr. Theoret. Phys. (Kyoto) **1**, 27 (1946); J. Schwinger, Phys. Rev. **74**, 1439 (1948); **75**, 651 (1949), **76**, 790 (1949); R. P. Feynman, Phys. Rev. **76**, 769 (1949); F. J. Dyson, Phys. Rev. **75**, 486, 1736 (1949).

The S matrix is given by

$$S=\sum_{n=0}^{\infty}\left(\frac{1}{i\hbar}\right)^n\frac{1}{n!}\int_{-\infty}^{\infty}d^4x_1\cdots d^4x_nP[H_I(x_1),\cdots,H_I(x_n)], \tag{5.16}$$

or more conveniently for our purpose,

$$S=1+\frac{1}{i\hbar}\int_{-\infty}^{\infty}H_I(x_1)d^4x_1+\left(\frac{1}{i\hbar}\right)^2\int_{-\infty}^{\infty}d^4x_1\int_{-\infty}^{t_1}d^4x_2H_I(x_1)H_I(x_2)+\cdots. \tag{5.17}$$

We quantize the radiation field too.

$$A_\mu(x)=\frac{1}{\sqrt{V}}\sum_{\mathbf{k}}\sum_{\lambda=1}^{2}\frac{e_\mu^{(\lambda)}}{\sqrt{2\omega_k}}(e^{ik\cdot x}a^{(\lambda)}(k)+e^{-ikx}a^{(\lambda)*}(k)). \tag{5.18}$$

We take only transversal fields because we already included longitudinal and scalar fields into the Coulomb potential. Then for the coherent scattering we get, to the lowest order,

$$\begin{aligned}
&\langle\mathbf{k}',\mathbf{e}_{\text{out}},0|S|\mathbf{k},\mathbf{e}_{\text{in}},0\rangle\\
&=\frac{e^2}{mc^2}\frac{\pi}{\hbar}\frac{1}{V\omega}\delta(\omega_{\text{out}}-\omega)\Bigg\{(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})\int n_0(\mathbf{x}_1)e^{i(\mathbf{k}-\mathbf{k}')}d^3x_1-\frac{1}{2}\sum_\alpha\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_\alpha)e^{-i(\mathbf{k}'\cdot\mathbf{x}_1)}d^3x_1\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha^*)e^{i(\mathbf{k}\cdot\mathbf{x}_2)}d^3x_2}{\int U_\alpha^*N_\alpha d\tau(\omega_\alpha-\omega-i\epsilon)}\\
&-\frac{1}{2}\sum_\alpha\frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_\alpha)e^{i(\mathbf{k}\cdot\mathbf{x}_1)}d^3x_1\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_\alpha^*)e^{-i(\mathbf{k}'\cdot\mathbf{x}_2)}d^3x_2}{\omega_\alpha\int U_\alpha^*N_\alpha d\tau(\omega_\alpha+\omega-i\epsilon)}-\frac{\hbar}{4m}\sum_{\alpha_1,\alpha_2}\frac{1}{\omega_{\alpha1}\omega_{\alpha2}\int U_{\alpha1}^*N_{\alpha1}d\tau\int U_{\alpha2}^*N_{\alpha2}d\tau}\\
&\times\int d^3x_1[\omega_{\alpha1}N_{\alpha1}(\mathbf{x}_1)(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha2}(\mathbf{x}_1))+\omega_{\alpha2}N_{\alpha2}(\mathbf{x}_1)(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha1}(\mathbf{x}_1))]e^{-i(\mathbf{k}'\mathbf{x}_1)}\\
&\times\frac{\int d^3x_2[\omega_{\alpha1}N_{\alpha1}^*(\mathbf{x}_2)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha2}^*(\mathbf{x}_2))+\omega_{\alpha2}N_{\alpha2}^*(\mathbf{x}_2)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha1}^*(\mathbf{x}_2))]e^{i(\mathbf{k}\cdot\mathbf{x}_2)}}{(\omega_{\alpha1}+\omega_{\alpha2}-\omega-i\epsilon)}\\
&-\frac{\hbar}{4m}\sum_{\alpha_1,\alpha_2}\frac{1}{\omega_{\alpha1}\omega_{\alpha2}\int U_{\alpha1}^*N_{\alpha1}d\tau\int U_{\alpha2}^*N_{\alpha2}d\tau}\\
&\times\int d^3x_1[\omega_{\alpha1}N_{\alpha1}(\mathbf{x}_1)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha2}(\mathbf{x}_1)+\omega_{\alpha2}N_{\alpha2}(\mathbf{x}_1)(\mathbf{e}_{\text{in}}\cdot\nabla U_{\alpha1}(\mathbf{x}_1))]e^{i(\mathbf{k}\cdot\mathbf{x}_1)}\\
&\times\frac{\int d^3x_2[\omega_{\alpha1}N_{\alpha1}^*(\mathbf{x}_2)(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha2}^*(\mathbf{x}_2))+\omega_{\alpha2}N_{\alpha2}^*(\mathbf{x}_2)(\mathbf{e}_{\text{out}}\cdot\nabla U_{\alpha1}^*(\mathbf{x}_2))]e^{-(\mathbf{k}'\cdot\mathbf{x}_2)}}{(\omega_{\alpha1}+\omega_{\alpha2}+\omega-i\epsilon)}\Bigg\}.
\end{aligned} \tag{5.19}$$

We are familiar with the first three terms. These terms agree with previous results in Sec. III. The last two terms are new. These terms arise from

$$(e/2c)n_1(x)(\mathbf{A}(x)\cdot\nabla u_1(x)+\nabla u_1(x)\cdot\mathbf{A}(x)) \tag{5.20}$$

in $H_I(x)$. Therefore, two hydrodynamical oscillations can be excited or annihilated simultaneously. From the consideration given in Sec. IV, we see that the state where two modes are excited is not stationary. Two modes in combination excite another mode. Therefore, such a state has an exponentially decaying time dependence. So our method of derivation is not correct in its treatment of these higher-order processes.

We can make the similar calculation for incoherent scattering. We write down the result

$$\langle \mathbf{k}', \mathbf{e}_{\text{out}}; \beta | S | \mathbf{k}, \mathbf{e}_{\text{in}}; 0 \rangle$$

$$= \frac{e^2\pi}{mc^2} \frac{(\omega_\beta+\omega'-\omega)}{V(\omega'\omega)^{\frac{1}{2}}} \frac{1}{(2B_\beta\omega_\beta\hbar)^{\frac{1}{2}}} \Bigg\{ -(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})\omega_\beta \int N_\beta^*(\mathbf{x}_1) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_1} d^3x_1$$

$$+\frac{1}{2}\sum_i \Bigg[\frac{1}{\omega_i \int U_i^* N_i d\tau} \Bigg\{ \frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_i^*) e^{i(\mathbf{k}\cdot\mathbf{x}_2)} d^3x_2 \int d^3x_1\, e^{-i(\mathbf{k}'\cdot\mathbf{x}_1)} \mathbf{e}_{\text{out}}\cdot(\omega_\beta N_\beta^*\nabla U_i - \omega_i N_i \nabla U_\beta^*)}{(\omega_i-\omega-i\epsilon)}$$

$$+\frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_i^*) e^{-i(\mathbf{k}'\cdot\mathbf{x}_2)} d^3x_2 \int d^3x_1\, e^{i(\mathbf{k}\cdot\mathbf{x}_1)} \mathbf{e}_{\text{in}}\cdot(\omega_\beta N_\beta^*\nabla U_i - \omega_i N_i \nabla U_\beta^*)}{(\omega_i+(\omega-\omega_\beta)-i\epsilon)}$$

$$+\frac{\int n_0(\mathbf{e}_{\text{in}}\cdot\nabla U_i) e^{i(\mathbf{k}\cdot\mathbf{x}_1)} d^3x_1 \int d^3x_2\, e^{-i(\mathbf{k}'\cdot\mathbf{x}_2)} \mathbf{e}_{\text{out}}\cdot(\omega_\beta N_\beta^*\nabla U_i^* + \omega_i N_i^*\nabla U_\beta^*)}{(\omega_i+\omega-i\epsilon)}$$

$$+\frac{\int n_0(\mathbf{e}_{\text{out}}\cdot\nabla U_i) e^{-i(\mathbf{k}'\cdot\mathbf{x}_1)} d^3x_1 \int d^3x_2\, e^{i(\mathbf{k}\cdot\mathbf{x}_2)} \mathbf{e}_{\text{in}}\cdot(\omega_\beta N_\beta^*\nabla U_i^* + \omega_i N_i^*\nabla U_\beta^*)}{(\omega_i+\omega_\beta-\omega-i\epsilon)} \Bigg\} \Bigg] \Bigg\}. \quad (5.21)$$

Again the first three terms are familiar and agree with previous results *provided* that those parts of the second and third terms in Eq. (4.14) which have a factor $(\omega_k N_k^*\nabla U_i - \omega_i N_i \nabla U_k^*)$ after rearranging these two terms such that they have the denominator which depends linearly on ω, can be neglected compared with other terms. The last two terms correspond to the simultaneous excitation of two modes i and β. However, from the reason given previously, our method for this particular process is not correct.

In this way, we can automatically compute other processes which involve the interaction between hydrodynamical electron gas and radiation field. However, we do not proceed further into these processes, but we turn attention to the Compton scattering because, for this process, we can easily compare hydrodynamical results with experiments and we can work out cross sections with the semi-classical JWKB approximation, both for the hydrodynamical model and for standard wave mechanics.

VI. COMPTON SCATTERING

Rederivation by Orbital Method Plus the Method of Stationary Phase of Heisenberg's Formula

The intensity for the Compton scattering is in the approximation of direct scattering, proportional to

$$|M_{if}|^2 = |\psi_i(\mathbf{x}_1, \sigma_1; \cdots; \mathbf{x}_{2n}, \sigma_{2n})\, \psi_f^*(\mathbf{x}_1, \sigma_1; \cdots; \mathbf{x}_{2n}, \sigma_{2n}) \sum_{j=1}^{2n} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}_j} d\tau_1 \cdots d\tau_{2n}|^2, \quad (6.1)$$

where $Z=2n$.

The ground state wave function of the atom is given in the familiar approximation by a determinant

$$\psi_i(\mathbf{x}_1,\sigma_1;\cdots;\mathbf{x}_{2n},\sigma_{2n})=\frac{1}{[(2n)!]^{\frac{1}{2}}}\begin{vmatrix}\phi_{1,+}(\mathbf{x}_1,\sigma_1) & \cdots & \phi_{s,+}(\mathbf{x}_1,\sigma_1) & \cdots & \phi_{n,-}(\mathbf{x}_1,\sigma_1)\\ \vdots & & & & \vdots\\ \phi_{1,-}(\mathbf{x}_{2n},\sigma_{2n}) & \cdots & \phi_{s,+}(\mathbf{x}_{2n},\sigma_{2n}) & \cdots & \phi_{n,-}(\mathbf{x}_{2n},\sigma_{2n})\end{vmatrix}. \tag{6.2}$$

Let one of the electrons be excited from the sth level with spin $+$ to the $(s+k)$th level with spin $+$. Then the wave function of the final state will again be a determinant of the form

$$\psi_f(\mathbf{x}_1,\sigma_1;\cdots;\mathbf{x}_{2n},\sigma_{2n})=\frac{1}{[(2n)!]^{\frac{1}{2}}}\begin{vmatrix}\phi_{1,+}(\mathbf{x}_1,\sigma_1) & \cdots & \phi_{s+k,+}(\mathbf{x}_1,\sigma_1) & \cdots & \phi_{n,-}(\mathbf{x}_1,\sigma_1)\\ \vdots & & & & \vdots\\ \phi_{1,+}(\mathbf{x}_{2n},\sigma_{2n}) & \cdots & \phi_{s+k,+}(\mathbf{x}_{2n},\sigma_{2n}) & \cdots & \phi_{n,-}(\mathbf{x}_{2n},\sigma_{2n})\end{vmatrix}. \tag{6.3}$$

Then we can easily show that

$$|M_{if}|^2=\left|\int\phi_s(\mathbf{x})\phi_{s+k}^*(\mathbf{x})e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}}d\tau\right|^2. \tag{6.4}$$

For the total intensity, we sum $|M_{if}|^2$ over f and separate off the coherent scattering term as Heisenberg did. We get

$$Z-2\int\int dVdV'\Big|\sum_{\substack{\text{occupied}\\ \text{states}}}\phi_i^*(\mathbf{r})\phi_i(\mathbf{r}')\Big|^2e^{i\mathbf{s}\cdot(\mathbf{r}-\mathbf{r}')}, \tag{6.5}$$

where

$$\mathbf{s}=\mathbf{k}-\mathbf{k}'.$$

To evaluate the second term in this equation, we use the JWKB approximation for $\phi_i(\mathbf{r})$ in the following way:

$$\phi_{n,l,m}(\mathbf{r})=R_{n,l}(r)Y_{l,m}(\theta,\phi), \tag{6.6}$$

$$R_{n,l}(r)=\frac{1}{r}\left(\frac{2m_e}{\pi\hbar}\omega_{,n}\right)^{\frac{1}{2}}\frac{\cos\left\{\int_{r_{\mathrm{T.P.}}}^{r}\left[\frac{2m_e}{\hbar^2}\left(E_{n,l}-V(r)-\frac{(l+\frac{1}{2})^2\hbar^2}{2m_er^2}\right)\right]^{\frac{1}{2}}dr-\frac{\pi}{4}\right\}}{\left[\frac{2m_e}{\hbar^2}\left(E_{n,1}-V(r)-\frac{(l+\frac{1}{2})^2\hbar^2}{2m_er^2}\right)\right]^{\frac{1}{4}}}, \tag{6.7}$$

where

$$\omega_{,n}=\frac{1}{\hbar}\left(\frac{\partial E_{n,l}}{\partial n}\right)_l. \tag{6.8}$$

And in the same approximation, the spherical harmonic takes the following form:

$$Y_{l,m}(\theta,\phi)=\frac{[(l+\frac{1}{2})]^{\frac{1}{2}}}{\pi[(l+\frac{1}{2})^2\sin^2\theta-m^2]^{\frac{1}{4}}}\cos\left\{\int_{\theta_{\mathrm{T.P.}}}^{\theta}\left((l+\tfrac{1}{2})^2-\frac{m^2}{\sin^2\theta}\right)^{\frac{1}{2}}d\theta-\frac{\pi}{4}\right\}e^{im\phi}. \tag{6.9}$$

Using these expressions for $R_{n,l}$ and $Y_{l,m}$, we compute Eq. (6.5).

$$\int\int dVdV'\Big|\sum_{\substack{\text{occupied}\\ \text{states}}}\phi_i^*(\mathbf{r})\phi_i(\mathbf{r}')\Big|^2e^{i\mathbf{s}(\mathbf{r}-\mathbf{r}')}$$
$$=2\sum_{\substack{n,l,m\\ n',l',m'}}\int\int dVdV'\phi_{n,l,m}^*(\mathbf{r})\phi_{n',l',m'}(\mathbf{r})\phi_{n,l,m}(\mathbf{r}')\phi_{n',l',m'}^*(\mathbf{r}')e^{i\mathbf{s}(\mathbf{r}-\mathbf{r}')}. \tag{6.10}$$

Let us perform the angular integration first,

$$\int\phi_{n,l,m}^*(\mathbf{r})\phi_{n',l',m'}(\mathbf{r})e^{i\mathbf{s}\cdot\mathbf{r}}dV=\int dr r^2R_{n,l}(r)R_{n',l'}(r)\sin\theta d\theta d\phi Y_{l,m}^*(\theta,\phi)Y_{l',m'}(\theta,\phi)e^{isr\cos\theta}, \tag{6.11}$$

where we choose the $\mathbf{s}$ direction as z direction.

Using Eq. (6.9) we get

$$\int \sin\theta d\theta d\phi Y_{l,m}^*(\theta,\phi) Y_{l',m'}(\theta,\phi) e^{isr\cos\theta}$$
$$= \delta_{m',m}\left(\frac{(l+\frac{1}{2})(l'+\frac{1}{2})}{2\pi sr}\right)^{\frac{1}{2}} \frac{1}{(\cos\theta_1)^{\frac{1}{2}}((l+\frac{1}{2})^2(l'+\frac{1}{2})^2-m^2s^2r^2)^{\frac{1}{4}}}\left\{\exp\left[i\int_{\theta_{T.P.}}^{\theta_1}(k_{l,m}(\theta)+k_{l',m}(\theta))d\theta+sr\cos\theta_1+\delta_1\right]\right.$$
$$+\exp\left[i\int_{\theta_{T.P.}}^{\theta_1}(k_{l,m}(\theta)-k_{l',m}(\theta))d\theta+sr\cos\theta_1+\delta_2\right]$$
$$\left.+\exp\left[i\int_{\theta_{T.P.}}^{\theta_1}(-k_{l,m}(\theta)-k_{l',m}(\theta))d\theta-sr\cos\theta_1+\delta_3\right]\right\}. \quad (6.12)$$

In the above derivation we used the method of stationary phase. Here

$$k_{l,m}(\theta)=[(l+\tfrac{1}{2})^2-(m^2/\sin^2\theta)]^{\frac{1}{2}}. \quad (6.13)$$

δ_1, δ_2, and δ_3 are constants independent of r. For $l>l'$, we have to drop the third term and for $l<l'$ we have to drop the second term. The stationary angle θ_1 is defined by

$$\frac{d}{d\theta}\left(\int_{\theta_{T.P.}}^{\theta}(k_{l,m}(\theta)+k_{l',m}(\theta))d\theta+sr\cos\theta\right)=0,$$
$$\frac{d}{d\theta}\left(\int_{\theta_{T.P.}}^{\theta}(k_{l,m}(\theta)-k_{l',m}(\theta))d\theta+sr\cos\theta\right)=0, \quad (6.14)$$
$$\frac{d}{d\theta}\left(\int_{\theta_{T.P.}}^{\theta}(-k_{l,m}(\theta)+k_{l',m}(\theta))d\theta+sr\cos\theta\right)=0,$$

where

$$0\leqslant\theta\leqslant\tfrac{1}{2}\pi.$$

All three equations give the same answer to θ_1.

$$\sin\theta_1=\frac{1}{sr}\{(l+\tfrac{1}{2})^2+(l'+\tfrac{1}{2})^2 \pm 2((l+\tfrac{1}{2})^2(l'+\tfrac{1}{2})^2-m^2s^2r^2)^{\frac{1}{2}}\}^{\frac{1}{2}}. \quad (6.15)$$

The physical interpretation of this equation in terms of vector summation of angular momenta appears in Fig. 3.

Next we make the approximation—based on the correspondence principle—to drop the rapidly varying terms in the product of JWKB radial wave functions at two nearby points,[20] finding

$$R_{n,l}(r)R_{n,l}(r') \approx \left(\frac{m_e}{\pi\hbar}\omega_{,n}\right)\frac{1}{R^2}\frac{1}{\left(\frac{2m_e}{\hbar^2}\left(E_{n,l}-V(R)-\frac{(l+\frac{1}{2})^2\hbar^2}{2m_eR^2}\right)\right)^{\frac{1}{2}}}$$
$$\times\cos\left(\frac{2m_e}{\hbar^2}\left(E_{n,l}-V(R)-\frac{(l+\frac{1}{2})^2\hbar^2}{2m_eR^2}\right)\right)^{\frac{1}{2}}r_1, \quad (6.16)$$

where

$$R=\tfrac{1}{2}(r+r'),$$
$$r_1=r-r'. \quad (6.17)$$

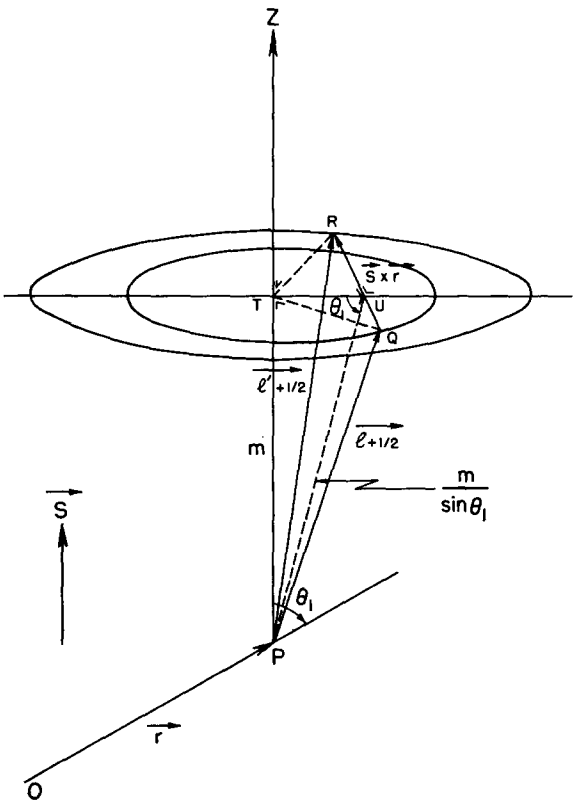

FIG. 3. Final angular momentum of the struck electron ($\langle l'+\frac{1}{2}\rangle_{av}$) (in units $\hbar$) represented as the vector sum of the original angular momentum ($\langle l+\frac{1}{2}\rangle_{av}$) and the transfer of angular momentum $\mathbf{s}\times\mathbf{r}$. Here $\mathbf{s}$ is the transfer of linear angular momentum. The Z axis has been chosen parallel to this vector. The angular momentum m about this axis is unchanged by the impact. The dihedral angle θ, between the two planes TRQ and PRQ, is important in analyzing the impact. It is identical with the angle between the lines TU and PU because TU and PU are perpendicular to QR. θ_1 is the angle for which the phase of the product of $P_{l,m}(\theta)$, $P_{l',m}(\theta)$ and $e^{isr\cos\theta}$ becomes stationary. For θ_1 the following equation holds:

$$[(l'+\tfrac{1}{2})^2-(m^2/\sin^2\theta_1)]^{\frac{1}{2}}+[(l+\tfrac{1}{2})^2-(m^2/\sin^2\theta_1)]^{\frac{1}{2}}=sr\sin\theta_1,$$

or solving with respect to θ_1,

$$\sin\theta_1=\frac{1}{sr}\{(l+\tfrac{1}{2})^2+(l'+\tfrac{1}{2})^2\pm 2((l+\tfrac{1}{2})^2(l'+\tfrac{1}{2})^2-m^2s^2r^2)^{\frac{1}{2}}\}^{\frac{1}{2}}.$$

[20] S. O. Hart, A. B. thesis, Princeton, 1942. The author could not find Hart's thesis. He knew Hart's work from Wheeler and Fireman's note.

We insert this relation into Eq. (6.10) and get the following result.

$$\int\int dVdV' \Big| \sum_{\substack{\text{occupied}\\ \text{states}}} \phi_i^*(\mathbf{r})\phi_i(\mathbf{r}') \Big|^2 e^{is(\mathbf{r}-\mathbf{r}')}$$

$$= \sum_{\substack{n,l,m\\ n',l',m'}} \frac{m_e^2\hbar}{\pi^2 s} \cdot \omega_{,n}\omega_{,n'}(l+\tfrac{1}{2})(l'+\tfrac{1}{2})$$

$$\times \int \frac{dR}{R} \frac{1}{p_R p_R' \cos\theta_1((l+\tfrac{1}{2})^2(l'+\tfrac{1}{2})^2 - m^2 s^2 R^2)^{\frac{1}{2}}}$$

$$\times \delta_{m,m'}\{\delta(p_R+p_R'-sh\cos\theta_1)+\delta(p_R-p_R'+sh\cos\theta_1)$$

$$+\delta(p_R-p_R'-sh\cos\theta_1)\}. \quad (6.18)$$

Using Eq. (6.14) we rewrite this expression into the following form:

$$\int dR \int_0^{\pi/2} d\theta \sum_{\substack{n,l,m\\ n',l',m'}} \left(\frac{m_e^2\hbar}{\pi^2}\right) \omega_{,n}\omega_{,n'}(l+\tfrac{1}{2})(l'+\tfrac{1}{2})$$

$$\times \frac{1}{p_R p_R' k_{l,m}(\theta) k_{l',m'}(\theta)} \delta_{m,m'}\delta(k_{l,m}(\theta) - k_{l',m'}(\theta)$$

$$+sR\sin\theta)\delta(p_R - p_R' + sh\cos\theta), \quad (6.19)$$

where

$$p_R = \left\{ 2m_e \left[E_{n,l} - V(R) - \frac{(l+\tfrac{1}{2})^2\hbar^2}{2m_e R^2} \right] \right\}^{\frac{1}{2}}.$$

Finally we change n, l, and m to p_R, p_θ, and p_ϕ. We get

$$\frac{1}{\pi^2\hbar^3} \int dR \int_0^{\pi/2} d\theta \int dp_R dp_\theta dp_\phi dp_R' dp_\theta' dp_\phi' \delta(p_\phi - p_\phi')$$

$$\times \delta(p_\theta - p_\theta' + shR\sin\theta)\delta(p_R - p_R' + sh\cos\theta)$$

$$= 2\int R^2 dR \int_0^{\pi} \sin\theta d\theta \int_0^{2\pi} d\phi \int \frac{dp_R dp_\theta dp_\phi dp_R' dp_\theta' dp_\phi'}{R^2\hbar^3 \sin\theta}$$

$$\times \delta(p_\phi - p_\phi')\delta(p_\theta - p_\theta' + s\hbar R\sin\theta)$$

$$\times \delta(p_R - p_R' + s\hbar\cos\theta). \quad (6.20)$$

We can easily see that the integration

$$\int dp_R dp_\theta dp_\phi dp_R' dp_\theta' dp_\phi' \delta(p_\phi - p_\phi')$$

$$\times \delta(p_\theta - p_\theta' + s\hbar R\sin\theta)\delta(p_R - p_R' + s\hbar\cos\theta) \quad (6.21)$$

is the area of intersection or "overlap" in (p_R, p_θ, p_ϕ) space when the Fermi sphere in momentum space at the spatial distance R from nucleus is transferred by $s\hbar$ in z direction. And since

$$dp_R dp_\theta dp_\phi / (R^2\sin\theta) = dp_x dp_y dp_z, \quad (6.22)$$

we get Heisenberg's result

$$2\int dV \frac{dp_x dp_y dp_z dp_x' dp_y' dp_z'}{h^3} \delta(\mathbf{p} - \mathbf{p}' + \mathbf{s}\hbar). \quad (6.23)$$

$$p, p' \leqslant p_F(R)$$

Thus we proved that the correspondence principle argument gives a right answer. We expect further that this same method also applies to hydrodynamics.

Predictions of the Hydrodynamical Formula Using the JWKB Approximation and the Method of Stationary Phase

In hydrodynamics the Compton scattering is, in the approximation of direct scattering, given by

$$\left(\frac{d\sigma}{d\Omega}\right)_{k,l,m} = (\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2 \left(\frac{e^2}{m_e c^2}\right)^2 \frac{\omega'(\omega-\omega')}{\omega} \frac{\hbar}{2m_e}$$

$$\times \frac{1}{\int U_{k,l,m}^* N_{k,l,m} d\tau} \left| \int N_{k,l,m}^* e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} d\tau \right|^2. \quad (6.24)$$

This differential cross section corresponds to the excitation of the hydrodynamical vibration of proper mode, k, l, and m.

In contrast with coherent scattering, the hydrodynamical vibration of angular momentum $l=0$ gives a nonzero contribution to the incoherent scattering. For a fixed momentum transfer $\hbar(\mathbf{k}'-\mathbf{k})$, we can show from this formula using the completeness theorem, that the total intensity of the Compton scattering is given by

$$\sum_i \left(\frac{d\sigma}{d\Omega}\right)_{0\to j} = Z\left(\frac{e^2}{m_e c^2}\right)^2 (\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2$$

$$\times \frac{\hbar^2(\mathbf{k}-\mathbf{k}')^2}{2m_e} \left(\frac{1}{\langle\hbar\omega_j\rangle} - \frac{1}{\hbar\omega_0} \right), \quad (6.25)$$

where we assumed that only modes of hydrodynamical vibration in the neighborhood of a most probable scattered frequency contribute strongly to the Compton scattering so that we could extract the average value $\langle\hbar\omega_j\rangle$ of the excitation energy from the summation over j. For a frequency sufficiently high to neglect $(1/\hbar\omega_0)$ term, we have

$$\sum_i \left(\frac{d\sigma}{d\Omega}\right)_{0\to j} = Z\left(\frac{e^2}{m_e c^2}\right)^2 (\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2$$

$$\times \frac{\hbar^2(\mathbf{k}-\mathbf{k}')^2}{2m_e} \Big/ \langle\hbar\omega_j\rangle. \quad (6.26)$$

$\langle\hbar\omega_j\rangle$ depends upon angle of scattering and incoming frequency. To make this situation clear, we use the

JWKB approximation for the hydrodynamical vibration amplitudes N_j and U_j.

Let us compute the JWKB approximation formula for $N_{k,l,m}$ and $U_{k,l,m}$, which satisfy Eqs. (2.15) and (2.16) (neglecting Coulomb interaction, for simplicity),

$$\nabla(n_0\nabla U_{k,l,m})+\omega'^2 N_{k,l,m}=0, \tag{6.27}$$

where

$$U_{k,l,m}=\frac{1}{m_e n_0}\frac{dp}{dn_0}N_{k,l,m}.$$

Solving this equation in the JWKB approximation, we get

$$U_{k,l,m}(r,\theta,\phi)=\frac{\text{const}}{rn_0^{\frac{1}{2}}}\frac{\cos\int_{r_{T.P.}}^{r}\left(\frac{m_e\omega'^2}{dp/dn_0}+\left(\frac{n_0'}{2n_0}\right)^2-\frac{n_0''}{2n_0}-\frac{n_0'}{rn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}dr}{\left(\frac{m_e\omega'^2}{dp/dn_0}+\left(\frac{n_0'}{2n_0}\right)^2-\frac{n_0''}{2n_0}-\frac{n_0'}{rn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{4}}}Y_{l,m}(\theta,\phi) \tag{6.28}$$

and

$$N_{k,l,m}=\frac{m_e n_0}{dp/dn_0}U_{k,l,m}. \tag{6.29}$$

Now

$$\frac{n_0'}{n_0}\sim\frac{1}{(\text{scale of potential})}\equiv\frac{1}{L} \tag{6.30}$$

and

$$\frac{m_e\omega'^2}{dp/dn_0}\sim\frac{1}{\lambda^2}, \tag{6.31}$$

where λ is the wavelength of sound wave. Therefore, the condition $L\gg\lambda$ justifies neglecting $(n_0'/n_0)^2$, n_0''/n_0, and n_0'/rn_0. But $L\gg\lambda$ is just the condition for the validity of the JWKB approximation. Therefore, under the conditions where one can legitimately use the JWKB approximation at all, one can neglect terms in $(n_0'/n_0)^2$, n_0'/rn_0, and n_0''/n_0.

Let us consider the normalization integral,

$$\int U_{k,l,m}{}^*N_{k,l,m}d\tau.$$

We put const$=1$ in Eq. (6.28) because, after all, this constant is cancelled by itself. Then we have

$$\int U_{k,l,m}{}^*N_{k,l,m}d\tau=\int_{r_{T.P.(1)}}^{r_{T.P.(2)}}dr\frac{m_e}{dp/dn_0}\frac{\cos^2\int_{r_{T.P.(1)}}^{r}\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}dr}{\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{4}}}$$

$$\approx\frac{1}{2\omega'}\int_{r_{T.P.(1)}}^{r_{T.P.(2)}}dr\frac{\hbar\omega'}{\frac{1}{m_e}\frac{dp}{dn_0}}\Bigg/\left[\left[\frac{\hbar\omega'}{\left(\frac{1}{m_e}\frac{dp}{dn_0}\right)^{\frac{1}{2}}}\right]^2-\frac{(l+\frac{1}{2})^2}{r^2}\right]^{\frac{1}{2}}. \tag{6.32}$$

We give a physical interpretation to this equation in correspondence with classical mechanics. To a phonon of energy $\hbar\omega'$ we ascribe momentum

$$\hbar\omega'\Big/\left(\frac{1}{m_e}\frac{dp}{dn_0}\right)^{\frac{1}{2}}$$

and inertial mass

$$\hbar\omega'\Big/\left(\frac{1}{m_e}\frac{dp}{dn_0}\right).$$

Then we can rewrite Eq. (6.32) into the following form,

$$\frac{1}{2\omega'}\int_0^{\frac{1}{2}T_{\text{hydro}}}dt=\frac{T_{\text{hydro}}}{4\omega'}, \tag{6.33}$$

where T_{hydro} is the period of orbital motion of a phonon in the region of binding.

After having thus solved the hydrodynamical wave equation, we have to apply the boundary conditions to find the proper modes of oscillation. We apply the semiclassical method of finding eigenvalues to sound wave. We get

$$\int_{r_{T.P.(1)}}^{r_{T.P.(2)}}dr\left(\left(\hbar\omega'\Big/\left(\frac{1}{m_e}\frac{dp}{dn_0}\right)^{\frac{1}{2}}\right)^2-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}dr$$

$$=(k+\tfrac{1}{2})\pi. \tag{6.34}$$

Taking the derivative of both sides with respect to k we get

$$\int_{r_{T.P.(1)}}^{r_{T.P.(2)}}dr\frac{\frac{m_e}{dp/dn_0}\omega'\left(\frac{\partial\omega'}{\partial k}\right)_l}{\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}}=\pi, \tag{6.35}$$

or

$$\left(\frac{\partial\omega'}{\partial k}\right)_l\int_0^{\frac{1}{2}T_{\text{hydro}}}dt=\pi, \tag{6.36}$$

hence

$$T_{\text{hydro}}=2\pi\Big/\left(\frac{\partial\omega'}{\partial k}\right)_l. \tag{6.37}$$

Though, generally, ω' depends on k and l, for a large value of $r_{\text{T.P.}}(2)$, ω' depends only on k. Eventually we extend the arbitrary surface to infinity, so that hencefourth we assume that ω' depends only on k. Thus we get for the normalization integral

$$\int U_{k,l,m}{}^{*}N_{k,l,m}d\tau=\frac{\pi}{2\omega'\omega'_k}, \tag{6.38}$$

where

$$\omega'_k=(\partial\omega'/\partial k)_l. \tag{6.39}$$

Inserting Eqs. (6.28), (6.29), and (6.38) into Eq. (6.24) we get

$$\left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l=(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2\left(\frac{e^2}{m_ec^2}\right)^2\frac{\omega'^2(\omega-\omega')}{\omega}8m_e\hbar(l+\tfrac{1}{2})$$
$$\times\left|\int dr r\frac{n_0^{\frac{1}{2}}}{dp/dn_0}\frac{\cos\left\{\int_{r_{\text{T.P.}}}^{r}dr\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}\right\}}{\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{4}}}j_l(|\mathbf{k}-\mathbf{k}'|r)\right|^2, \tag{6.40}$$

where we have defined $(d^2\Sigma/d\Omega d\omega')_l$ such that

$$\left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l=\left(\frac{d\sigma}{d\Omega}\right)_l\left(\frac{\partial k}{\partial\omega'}\right). \tag{6.41}$$

To evaluate the integration in Eq. (6.40) again, we use the JWKB approximation for $j_l(|\mathbf{k}-\mathbf{k}'|r)$:

$$j_l(|\mathbf{k}-\mathbf{k}'|r)=\frac{1}{|\mathbf{k}-\mathbf{k}'|r}\frac{1}{((\mathbf{k}-\mathbf{k}')^2-(l+\frac{1}{2})^2/r^2)^{\frac{1}{4}}}$$
$$\times\cos\left\{\int_{r_{\text{T.P.}}}^{r}\left((\mathbf{k}-\mathbf{k}')^2-\frac{(l+\frac{1}{2})^2}{r^2}\right)^{\frac{1}{2}}dr-\frac{\pi}{4}\right\}. \tag{6.42}$$

We make another approximation—correspondence principle argument—for the product of two Bessel functions $j_l(|\mathbf{k}-\mathbf{k}'|r)$ and $j_l(|\mathbf{k}-\mathbf{k}'|r')$ (dropping rapidly varying parts!):

$$j_l(|\mathbf{k}-\mathbf{k}'|r)\cdot j_l(|\mathbf{k}-\mathbf{k}'|r')$$
$$\approx\frac{1}{2|\mathbf{k}-\mathbf{k}'|((\mathbf{k}-\mathbf{k}')^2-(l+\frac{1}{2})^2/R^2)^{\frac{1}{2}}R^2}$$
$$\times\cos(((\mathbf{k}-\mathbf{k}')^2-(l+\tfrac{1}{2})^2/R^2)^{\frac{1}{2}}r_1). \tag{6.43}$$

Further, we make this approximation for the product of two radial parts of $N_{k,l,m}(\mathbf{r})$ and $U_{k,l,m}(\mathbf{r}')$. Then we get

$$\left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l=(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2\left(\frac{e^2}{m_ec^2}\right)^2\frac{\omega'^2(\omega-\omega')}{\omega|\mathbf{k}-\mathbf{k}'|}(2m_e\hbar)(l+\tfrac{1}{2})\int dR\frac{n_0(R)}{\left(\frac{dp}{dn_0}\right)^2}\frac{1}{\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}\left((\mathbf{k}-\mathbf{k}')^2-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}}$$
$$\times\int dr_1\cos\left(\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}r_1\right)\cos\left(\left((\mathbf{k}-\mathbf{k}')^2-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}r_1\right). \tag{6.44}$$

We perform the integration over r_1 from $-\infty$ to $+\infty$. We have

$$\left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l=(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2\left(\frac{e^2}{m_ec^2}\right)^2\frac{\omega'^2(\omega-\omega')}{\omega|\mathbf{k}-\mathbf{k}'|}(2\pi m_e\hbar)(l+\tfrac{1}{2})\int dR\frac{n_0(R)}{\left(\frac{dp}{dn_0}\right)^2\left((\mathbf{k}-\mathbf{k}')^2-\frac{(l+\frac{1}{2})^2}{R^2}\right)}$$
$$\times\delta\left(\left(\frac{m_e\omega'^2}{dp/dn_0}-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}-\left((\mathbf{k}-\mathbf{k}')^2-\frac{(l+\frac{1}{2})^2}{R^2}\right)^{\frac{1}{2}}\right). \tag{6.45}$$

TABLE I.

	Electrons in atom	Hydrodynamical disturbances
Global analysis	Wave equations; boundary conditions, eigenvalues and eigenfunctions ψ_j	Hydrodynamical equation of motion, equation of continuity, Poisson's equation, equation of state, boundary conditions, eigenvalues, density deviation N_j, and velocity potential U_j for hydrodynamical vibration of proper mode j.
Local analysis. Approximation the better the higher one is in the spectrum of proper modes	Electrons moving in any small region of atom characterized by *both* position *and* momentum. Two electrons per cell in phase space.	Phonons moving in any small region of atom characterized by *both* position *and* momentum. Spectrum becomes continuous when artificial boundary surface is brought to infinity.

From this equation we see that only when

$$\hbar\omega' \Big/ \left(\frac{1}{m_e}\frac{dp}{dn_0}\right)^{\frac{1}{2}} = |\mathbf{k}-\mathbf{k}'|$$

does the δ function differ from zero. In other words, at the distance where the momentum

$$\hbar\omega' \Big/ \left(\frac{1}{m_e}\frac{dp}{dn_0}\right)^{\frac{1}{2}}$$

of the sound wave takes the value of the momentum transfer from the photon, there the phonon is produced. Just as electron states (extended functions of position) are analyzed on a *local* basis in the statistical atom model (Table I), so characteristic modes of hydrodynamical vibration—which are solutions of a *global* eigenvalue problem—can be replaced in a certain approximation by the *local* concept of phonons.

Integrating $(d^2\Sigma/d\Omega d\omega')_l$ over the modified circular frequency ω' and over the angular momentum of hydrodynamical vibration l, we get the total intensity.

$$\frac{d\sigma_{\text{Compton}}}{d\Omega} = \int dl \int d\omega' \left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l = (\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2 \left(\frac{e^2}{m_e c^2}\right)^2 \times \frac{2\pi\hbar}{m_e^{\frac{1}{2}}} |\mathbf{k}-\mathbf{k}'| \int R^2 dR \frac{n_0(R)}{(dp/dn_0)^{\frac{1}{2}}}. \quad (6.46)$$

To make this integration, we perform the transformation of variable

$$R = \mu x = \frac{1}{4}\left(\frac{9\pi^2}{2Z}\right)^{\frac{1}{3}} a_0 x,$$

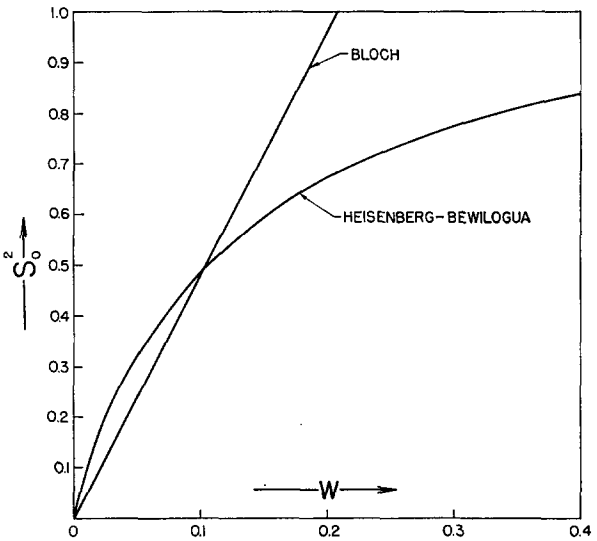

FIG. 4. Intensity factor S_0^2 for Compton scattering defined as ratio between differential cross section and the "ideal" value $(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2(e^2/m_e c^2)^2 Z$. The curved line shows Heisenberg-Bewilogua's result

$$S_0^2 = 1 - \int_0^{\xi_0} \left(\left(\frac{\phi(\xi)}{\xi}\right)^{\frac{1}{2}} - W^2\right)^2 \left(\left(\frac{\phi(\xi)}{\xi}\right)^{\frac{1}{2}} + \tfrac{1}{2}W\right) \xi^2 d\xi,$$

where

$$W = \frac{0.176\times 10^{-8}}{Z^{\frac{1}{3}}} \frac{4}{\lambda} \sin\frac{\theta}{2}$$

and

$$(\phi(\xi_0)/\xi_0)^{\frac{1}{2}} = W.$$

The straight line shows the correspondence theoretical expression for Bloch's hydrodynamical formula. This expression agrees with Heisenberg's formula except for a numerical factor $3^{\frac{1}{2}}/2$, provided that we take the term linearly dependent on in Heisenberg's formula.

and

$$n_0(R) = \frac{Z}{4\pi\mu^3}\left(\frac{\phi}{x}\right)^{\frac{3}{2}}. \quad (6.47)$$

Inserting this into Eq. (6.46) we have

$$\int dl \int d\omega' \left(\frac{d^2\Sigma}{d\Omega d\omega'}\right)_l = (\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2 \left(\frac{e^2}{m_e c^2}\right)^2 Z \frac{0.176\times 10^{-8}}{Z^{\frac{1}{3}}} \times \frac{4}{\lambda}\sin\frac{\theta}{2}\left(1.73\int_0^\infty \phi(x)dx\right). \quad (6.48)$$

It is interesting to compare this new result with Heisenberg-Bewilogua's result. We take the computation of Miranda[21] for $-\phi'(0)=1.5882$ for a neutral atom. From Fig. 4 we see that for low values of momentum transfer both results give nearly, at least qualitatively, the same answer. In fact, if we take the term linearly dependent on momentum transfer in Heisenberg's expression

$$(\mathbf{e}_{\text{in}}\cdot\mathbf{e}_{\text{out}})^2(e^2/m_e c^2)^2 Z S_0^2$$

[21] C. Miranda, Mem. Acc. Italia **5**, 283 (1934).

and

$$S_0^2=1-\left(\left(\frac{\phi(\xi)}{\xi}\right)^{\frac{1}{2}}-W\right)^2 \times\left(\left(\frac{\phi(\xi)}{\xi}\right)^{\frac{1}{2}}+\tfrac{1}{2}W\right)\xi^2 d\xi, \quad (6.49)$$

where

$$W=\frac{0.176\times10^{-8}}{Z^{\frac{1}{3}}}\frac{4}{\lambda}\sin\frac{\theta}{2}$$

and

$$\left[\frac{\phi(\xi_0)}{\xi_0}\right]^{\frac{1}{2}}=W,$$

we find that hydrodynamics and wave mechanics give the same answer within the Bloch factor of $3^{\frac{1}{2}}/2$. This fact seems to indicate that our method of correspondence principle argument is right.

VII. CONCLUSION

The principal results of this paper are Eq. (2.42) for the absorption cross section, Eq. (3.4) for the coherent scattering cross section, and Eq. (4.14) for the incoherent scattering cross section. Some understanding is won about the connection between the hydrodynamical method and the wave mechanical method. In other words, we showed that, for the total intensity of the Compton scattering, the hydrodynamical treatment gives the same results as does the linear term in the momentum transfer in Heisenberg's expression except for Bloch's factor of $3^{\frac{1}{2}}/2$.

Equation (6.24) for the Compton scattering, in the approximation of direct scattering, can be used for comparison with experiment if we know the charge density deviation for single mode k, l, and m of hydrodynamical vibration. Equation (2.42) for the absorption cross section can also be used for the same purpose with almost the same knowledge. Others, Eqs. (3.4) and (4.14) require—before application—electronic machine calculations of all proper modes of oscillation of the gas model of the atom, analogous to those which Wheeler and Fireman made for $l=1$.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Professor John A. Wheeler for his suggesting this problem, his guidance, his continuous encouragement, and his invaluable assistance in clarifying the presentation of this paper.

Symmetric Expansion of One- and Two-Center Coulomb Potentials*

PETER R. FONTANA
Laboratory for Molecular Structure and Spectra, Department of Physics, The University of Chicago, Chicago, Illinois
(Received March 16, 1961)

A one-center expansion of the electrostatic interaction energy of a discrete charge distribution is developed by making use of the algebra of irreducible tensors. The result is completely symmetric in the coordinates of the particles, and the relative magnitude of the vectors need not be specified. It is shown that a suitable interaction representation provides useful formulas for electrostatic and quantum mechanical applications. In addition, some transformation equations make it possible to refer any arbitrary number of vectors to a second origin, thus yielding general two-center expansions for overlapping charge distributions.

I. INTRODUCTION

IN many electrostatic and quantum mechanical problems, an expansion of the interaction energy

$$\sum_{i,j}{}' \frac{e_i e_j}{|\mathbf{r}_i-\mathbf{r}_j|} \tag{1}$$

between n particles in terms of the individual particle coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$ is needed. In atomic problems, these vectors are referred to a single origin; whereas in molecular applications, the origin of the $\mathbf{r}_i$ vectors may be different from that of the $\mathbf{r}_j$'s.

In the usual one-center expansion of (1) using the generating function of Legendre polynomials one has to distinguish between the magnitudes $|\mathbf{r}_i|$ and $|\mathbf{r}_j|$ of the two vectors, and one obtains two expansions; one in which $r_i>r_j$ and another one which satisfies the condition $r_j>r_i$. Quite often, and especially when (1) appears in differential equations, this limitation of the expansion is a very cumbersome one. A similar situation exists in the bipolar expansion of the Coulomb potential between two charge distributions. The range of validity of the formulas depends on the relative magnitudes of r_i, r_j, and R, the separation of the two centers.[1-3]

The method outlined here provides formulas for the one-center expansion of (1) which are symmetric in the $\mathbf{r}_i$ and $\mathbf{r}_j$ coordinates and which do not require knowledge of the relative magnitudes of the two vectors. Some transformation equations will make it possible to refer any number of the $\mathbf{r}_i$ to one, two, and in principle to any number of centers, and one obtains general expressions for the interaction energy between overlapping charge distributions. This paper is restricted to a detailed discussion of the one- and two-center cases.

II. ONE-CENTER EXPANSION OF THE ELECTROSTATIC INTERACTION ENERGY

The interaction energy of (1) is a scalar invariant and thus it is possible to write it as a suitable contraction of irreducible tensors of the same rank.[4] To see how these tensors can be generated, let us consider the first term of (1) which can be written as

$$\frac{e_1e_2}{r_{12}}=e_1e_2\left[r_1^2+r_2^2-\frac{8\pi}{3}\sum_\alpha \mathcal{Y}_1^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2)\right]^{-\frac{1}{2}}, \tag{2}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ refer to charges e_1 and e_2, respectively, and the solid spherical harmonic $\mathcal{Y}_a^\alpha(\mathbf{r})$ is defined as

$$\mathcal{Y}_a^\alpha(\mathbf{r})=r^a Y_a^\alpha(\theta,\varphi),\ \mathcal{Y}_a^{\alpha*}(\mathbf{r})=(-)^\alpha \mathcal{Y}_a^{-\alpha}(\mathbf{r}). \tag{3}$$

Setting $r=[r_1^2+r_2^2]^{\frac{1}{2}}$, we can rewrite (2) as

$$\frac{e_1e_2}{r}\left[\left(1-\frac{8\pi}{3r^2}\sum_\alpha \mathcal{Y}_1^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2)\right)^{-\frac{1}{2}}\right]. \tag{4}$$

The term in the square brackets of (4) can be expanded since

$$\frac{8\pi}{3r^2}\sum_\alpha \mathcal{Y}_1^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2)\leqslant 1$$

and we get[5]

$$\frac{e_1e_2}{r_{12}}=\frac{e_1e_2}{r}\sum_n \frac{(2n-1)!!}{(2n)!!}x^n, \tag{5}$$

where we have set

$$x=\frac{8\pi}{3r^2}\sum_\alpha \mathcal{Y}_1^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2). \tag{6}$$

Writing

$$x^n=(\{[x\cdot x]\cdot x\}\cdot x)\cdots, \tag{7}$$

then x^n can be reduced by $n-1$ contractions with the use of the algebra of irreducible tensors. The first

* Work supported by a grant from the National Science Foundation.

[1] B. C. Carlson and G. S. Rushbrooke, Proc. Cambridge Phil. Soc. **46**, 626 (1950).

[2] R. J. Buehler and J. O. Hirschfelder, Phys. Rev. **83**, 628 (1951); **85**, 149 (1952).

[3] M. E. Rose, J. Math. and Phys. **37**, 215 (1958).

[4] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957), Chap. V.

[5] $(2n-1)!!=1\cdot3\cdot5\cdots(2n-1)$; $(2n)!!=2\cdot4\cdot6\cdots(2n)$.

contraction in (7) is obtained in the following way:

$$x\cdot x=\left(\frac{8\pi}{3r^2}\right)^2\sum_{\alpha,\beta}(-)^{\alpha+\beta}\mathcal{Y}_1^{-\alpha}(\mathbf{r}_1)\mathcal{Y}_1^{-\beta}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2)\mathcal{Y}_1^{\beta}(\mathbf{r}_2)$$

$$=\left(\frac{8\pi}{3r^2}\right)^2\sum_{\alpha,\beta}\sum_{p,q}\frac{(-)^{\alpha+\beta}9}{4\pi[(2p+1)(2q+1)]^{\frac{1}{2}}}$$

$$\times C(11p;-\alpha,-\beta)C(11q;\alpha,\beta)C(11p;00)$$

$$\times C(11q;00)r_1^{2-p}r_2^{2-q}\mathcal{Y}_p^{-\alpha-\beta}(\mathbf{r}_1)\mathcal{Y}_q^{\alpha+\beta}(\mathbf{r}_2). \quad (8)$$

In the second line of (8) we have used the coupling rule for spherical harmonics which can be obtained from the Clebsch-Gordan series. The parity coefficients $C(11p;00)$ and $C(11q;00)$ are zero unless p, $q=2$, 1, or 0, and the sums $1+1+p$ and $1+1+q$ are even. Thus p and q can assume the values 0 and 2 only.

The substitution $\gamma=\alpha+\beta$ transforms (8) into

$$x\cdot x=\left(\frac{8\pi}{3r^2}\right)^2\sum_{p,q}\sum_{\gamma}\frac{(-)^{\gamma}9}{4\pi[(2p+1)(2q+1)]^{\frac{1}{2}}}$$

$$\times\left[\sum_{\alpha}C(11p;\alpha,\gamma-\alpha)C(11q;\alpha,\gamma-\alpha)\right]$$

$$\times C(11p;00)C(11q;00)$$

$$\times r_1^{2-p}r_2^{2-q}\mathcal{Y}_p^{-\gamma}(\mathbf{r}_1)\mathcal{Y}_q^{\gamma}(\mathbf{r}_2). \quad (9)$$

The orthogonality condition of the C coefficients, however, requires that

$$\sum_{\alpha}C(11p;\alpha,\gamma-\alpha)C(11q;\alpha,\gamma-\alpha)=\delta_{p,q} \quad (10)$$

and $x\cdot x$ reduces to

$$\left(\frac{8\pi}{3r^2}\right)^2\sum_{p,\gamma}\frac{9}{4\pi(2p+1)}$$

$$\times C(11p;00)^2r_1^{2-p}\mathcal{Y}_p^{\gamma*}(\mathbf{r}_1)r_2^{2-p}\mathcal{Y}_p^{\gamma}(\mathbf{r}_2). \quad (11)$$

Now,

$$C(11p;00)^2=\begin{cases}\frac{2}{3}; & p=2\\ \frac{1}{3}; & p=0,\end{cases} \quad (12)$$

and we get for the first contraction in (7)

$$x\cdot x=\frac{4\pi}{r^4}\left[\frac{4}{3}r_1^2r_2^2\mathcal{Y}_0^0(\mathbf{r}_1)\mathcal{Y}_0^0(\mathbf{r}_2)+\frac{8}{15}\sum_{\gamma}\mathcal{Y}_2^{\gamma*}(\mathbf{r}_1)\mathcal{Y}_2^{\gamma}(\mathbf{r}_2)\right]. \quad (13)$$

Equation (13) expresses $x\cdot x$ in terms of multiples of order 2 and 0, and one can show in a straightforward way that x^n generates multipoles of order n, $n-2$, $n-4\cdots1$, or 0. In particular,

$$x^3=\frac{4\pi}{r^6}\left[\frac{8}{5}r_1^2r_2^2\sum_{\alpha}\mathcal{Y}_1^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_1^{\alpha}(\mathbf{r}_2)+\frac{16}{35}\sum_{\beta}\mathcal{Y}_3^{\beta*}(\mathbf{r}_1)\mathcal{Y}_3^{\beta}(\mathbf{r}_2)\right] \quad (14)$$

$$x^4=\frac{4\pi}{r^8}\left[\frac{16}{5}r_1^4r_2^4\mathcal{Y}_0^0(\mathbf{r}_1)\mathcal{Y}_0^0(\mathbf{r}_2)+\frac{64}{35}r_1^2r_2^2\sum_{\alpha}\mathcal{Y}_2^{\alpha*}(\mathbf{r}_1)\mathcal{Y}_2^{\alpha}(\mathbf{r}_2)+\frac{128}{315}\sum_{\beta}\mathcal{Y}_4^{\beta*}(\mathbf{r}_1)\mathcal{Y}_4^{\beta}(\mathbf{r}_2)\right], \quad (15)$$

and by induction

$$x^n=\sum_{\lambda}^{n}\frac{4\pi 2^n n!r_1^{n-\lambda}r_2^{n-\lambda}}{r^{2n}(n+\lambda+1)!!(n-\lambda)!!}\sum_{\mu}\mathcal{Y}_\lambda^{\mu*}(\mathbf{r}_1)\mathcal{Y}_\lambda^{\mu}(\mathbf{r}_2)$$

$$\lambda=0,\,2,\,\cdots n-2,\,n \quad \text{for } n \text{ even}$$

$$\lambda=1,\,3,\,\cdots n-2,\,n \quad \text{for } n \text{ odd}. \quad (16)$$

Substitution of (16) into (5) yields for the electrostatic interaction energy between the two charges e_1 and e_2

$$\frac{e_1e_2}{r_{12}}=4\pi e_1e_2\sum_{\lambda}\left[\sum_{n}\frac{(2n-1)!!r_1^{n-\lambda}r_2^{n-\lambda}}{(n+\lambda+1)!!(n-\lambda)!!r^{2n+1}}\right]\times\sum_{\mu}\mathcal{Y}_\lambda^{\mu*}(\mathbf{r}_1)\mathcal{Y}_\lambda^{\mu}(\mathbf{r}_2),\quad n=\lambda,\lambda+2,\lambda+4,\cdots. \quad (17)$$

This is a symmetric one-center expansion and has the advantage over the Laplace expansion that one does not have to specify the relative magnitudes of $\mathbf{r}_1$ and $\mathbf{r}_2$.

A comparison of (17) with the Laplace expansion

$$\frac{e_1e_2}{r_{12}}=4\pi e_1e_2\sum_{\lambda,\mu}\frac{1}{r_>^{2\lambda+1}}\mathcal{Y}_\lambda^{\mu*}(\mathbf{r}_<)\mathcal{Y}_\lambda^{\mu}(\mathbf{r}_>) \quad (18)$$

where $\mathbf{r}_<$ is the lesser and $\mathbf{r}_>$ the greater of the two vectors, yields the following useful results:

(1) After setting $r_<=r_>$, $r_1=r_2$ and comparing the monopole terms of (17) and (18), we get the expansion

$$\sqrt{2}=\sum_{\substack{n=0\\ \text{even}}}\frac{(2n-1)!!}{(2n)!!(n+1)}. \quad (19)$$

(2) A term by term comparison of (17) and (18) yields the result

$$\frac{r_<^{\lambda}}{r_>^{\lambda+1}}=(2\lambda+1)\sum_{n}\frac{r_1^nr_2^n(2n-1)!!}{(n+\lambda+1)!!(n-\lambda)!!r^{2n+1}},\quad n=\lambda,\lambda+2,\lambda+4,\cdots, \quad (20)$$

where $r_{<,>}=r_{1,2}$ and $r=[r_1^2+r_2^2]^{\frac{1}{2}}$.

Equation (20) can be proved directly by setting $r_1<r_2$, $y=r_</r_>$ and then expanding r^{-2n-1} in a power series.

$$y^\lambda=\sum_n \frac{(2\lambda+1)(2n-1)!!y^n}{(n+\lambda+1)!!(n-\lambda)!!}\left\{\sum_{m=0}^{\infty}(-)^m\frac{(n+m-\tfrac{1}{2})!}{m!(n-\tfrac{1}{2})!}y^{2m}\right\}$$
$$=y^\lambda+[-(\lambda+\tfrac{1}{2})+\tfrac{1}{2}(2\lambda+1)]y^{\lambda+2}+[\tfrac{1}{2}(\lambda+\tfrac{1}{2})(\lambda+\tfrac{3}{2})$$
$$-(\lambda+\tfrac{1}{2})(\lambda+\tfrac{5}{2})+\tfrac{1}{2}(\lambda+\tfrac{1}{2})(\lambda+\tfrac{7}{2})]y^{\lambda+4}+\cdots. \quad (21)$$

It is easily shown that all the coefficients of the $y^{\lambda+2k}$ terms with $k=1, 2, \cdots$ vanish, and thus (20) is verified.

In some cases, it is convenient to rewrite (17) by making the following substitutions: $r_1=r\sin\chi$; $r_2=r\cos\chi(0\leqslant\chi\leqslant\frac{1}{2}\pi; r=[r_1^2+r_2^2]^{\frac{1}{2}})$.

We then have

$$\frac{e_1e_2}{r_{12}}=\frac{4\pi e_1e_2}{r}\sum_{\lambda,\mu}X_\lambda(\chi)Y_\lambda^{\mu*}(\theta_1,\varphi_1)Y_\lambda^{\mu}(\theta_2,\varphi_2), \quad (22)$$

where

$$X_\lambda(\chi)=\sum_n\frac{\sin^n\chi\cos^n\chi(2n-1)!!}{(n+\lambda+1)!!(n-\lambda)!!},$$
$$n=\lambda, \lambda+2, \lambda+4, \cdots. \quad (23)$$

This "interaction representation" of the one-center expansion of the electrostatic interaction energy has the advantage over (17) that all the summations extend over angle variables.

In the r, χ representation we also have from (5)

$$\frac{e_1e_2}{r_{12}}=\frac{e_1e_2}{r[1-2\sin\chi\cos\chi\cos\theta_{12}]^{\frac{1}{2}}}$$
$$=\sum_n\frac{e_1e_2 2^n(2n-1)!!}{r(2n)!!}(\sin\chi\cos\chi)^n(\cos\theta_{12})^n, \quad (24)$$

where θ_{12} is the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. Comparing (24) with (22) yields

$$(\cos\theta_{12})^n=4\pi\sum_{\lambda\mu}\frac{n!}{(n+\lambda+1)!!(n-\lambda)!!}$$
$$\times Y_\lambda^{\mu*}(\theta_1,\varphi_1)Y_\lambda^{\mu}(\theta_2,\varphi_2),$$
$$\lambda=0, 2, \cdots n-2, n \quad \text{for } n \text{ even}$$
$$\lambda=1, 3, \cdots n-2, n \quad \text{for } n \text{ odd}, \quad (25)$$

and by applying the addition theorem of spherical harmonics to (25) we get the well known result[6]

$$(\cos\theta)^n=\sum_\lambda^n\frac{n!(2\lambda+1)}{(n+\lambda+1)!!(n-\lambda)!!}P_\lambda(\cos\theta),$$
$$\lambda=0, 2\cdots n-2, n \quad \text{for } n \text{ even}$$
$$\lambda=1, 3\cdots n-2, n \quad \text{for } n \text{ odd}. \quad (26)$$

These results make it possible to express $X_\lambda(\chi)$ of (23) in terms of Legendre polynomials by setting $\eta=\frac{1}{2}\pi-2\chi$. Use of (26) then transforms (23) into

$$X_\lambda(\eta)=\sum_{n,\nu}\frac{n!(2n-1)!!(2\nu+1)}{2^n(n+\lambda+1)!!(n+\nu+1)!!(n-\lambda)!!(n-\nu)!!}$$
$$\times P_\nu(\cos\eta),$$
$$n=\lambda, \lambda+2\cdots,$$
$$\nu=0 \quad \text{or} \quad 1, \cdots n-2, n, \quad (27)$$

which will be useful when (22) appears in integrals.

This method is of course also applicable to potentials which are of the form $(r_{ij})^{-n}$ where n is a positive integer. The general derivation is analogous to the one outlined here; the only change occurs in Eq. (5).

It may be worth noting that (17) is proportional to the Green's function for the Laplace equation.[7] If r_1, θ_1, φ_1 are the coordinates of the observer point, and r_2, θ_2, φ_2 the coordinates of the source point, then

$$G(\mathbf{r}_1|\mathbf{r}_2)=4\pi\sum_{\lambda\mu}\left[\sum_n\frac{(2n-1)!!r_1^n r_2^n}{(n+\lambda+1)!!(n-\lambda)!!r^{2n+1}}\right]$$
$$\times Y_\lambda^{\mu*}(\theta_1,\varphi_1)Y_\lambda^{\mu}(\theta_2,\varphi_2)$$
$$=G(\mathbf{r}_2|\mathbf{r}_1), \quad (28)$$

and the reciprocity theorem is immediately satisfied.

III. TWO-CENTER EXPANSION OF THE ELECTROSTATIC INTERACTION ENERGY

As a straightforward extension of the methods outlined in Sec. II, let us transform $\mathbf{r}_2$ to a second center such that $\mathbf{r}_2=\mathbf{R}+\boldsymbol{\varrho}_2$, where $\mathbf{R}$ connects the two centers and $\boldsymbol{\varrho}_2$ originates at the second center and points to charge 2 (Fig. 1). The corresponding transformation of the one-center expansion (17) can be

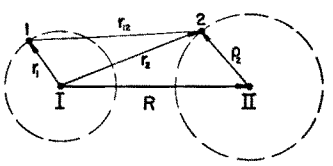

FIG. 1. Vector diagram for the transformation from the one-center to the two-center expansion.

[6] See for instance, P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), p. 1326.

[7] Reference 6, p. 1273.

accomplished in three steps:

(1)

$$r_2=\left[R^2+\rho_2^2+\frac{8\pi}{3}\sum_\alpha \mathcal{Y}_1{}^{\alpha*}(\mathbf{R})\mathcal{Y}_1{}^{\alpha}(\boldsymbol{\rho}_2)\right]^{\frac{1}{2}} \quad (29)$$

and

$$r_2{}^{2k}=\sum_{\xi,a}\frac{4\pi k!2^\xi R^{\xi-a}\rho_2{}^{\xi-a}}{(k-\xi)!(R^2+\rho_2{}^2)^{\xi-k}(\xi+a+1)!!(\xi-a)!!}\times\sum_\alpha \mathcal{Y}_a{}^{\alpha*}(\mathbf{R})\mathcal{Y}_a{}^{\alpha}(\boldsymbol{\rho}_2), \quad (30)$$

where we have set $2k=n-\lambda$.

(2)

$$r=\left[r_1{}^2+\rho_2{}^2+R^2+\frac{8\pi}{3}\sum_\alpha \mathcal{Y}_1{}^{\alpha*}(\mathbf{R})\mathcal{Y}_1{}^{\alpha}(\boldsymbol{\rho}_2)\right]^{\frac{1}{2}} \quad (31)$$

and

$$r^{-2n-1}=\sum_{\eta,b}(-)^\eta\frac{(2\eta+2n-1)!!R^{\eta-b}\rho_2{}^{\eta-b}}{\mathfrak{R}^{2\eta+2n+1}(2n-1)!!(\eta+b+1)!!(\eta-b)!!}\times\sum_\beta \mathcal{Y}_b{}^{\beta*}(\mathbf{R})\mathcal{Y}_b{}^{\beta}(\boldsymbol{\rho}_2), \quad (32)$$

where $\mathfrak{R}=[r_1{}^2+\rho_2{}^2+R^2]^{\frac{1}{2}}$.

(3) The solid spherical harmonics $\mathcal{Y}_\lambda{}^\mu(\mathbf{r}_2)$ can be transformed to the second center by applying the addition theorem of Rose.[3]

$$\mathcal{Y}_\lambda{}^\mu(\mathbf{r}_2)=\mathcal{Y}_\lambda{}^\mu(\mathbf{R}+\boldsymbol{\rho}_2)=[4\pi(2\lambda+1)!]^{\frac{1}{2}}\times\sum_{c=0}\sum_\gamma\frac{C(c,\lambda-c,\lambda;\gamma,\mu-\gamma)}{[(2c+1)!(2\lambda-2c+1)!]^{\frac{1}{2}}}\times\mathcal{Y}_c{}^\gamma(\mathbf{R})\mathcal{Y}_{\lambda-c}{}^{\mu-\gamma}(\boldsymbol{\rho}_2). \quad (33)$$

The range of γ is restricted by the C coefficient to $-c\leqslant\gamma\leqslant c$.

The transformation equations when combined with (17) provide the general two-center expansion of the electrostatic interaction energy. Setting either $\mathbf{r}_1$, $\boldsymbol{\rho}_2$, or $\mathbf{R}$ to zero yields a one-center expansion in $(\boldsymbol{\rho}_2,\mathbf{R})$, $(\mathbf{r}_1,\mathbf{R})$, or $(\mathbf{r}_1,\boldsymbol{\rho}_2)$, respectively.

The structure of the resulting two-center expansion is not nearly as compact as the form of the one-center expansion. The only advantage attained here is that the method does not require the knowledge of the relative magnitudes of $\mathbf{r}_1$, $\boldsymbol{\rho}_2$, and $\mathbf{R}$. The formulas can be considerably simplified in the region where $R>r_1$ and $R>\rho_2$ since here the shell formed by r_1 does not intersect the shell of ρ_2 (see dashed curves in Fig. 1). This process of transforming vectors to different centers could of course be continued by setting $\boldsymbol{\rho}_2=\mathbf{R}'+\boldsymbol{\rho}_2'$ in (29)–(33). The form of the resulting three-center expansion, however, while possible in principle is complicated in practice.

ACKNOWLEDGMENT

The author wishes to thank Professor R. S. Mulliken for making his stay at the University of Chicago possible.

APPENDIX

In second-order perturbation theory one often has to evaluate the matrix element

$$\langle 0|1/r_{ij}{}^2|0\rangle, \quad (A1)$$

where

$$|0\rangle=u_0u_1\cdots u_iu_j\cdots u_m \quad (A2)$$

and

$$u_i=N_i r_i{}^{n_i-1}e^{-\xi_i r_i}Y_{l_i}{}^{m_i}(\theta_i,\varphi_i). \quad (A3)$$

Using the results of Sec. II, the operator in (A1) can be written as

$$\frac{1}{r_{ij}{}^2}=\sum_{\lambda\mu}\left[\sum_\eta\frac{4\pi 2^\eta\eta!r_i{}^\eta r_j{}^\eta}{r^{2\eta+2}(\eta+\lambda+1)!!(\eta-\lambda)!!}\right]\times Y_\lambda{}^{\mu*}(\theta_i,\varphi_i)Y_\lambda{}^\mu(\theta_j,\varphi_j),\quad \eta=\lambda,\lambda+2,\cdots, \quad (A4)$$

which in hyperspherical coordinates r, χ takes on the form

$$\frac{1}{r_{ij}{}^2}=\frac{4\pi}{r^2}\sum_{\lambda\mu}\left[\sum_\eta\frac{2^\eta\eta!\sin^\eta\chi\cos^\eta\chi}{(\eta+\lambda+1)!!(\eta-\lambda)!!}\right]\times Y_\lambda{}^{\mu*}(\theta_i,\varphi_i)Y_\lambda{}^\mu(\theta_j,\varphi_j). \quad (A5)$$

On the other hand, if one contracts $(1/r_{ij})\cdot(1/r_{ij})$ by using the Laplace expansion of (18), one gets the following expression:

$$\frac{1}{r_{ij}{}^2}=\sum_{\substack{\lambda,\mu\\ \lambda_1,\lambda_2}}\frac{4\pi r_<{}^{\lambda_1+\lambda_2}}{r_>{}^{\lambda_1+\lambda_2+2}(2\lambda+1)}C(\lambda_1\lambda_2\lambda;00)^2\times Y_\lambda{}^{\mu*}(\theta_i,\varphi_i)Y_\lambda{}^\mu(\theta_j,\varphi_j), \quad (A6)$$

where the summations over λ, λ_1, and λ_2 are limited by the condition $\lambda=\lambda_1+\lambda_2,\lambda_1+\lambda_2-1,\cdots,|\lambda_1-\lambda_2|$, and the sum of $\lambda_1+\lambda_2+\lambda$ must be even. The two expansions of (A4) and (A6) can be compared by the method used to prove Eq. (20).

In evaluating the matrix element of (A1), the use of the Laplace expansion (A6) yields a sum of diverging terms and thus is untractable. In hyperspherical coordinates, however, the matrix element of (A1) can be calculated in a straightforward way yielding a series which consists of finite terms.

Isoperimetric and Other Inequalities in the Theory of Neutron Transport

LAWRENCE DRESNER*
Kernforschungszentrum Karlsruhe, Karlsruhe, Germany
(Received April 28, 1961)

Some isoperimetric and other inequalities occurring in the one-velocity theory of neutron transport are derived. The quantities involved in these inequalities all refer to bare solids with isotropic scattering and are: the critical multiplication, the first-collision probability, the non-escape probability, and the buckling. The inequalities proved provide upper and lower bounds for the quantities considered, and numerous examples of the estimation of these quantities in cases not readily amenable to direct calculation are given.

1. INTRODUCTION

1.1. In a mathematically complicated subject like the theory of neutron transport, simple, exact, and general formulas are usually not obtainable. In order to calculate quantities of interest, recourse must generally be had either to numerical calculation or to the introduction of simplifying but untrue assumptions. The introduction of such assumptions usually provides explicit and easily evaluated formulas but also usually results in errors of indeterminate sign and magnitude. It must be fairly said that situations in which these errors are small can generally be recognized when one understands the "physical" content of the theory, but the intuitive nature of this approach nevertheless involves an unavoidable, and furthermore itself uncertain, extent of error.

An elegant and also useful way out of this dilemma consists of enlarging the class of acceptable results to include inequalities. Thereby is one often provided with relationships involving the quantities of interest which are again simple and general, and which are furthermore exact at least in the sense of involving no mutilation of the theory. Two such inequalities providing, respectively, an upper and lower bound will furthermore yield estimates *whose maximum possible error is known.* If these bounds are close ones, as often happens, the numerical accuracy of the estimates may suffice for practical purposes.

1.2. The quantities of interest which we shall consider in this paper are all set functions which arise in the theory of neutron transport and which refer to bare, homogeneous, convex solids with isotropic scattering.

They are: the critical multiplication, the non-escape (absorption) probability of neutrons from a uniform isotropic source inside the solid, the corresponding first-collision probability, the buckling, and the diffusion-theoretic non-escape probability. The first three of these set-functions belong to strict one-velocity transport theory; the fourth and fifth to the simpler diffusion theory.

The transport-theoretic quantities have been calculated accurately in terms of simple formulas or as the result of not prohibitively great numerical labor only for the simplest geometrical shapes. The critical multiplication[1] has been calculated only for slabs and spheres, the non-escape probability[2] only for slabs, while the first-collision probability[3] has been calculated only for slabs, spheres, infinite right circular cylinders, hemispheres, and some oblate spheroids. For such a simple solid as a cube, however, no exact values for any of these quantities are available. Even the diffusion-theoretic quantities can only be calculated easily for spheres, rectangular parallelepipeds, and finite and infinite right circular cylinders. For more exotic shapes than those just mentioned, straightforward calculation can be very tedious. To avoid this tedium we can try to bound the quantities of interest using the inequalities developed in the body of this paper, and hence estimate them for solids of irregular shape.

1.3. Inequalities for set functions can be derived in several ways. The first and simplest way is just to compare, when possible, the values of the same set function for two solids, one of which can be totally included in the other. A second and more subtle way is to compare the values of the same set function for two solids which are related to each other by some process of symmetrization. (Symmetrization is the name given to a class of geometric transformations by which a solid is transformed into another which in some sense (depending on the precise nature of the transformation) is more symmetrical than its ancestor.) The first process of this kind was invented in 1836 by J. Steiner[4] who showed that this symmetrization leaves the volume of the solid unchanged while diminishing its surface area. Since constant reapplication of Steiner's symmetrization reduces all finite solids to spheres, Steiner was able to prove the classical isoperimetric theorem: Of all solids of a given volume, the sphere has minimum surface

* On leave from Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A., operated by Union Carbide Corporation for the U. S. Atomic Energy Commission.

[1] E. Inönü, Nuclear Sci. and Eng. **5**, 248 (1959); M. H. L. Pryce, MSP-2A (declassified 1947), H.M. Stationery Office, London; E. Inönü, USAEC Report ORNL-2842, p. 134, 1959.

[2] N. C. Francis, J. C. Stewart, L. S. Bohl, and T. J. Krieger, *Proceedings of the Second United Nations International Conference on the Peaceful Use of Atomic Energy*, Vol. 16, p. 517, 1958.

[3] K. M. Case, G. Placzek, and F. de Hoffmann, *Introduction to the Theory of Neutron Diffusion* (U. S. Government Printing Office, Washington, D. C., 1953), Vol. I.

[4] G. Pólya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics* (Princeton University Press, Princeton, New Jersey, 1951).

area.[5] It follows from this theorem that $S^3 \geqq 36\pi V^2$ for any arbitrary solid of surface S and volume V; this "isoperimetric" inequality can now be used to bound the surface of any solid from below. Similarly to the surface area many other set functions, including those which interest us here, vary monotonically under symmetrizing transformations. Thus for each an isoperimetric inequality holds from which a bound may be derived.

Another rich source of inequalities are the variational expressions which exist for many set functions. These expressions, *when they are either of the maximum or minimum type*, can be used directly to obtain bounds by the appropriate choice of trial functions. Indirectly, they can be used as very convenient starting points for the derivation of the inclusion and isoperimetric inequalities mentioned above.

A third source of inequalities arises from the application of what may be termed the "standard" inequalities of analysis to the sum or integral representations of the quantities of interest. In particular, the law of the mean, the inequality connecting the geometric and arithmetic means of a function, some more general inequalities involving convex functions, and the classical inequality of Schwarz are all used later in just this connection.

1.4. Research of the type described above has had a very long history. The isoperimetric theorems connecting the perimeter and area of a circle and the surface area and volume of a sphere were known to the Greeks. The powerful concept of symmetrization, by whose use many more isoperimetric theorems can be proven, was invented by Steiner more than a century ago, and only shortly thereafter a number of interesting isoperimetric inequalities concerning certain physical rather than purely geometric quantities were announced. In 1856, B. Saint Venant conjectured an isoperimetric inequality involving the torsional rigidity of elastic prisms on inductive grounds. In 1877, several isoperimetric theorems concerning the principal frequency of vibration of plates and membranes were stated without proof by Lord Rayleigh, who also developed the variational method of obtaining bounds to a high degree of refinement. In 1903 a famous isoperimetric theorem regarding the electrical capacity of solids was stated by Poincaré, but accompanied by an incomplete proof.

In the years between about 1900 and the present, effort was given to the proof and elaboration of these conjectures by T. Carleman, G. Faber, E. Krahn, R. Courant, G. Szegö, G. Pólya, and others. These workers confined their attention largely to inequalities involving purely geometric quantities and those physical quantities arising from Laplace's, Helmoltz's, or related equations (i.e., arising in electrostatics, the study of vibrations of plates and membranes, hydrodynamics, the theory of elasticity, the theory of heat conduction, etc.). In 1951, G. Szegö and G. Pólya published a book[4] in which all the old results and many new ones are systematically described, thus elevating this research, which in these authors' words "moves somewhat outside the usual channels," to the level of a discipline.

The mathematics of neutron diffusion theory is extremely similar to the mathematics of the studies mentioned parenthetically in the last paragraph. The methods described by Pólya and Szegö can thus be systematically applied to diffusion theory. Indeed, in some cases all that is required is a simple reinterpretation of Pólya and Szegö's results. Regrettably, only those quantities can be really effectively treated which admit of a variational representation of the maximum or minimum type; this limitation is probably a fundamental one. A single paper pointed in this direction has already been written by Ackroyd and Ball[6] who studied the effect of Steiner symmetrization on critical mass in diffusion theory.

The mathematics of strict transport theory is, however, essentially different from that involved in electrostatics, hydrodynamics, etc., since it is governed not by second-order partial differential equations but rather by integral (or integro-differential) equations. To obtain inequalities from these equations the author previously developed several new techniques[7] whose application is extended in the present work.

1.5. The arrangement of this paper is as follows: In the next short section, the quantities of interest are precisely defined, and in the following section, the process of symmetrization is defined. Following that, in the fourth section, the subsequently used representations of the quantities of interest are derived. In the fifth section, the various theorems are stated and proved. In the sixth section, a discussion and some examples are given. The reader who wishes to avoid the laborious details of the proofs may read Secs. 2 and 3, the *statements* of the theorems in Sec. 5, and Sec. 6 without difficulty.

2. DEFINITIONS

2.1. In one-velocity transport theory the criticality of a bare, homogeneous reactor with isotropic scattering is governed by the integral equation

$$\phi(\mathbf{r}) = c\int_V K(|\mathbf{r}-\mathbf{r}'|)\phi(\mathbf{r}')d^3\mathbf{r}', \tag{1a}$$

where

$$K(r) = e^{-r}/(4\pi r^2). \tag{1b}$$

[5] The word "isoperimetric" is actually a misnomer since the solids have the same volume (area) not the same surface area (perimeter). However, the theorem stated is a trivial deduction from the truly isoperimetric theorem: Of all solids of a given surface area, the sphere has maximum volume.

[6] R. T. Ackroyd and J. M. Ball, "On the conjecture that Steiner symmetrization reduces critical mass," UKAEA, Risley, Declassified Reprint WHC-(C)P-36, No. 8135, 1955.

[7] L. Dresner, Nuclear Sci. and Eng. **6**, 63 (1959); **7**, 260 (1960); **9**, 151 (1961).

Here $\phi(\mathbf{r})$ is the flux of neutrons at $\mathbf{r}$, defined as the product of the neutron density at $\mathbf{r}$ and the (single) neutron velocity, c is an eigenvalue whose physical significance is that it is the average number of secondary neutrons emerging from each collision which will make the reactor just critical, and $K(|\mathbf{r}-\mathbf{r}'|)$ is an integral kernel which represents the probability density that a neutron originating at $\mathbf{r}'$ will have its next collision in a differential volume element at $\mathbf{r}$. V finally is the reactor volume. *Here, as in the rest of the paper, the mean free path* (m.f.p.) *in the reactor has been chosen as the unit of length.* The spectrum of eigenvalues of (1) is discrete and characterized by the fact that there is a lowest one c_0.[8] To this lowest eigenvalue corresponds an eigenfunction (flux) which is positive everywhere in the reactor interior, while to all other higher eigenvalues correspond eigenfunctions which change sign somewhere in the reactor interior. Thus the lowest eigenvalue alone corresponds to a physically realizable persisting state, and henceforth, only it will be called the critical multiplication. Its reciprocal P, which will prove more convenient to consider in what follows, is just the average first-collision probability of neutrons spatially distributed in the persisting flux mode. It will therefore simply be called the critical first-collision probability.

2.2. Another average first-collision probability of interest is that of the neutrons originating from a uniform, isotropic source inside V. It will henceforth just be called the first-collision probability and will be denoted by P_c.

2.3. If V is filled with a non-multiplying medium capable of scattering and absorption only, one can consider a second probability referring to a uniform, isotropic source, viz., the average absorption or non-escape probability P_a. This quantity is defined as the average probability that a neutron will be absorbed in V, rather than leak out of it, irrespective of how many scattering collisions it has had.

2.4. The diffusion theoretic calculation of criticality is governed not by (1) but rather by the much simpler second-order partial differential equation

$$D\nabla^2\phi(\mathbf{r})+(c-1)\phi(\mathbf{r})=0, \tag{2}$$

where D is the diffusion constant and c and ϕ are defined as before. The most common boundary condition used with (2) is that the flux ϕ shall vanish on some prescribed surface, usually lying just outside the actual reactor surface. For the considerations of this paper no formal distinction exists between this extrapolated surface and the actual reactor surface. Hence, to (2) we shall add the condition that ϕ vanish on the reactor surface S.

The eigenvalue problem

$$\nabla^2\phi(\mathbf{r})+B^2\phi(\mathbf{r})=0 \quad \text{in } V, \tag{3a}$$

$$\phi(\mathbf{r})=0 \quad \text{on } S, \tag{3b}$$

which arises from (2) has infinitely many discrete eigenvalues B_n^2, of which there is a lowest, B_0^2.[8] This lowest eigenvalue alone corresponds to a flux ϕ which does not change sign inside V. We call it the buckling. It is a purely geometric quantity depending only on the size and shape of V. In terms of it the criticality condition may be expressed as $c=1+DB_0^2$.

2.5. It is not possible to define an average first-collision probability in pure diffusion theory since the individual flights of the neutrons do not appear in the theory. On the other hand, the average absorption probability can be defined simply as the ratio of the total absorption rate in V to the total source rate in V. When the source is a uniform, isotropic one, we shall denote the corresponding diffusion-theoretic average absorption probability by P_{ad}.

3. SYMMETRIZATION

3.1. The process of Steiner symmetrization can be succinctly defined as follows: Symmetrization with respect to a plane Q changes the solid V into a solid V^* such that:

(i) V^* is reflection symmetric with respect to Q.

(ii) Any straight line perpendicular to Q that intersects one of the solids V and V^* intersects the other also. Both intersections have the same length.

(iii) The intersection with V^* consists of just one line segment. The plane Q is called the plane of symmetrization.

A simple picture of the process of symmetrization is this: The solid is broken into paraxial differential cylinders, all of which are perpendicular to Q. These cylinders are then slid parallel to their axes until their midpoints all lie in Q. In case any of the cylinders consists of several pieces these are slid together and then the resulting single cylinder is slid so that its midpoint lies in Q.

3.2. It is clear from the definition of Steiner symmetrization that it leaves the volume of the solid unchanged. The surface area, on the other hand, is either decreased or remains the same. This last result is not at all obvious; it was first proved by Steiner. A little thought will convince the reader that repeated Steiner symmetrization with respect to a suitably chosen infinitude of planes will change any finite solid into a sphere of equal volume. Furthermore, repeated symmetrization in a suitable infinitude of planes all containing a common line L will reduce any infinite cylinder to a right circular cylinder with axis L. From these last two statements isoperimetric theorems follow for any quantities which never increase (decrease) under Steiner symmetrization.

[8] A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chain Reactors* (University of Chicago Press, Chicago, Illinois, 1958), pp. 406–10. See also the remarks by B. Davison, *Neutron Transport Theory* (Oxford University Press, London, 1957), pp. 195–6.

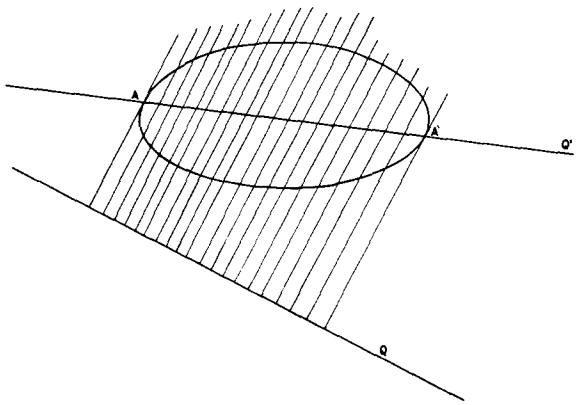

FIG. 1. Steiner symmetrization of a right elliptical cylinder.

3.3. Symmetrization of a right elliptical cylinder with respect to a plane containing its axis produces another right elliptical cylinder of the same base area but with the same or a smaller eccentricity. This fact can be used to show that any quantity which never increases (decreases) under Steiner symmetrization is a monotone increasing (decreasing) function of eccentricity, the base area being held fixed. It can be proven as follows: The midpoints of the differential cylindrical elements already lie in a plane Q' by a well-known property of the ellipse (see Fig. 1). Steiner symmetrization is then equivalent to rotating the plane Q' around the center of the ellipse until it is parallel to Q, all points sliding on lines perpendicular to Q, as though they were beads sliding on wires. This transformation is affine, and hence carries the original ellipse into another (in this case also of the same area). From the fact that one of the *new* axes AA' cannot be larger than the old major axis nor smaller than the old minor axis, but can be chosen arbitrarily in between, the desired conclusion follows.

3.4. Similar conclusions hold for spheroids but to prove them we must introduce the notion of Schwarz symmetrization. A solid V and a *solid of revolution* V^* can be related by Schwarz symmetrization as follows:

(i) Any plane perpendicular to the axis of revolution of V^* which intersects one of the solids V and V^* also intersects the other.

(ii) Both intersections have the same area.

Clearly, Schwarz symmetrization leaves the volume invariant. As it happens, the solid V^* which results from Schwarz symmetrization can also be obtained by an appropriately chosen infinitude of Steiner symmetrizations. We choose this infinitude as follows: All the planes of symmetrization contain the axis of the Schwarz symmetrization, but are otherwise distributed randomly in azimuth. This series of symmetrizations reduces all cross sections perpendicular to the common axis to circles, and thus has the same effect as Schwarz symmetrization.

If we first Steiner symmetrize a prolate spheroid we get, in general, an ellipsoid of equal volume whose largest principal axis lies in magnitude between the principal axes of the initial spheroid (again by an affine transformation). Schwarz symmetrization of the resulting ellipsoid with respect to largest principal axis gives another prolate spheroid of the same volume, but smaller eccentricity. If the original spheroid was oblate, the only difference is that the Schwarz symmetrization must be carried out around the smallest principal axis of the resulting ellipsoid. Since the largest (smallest) principal axis of the ellipsoid can be made as close to that of the original prolate (oblate) spheroid as desired, the eccentricity of the resulting spheroid can be made arbitrarily close to that of the original spheroid, from which the desired conclusion follows: Any quantity which never increases (decreases) under Steiner symmetrization is a monotone increasing (decreasing) function of the spheroid eccentricity, *the volume being held fixed.* (The spheroids being compared, however, must either be all prolate or all oblate.)

3.5. Steiner symmetrization of a hemisphere with respect to its diametral plane gives a volume equivalent oblate spheroid with a ratio of principal axes of $1:1:\frac{1}{2}$. Schwarz symmetrization with respect to a diameter gives a volume equivalent prolate spheroid with a ratio of principal axes of $1:\sqrt{2}/2:\sqrt{2}/2$.

Further discussion of Steiner symmetrization can be found in Pólya and Szegö's book.[4]

4. REPRESENTATIONS OF QUANTITIES OF INTEREST

4.1. A variational representation of the lowest eigenvalue c_0 of (**1**) is given by the Rayleigh quotient

$$P=\frac{1}{c_0}\geqq\frac{\int_V d^3r\int_V d^3r'\phi(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi(\mathbf{r}')}{\int_V\phi^2(\mathbf{r})d^3r}, \tag{4}$$

where $\phi(\mathbf{r})$ is any function. Equality in (4) occurs if and only if $\phi(\mathbf{r})=\phi^*(\mathbf{r})$, the true solution of (**1**). The sense of the inequality in (**4**) is related to the nature of the eigenvalue spectrum of (**1**), which we prove following the method of Davison[9]: Let $\phi_n(\mathbf{r})$ be the normalized, orthogonal eigenfunctions of (**1**) corresponding to the eigenvalues c_n. In terms of them, Davison writes the kernel $K(|\mathbf{r}-\mathbf{r}'|)$ in a bilinear Hilbert-Schmidt series[10]

$$K(|\mathbf{r}-\mathbf{r}'|)=\sum_{n=0}^{\infty}\frac{\phi_n(\mathbf{r})\phi_n(\mathbf{r}')}{c_n}, \tag{5}$$

[9] B. Davison, reference 8.

[10] See, for example, S. G. Mikhlin, *Integral Equations* (Pergamon Press, London, 1957), Chap. II, especially pp. 88–92.

indicating that (5) holds irrespective of whether the $\phi_n(\mathbf{r})$ form a complete set or not. Furthermore, $\phi(\mathbf{r})$ may be written

$$\phi(\mathbf{r})=\sum_{n=0}^{\infty} a_n\phi_n(\mathbf{r})+p(\mathbf{r}), \tag{6}$$

where $p(\mathbf{r})$ is orthogonal to every $\phi_n(\mathbf{r})$ and vanishes if the latter are complete. Then the right-hand side of (4) is given by

$$\sum_{n=0}^{\infty}(a_n^2/c_n)\Big/\left(\sum_{n=0}^{\infty}a_n^2+\int_V p^2(\mathbf{r})d^3r\right)\leqq c_0^{-1}, \tag{7}$$

since $c_0^{-1}\geqq c_1^{-1}\geqq c_2^{-1}\geqq\cdots$.

4.2. A useful and obvious representation of P_c is

$$P_c=V^{-1}\int_V d^3r\int_V d^3r'K(|\mathbf{r}-\mathbf{r}'|). \tag{8}$$

Another useful representation for P_c, whose derivation is outside the scope of this paper, is

$$P_c=1-\bar{l}^{-1}\int(1-e^{-l})f(l)dl, \tag{9}$$

where $f(l)$ is a certain normalized distribution of chord lengths l, whose mean $\bar{l}$ is equal to $4V/S$, i.e., to four times the volume-to-surface ratio.[3]

4.3. For P_a the only representation we shall use is a variational one. To derive it we must proceed as follows[11]: First consider the equation

$$H\psi^*=S, \tag{10}$$

where H is a positive, hermitian operator, and the star denotes the true solution of (10). A variational expression for the inner product (S,ψ^*) can be obtained by noting that for any ψ

$$\begin{aligned}|(S,\psi)|^2&=|(H\psi^*,\psi)|^2\\&\leqq(H\psi^*,\psi^*)(\psi,H\psi)=(S,\psi^*)(\psi,H\psi).\end{aligned} \tag{11}$$

Here, the inequality has been obtained from an obvious generalization of the Schwarz inequality in which (f,Hg) plays the role of the inner product between f and g. That this is possible depends on the positiveness of the operator H; for this property of H ensures that all norms (f,Hf) are non-negative. The proof that the operators H to which (11) is applied in this paper are positive as well as a proof of the generalized Schwarz inequality are to be found in the Appendix. From (11) it follows that

$$(S,\psi^*)\geqq|(S,\psi)|^2/(\psi,H\psi) \tag{12}$$

for any ψ, with equality if and only if $\psi=\psi^*$.

Now, when a uniform, isotropic source of unit total strength exists in a *non-multiplying* medium V, the flux is determined by the inhomogeneous equation

$$\begin{aligned}\phi^*(\mathbf{r})=c\int_V K(|\mathbf{r}-\mathbf{r}'|)\phi^*(\mathbf{r}')d^3r'\\+V^{-1}\int_V K(|\mathbf{r}-\mathbf{r}'|)d^3r',\end{aligned} \tag{13}$$

where c is now just the ratio of scattering to total cross section in V. The terms on the rhs are contributions to ϕ^* from collided and uncollided neutrons. The total absorption rate in V is given by

$$\begin{aligned}&(1-c)\int_V\phi^*(\mathbf{r})d^3r\\&=(1-c)\left[c\int_V d^3r\int_V d^3r'K(|\mathbf{r}-\mathbf{r}'|)\phi^*(\mathbf{r}')\right.\\&\qquad\left.+V^{-1}\int_V d^3r\int_V d^3r'K(|\mathbf{r}-\mathbf{r}'|)\right],\qquad(14a)\\&=(1-c)\left[cV\int_V S(\mathbf{r})\phi^*(\mathbf{r})d^3r+P_c\right].\qquad(14b)\end{aligned}$$

The second line follows from (8) and the identification of $S(\mathbf{r})$ with the last term on the rhs of (13). Furthermore, H must then be given by

$$H=\int\cdots[\delta(\mathbf{r}-\mathbf{r}')-cK(|\mathbf{r}-\mathbf{r}'|)]d^3r'. \tag{15}$$

Now, applying (12) to the first term on the rhs of (14b), we have that for any function ϕ

$$P_a=(1-c)\int_V\phi^*(\mathbf{r})d^3r\geqq(1-c)\left[\frac{c\left\{V^{-1}\int_V d^3r\int_V d^3r'\phi(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\right\}^2}{V^{-1}\int_V\phi^2(\mathbf{r})d^3r-cV^{-1}\int_V d^3r\int_V d^3r'\phi(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi(\mathbf{r}')}+P_c\right]. \tag{16}$$

Since the source in V is of unit total strength, the lhs of (16) is equal to P_a and has been so denoted. Equality again occurs if and only if $\phi=\phi^*$.

[11] T. Kahan, G. Rideau, and P. Roussopoulos, *Memorial des Sciences Mathematiques, Fascicule CXXXIV* (Gauthier-Villars, Paris, 1956); N. C. Francis *et al.*, reference 2.

4.4. For B_0^2 we also employ a variational representation based on a Rayleigh quotient, viz.[12]:

$$B_0^2 \leqq -\int_V \phi\nabla^2\phi d^3r \Big/ \int_V \phi^2 d^3r = \int_V |\nabla\phi|^2 d^3r \Big/ \int_V \phi^2 d^3r \quad (17)$$

for any (suitably well behaved) function ϕ which vanishes on S, the surface of V. The sense of the inequality follows from the eigenvalue structure of (3), that is from the fact that B_0^2 is the *lowest* eigenvalue. Equality again occurs if and only if ϕ equals the exact flux given by (3).

4.5. For eigenvalue problems with Hermitian operators another variational representation of the eigenvalue due to Weinstein[13] exists which gives both upper and lower bounds. However, the most forceful application of Weinstein's method unfortunately involves a number of assumptions which render it fundamentally unsound.

Let us begin by considering the quantity

$$M = (\{H-R\}\phi, \{H-a\}\phi)/(\phi,\phi), \quad (18)$$

where H is an hermitian operator, R is the Rayleigh quotient associated with ϕ, i.e., $(\phi,H\phi)/(\phi,\phi)$, and a is any number. Multiplying the numerator out shows that

$$M = W - R^2, \quad (19)$$

where

$$W = (H\phi,H\phi)/(\phi,\phi). \quad (20)$$

If now we set $\phi = \sum_{n=0}^{\infty} a_n\phi_n$, where ϕ_n are the eigenfunctions of H corresponding to eigenvalues λ_n *and are now assumed to form a complete set*, it can easily be shown that

$$M = \sum_{n=0}^{\infty} |a_n|^2(\lambda_n-R)(\lambda_n-a) \Big/ \sum_{n=0}^{\infty} |a_n|^2. \quad (21)$$

Now we designate by λ_m the eigenvalue to which R lies closest. Furthermore we choose $a=R$. In this case, it follows from (21) that

$$M \geqq (\lambda_m-R)^2, \quad (22)$$

from which it follows that

$$R+\sqrt{M} \geqq \lambda_m \geqq R-\sqrt{M}. \quad (23)$$

Now if in fact $\lambda_m=\lambda_0$, the lowest eigenvalue, (23) will give bounds for it. The Rayleigh quotient R itself is clearly a better upper bound than $R+\sqrt{M}$, but for the lower bound $R-\sqrt{M}$, there is as yet no competitor. Two remarks need to be made: First, of the identification $\lambda_m=\lambda_0$ we cannot in general be sure. Second, if the trial function ϕ differs from the true lowest eigenfunction of H by a small quantity of order ζ, λ_0 will differ from R by a quantity of order ζ^2 as is well known, but it will differ as we shall see presently from $R-\sqrt{M}$ by a quantity of order $|\zeta|$. Thus, (23) will provide only very rough bounds. This situation can be improved if we choose $a=\lambda_m+\lambda_{m+1}-R$ or $a=\lambda_m+\lambda_{m-1}-R$ according as R is $>$ or $<\lambda_m$. In these cases, respectively,

$$M \geqq (\lambda_{m+1}-R)(R-\lambda_m), \quad R>\lambda_m, \quad (24a)$$

$$M \geqq (R-\lambda_{m-1})(\lambda_m-R), \quad R<\lambda_m. \quad (24b)$$

If $\lambda_m=\lambda_0$, we can use (24a) and obtain

$$\lambda_0 \geqq R-(W-R^2)(\lambda_1-R)^{-1}. \quad (25)$$

In case some simple estimate of λ_1 can be made, (25) may provide a much sharper estimate of λ_0 than (23). To see how this may happen let us consider a trial function ϕ, which differs from the true lowest eigenfunction ϕ_0, by a quantity of order ϵ. Then, for small ϵ, it can easily be shown that $W-R^2$ is of order ϵ^2. Since $R-\lambda_0$ is also of order ϵ^2, it follows that the rhs of (25) differs from λ_0 by a quantity of the order of ϵ^2 at most. In the case of Eq. (23), however, the rhs and lhs both differ from λ_0 by a quantity of order ϵ.

In principle, Weinstein's method may be used to bound B_0^2; we shall say more of this application later.

4.6. Finally, we derive a variational representation for P_{ad} using (12) exactly as we did in treating P_a. This we do as follows: When a uniform, isotropic source of unit total strength exists in V, the flux ϕ is given according to diffusion theory by

$$-\nabla^2\phi^*(\mathbf{r})+\kappa^2\phi^*(\mathbf{r}) = (VD)^{-1} \quad \text{in } V, \quad (26a)$$

$$\phi^*(\mathbf{r}) = 0 \quad \text{on } S, \quad (26b)$$

where κ^2 is the inverse squared diffusion length and is given by $(1-c)/D$. The total absorption rate, equal here to P_{ad}, is given by

$$P_{ad} = (1-c)\int_V \phi^*(\mathbf{r})d^3r = (1-c)VD\int_V S(\mathbf{r})\phi^*(\mathbf{r})d^3r, \quad (27)$$

where here $S(\mathbf{r})=(VD)^{-1}$. Using (12) plus the identification $H=-\nabla^2+\kappa^2$, we have that for any function ϕ

$$P_{ad} \geqq \frac{(1-c)VD\left(\int_V (VD)^{-1}\phi(\mathbf{r})d^3r\right)^2}{\int_V \phi(\mathbf{r})(-\nabla^2+\kappa^2)\phi(\mathbf{r})d^3r}, \quad (28a)$$

[12] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik*, Erster Band (Springer Verlag, Berlin, 1931), sechstes Kapitel.

[13] D. H. Weinstein, Proc. Nat. Acad. Sci. **20**, 529 (1934); G. Goertzel and N. Tralli, *Some Mathematical Methods of Physics* (McGraw-Hill Book Company, Inc., New York, 1960), pp. 213–15.

or

$$P_{ad}\geqq\frac{1}{V}\frac{\left(\int_V\phi(\mathbf{r})d^3r\right)^2}{\kappa^{-2}\int_V|\nabla\phi|^2d^3r+\int_V\phi^2d^3r}, \tag{28b}$$

if we use Green's theorem. Equality occurs if and only if $\phi=\phi^*$.

5. THEOREMS AND PROOFS

5.1. A versatile class of results which will prove extremely useful in estimating all of the quantities P, P_a, P_c, B_0^2, and P_{ad} is the class of inclusion theorems. Our results along this line are expressed in the following theorem:

Theorem 1. If V_1 can be entirely included in V_2, then $P(V_2)\geqq P(V_1)$, $B_0^2(V_2)\leqq B_0^2(V_1)$, $V_2P_{ad}(V_2)\geqq V_1P_{ad}(V_1)$, $V_2P_c(V_2)\geqq V_1P_c(V_1)$, and $V_2P_a(V_2)\geqq V_1P_a(V_1)$.

Proof. The proofs of all parts of the theorem follow the same general rationale and are based on the expressions (4), (17), (28b), (8), and (16). The details for the first three parts of the theorem are very similar and we shall only carry them through for P: Let $\phi_1^*(\mathbf{r})$ be the exact flux in V_1, i.e., the exact solution of (1) in V_1. Let us define a trial function $\phi_2(\mathbf{r})$ in V_2 by the stipulation: $\phi_2(\mathbf{r})=\phi_1^*(\mathbf{r})$ in V_1, $\phi_2(\mathbf{r})=0$ otherwise. Then

$$P(V_1)=\frac{\int_{V_1}d^3r\int_{V_1}d^3r'\phi_1^*(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi_1^*(\mathbf{r}')}{\int_{V_1}[\phi_1^*(\mathbf{r})]^2d^3r}$$

$$=\frac{\int_{V_2}d^3r\int_{V_2}d^3r'\phi_2(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi_2(\mathbf{r}')}{\int_{V_2}\phi_2^2(\mathbf{r})d^3r}$$

$$\leqq P(V_2), \tag{29}$$

Q.E.D. The fourth part of the theorem, that for P_c, follows trivially from (8), the last part for P_a results from proving an inclusion theorem for the quantity $V(P_a-P_c)$ from (16), and then using the already proven result for P_c.

5.2. Another class of results arising from the comparison of different solids is expressed by:

Theorem 2. P, P_a, P_c, and P_{ad} all increase under Steiner symmetrization. B_0^2 decreases under Steiner symmetrization.

Proof. The proof of this theorem must be accomplished by two separate techniques. The first suffices to prove the theorem for the transport-theoretic quantities P, P_c, and P_a, while the second is reserved for the diffusion-theoretic quantities B_0^2 and P_{ad}. Let us begin with P_c in the form (8). Let us break the solid V up into paraxial differential cylindrical volume elements all of which are perpendicular to the plane of symmetrization, which for simplicity we take to be the xy plane. Let two of these cylindrical volume elements lie at x_1y_1 and x_2y_2 and have base areas dx_1dy_1 and dx_2dy_2, respectively. Let them intersect the surface of the (convex) solid V in z_1' and z_1'', z_2', and z_2'', respectively. The contribution of this pair to the multiple integral in (8) is

$$dP_c=\frac{dx_1dy_1dx_2dy_2}{V}\int_{z_1'}^{z_1''}dz_1\int_{z_2'}^{z_2''}dz_2$$
$$\times K([(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{\frac{1}{2}}). \tag{30a}$$

This can be rewritten as

$$dP_c=\frac{dx_1dy_1dx_2dy_2}{V}\int_{-\infty}^{+\infty}dz_1\int_{-\infty}^{+\infty}dz_2f_1(z_1)f_2(z_2)$$
$$\times K([(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{\frac{1}{2}}), \tag{30b}$$

where $f_1(z_1)=1$ for $z_1'\leqq z_1\leqq z_1''$ and zero otherwise, and similarly for f_2. Now let us rearrange the functions f_1 and f_2 in symmetrical decreasing order, i.e., let us replace $f_1(z_1)$ by a new function $\tilde{f}_1(z_1)$ which (i) is symmetric around $z_1=0$; (ii) is monotone decreasing; and (iii) has values between w and $w+dw$ over a set of the same measure as that for which f_1 has values between w and $w+dw$; and similarly with f_2. Since in (30b) K is a monotone decreasing function of $|z_1-z_2|$, by theorem 380 of Hardy *et al.*,[14] this rearrangement increases the integral of (30b). The result of this rearrangement can also be seen to be just the integral

$$d\bar{P}_c=\frac{dx_1dy_1dx_2dy_2}{V}\int_{-a/2}^{+a/2}dz_1\int_{-b/2}^{+b/2}dz_2$$
$$\times K([(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2]^{\frac{1}{2}}), \tag{31}$$

where $a=z_1''-z_1'$ and $b=z_2''-z_2'$. But this is precisely the result of Steiner symmetrization, since the infinitesimal cylinders now have their midpoints in the plane of symmetrization. Thus, P_c increases under Steiner symmetrization.

The proofs for P and P_a follow similar lines. For P this is the procedure: If $\phi^*(\mathbf{r})$ is the exact solution of (1) in V, then

$$P(V)=\frac{\int_V d^3r\int_V d^3r'\phi^*(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi^*(\mathbf{r}')}{\int_V[\phi^*(\mathbf{r})]^2d^3r}. \tag{32}$$

[14] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Cambridge University Press, London and New York, 1934).

On the other hand, for any arbitrary function $\phi^\dagger(\mathbf{r})$ in $V^\dagger$, the Steiner symmetrized solid

$$P(V^\dagger)\geqq\frac{\int_{V^\dagger}d^3r\int_{V^\dagger}d^3r'\phi^\dagger(\mathbf{r})K(|\mathbf{r}-\mathbf{r}'|)\phi^\dagger(\mathbf{r}')}{\int_{V^\dagger}[\phi^\dagger(\mathbf{r})]^2d^3r}. \quad (33)$$

To obtain $\phi^\dagger(\mathbf{r})$ we again divide the solid V into infinitesimal cylinders whose axes are perpendicular to the plane of symmetrization Q. $\phi^\dagger(\mathbf{r})$ is obtained by sliding each of these cylinders with the value of $\phi(\mathbf{r})$ fixed in the cylinder until all the midpoints lie in Q, and then rearranging $\phi^*(\mathbf{r})$ along each of these cylinders in symmetrical decreasing order with the midpoints of the cylinders as the centers of symmetry. Since rearranging of a function in symmetrical decreasing order does not alter the measure of the set over which it lies between specified values, the denominators of the rhs's of (32) and (33) are clearly equal. However (since ϕ^* and $\phi^\dagger$ are $\geqq 0$), by a repetition of the argument given above in connection with P_c, it can be shown that the numerator of the rhs of (33) exceeds that of (32). Thus $P(V^\dagger)\geqq P(V)$, Q.E.D.

A very similar method applied to (16), together with the already proven result for P_c, yields the announced result for P_a.

For B_0^2 and P_{ad} the approach is roughly similar but does not involve the notion of "symmetrical rearrangement in decreasing order." Instead a function $\phi^\dagger$ in V is used in (17) and (28b) which is obtained from ϕ^*, the exact flux in V, by symmetrizing its level surfaces. That is to say, if $\phi^*=C$ on the surface S_1 of V_1 then $\phi^\dagger=C$ on the surface $\tilde{S}_1$ of the symmetrized solid $V_1^\dagger$. From this definition it easily follows that

$$\int_V G(\phi^*(\mathbf{r}))d^3r=\int_{V^\dagger}G(\phi^\dagger(\mathbf{r}))d^3r, \quad (34)$$

where G is any function. However, as we shall presently see

$$\int_V|\nabla\phi^*|^2d^3r\geqq\int_{V^\dagger}|\nabla\phi^\dagger|^2d^3r. \quad (35)$$

From these last two equations the desired results for B_0^2 and P_{ad} easily follow.

Rather than (35) we shall prove a more general theorem, due to Pólya and Szegö,[4] whose method we follow without change: Let $F(x)$ be a concave-upwards, monotone increasing function of x. Then, with $\phi^\dagger$ and ϕ^* related as above,

$$I\equiv\int_V F(|\nabla\phi^*|)d^3r\geqq\int_{V^\dagger}F(|\nabla\phi^\dagger|)d^3r\equiv\tilde{I}. \quad (36)$$

To prove (36) we proceed as follows: Let S be a level surface of ϕ^* on which $\phi^*=C$; on $\tilde{S}$ let $\phi^\dagger=C$. Let an infinitesimal cylinder perpendicular to the plane of symmetrization (now chosen as the xy plane) and with base area $dA=dxdy$ intersect S at z_1 and z_2, and $\tilde{S}$ at $\pm z_0$. Let us compare the contributions to I and $\tilde{I}$ from the respective volumes lying inside $dxdy$ and corresponding, respectively, to values of ϕ^* and $\phi^\dagger$ between C and $C+dC$. In V there are two such volumes, one at z_1 of volume $dV=dAdC|dz_1/dC|$, and one at z_2 of volume $dV=dAdC|dz_2/dC|$. At z_1 the value of $|\nabla\phi^*|$ is

$$\left|\frac{dC}{dz_1}\right|\frac{1}{n_{z_1}},$$

where n_{z_1} is the z component of the outward normal to S at (x,y,z_1), and similarly at z_2. The contribution of the two volumes to I is then just

$$dI=\left[\left|\frac{dz_1}{dC}\right|F\left(\left|\frac{dC}{dz_1}\right|\frac{1}{n_{z_1}}\right)+\left|\frac{dz_2}{dC}\right|F\left(\left|\frac{dC}{dz_2}\right|\frac{1}{n_{z_2}}\right)\right]dAdC. \quad (37)$$

Since F is concave upwards, we may write according to theorem 204 of Hardy *et al.*,[14]

$$dI\geqq F\left(\left[\frac{1}{n_{z_1}}+\frac{1}{n_{z_2}}\right]\Big/\left[\left|\frac{dz_1}{dC}\right|+\left|\frac{dz_2}{dC}\right|\right]\right)\times\left(\left|\frac{dz_1}{dC}\right|+\left|\frac{dz_2}{dC}\right|\right)dAdC. \quad (38)$$

Next we note that $2z_0=z_2-z_1$. Hence,

$$2\left|\frac{dz_0}{dC}\right|=\left|\frac{dz_2}{dC}-\frac{dz_1}{dC}\right|. \quad (39a)$$

Since dz_2/dC and dz_1/dC must have *opposite* signs, (39a) can be rewritten

$$2\left|\frac{dz_0}{dC}\right|=\left|\frac{dz_2}{dC}\right|+\left|\frac{dz_1}{dC}\right|. \quad (39b)$$

Furthermore, since

$$2\frac{\partial z_0}{\partial x}=\frac{\partial z_2}{\partial x}-\frac{\partial z_1}{\partial x}, \quad (40a)$$

$$2\frac{\partial z_0}{\partial y}=\frac{\partial z_2}{\partial y}-\frac{\partial z_1}{\partial y}, \quad (40b)$$

we have by application of Minkowski's inequality

$$2\left(1+\left(\frac{\partial z_0}{\partial y}\right)^2+\left(\frac{\partial z_0}{\partial y}\right)^2\right)^{\frac{1}{2}}$$

$$=\left(4+\left(\frac{\partial z_2}{\partial x}-\frac{\partial z_1}{\partial x}\right)^2+\left(\frac{\partial z_2}{\partial y}-\frac{\partial z_1}{\partial y}\right)^2\right)^{\frac{1}{2}} \quad (41a)$$

$$\leqq\left(1+\left(\frac{\partial z_2}{\partial x}\right)^2+\left(\frac{\partial z_2}{\partial y}\right)^2\right)^{\frac{1}{2}}$$

$$+\left(1+\left(\frac{\partial z_1}{\partial x}\right)^2+\left(\frac{\partial z_1}{\partial y}\right)^2\right)^{\frac{1}{2}}. \quad (41b)$$

But (41b) is simply the equation

$$\frac{2}{n_{z0}}\leqq\frac{1}{n_{z1}}+\frac{1}{n_{z2}}, \quad (42)$$

since for any surface

$$(n_z)^{-1}=\left(1+\left(\frac{\partial z}{\partial x}\right)^2+\left(\frac{\partial z}{\partial y}\right)^2\right)^{\frac{1}{2}}.$$

Substituting (39b) and (42) in (38), and using the monotonicity of F, we obtain

$$dI\geqq 2F\left(\left|\frac{dC}{dz_0}\right|\frac{1}{n_{z0}}\right)\left|\frac{dz_0}{dC}\right|dA\,dC. \quad (43)$$

Comparing the rhs of (43) with that of (37), we see it is just the contribution to dI from the volume dV lying inside dA for which $C\leqq\phi^{\dagger}\leqq C+dC$. Thus, $dI\geqq d\bar{I}$ and $I\geqq\bar{I}$, Q.E.D.

It is worth noting that (41) and (42) express the essential step in showing that the surface is decreased by Steiner symmetrization; for further discussion of this as well as the preceding proof the reader is referred to reference 4.

5.3. In cylindrical and rectangular coordinates, among others, the diffusion equation is separable, so that solutions to problems involving right cylinders and rectangular parallelepipeds may be expressed in terms of results applicable to slabs and infinite cylinders. For two of the transport-theoretic quantities considered in this paper, viz., P and P_c, results are available which have to some extent the same effect. The first of these is:

Theorem 3. If V is the volume common to (i) two perpendicular slabs S_1 and S_2, or (ii) three mutually perpendicular slabs S_1, S_2, and S_3, or (iii) an infinite right cylinder C and a slab S perpendicular to it, then

(i) $P_c(S_1)$ *and* $P_c(S_2)\geqq P_c(V)\geqq P_c(S_1)P_c(S_2)$.

(ii) $P_c(S_1)$ *and* $P_c(S_2)$ *and* $P_c(S_3)\geqq P_c(V)\geqq P_c(S_1)\times P_c(S_2)P_c(S_3)$.

(iii) $P_c(C)$ *and* $P_c(S)\geqq P_c(V)\geqq P_c(C)P_c(S)$; *where* $P_c(S_1)$ *is the value of* P_c *for the slab* S_1, *etc. The same results also hold for* P.

Proof. We shall first prove the theorem for P_c: Let us introduce the characteristic function $f(\mathbf{r})$ of V defined by

$$f(\mathbf{r})=1 \quad \mathbf{r} \text{ in } V$$
$$=0 \quad \text{otherwise} \quad (44)$$

and its Fourier transform

$$f(\mathbf{k})=\int_V e^{i\mathbf{k}\cdot\mathbf{r}}d^3r;\quad f(\mathbf{r})=(2\pi)^{-3}\int_\infty f(\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{r}}d^3k. \quad (45a)$$

Introducing these into (8) one can show after some simple manipulation that

$$VP_c=(2\pi)^{-3}\int_\infty |f(\mathbf{k})|^2\frac{\arctan k}{k}d^3k. \quad (46a)$$

Here, use has been made of the fact that

$$K(\mathbf{k})=\int_\infty \frac{e^{-r}}{4\pi r^2}e^{i\mathbf{k}\cdot\mathbf{r}}d^3r=\frac{\arctan k}{k}. \quad (45b)$$

If in (8) one replaces $K(\mathbf{r})$ of $\delta(\mathbf{r})$, the Dirac delta function, (46a) becomes

$$V=(2\pi)^{-3}\int_\infty |f(\mathbf{k})|^2 d^3k. \quad (46b)$$

Let us first consider part (i) of the theorem. Let V be a rectangular parallelepiped of sides b_1, b_2, and b_3. Then,

$$f(\mathbf{k})=\prod_{j=1}^{3}\int_{-b_j/2}^{+b_j/2}e^{ik_jx_j}dx_j=\prod_{j=1}^{3}f_j(k_j), \quad (47)$$

where x_j are the cartesian coordinates of $\mathbf{r}$, and k_j are the cartesian coordinates of $\mathbf{k}$. Now if b_2 and b_3 become very large, $f(\mathbf{k})$ is only appreciable when k_2 and k_3 are near zero. Hence, for large b_2 and b_3, $k\approx k_1$, and

$$(2\pi)^3VP_c=\prod_{j=1}^{3}\int_{-\infty}^{+\infty}|f_j(k_j)|^2\frac{\arctan k_1}{k_1}dk_j \quad (48a)$$

$$=(2\pi)b_2(2\pi)b_3\int_{-\infty}^{+\infty}|f_1(k_1)|^2\frac{\arctan k_1}{k_1}dk_1. \quad (48b)$$

The second equality follows from the one-dimensional analog of (46b). If we let b_2 and b_3 approach infinity, we then have

$$(2\pi)b_1P_c(b_1)=\int_{-\infty}^{+\infty}|f_1(k)|^2\frac{\arctan k}{k}dk, \quad (49a)$$

where $P_c(b_1)$ is the value of P_c for a slab of thickness b_1. On the other hand, if only b_3 becomes infinite,

$$(2\pi)^2b_1b_2P_c=\prod_{j=1}^{2}\int_{-\infty}^{+\infty}|f_j(k_j)|^2\frac{\arctan k}{k}dk_j, \quad (49b)$$

where $k^2=k_1^2+k_2^2$ and P_c is appropriate to the volume common to two perpendicular slabs. With k so defined it is true that

$$\frac{\arctan k_1}{k_1} \quad \text{and} \quad \frac{\arctan k_2}{k_2} \geqq \frac{\arctan k}{k} \geqq \frac{\arctan k_1}{k_1} \cdot \frac{\arctan k_2}{k_2}. \tag{50}$$

Substitution of (50) in (49b) and use of (49a) yields the conclusion $P_c(b_1)$ and $P_c(b_2) \geqq P_c \geqq P_c(b_1) \cdot P_c(b_2)$, Q.E.D. Parts (ii) and (iii) of the theorem for P_c are treated similarly.

For P we proceed as follows: If we introduce the Fourier transforms

$$\phi(\mathbf{k}) = \int_V \phi(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} d^3r; \qquad \phi(\mathbf{r}) = (2\pi)^{-3} \int_\infty \phi(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k \tag{51}$$

of any trial function $\phi(\mathbf{r})$ *which vanishes outside of* V, then it follows from (4) that

$$P \geqq \int_\infty |\phi(\mathbf{k})|^2 \frac{\arctan k}{k} d^3k \Big/ \int_\infty |\phi(\mathbf{k})|^2 d^3k. \tag{52}$$

Let us choose V to be the same rectangular parallelepiped as before. As a trial function, $\phi(\mathbf{r})$, let us choose $\prod_{j=1}^{3} \phi_j(x_j)$ where $\phi_j(x_j)$ is the *exact* solution of (1) in a slab of thickness b_j, and therefore vanishes for $|x_j| > \frac{1}{2}b_j$. Then

$$\phi(\mathbf{k}) = \prod_{j=1}^{3} \phi_j(k_j). \tag{53}$$

It now follows from (52) and (53) by reasoning quite similar to that used previously in connection with P_c that

$$P \geqq P(b_1)P(b_2)P(b_3), \tag{54}$$

where $P(b_j)$ is the critical first-collision probability for a slab of thickness b_j. This is the second inequality of part (ii) for P. Since $P(b_3)$ approaches unity as b_3 becomes infinite, (54) also gives the second inequality of part (i). The second inequality of part (iii) follows similarly. For P the first inequality follows from theorem 1.

5.4. The potentialities of this method are not yet exhausted and a slightly more subtle application of it yields the following very beautiful and powerful theorem for P_c:

Theorem 4. Part 1. Consider a convex solid V and an arbitrary line L in space. Let the position of a point on L be measured by a coordinate z. Let the intersection of V and a plane Q perpendicular to L at z be denoted by $A(z)$. Let $P_c(z)$ be the collision probability of an infinite right cylinder with $A(z)$ as base and L as axis. Then

$$P_c(V) \leqq V^{-1} \int_a^b dz A(z) P_c(z),$$

where a and b are the limits determined by the planes tangent to V and perpendicular to L.

Theorem 4. Part 2. Consider a convex solid V and an arbitrary plane Q in space. Let the position of a point on Q be measured by a coordinate two-vector $\boldsymbol{\varrho}$. Let the intersection of V and a normal L to Q at $\boldsymbol{\varrho}$ be a line segment of length $t(\boldsymbol{\varrho})$. Let $P_c(\boldsymbol{\varrho})$ be the collision probability of a slab of thickness $t(\boldsymbol{\varrho})$. Then

$$P_c(V) \leqq V^{-1} \int_A d\boldsymbol{\varrho} t(\boldsymbol{\varrho}) P_c(\boldsymbol{\varrho}),$$

where A is the projected area of V on Q.

Proof. To prove this theorem we again employ the method of characteristic functions introduced above. For the first part of the theorem let us write the Fourier transform $f(\mathbf{k})$ of the characteristic function of V as

$$f(\mathbf{k}) = \int_a^b e^{ik_3 z} dz \int_{A(z)} d\boldsymbol{\varrho} \exp(i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}), \tag{55a}$$

where $\boldsymbol{\varrho}$ is the coordinate two-vector in the plane Q. Substituting (55a) in (46a) and rearranging the order of integration we obtain

$$(2\pi)^3 V P_c(V) = \int_{-\infty}^{+\infty} dk_3 \int_a^b e^{ik_3 z} dz \int_a^b e^{-ik_3 z'} dz' \int_\infty d\mathbf{k}_\rho \times \int_{A(z)} d\boldsymbol{\varrho} \exp(i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}) \times \int_{A(z')} d\boldsymbol{\varrho}' \exp(-i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}') K(\mathbf{k}). \tag{56}$$

Since the integrand with respect to k in (56) is positive for all $\mathbf{k}$ we can insert $K(\mathbf{k}_\rho) \geqq K(\mathbf{k})$ on the right-hand side, perform the k_3 and z' integrations in that order, and obtain

$$(2\pi)^3 V P_c(V) \leqq (2\pi) \int_a^b dz \Big[\int_\infty d\mathbf{k}_\rho \int_{A(z)} d\boldsymbol{\varrho} \exp(i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}) \times \int_{A(z)} d\boldsymbol{\varrho}' \exp(-i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}') K(\mathbf{k}_\rho) \Big]. \tag{57}$$

If we let a approach $-\infty$ and b approach $+\infty$ and imagine $A(z)$ is a fixed area, we obtain, by now familiar reasoning, the result that the square bracket in (57) is

just $(2\pi)^2 A(z)P_c(z)$. But then

$$VP_c(V) \leqq \int_a^b dz A(z)P_c(z) \tag{58}$$

Q.E.D. The second part of theorem 3 is proven in an exactly similar manner, except we write $f(\mathbf{k})$ as

$$f(\mathbf{k}) = \int_A \exp(i\mathbf{k}_\rho \cdot \boldsymbol{\varrho}) d\boldsymbol{\varrho} \int_{a(\rho)}^{b(\rho)} e^{ik_3 z} dz, \tag{55b}$$

where $a(\boldsymbol{\varrho})$ and $b(\boldsymbol{\varrho})$ are the intersections of the normal L to Q at $\boldsymbol{\varrho}$ with V, interchange the order of integration so that $\mathbf{k}_\rho$, $\boldsymbol{\varrho}$ and, $\boldsymbol{\varrho}'$ are last, and substitute $K(k_3) \geqq K(\mathbf{k})$ for $K(\mathbf{k})$.

5.5. With a somewhat different use of Fourier transforms one can prove the following theorem:

Theorem 5. $P \geqq B_0^{-1} \arctan B_0$, where B_0^2 defined by (3) is the buckling of the solid V to which P refers.

This theorem is related to the so-called "second fundamental theorem of reactor physics" [15]; more will be said about this connection in Sec. 6.

Proof. If the angular integrals in (52) are performed it becomes

$$P \geqq \int_0^\infty W(k) \frac{\arctan k}{k} dk, \tag{59}$$

where $W(k)$ is a positive normalized weighting function of k only. Now $k^{-1} \arctan k$ is a convex-downwards decreasing function of k^2; thus by theorem 204 of Hardy *et al.*[14]

$$\int_0^\infty W(k) \frac{\arctan k}{k} dk \geqq \frac{\arctan k_0}{k_0}, \tag{60}$$

where

$$k_0^2 = \int_0^\infty W(k) k^2 dk \tag{61a}$$

$$= \int_\infty |\phi(\mathbf{k})|^2 k^2 d^3k \Big/ \int_\infty |\phi(\mathbf{k})|^2 d^3k. \tag{61b}$$

If we invert the transforms in Eq. (61b), noting that

$$i\nabla\phi(\mathbf{r}) = (2\pi)^{-3} \int_\infty \mathbf{k}\phi(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k, \tag{62}$$

we find

$$k_0^2 = \int_V [\nabla\phi(\mathbf{r})]^2 d^3r \Big/ \int_V \phi^2(\mathbf{r}) d^3r. \tag{63}$$

Equation (62) can only converge if $\phi(\mathbf{r}) = 0$ on S the surface of V; for, otherwise, since $\phi(\mathbf{r}) = 0$ outside of V [see (51)], $\nabla\phi(\mathbf{r})$ will have an infinite singularity on S. Now, since k^{-1} arctan k is a monotone decreasing function, the best value for k_0^2 will be the smallest possible. But it follows from (63) and (17) that $(k_0^2)_{\min} = B_0^2$. Thus combining this result, (59), and (60) we have $P \geqq B_0^{-1}$ arctan B_0, Q.E.D.

5.6. As noted in the introduction, a rich source of inequalities are the variational representations of the different quantities; e.g., by the simple choice of a constant trial function $\phi = 1$ in (4) and (16) one obtains the following two theorems immediately[16]:

Theorem 6. $P \geqq P_c$.

Theorem 7. $P_a \geqq (1-c)P_c/(1-cP_c)$.

By combining the variational technique with an application of Schwarz's inequality, one can furthermore prove:

Theorem 8. $P_a \leqq (1-c)P_c/(1-cP)$.

Proof. Let $\phi = \phi^*$, the exact solution of (13). Then

$$P_a = (1-c)\left\{ \frac{c\left\{ V^{-1} \int_V d^3r \int_V d^3r' \phi^*(\mathbf{r}) K(|\mathbf{r}-\mathbf{r}'|) \right\}^2}{V^{-1} \int_V [\phi^*(\mathbf{r})]^2 d^3r - cV^{-1} \int_V d^3r \int_V d^3r' \phi^*(\mathbf{r}) K(|\mathbf{r}-\mathbf{r}'|) \phi^*(\mathbf{r}')} + P_c \right\} \tag{64a}$$

$$\leqq (1-c)\left\{ \frac{cV^{-2}\left\{ \int_V d^3r \int_V d^3r' K(|\mathbf{r}-\mathbf{r}'|) \right\}\left\{ \int_V d^3r \int_V d^3r' \phi^*(\mathbf{r}) K(|\mathbf{r}-\mathbf{r}'|) \phi^*(\mathbf{r}') \right\}}{V^{-1} \int_V [\phi^*(\mathbf{r})]^2 d^3r - cV^{-1} \int_V d^3r \int_V d^3r' \phi^*(\mathbf{r}) K(|\mathbf{r}-\mathbf{r}'|) \phi^*(\mathbf{r}')} + P_c \right\}. \tag{64b}$$

The application of Schwarz's inequality here is made in the same way as in (11). Dividing the numerator and denominator in (64b) by

$$cV^{-1} \int_V d^3r \int_V d^3r' \phi^*(\mathbf{r}) K(|\mathbf{r}-\mathbf{r}'|) \phi^*(\mathbf{r}')$$

and using (4) and (8) one finds $P_a \leqq (1-c)P_c/(1-cP)$, Q.E.D. By exactly the same technique as above applied to (28b) one can prove:

[15] A. M. Weinberg and E. P. Wigner, reference 8, pp. 397–406.

[16] Theorem 6 is due to P. A. M. Dirac, "Approximate rate of neutron multiplication for a solid of arbitrary shape and uniform density," declassified British Report MS-D-5, Part I, 1943. Theorem 7 is originally due to H. Hurwitz, Jr., according to N. C. Francis *et al.* (reference 2); see also: G. W. Stuart, Nuclear Sci. and Eng. 2, 617 (1957).

Theorem 9. $P_{ad} \leqq (1+B_0^2/\kappa^2)^{-1}$.

Noting that in diffusion theory the critical first-collision probability

$$P_d = (1+DB_0^2)^{-1} = (1+(1-c)(B_0^2/\kappa^2))^{-1},$$

the last theorem can be rewritten as

$$P_{ad} \leqq (1-c)P_d/(1-cP_d).$$

(Here c means only the fraction of scattering per collision.) In a sense this statement is weaker than that of theorem 8, since there P is replaced by $P_c \leqq P$ in the numerator resulting in a lower upper-bound. Also somewhat weaker than the theorems already proved and a consequence of them are the physically obvious inequalities $P \geqq P_c \geqq P_a$ and $P_d \geqq P_{ad}$.

5.7. For the strictly transport-theoretic quantities which can be variationally represented, viz., P and P_a, use of more complicated trial functions than those mentioned above leads to great difficulties [although theorem 5 results in a manner of speaking from the use of a diffusion-theoretic trial function in (4)]. For the diffusion-theoretic quantities the case is otherwise, and some elegant and useful results can be obtained by suitable choice of trial fluxes. These results can most easily be expressed in terms of a certain "effective radius" of a solid R_0, which is defined by

$$R_0^{-2} = \frac{1}{3V}\int\int (\mathbf{r}\cdot\mathbf{n})^{-1} dS, \tag{65}$$

where V is the volume of the solid, $\mathbf{r}$ is the radius vector from some fixed point O in the interior of V to any point Q on the surface, $\mathbf{n}$ is the outward normal at Q, and dS is the infinitesimal element of surface at Q. When V has a center of symmetry it will be chosen as O, otherwise the choice is left open and R_0 will be a function of O. In terms of this "radius" one can prove:

Theorem 10. For any finite solid V, $B_0^2 \leqq \pi^2/R_0^2$.

Proof. We use the method of prescribed level surfaces described by Pólya and Szegö[4]: With O as origin let the equation of the surface S of V be $r = R(\omega)$, where ω is a unit vector giving the direction of $\mathbf{r}$, and r is the latter's magnitude. Let us choose the level surfaces of the trial flux $\phi(\mathbf{r})$ to be the surfaces $r = uR(\omega)$ where $0 \leqq u \leqq 1$. (The point $u=0$ is O, the surface $u=1$ is S.) Furthermore, let us set $\phi(uR(\omega)) = f(u)$, where $f(1)=0$ and $f(u)$ is as yet otherwise undetermined.

Now the volume dV between the surfaces u and $u+du$ and lying inside an infinitesimal cone whose apex is at O and whose intersection with S is dS, is given by $dV = u^2 du(\mathbf{r}\cdot\mathbf{n})dS$. Furthermore, at Q, $|\nabla\phi|$ is given by $|df/du|\,(\mathbf{r}\cdot\mathbf{n})^{-1}$. Using these relations in (17) gives

$$B_0^2 \leqq \frac{\int_0^1 (df/du)^2 u^2 du \int\int (\mathbf{r}\cdot\mathbf{n})^{-1} dS}{\int_0^1 f^2 u^2 du \int\int (\mathbf{r}\cdot\mathbf{n}) dS} = \frac{\int_0^1 (df/du)^2 u^2 du}{\int_0^1 f^2 u^2 du} \cdot R_0^{-2} \tag{66}$$

since

$$V = \int_0^1 u^2 du \int\int (\mathbf{r}\cdot\mathbf{n}) dS = \tfrac{1}{3}\int\int (\mathbf{r}\cdot\mathbf{n}) dS.$$

The best choice of $f(u)$ is that function which will make the rhs of (66) a minimum. We can formulate the requirements on $f(u)$ conveniently through the variational equations

$$\delta \int_0^1 (df/du)^2 u^2 du = 0, \tag{67a}$$

$$\int_0^1 u^2 f^2 du = 1, \tag{67b}$$

$$f(1) = 0. \tag{67c}$$

The corresponding Euler-Lagrange differential equation for $f(u)$ is

$$\frac{d^2 f}{du^2} + \frac{2}{u}\frac{df}{du} + \gamma^2 f = 0, \tag{68}$$

where γ is an undetermined Lagrange multiplier. The regular solution of (68) is

$$f(u) = u^{-\frac{1}{2}} J_{\frac{1}{2}}(\gamma u) \propto u^{-1}\sin\gamma u. \tag{69}$$

To satisfy the requirement $f(1)=0$, γ must be chosen as π. Furthermore, by a partial integration the ratio $\int_0^1 (df/du)^2 u^2 du / \int_0^1 f^2 u^2 du$ can be shown to equal $\gamma^2 = \pi^2$ if $f(u)$ satisfies (68). Substituting this value in (66) gives the desired result. Finally, equality occurs when V is a sphere. By an exact repetition of the foregoing argument one can prove:

Theorem 11. For any infinite right cylinder,

$$B_0^2 = \alpha^2/(R_0')^2,$$

where α is the first root of the Bessel function J_0 ($=2.405$), and R_0' is defined by

$$(R_0')^{-2} = (2A)^{-1}\int (\mathbf{r}\cdot\mathbf{n})^{-1} ds. \tag{70}$$

Here A is the base area of the cylinder, $\mathbf{r}$ is the two-dimensional radius vector from some arbitrary fixed point O in the interior of A to a point Q on the perimeter

of A, $\mathbf{n}$ is the outward normal at Q, and ds is the infinitesimal element of perimeter at Q. Equality occurs here for right circular cylinders.

5.8. With a very similar technique one can prove the theorems:

Theorem 12. For any finite solid,

$$P_{ad}\geqq 1-\frac{3}{\kappa R_0}\left\{\coth\kappa R_0-\frac{1}{\kappa R_0}\right\}.$$

Theorem 13. For any infinite right cylinder,

$$P_{ad}\geqq 1-\frac{2}{\kappa R_0'}\frac{I_1(\kappa R_0')}{I_0(\kappa R_0')}.$$

Proof. By using the same level lines as in the last section (5.7) in (28b) we can rewrite it in the case of a finite solid as

$$P_{ad}\geqq \frac{3\left(\int_0^1 fu^2du\right)^2}{\int_0^1 f^2u^2du+(\kappa R_0)^{-2}\int_0^1 (df/du)^2u^2du}, \quad (71)$$

where $f(u)$ is again undetermined, save $f(1)=0$. The requirement that $f(u)$ be so chosen as to make the rhs of (71) a maximum leads to the Euler-Lagrange equation

$$\frac{d^2f}{du^2}+\frac{2}{u}\frac{df}{du}-(f-\gamma)\kappa^2R_0^2=0;\quad f(1)=0, \quad (72)$$

where γ is again an undetermined constant. It enters $f(u)$ however, only as a multiplicative factor and hence does not affect the rhs of (71). Indeed,

$$f(u)=\gamma\left\{1-\frac{\sinh(\kappa R_0u)}{u\sinh(\kappa R_0)}\right\}. \quad (73)$$

With the help of this expression and a partial integration, the rhs of (71) can be evaluated and yields the theorem as stated. Equality occurs again for spheres. When the solid is an infinite right cylinder the proof is similar. In theorem 13 equality occurs for right circular cylinders.

5.9. For the buckling, the variational treatment can be extended by application of the method of Weinstein. To carry this method through we must not only calculate the Rayleigh quotient R, as is done in the last section, but also the quantity W of (20). Choosing ϕ exactly as in Sec. 5.7 [i.e., choosing $f(u)$ given by (69) in the case of finite solids, etc.], we find that for finite solids and infinite right cylinders, respectively,

$$W=\pi^4/R_1^4 \quad \text{(finite solids)}, \quad (74a)$$

$$W=\alpha^4/(R_1')^4 \quad \text{(infinite right cylinders)}, \quad (74b)$$

where

$$R_1^{-4}=\tfrac{1}{3}V^{-1}\int\int(\mathbf{r}\cdot\mathbf{n})^{-3}dS \quad (75a)$$

and

$$(R_1')^{-4}=\tfrac{1}{2}A^{-1}\int(\mathbf{r}\cdot\mathbf{n})^{-3}dS \quad (75b)$$

We have not stated these results in the form of a theorem because of the somewhat uncertain nature of our application of Weinstein's method. The proof of (74) is as follows:

Proof. We consider only finite solids; the proof for cylinders is very similar. Since $H=-\nabla^2$, we need an expression for ∇^2 in terms of the variable u. This we obtain by noting first that, from our previous expressions for dV and $|\nabla\phi|$, it follows that

$$\int_V\nabla\psi\cdot\nabla\phi dV=\int\int(\mathbf{r}\cdot\mathbf{n})^{-1}dS\int_0^1\frac{df}{du}\frac{dg}{du}u^2du, \quad (76)$$

where $\phi(\mathbf{r})=f(u)$ and $\psi(\mathbf{r})=g(u)$. An integration by parts in both sides of (76) gives [since both $\phi(\mathbf{r})$ and $\psi(\mathbf{r})$ vanish on S]:

$$-\int_V\psi\nabla^2\phi dV$$

$$=-\int\int dS(\mathbf{r}\cdot\mathbf{n})^{-1}\int_0^1 g(u)\left\{\frac{d^2f}{du^2}+\frac{2}{u}\frac{df}{du}\right\}u^2du \quad (77a)$$

$$=-\int dS(\mathbf{r}\cdot\mathbf{n})u^2du\cdot g(u)\cdot\left\{\frac{d^2f}{du^2}+\frac{2}{u}\frac{df}{du}\right\}(\mathbf{r}\cdot\mathbf{n})^{-2}. \quad (77b)$$

Since the first factor in the integrand on the rhs of (77b) is dV, and the second factor is ψ, *which is arbitrary*, it must be that

$$\nabla^2\phi=\left\{\frac{d^2f}{du^2}+\frac{2}{u}\frac{df}{du}\right\}(\mathbf{r}\cdot\mathbf{n})^{-2}. \quad (78)$$

Then

$$W=\frac{\int_0^1\left\{\frac{d^2f}{du^2}+\frac{2}{u}\frac{df}{du}\right\}^2u^2du\int\int(\mathbf{r}\cdot\mathbf{n})^{-3}dS}{\int_0^1u^2f^2du\int\int(\mathbf{r}\cdot\mathbf{n})dS}. \quad (79)$$

If we furthermore require f to satisfy (68) with $\gamma=\pi$, we get

$$W=\pi^4\int\int(\mathbf{r}\cdot\mathbf{n})^{-3}dS\Big/3V, \quad (80)$$

which is identical with (74a), Q.E.D.

5.10. The variational theorems of the last three sections explicitly state relations between B_0^2 and P_{ad} and certain effective "radii." A similar theorem for P_c which has an origin quite different from a variational principle is:

Theorem 14. $P_c \geqq 1-\bar{l}^{-1}(1-\exp(-\bar{l}))$; $\bar{l}-4V/S$.

Proof. The proof is based on the use of the following inequality of Hardy *et al.*[14] (theorem 184) in (9):

$$\int f(l)e^{-l}dl \geqq \exp\left[-\int f(l)ldl\right] \tag{81}$$

if $f(l)$ is a normalized probability density.

6. DISCUSSION AND EXAMPLES

6.1. A number of remarks will be given below concerning the question of when a particular theorem can be expected to yield a limit close to the actual value of the quantity being estimated and when not. All of these remarks, however, only apply in extreme cases and, in general, the limits supplied by the various theorems must be compared to see which are best.

Theorem 1. This "inclusion" theorem works best when the volumes of the solids being compared are not too different. Thus inscribing a sphere in a cube may give fairly good limits while inscribing one in a long, thin cylinder should give rather bad limits.

Theorem 2. When using the isoperimetric corollary to theorem 2, viz., "Of all finite solids of a given volume the sphere has maximum P, P_a, P_c, P_{ad}, and minimum B_0^2," the best results will be obtained with equilateral or "sphere-like" solids. Thus cubes, cylinders with height and diameter equal, or ellipsoids of low eccentricity are all suitable for the application of theorem 2, while solids which are much longer in some directions than in others are quite unsuitable. Similar remarks refer to the isoperimetric corollary for cylinders.

Theorem 3. In discussing this theorem let us consider for the sake of argument situation (iii) of the hypothesis, viz., the perpendicular intersection of a cylinder C and a slab S. Furthermore, although we only discuss P_c in what follows, similar remarks apply to P. If the radius of the cylinder C is large, then $P_c(S)$ will be a very good upper limit and $P_c(S)P_c(C)$ a very good lower limit for P_c of the intersection solid. This is simply because $P_c(C)$ does not differ very much from unity, and thus the upper and lower limits do not differ very much from each other. Quite a similar conclusion holds if the slab is very thick. On the other hand, if the thickness and radius are both small, then it immediately follows that $P_c(C)\cdot P_c(S)$ will be a very bad lower limit since the value of the product of the two P_c's falls much more rapidly with decreasing size than the P_c of the intersection solid. One expects that the upper limit in this latter case will also not be very close to the exact value for the following reason: In the intersection solid neutrons born at any point and with any direction of velocity are within a short flight of the edge. In the bounding solids, i.e., cylinder or slab, however, neutrons whose velocity is nearly parallel to the elements of the surface are removed by a long flight from the edge. In sum, theorem 3 will work best for large solids.

Theorem 4. This theorem essentially generalizes the *upper* limits obtainable from theorem 3, and roughly similar remarks apply to it.

Theorem 5. This theorem is based essentially on the choice of a diffusion-theoretic trial function in the variational expression for P. Thus it ought to be a close underestimate for relatively large reactors where diffusion theory is approximately valid. This tendency is reinforced by the fact that for large reactors both P and B_0^{-1} arctan B_0 approach unity.

Theorem 5 is related to the so-called "second fundamental theorem of reactor theory" [15] which, for a one-velocity, bare reactor with isotropic scattering, equates P and B_0^{-1} arctan B_0, but which permits adjustment of B_0 through the introduction of an extrapolated surface. The requirement that the flux vanish on an extrapolated surface has the effect of decreasing B_0 and raising the value of B_0^{-1} arctan B_0. This will, in general, prolong agreement of this latter formula with P to much smaller sizes than otherwise, but render the sign of the error uncertain. Moreover the choice of an extrapolated surface is arbitrary although quite reasonable procedures can be worked out based on the extrapolation distance one obtains in Milne's problem. This arbitrariness renders the existence of any general inequality involving P and an extrapolated buckling unlikely, so that theorem 5 seems the strongest theorem we can prove in this direction.

Theorem 6. Theorem 6 is based on the choice of a flat trial flux and therefore should be best for small solids, in which the curvature of the true flux is not too large. Furthermore, since both P and P_c must both approach unity for large solids, this theorem may even provide useful estimates for large solids.

Theorems 7, 8. These two theorems are discussed together because: If P and P_c are close to one another, then *ceteris paribus* the upper and lower limits provided by these theorems should also be close. This will occur particularly for small solids as mentioned in the last paragraph although it should be pointed out, for example, that for slabs of any thickness P and P_c *never* differ by more than 3%. Trouble can develop, however, when c, the scattering fraction, is near unity and the solid is large, so that P is near unity, too. Then the upper limit provided by theorem 8 may grow inconveniently large.

Theorem 9. Theorem 9 has the same meaning in diffusion theory as theorem 8 has in the strict transport theory. One expects therefore, that for small solids the two sides of the inequality are not widely different in analogy with the discussion above. This can be directly supported as follows: Aside from the use of Schwarz's inequality, the chief step in the derivation of theorem 9 (or for that matter 8 too) is the use of the flux originating from a uniform isotropic source as a trial value for the critical flux (i.e., as a trial flux in the variational principle for B_0^2 or P). Since the first of these fluxes is concave upwards and the second concave downwards,

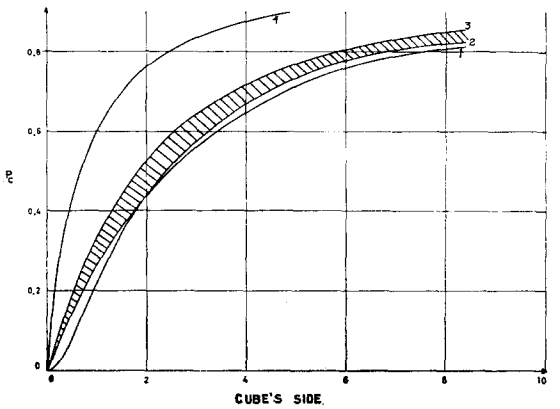

FIG. 2. Limits for P_c of cubes as a function of the length of the side. The limits shown are: 1, the lower limit from theorem 14; 2, the cube of P_c of the circumscribing slab; 3, P_c of the volume equivalent sphere; and 4, P_c of the circumscribing slab. The true value must lie in the cross-hatched area.

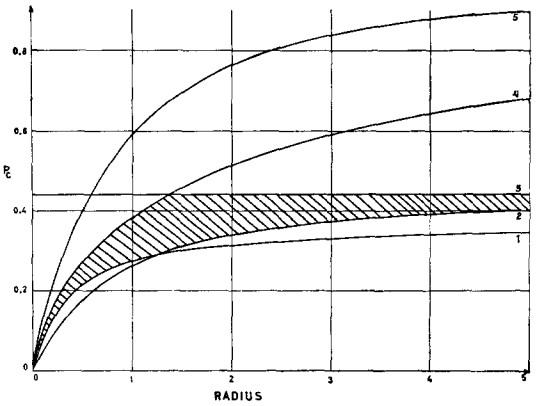

FIG. 3. Limits for P_c of $\frac{1}{2}$ m.f.p. thick disks of various radii. The limits shown are: 1, the lower limit from theorem 14; 2, the product of the P_c's of the circumscribing slab and cylinder; 3, P_c of the circumscribing slab; 4, P_c of the volume equivalent sphere; 5, P_c of the circumscribing cylinder. The true value must lie in the cross-hatched area.

one can be a good trial value for the other only when the curvature of both is negligible. This happens, however, only in small solids. Furthermore, when the flux is essentially flat, the use made of Schwarz's inequality also entails little error.

Theorems 10–13. Since in theorems 10 and 12 equality holds for spheres, these theorems should give very close limits for "sphere-like" solids. However, these need not be the only solids for which they give close limits, since the trial fluxes we have used are quite reasonable for many solids. Similar remarks apply to theorems 11 and 13.

Theorem 14. Regrettably little else can be said about when to expect close estimates from this theorem, save that it has the right values for very large and very small bodies.

6.2. As our first example let us consider the estimation of P_c for cubes. According to theorem 2 the value of P_c for the volume equivalent sphere is an upper limit for P_c of a cube. Another upper limit is P_c for a circumscribing slab, that is a slab of thickness equal to the cube's side. This follows from theorem 3. Theorem 3 also gives a lower limit, namely the cube of P_c for the circumscribing slab. Finally, theorem 14 gives a lower limit. These limits are plotted in Fig. 2 as a function of the cube's side; the true value of P_c for a cube must lie in the shaded region.

A number of observations concerning this figure are relevant. In the first place, for cubes, P_c of the circumscribing slab is a very bad overestimate as one might originally have expected. Indeed, in the range of sides from 0.2 to 8.0 m.f.p., this upper limit is much larger than that given by the volume equivalent sphere. On the other hand, for cubes one expects P_c for the volume equivalent sphere to be a fairly close over-estimate and this is borne out in the case at hand by its nearness to the lower limits in Fig. 2. For large cubes the underestimate provided by the cube of P_c of the circumscribing slab is the better of the two considered; however, for small cubes for which P_c of the circumscribing slab becomes small, its cube becomes extremely small and provides a rather useless limit. Thus, for cubes whose side is less than 2.0 m.f.p., the better lower limit is that of theorem 14.

For solids which are not "sphere-like" P_c of the volume equivalent sphere is usually a gross overestimate. This can be clearly seen in Fig. 3 where limits for P_c of disks $\frac{1}{2}$ m.f.p. thick and of various radii are plotted. These limits are: P_c for a slab $\frac{1}{2}$ m.f.p. thick (an upper limit by theorem 3); P_c for an infinite cylinder of radius equal to the disk radius (an upper limit by theorem 3); the product of these two numbers (a lower limit by theorem 3); P_c for the volume equivalent sphere (an upper limit by theorem 2); and the lower limit given by theorem 14. For disks for which the radius is very much greater than the thickness, P_c for the volume equivalent sphere is much larger than P_c for the circumscribing slab, whereas when the thickness and radii are comparable this situation is reversed. Not surprising is the further fact that the lower limit from theorem 14 is better than that from theorem 3 when the disk radius is small and worse when the disk radius is large.

Cubes and disks belong to that special class of solids which can be formed by the orthogonal intersection of slabs and cylinders. When we consider solids not belonging to this class we can no longer use theorem 3; however, we can use theorem 4 instead. The latter, however, supplies only an upper limit; hence, we have for the upper limit but two choices: the one just mentioned (theorem 4) and P_c of the volume equivalent sphere. For a lower limit we can use only theorem 14 in general.

Oblate spheroids are an excellent example of solids not belonging to this special class. A simple and useful upper limit for P_c for them can be obtained from theorem 4, part 2 by choosing the plane Q perpendicular

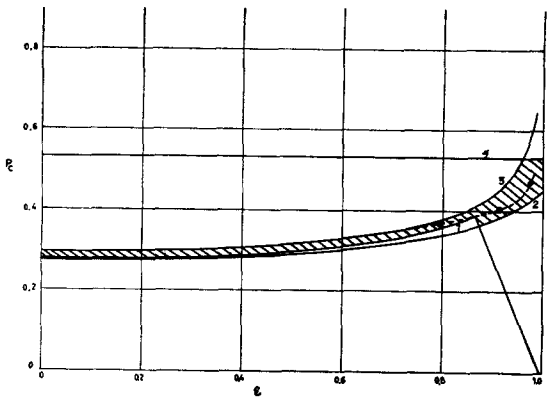

FIG. 4. Limits for P_c of oblate spheroids whose minor axes are 1 m.f.p. long as a function of eccentricity. The limits shown are: 1, a lower limit based on the P_c of certain hemispheres related to the spheroids by Steiner symmetrization (see text); 2, the lower limit from theorem 14; 3, P_c of the volume equivalent sphere; 4, an upper limit from the factorization theorem 4. The true value must lie in the cross-hatched area. Given also are three exact values available from reference 3.

to the minor axis of the spheroid. If the half-length of the minor axis is b it can then be shown from theorem 4 that

$$P_c \leqq 3\int_0^1 u^2 P_{cs}(2bu)du, \tag{82}$$

where P_{cs} is the value of P_c for a slab of thickness $2bu$. Interestingly enough, this limit depends only on the length $2b$ of the minor axis and not at all on the eccentricity of the spheroid!

In Fig. 4 the limit (82), the value of P_c for the volume equivalent sphere, and the limit from theorem 14 have been plotted as functions of the eccentricity ϵ for oblate spheroids with $2b=1$ m.f.p. ϵ is defined by

$$\epsilon^2 = 1-b^2/a^2, \tag{83}$$

where a is the semi-major axis of the spheroid. The limit

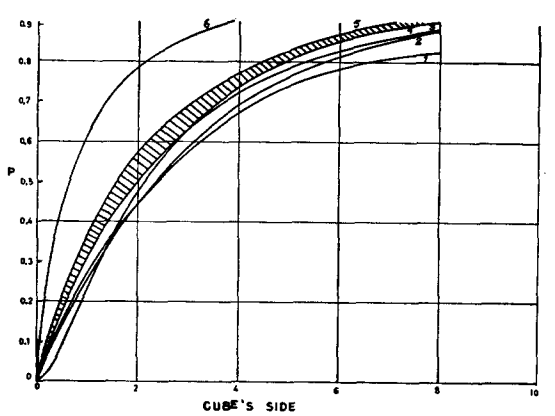

FIG. 5. Limits for P of cubes as a function of the length of the side. The limits shown are: 1, the lower limit for P_c taken from Fig. 2; 2, B_0^{-1} arctan B_0; 3, P of the inscribed sphere; 4, the cube of P of the circumscribing slab; 5, P of the volume equivalent sphere; and 6, P of the circumscribing slab. The true value must lie in the cross-hatched area.

of theorem 14 has been calculated with the aid of the relations[3]:

$$\bar{l} = (8/3)b/F(\epsilon), \tag{84a}$$

$$F(\epsilon) = 1+\epsilon^{-1}(1-\epsilon^2)\tanh^{-1}\epsilon. \tag{84b}$$

In addition to these limits one other can be derived which is applicable only to spheroids. If $\bar{P}_c(V)$ is the first-collision probability for a *hemisphere* of volume V, we can write for oblate spheroids:

$$P_c(b=\tfrac{1}{2};\epsilon) = P_c(V;\epsilon) \geqq P_c(V;\epsilon=\sqrt{3}/2) \geqq \bar{P}_c(V) \quad \text{if } \epsilon \leqq \sqrt{3}/2. \tag{85}$$

The first inequality follows from the monotonic decreasing behavior of the first-collision probability for spheroids with eccentricity proved in Sec. 3.4; the second from the fact that an oblate spheroid of eccentricity $\sqrt{3}/2$ results from Steiner symmetrizing a hemisphere in its diametral plane (Sec. 3.5). When $\epsilon \geqq \sqrt{3}/2$

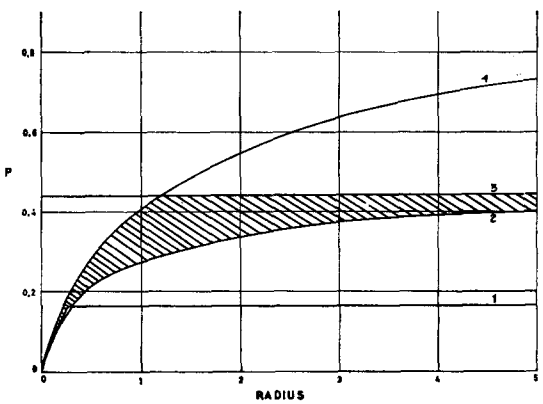

FIG. 6. Limits for P of $\frac{1}{2}$ m.f.p. thick disks of various radii. The limits shown are: 1, P of the inscribed sphere; 2, the lower limit for P_c taken from Fig. 3; P of the volume equivalent sphere; and P of the circumscribing slab. The true value must lie in the cross-hatched area.

we can furthermore write

$$\frac{P_c(b=\frac{1}{2};\epsilon)}{4(1-\epsilon^2)} \geqq P_c(b=\tfrac{1}{2};\sqrt{3}/2) \geqq \bar{P}_c(V') \quad \text{if } \epsilon \geqq \sqrt{3}/2, \tag{86}$$

where V' is the volume of an oblate spheroid with $b=\frac{1}{2}$ and $\epsilon=\sqrt{3}/2$. Here, the first inequality comes from the inclusion theorem (theorem 1), and the second again from the Steiner symmetrization process. Since $\bar{P}_c$ is tabulated,[3] this limit can be realized and is also plotted in Fig. 4.

Included in the diagram are three exact values of P_c corresponding to ratios a/b equal to 5/3, 5/2, and 5 which have been taken from the work of Case *et al.*[3] They indicate that for values of $\epsilon \lesssim 0{,}7$ at least the value of P_c is very close to that of the volume equivalent sphere. The upper limit of theorem 4 under these circumstances (i.e. $\epsilon \lesssim 0{,}7$) is much too high. However, when the eccentricity approaches 1 with the minor axis

remaining fixed, the volume increases rapidly and P_c for the volume equivalent sphere rapidly approaches 1. Finally for the most eccentric spheroids the upper limit of theorem 4 becomes applicable. Since when $1-\epsilon \ll 1$, $F(\epsilon)$ is very close to unity in this case the upper and lower limits depend only on b.

6.3. In the estimation of P, somewhat similar indications apply. Shown in Fig. 5 are the following limits for the P of cubes: (i) P of the circumscribing slab (upper limit by theorem 3 or theorem 1); (ii) the cube of P of the circumscribing slab (lower limit by theorem 3); (iii) P of the volume equivalent sphere (upper limit by theorem 2); (iv) B_0^{-1} arctan B_0 (lower limit by theorem 5); (v) the lower limit given by theorems 6 and 14; and (vi) P for the inscribed sphere (lower limit by theorem 1). (i) and (iii) are upper limits of which (iii) is much the lower of the two due to the equilateral nature of the cube. For large cubes (ii) gives the best lower limit; for small ones it is a gross underestimate

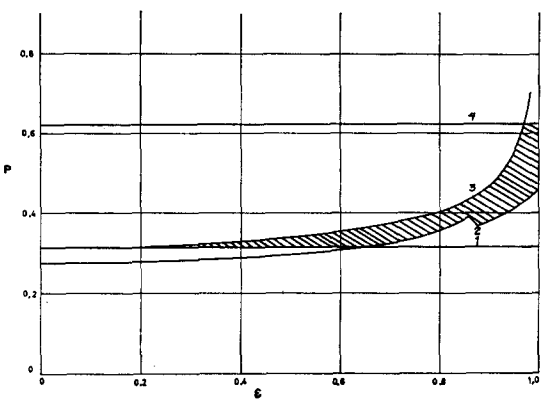

FIG. 7. Limits for P of oblate spheroids whose minor axis is 1 m.f.p. long as a function of eccentricity. The limits shown are: 1, P of the inscribed sphere; 2, the lower limit for P_c taken from Fig. 4; 3, P of the volume equivalent sphere; and 4, P of the circumscribing slab. The true value must lie in the cross-hatched area.

an (vi) is the best lower limit. (ii) and (vi) together cover the range plotted and for the best estimate in this case it is not necessary to use (iv) and (v).

In Figs. 6 and 7 are shown the following limits for P of the disks and spheroids we discussed in the last section: (i) P of the circumscribing slab, (ii) P of the volume equivalent sphere, (iii) P of the inscribed sphere, and (iv) the previously calculated lower limit to P_c. The first two are upper limits, the second two are lower limits.

In Fig. 6, both of the limits (i) and (ii) are used, (i) for the larger radii and (ii) for the smaller exactly as in Fig. 3. The lower limit consists mainly of (iv) except for the smallest cylinders where (iii) was used. In Fig. 7, the situation is quite similar to that of Fig. 4. For small eccentricities, (ii) was used for the upper limit; while for large eccentricities (i), which is the

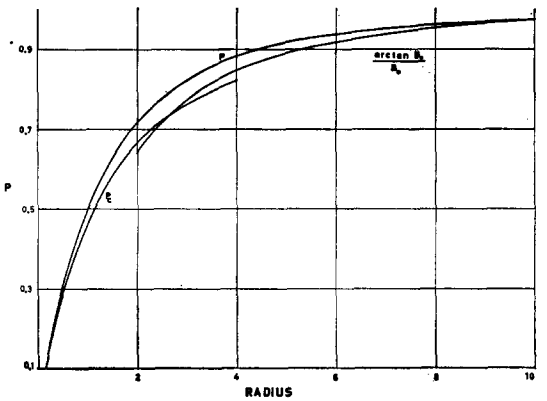

FIG. 8. A comparison of P with P_c and B_0^{-1} arctan B_0 for spheres.

analog of the upper limit in Fig. 4 from theorem 4, was used. For large eccentricities, (iv) was used for a lower limit; while for small eccentricities, (iii) was used. Figure 7 shows clearly the extremely slow variation of P with eccentricity for small eccentricity.

In neither of these last two figures was B_0^{-1} arctan B_0 used as a lower limit for P. In the case of the oblate spheroids, this is because the calculation of B_0 presents difficulties, and indeed the estimation of B_0^2 for spheroids forms the subject of one of the later paragraphs of this paper. In the case of the disks, however, B_0^{-1} arctan B_0 was calculated and found always to be less than limit (iv) above. This is due to the fact that theorem 5 is always unsuitable for a disk of thickness $\frac{1}{2}$ m.f.p., since such small dimensions preclude the use of diffusion theory.

The expectation that P_c should be the closer lower limit to P for small solids and B_0^{-1} arctan B_0 the closer lower limit to P for large solids has already been alluded to in the discussion of paragraph 6.1. Presented in Figs. 8 and 9 are comparisons of these two limits with the exact values of P for spheres and slabs, at once confirming this expectation and showing the rather good accuracy attainable with these variationally derived limits.

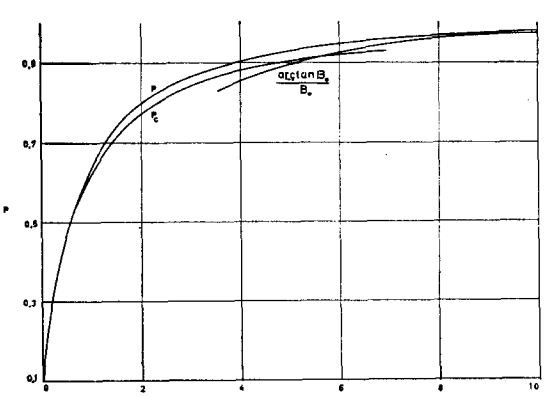

FIG. 9. A comparison of P with P_c and B_0^{-1} arctan B_0 for slabs.

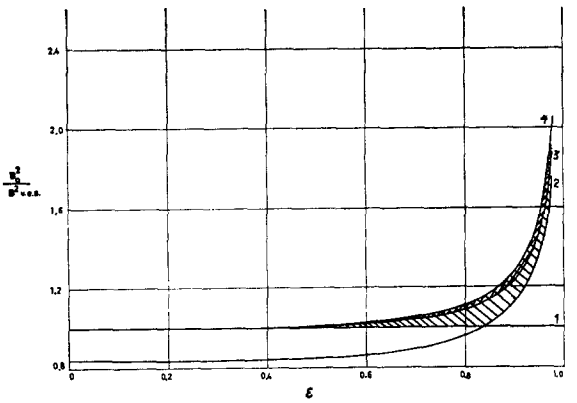

FIG. 10. Limits for the buckling of prolate spheroids as a function of eccentricity. Plotted as ordinate is the ratio of the buckling of the spheroid to that of the volume equivalent sphere. The limits shown are: 1, that derived from the volume equivalent sphere, i.e., unity; 2, that derived from the circumscribed cylinder; 3, that derived from Weinstein's method, i.e., from (25) and (87); and 4, that derived from theorem 10. The true value must lie in the cross-hatched area; if curve 3 is admitted as a lower limit the true value must then lie in the smaller doubly cross-hatched area.

6.4. We shall study but one more example, this one chosen to illustrate the technique of estimating the diffusion-theoretic quantities. We consider estimating the buckling of a prolate spheroid: From theorem 10 and (65) a short calculation shows that

$$\frac{B_0^2}{B_{\text{ves}}^2} \leqq \frac{1-\frac{1}{3}\epsilon^2}{(1-\epsilon^2)^{\frac{1}{3}}}, \tag{87}$$

where B_{ves} is the buckling of the volume equivalent sphere and ϵ is the eccentricity again defined by (83). From theorem 2 it follows that $B_0 \geqq B_{\text{ves}}$, so that the lhs of (87) is always greater than or equal to unity. This last limit we expect to be a good one near $\epsilon=0$, but to become quite useless for highly eccentric spheroids. The only remedy we have for this situation must be found in theorem 1, the inclusion theorem, since no other one can be directly applied to the estimation of B_0. For eccentric prolate spheroids a suitable solid for comparison is the circumscribed cylinder, i.e., that one with the semi-minor axis as radius and the major axis as height. It follows from the properties of this solid and theorem 1 that

$$\frac{B_0^2}{B_{\text{ves}}^2} > \frac{4\alpha^2/\pi^2+(1-\epsilon^2)}{4(1-\epsilon^2)^{\frac{1}{3}}}. \tag{88}$$

The limits (87) and (88) for the ratio B_0/B_{ves} have been plotted in Fig. 10 as functions of the eccentricity. Also plotted is the lower limit unity. Finally, a lower limit based on the method of Weinstein is plotted. This curve was obtained by using (25) with the further assumption that $\lambda_1=4R$ (*correct when* $\epsilon=0$). W, which is given by (74) and (75), can easily be evaluated explicitly and is given by

$$\frac{W}{B_{\text{ves}}^4} = \frac{1-\frac{2}{3}\epsilon^2+\frac{1}{5}\epsilon^4}{(1-\epsilon^2)^{\frac{1}{3}}}. \tag{89}$$

Under these circumstances the estimate given by (25) differs from B_0^2 by a quantity of order ϵ^4. This high accuracy is reflected in the fact that for small ϵ the curve based on (89) and that on theorem 10 (which also differs from B_0^2 by order ϵ^4) nearly coincide. Of course, the limit based on (88) is not a proven lower limit because of the inexact value of λ_1 used to obtain it, and its inclusion in Fig. 10 is to some extent contrary to the spirit of the rest of the paper.

6.5. What remains to be done? Very much indeed the author believes, so that when it is all mentioned the present paper will appear, as it properly should only as a beginning.

In the first place, the notion of seeking inequalities rather than equalities, and the rather exotic techniques (at least for physicists) this notion brings with it, such as Steiner symmetrization or rearrangement of a function in symmetrical decreasing order, have only been very slightly applied to physical problems. It is doubtless true that this enlargement of the conventional point of view will be a very fruitful one.

In the second place, even if we confine ourselves to the framework of neutron transport phenomena, the present paper is little more than a start. For example, our considerations here have been based on the presupposition that the scattering process is isotropic. But surely it is true that Steiner symmetrization decreases the critical multiplication even in a solid in which scattering is anisotropic. And quite probably there is some inequality similar to theorem 5 in media with anisotropic scattering, too.

Not only must a generalization to anisotropic scattering be made, but reflected media must be considered as well. Indeed, a start in this direction has already been made by Ackroyd and Ball,[6] who essentially consider the effect of Steiner symmetrization on critical multiplication for reflected systems.

Finally, even within the restricted milieu of bare, one-velocity reactors with isotropic scattering there are a number of open questions. For example: Does a factorization theorem like theorem 3 hold for P_a or not? Are the multiple collision probabilities from a uniform, isotropic source increased by Steiner symmetrization? (The answer here seems intuitively clear; the basic difficulty is generalizing theorem 380 of Hardy *et al.*[14]) Does a factorization theorem hold for these multiple collision probabilities or not?

ACKNOWLEDGMENTS

The author wishes to thank Dr. F. Stummel and Dr. J. Merkwitz for many illuminating discussions.

APPENDIX

By a positive operator is meant one for which $(f,Hf)\geqq 0$ for all f. If H is both positive and Hermitian then for *any* λ

$$0\leqq (f+\lambda g, H\{f+\lambda g\}) = (f,Hf)+2\ \mathrm{Re}\,[\lambda(f,Hg)]+|\lambda|^2(g,Hg). \quad \text{(A1)}$$

If we now choose $\arg\lambda=-\arg(f,Hg)$, it is easy to verify that

$$2\ \mathrm{Re}[\lambda(f,Hg)]=2|\lambda|\,|(f,Hg)|. \quad \text{(A2)}$$

Combining (A1) and (A2) gives the identity in $|\lambda|$:

$$|\lambda|^2(g,Hg)+2|(f,Hg)|\,|\lambda|+(f,Hf)\geqq 0. \quad \text{(A3)}$$

For (A3) to hold for all values of the modulus $|\lambda|$, the discriminant must never be positive, i.e.,

$$|(f,Hg)|^2-(f,Hf)(g,Hg)\leqq 0, \quad \text{(A4)}$$

which is a generalization of the usual Schwarz inequality.

The first of the operators for which (A4) is to be applied is (15). To prove it is positive we first expand $f(\mathbf{r})$ as

$$f(\mathbf{r})=\sum_{n=0}^{\infty} a_n\phi_n(\mathbf{r})+p(\mathbf{r}), \quad \text{(A5)}$$

where $p(\mathbf{r})$ is orthogonal to all the $\phi_n(\mathbf{r})$. Then we use the Hilbert-Schmidt series (5) for $K(|\mathbf{r}-\mathbf{r}'|)$ to obtain

$$\begin{aligned}(f,Hf)&=\int_V |f(\mathbf{r})|^2 d^3r - c\sum_{n=0}^{\infty}|a_n|^2/c_n\\ &\geqq \int_V |f(\mathbf{r})|^2 d^3r - \sum_{n=0}^{\infty}|a_n|^2\\ &=\int_V |p(\mathbf{r})|^2 d^3r \geqq 0.\end{aligned} \quad \text{(A6)}$$

Here use has been made of the fact that c, the fraction of scattering, is by definition less than one, while c_n by virtue of its definition as a critical multiplication, must be larger than one.

The second operator to which (A4) is to be applied is $K(|\mathbf{r}-\mathbf{r}'|)$ itself; from (5) and (A5) is trivially follows that

$$(f,Kf)=\sum_{n=0}^{\infty}|a_n|^2/c_n\geqq 0. \quad \text{(A7)}$$

The third operator to which (A4) is to be applied is the operator $-\nabla^2+\kappa^2$ in the volume V, with vanishing boundary condition on the surface S of V. Then by a simple application of Green's theorem,

$$\begin{aligned}(f,Hf)&=\int_V f^*(\mathbf{r})\{-\nabla^2+\kappa^2\}f(\mathbf{r})d^3r\\ &=\int_V \{|\nabla f(\mathbf{r})|^2+\kappa^2|f(\mathbf{r})|^2\}d^3r\geqq 0.\end{aligned} \quad \text{(A8)}$$

Probability Distribution for Classical Fluids*

A. A. BROYLES
Physics Department, University of Florida, Gainesville, Florida
(Received April 28, 1961)

The probability distribution in phase space is obtained for particles in a small part of the total volume of a classical monatomic fluid in thermal equilibrium. It is shown that the distribution reduces to that obtained from a grand canonical ensemble as this part of the volume increases in size. The Debye-Huckel pair distribution function is obtained in the proper limit for the Coulomb case. The distribution is in the form of a first approximation with an infinite series of correction terms.

I. INTRODUCTION

IN classical statistical mechanics, the assumption is generally made that, for an isolated system in equilibrium, the representative point is equally likely to be found in any of equal elements of volume of the phase space that is available to the system. An argument for this assumption has been made by Kinchin from the work of Birkhoff[1] and has been used to justify the use of the grand canonical distribution for subsystems of such a system when they are weakly interacting.[2,3]

We shall not assume weakly interacting subsystems here but shall take, instead, an isolated system made up of a fluid composed of atoms of one kind whose interactions are pairwise. Although it is not clear that this system can be broken up into weakly interacting subsystems, many quantities, such as the radial distribution function, depend only upon the probability distribution of particles in a small region of the fluid. It is the purpose of this paper to develop a method for obtaining this probability distribution function for a given small region which will be referred to as the "inner region" with volume τ_i. The remainder of the volume of the fluid will be called the "outer region" with volume τ.

Since the particles of the fluid can move between the two regions, the probability distribution that we shall obtain corresponds to the grand canonical distribution for weakly interacting systems. The procedure for obtaining it must involve an integration over the positions and momenta of all the N particles in the outer region for each given number M in the inner region where the total number $N_t=M+N$ is fixed. Methods for carrying out such integrations have been developed from the theory of collective coordinates proposed by Pines and Bohm.[4] Such a method was worked out by the author[5] in connection with the problem of obtaining the probability distribution of electric fields on ions in a plasma. The procedure we shall use here is not identical to the one used there[5] but is similar enough so that some of the results of that paper indicate what we might expect for calculations using the method described here. It will become apparent later that the theory presented here reduces to that of Debye and Huckel for electrolytes in the Coulomb case and in the proper limit.

It is necessary for this treatment to assume that any singularity in the pair interaction potential is integrable. The Coulomb potential, of course, satisfies this criterion. The Lennard-Jones and hard sphere potentials do not. Where unallowable singularities occur, they must be eliminated by an artificial cutoff or by a procedure of separation into long and short range interactions such as that in reference 5.

With integrable pair interaction potentials, a small constant or integrable function of the order of τ^{-1} can be subtracted to make the integral of the pair potential over the whole volume be zero. This is often done in the case of Coulomb particles by providing a uniform background charge of opposite sign to that of the particles. It will be assumed that something of this sort has been done for all pair potentials.

At the end of the derivation presented here, the limit should be taken such that the total number of particles and the total volume of the system approaches infinity with the average number per unit volume remaining constant.

II. NOTATION

$\mathbf{A}$ = matrix in second degree term in expansion of f in Eq. (28).
$\mathbf{a}$ = column submatrix of $\mathbf{A}$ in Eq. (II-2).
$\mathbf{B}$ = submatrix of $\mathbf{A}$ in Eq. (II-3).
$\mathbf{b}$ = column matrix involving $\sum_{j=1}^{M} w(\mathbf{R}-\mathbf{R}_j)$ described following Eq. (10).
$\mathbf{b}_u$ = column matrix involving $\sum_{j=1}^{M} U^*(\mathbf{R}-\mathbf{R}_j)$ in Eq. (46).
$\mathbf{C}$ = arbitrary square matrix.
E = energy of the entire system.
$f=\ln_e$ of the integrand of J defined in Eq. (17).
g = definition in Eq. (15).
$g_0=g$ at the point of steepest descents.
H_0 = Hamiltonian for the outer region.
H_1 = Hamiltonian for the inner region.
H_{01} = interaction energy between inner and outer regions.

* This work was supported in part by the National Science Foundation.

[1] A. I. Kinchin, *Mathematical Foundations of Statistical Mechanics* (Dover Publications, Inc., New York, 1949), pp. 28–29.
[2] Terrell L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956).
[3] D. J. Candlin, Nuovo cimento **15**, 856 (1960).
[4] D. Pines and D. Bohm, Phys. Rev. **85**, 338 (1952).
[5] A. A. Broyles, Z. Physik **151**, 187 (1958).

H_{01}' = effective interaction energy in Eq. (60).
I = integral over outer region in Eq. (4).
I_t = integral over whole volume in Eq. (5).
i = subscript indicating inner region.
J = integral in Eq. (14).
$\mathbf{k}$ = Fourier transform variable.
l = Fourier transform variable.
M = number of particles in the inner region.
m = mass of particles in the fluid.
N = number of particles in the outer region.
N_t = number of particles in the entire system.
$n_\mathbf{r}$ = number of particles in cell centered at $\mathbf{r}$.
o = order of $\mathbf{t}$ and $\mathbf{C}$.
$\mathbf{P}_j$ = momentum of particle j.
$P(\mathbf{R}_1,\mathbf{P}_1,\cdots,\mathbf{R}_M,\mathbf{P}_M)$ = probability distribution in phase space of the inner region.
$\mathbf{Q}$ = position vector in configuration space.
$Q_1\cdots Q_9$ = quantities defined in Eq. (II-4).
$\mathbf{q}$ = center of a cell in configuration space.
$\mathbf{R}$ = position vector in configuration space.
$\mathbf{r}$ = center of a cell in configuration space.
$s = g_0$.
$\mathbf{t}$ = arbitrary matrix.
t = subscript referring to total volume.
$\mathbf{U}$ = shielded potential in Eq. (42).
$\mathbf{U}^*$ = same as $\mathbf{U}$ except $e^{-\phi}$ is omitted from Eq. (42).
$\mathbf{U}_t$ = shielded potential defined over the whole volume in Eq. (43).
$\mathbf{U}_\lambda$ = same as $\mathbf{U}$ with (N/τ) reduced by a factor λ.
$\mathbf{W}$ = matrix obtained from w by averaging over cells and setting diagonal elements equal to zero.
w = pair potential energy.
$\mathbf{y}$ = column matrix defined in Eq. (25).
β = i times the value of l at the point of steepest descents.
γ = coefficient of M in P in Eq. (59).
$\delta(x)$ = Dirac delta function.
$\delta(N,M)$ = Kronecker delta.
$\mathrm{E} = E - H_1$.
ϵ = size of elementary cell.
λ = parameter in Eq. (II-29).
$\rho(\mathbf{R})$ = function defined in Eq. (49).
τ = volume of the outer region.
τ_i = volume of the inner region.
ϕ = shielded potential given by Eq. (24).
$\square_0$ = differential operation defined in Eq. (27).

A dot over a letter indicates a diagonal matrix. See footnote 7.

III. MATHEMATICAL DEVELOPMENT

We shall assume that the entire fluid is confined to an energy shell of energy E in phase space and that its representative point is equally likely to be found at any point in this shell.[6] The total energy of the system is equal to the Hamiltonian which we shall split into three parts, namely

$$H_1 = \sum_{j=1}^{M} (P_j^2/2m) + \sum_{1\leqslant i<j\leqslant M} w(\mathbf{R}_i - \mathbf{R}_j), \tag{1}$$

$$H_0 = \sum_{j=M+1}^{M+N} (P_j^2/2m) + \sum_{M<i<j\leqslant M+N} w(\mathbf{R}_i - \mathbf{R}_j), \tag{2}$$

and

$$H_{01} = \sum_{i=1}^{M}\sum_{j=M+1}^{N+M} w(\mathbf{R}_i - \mathbf{R}_j), \tag{3}$$

where $w(\mathbf{R}_i - \mathbf{R}_j)$ is the potential energy of the interaction between particles i and j. Here, H_1 is the energy of the particles in the inner region, H_0 is the energy of those in the outer region, and H_{01} is the interaction between the two.

We would like to obtain the probability distribution, $P(\mathbf{R}_1,\mathbf{P}_1,\cdots\mathbf{R}_M\mathbf{P}_M)$, that there will be M particles in the inner region at the positions $\mathbf{R}_1, \cdots, \mathbf{R}_M$ and momenta $\mathbf{P}_1, \cdots, \mathbf{P}_M$. We do not care which particle is at each position but only that some particle is there. To obtain this probability, it is necessary to evaluate the integrals,

$$I(\mathbf{R}_1,\mathbf{P}_1\cdots\mathbf{R}_M\mathbf{P}_M) = \int\cdots\int \delta(\mathrm{E}-H_0-H_{01}) \prod_{j=M+1}^{N+M} d\mathbf{R}_j d\mathbf{P}_j \tag{4}$$

and

$$I_t = \int\cdots\int_t \delta(E-H_1-H_0-H_{01}) \prod_{j=1}^{N+M} d\mathbf{R}_j d\mathbf{P}_j, \tag{5}$$

where E is defined by

$$\mathrm{E} = E - H_1. \tag{6}$$

The delta functions are those used by Dirac and the subscript t indicates that the $\mathbf{R}$ integrations are to be taken over the entire volume of the system. No subscripts means the $\mathbf{R}$ integrations are to be taken only over the outer region. The integration over each component of the $\mathbf{P}$'s extends from minus infinity to plus infinity.

The ratio $I(\mathbf{R}_1,\mathbf{P}_1\cdots\mathbf{R}_M\mathbf{P}_M)\prod_{j=1}^{M} d\mathbf{R}_j d\mathbf{P}_j/I_t$ is the fraction of the total volume of the energy shell allowed to the system when particles 1 through M are specified to lie in the elements of volume and momentum around $\mathbf{R}_1, \mathbf{P}_1$ to $\mathbf{R}_M\mathbf{P}_M$.

Since we are not concerned with which particle lies at each position, we must permute these M particles among the M positions and add the volumes in phase space. This is equivalent to multiplying the above ratio by $M!$ to obtain the volume fraction. Exchanging particles between the inner and outer regions increases the fraction of the shell volume allowed by a factor $(N+M)!/(N!)(M!)$. Since N is very large, Sterling's approximation may be used to replace $(N+M)!/N!$

[6] For a discussion of the reasonableness of this assumption, see Richard C. Tolman, *The Principles of Statistical Mechanics* (Oxford University Press, New York, 1938), first edition, Sec. 26.

by N^M. This fraction of the shell volume allowed is the required probability, so we have

$$P(\mathbf{R}_1,\mathbf{P}_1,\cdots\mathbf{R}_M\mathbf{P}_M)=N^M I(\mathbf{R}_1,\mathbf{P}_1,\cdots\mathbf{R}_M,\mathbf{P}_M)/I_t. \quad (7)$$

Instead of evaluating the constant I_t, we shall determine the dependence of P on the $\mathbf{R}$'s, $\mathbf{P}$'s, and M and leave the normalization to be done in connection with any specific problem.

If we replace the Dirac delta function in Eq. (4) by its Fourier transform, we have

$$I=(2\pi)^{-1}\int_{-\infty}^{\infty}e^{il\mathrm{E}}\int\cdots\int\exp\{-il[H_0+H_{01}]\}\times\prod_{j=M+1}^{M+N}d\mathbf{P}_j d\mathbf{R}_j dl. \quad (8)$$

The integration over momenta may be performed in Eq. (8). If we also divide all space into cells of volume ϵ, Eq. (8) becomes

$$I=(2\pi)^{-1}(2\pi m/i)^{3N/2}\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}l^{-3N/2}e^{il\mathrm{E}}\times\sum_{\{n\}=0}^{\infty}\exp\{-il[\tfrac{1}{2}\tilde{n}\mathbf{W}\mathbf{n}+\tilde{b}\mathbf{n}]\}\times(N!/\textstyle\prod_r n_r!)\delta(N,\sum_r n_r)dl\epsilon^N, \quad (9)$$

where

$$\epsilon^2 W_{rq}=\int_r\int_q w(\mathbf{R}-\mathbf{Q})d\mathbf{Q}d\mathbf{R}-\delta_{r,q}\int_r\int_r w(\mathbf{R}-\mathbf{Q})d\mathbf{Q}d\mathbf{R}, \quad (10)$$

$\mathbf{r}$ specifies the center of each cell, $\mathbf{n}$ is a column matrix whose components n_r are the number of particles in the cell, $\mathbf{W}$ is a matrix whose components W_{rq} are the averages of the potential energy $w(\mathbf{R}-\mathbf{Q})$ of a particle in cell $\mathbf{r}$ over the cell and a particle in cell $\mathbf{q}$ over that cell except that the diagonal elements are zero, $\mathbf{b}$ is a column matrix whose element b_r is the average of $\mathbf{R}$ in $\sum_{j=1}^{M}w(\mathbf{R}-\mathbf{R}_j)$ over the cell whose center is at $\mathbf{r}$, and the delta is the Kronecker delta. The diagonal elements of $\mathbf{W}$ are zero to account for the $i=j$ term omission in Eq. (2). The symbol $\{n\}$ under the summation sign in Eq. (9) means that the summation is a multiple one over all possible values of the n_r's. Each n_r can vary from zero to infinity.

The multiple sum in Eq. (9) would separate into a product of sums if the exponent were linear in $\mathbf{n}$. A Fourier transform will linearize it with the help of the equation

$$\exp(-\tfrac{1}{2}\tilde{t}\mathbf{C}^{-1}\mathbf{t})=(2\pi)^{-o/2}|\mathbf{C}|^{1/2}\int_{-\infty}^{\infty}\cdots\int\exp(i\tilde{t}\mathbf{k}-\tfrac{1}{2}\tilde{k}\mathbf{C}\mathbf{k})d\mathbf{k}. \quad (11)$$

Here $\mathbf{t}$, $\mathbf{k}$, and $\mathbf{C}$ are matrices of order o and $|\mathbf{C}|$ is the determinant of $\mathbf{C}$. To apply this to Eq. (9), we must make the identifications,

$$\mathbf{t}=(il)^{\frac{1}{2}}\mathbf{n} \quad (12)$$

and

$$\mathbf{C}^{-1}=\mathbf{W}. \quad (13)$$

Now the order o of these matrices is τ/ϵ since this is the number of cells in the outside region, and the matrix multiplications in Eq. (9) involve summations only over cells in this region.

If we substitute Eqs. (11)–(13) into (9), the multiple sum factors and Eq. (9) becomes

$$I=(2\pi)^{(3N/2)-(\tau/2\epsilon)-1}i^{-3N/2}m^{3N/2}|\mathbf{W}|^{-\frac{1}{2}}\lim_{\epsilon\to 0}\epsilon^N J, \quad (14)$$

where

$$J=\int_{l=-\infty}^{\infty}l^{-3N/2}e^{il\mathrm{E}}\int_{-\infty}^{\infty}\cdots\int\exp(-\tfrac{1}{2}\tilde{k}\mathbf{W}^{-1}\mathbf{k})g^N d\mathbf{k}dl,$$

and

$$g=\textstyle\sum_r\exp\{-ilb_r+i^{\frac{1}{2}}l^{\frac{1}{2}}k_r\}. \quad (15)$$

IV. STEEPEST DESCENTS APPROXIMATION

The integral J may be approximately evaluated by the method of steepest descents. To apply this method, we write J in the form,

$$J=\int\int\cdots\int e^f d\mathbf{k}dl, \quad (16)$$

where

$$f=-(3N/2)\ln(l)+il\mathrm{E}-\tfrac{1}{2}\tilde{k}\mathbf{W}^{-1}\mathbf{k}+N\ln(g), \quad (17)$$

and expand f in a Taylor's series around the point of steepest descents. This point is located in the usual manner by setting the first partial derivatives equal to zero. In this way, we obtain the equations

$$\left(\frac{\partial f}{\partial l}\right)_0=-(3N/2l_0)+i\mathrm{E}+(N/g_0)\left(\frac{\partial g}{\partial l}\right)_0=0, \quad (18)$$

$$\left(\frac{\partial f}{\partial \mathbf{k}}\right)_0=-\mathbf{W}^{-1}\mathbf{k}_0+(N/g_0)\left(\frac{\partial g}{\partial \mathbf{k}}\right)_0=0. \quad (19)$$

Here $\partial/\partial\mathbf{k}$ is a column matrix operator with elements $\partial/\partial k_r$.

It is useful at this point to introduce new quantities to replace l_0 and $\mathbf{k}_0$. They are

$$\beta=il_0, \quad (20)$$

$$g_0=s, \quad (21)$$

and

$$\boldsymbol{\phi}=il_0\mathbf{b}-i^{\frac{1}{2}}l_0^{\frac{1}{2}}\mathbf{k}_0. \quad (22)$$

We shall see later that β corresponds to $(kT)^{-1}$ and $\boldsymbol{\phi}$ is a shielded interaction which, for example, becomes, in certain limits, the solution to the nonlinear equation for the potential of mean force (multiplied by β) in

the Debye-Huckel theory for particles interacting with a Coulomb force.

Eliminating l_0 and $\mathbf{k}_0$ from Eqs. (18) and (19) with the aid of Eqs. (20)–(22) leaves us with[7]

$$(3N/2\beta)-\mathbf{E}+(N/2\beta s)(\beta\tilde{b}+\tilde{\phi})\mathbf{e}^{-\phi}=0 \tag{23}$$

and

$$\phi=\beta\mathbf{b}+(N/s)\beta\mathbf{W}\mathbf{e}^{-\phi}, \tag{24}$$

where the derivatives of g have been obtained and substituted. (See Appendix I.) The matrix $\mathbf{e}^{-\phi}$ has elements $e^{-\phi_r}$.

It is convenient at this point to define a column matrix $\mathbf{y}$ with $(\tau/\epsilon)+1$ elements with the relations

$$\begin{aligned} y_r&=k_r-k_{0r},\\ y_l&=l-l_0. \end{aligned} \tag{25}$$

The Taylor's expansion of f around the point of steepest descent now becomes

$$f=\exp(\tilde{y}\square_0)f(l_0,\mathbf{k}_0), \tag{26}$$

where $\square_0$ is the column matrix of order $(\tau/\epsilon)+1$ defined by the relations

$$\begin{aligned} \square_{0r}&=\partial/\partial k_{0r},\\ \square_{0l}&=\partial/\partial l_0. \end{aligned} \tag{27}$$

The first degree terms in the Taylor's expansion of f must vanish because we are expanding around the point of steepest descents. The second degree term may be more easily written with the aid of the matrix $\mathbf{A}$ defined as the negative of the direct product of two $\square_0$'s operating on f,

$$\mathbf{A}=-\square_0\times\square_0 f(l_0,\mathbf{k}_0). \tag{28}$$

Expanding the exponential operator in Eq. (26) and substituting into Eq. (16), we have

$$J=\exp\{f(l_0,\mathbf{k}_0)\}\times\int_{-\infty}^{\infty}\cdots\int\exp\{\sum_{j=3}^{\infty}(j!)^{-1}(\tilde{y}\square_0)^j f(l_0,\mathbf{k}_0)\}\times\exp(-\tfrac{1}{2}\tilde{y}\mathbf{A}\mathbf{y})d\mathbf{y}. \tag{29}$$

The first two operations by $\square_0$ on f in Eq. (29) will yield $-\mathbf{A}$ according to Eq. (28). We can write, therefore,

$$\sum_{j=3}^{\infty}(j!)^{-1}(\tilde{y}\square_0)^j f(l_0,\mathbf{k}_0)=-\sum_{j=3}^{\infty}(j!)^{-1}(\tilde{y}\square_0)^{j-2}\tilde{y}\mathbf{A}\mathbf{y}. \tag{30}$$

The last exponential under the integral in Eq. (29) is symmetric with respect to the simultaneous change of sign of all components of $\mathbf{y}$. If the preceding exponential is expanded in series, only terms having an even number of $\mathbf{y}$'s will survive the integrations. Such terms can be obtained by differentiating the last exponential with respect to components of the matrix $\mathbf{A}$. Thus

$$y_r y_q\exp(-\tfrac{1}{2}\tilde{y}\mathbf{A}\mathbf{y})=-2\frac{\partial}{\partial A_{rq}}\exp(-\tfrac{1}{2}\tilde{y}\mathbf{A}\mathbf{y}). \tag{31}$$

If we substitute Eqs. (30) and (31) into (29), we find the equation

$$J=\exp\{f(l_0,\mathbf{k}_0)\}\times\exp\left\{\sum_{j=3}^{\infty}\frac{(-2)^{j/2}}{j!}\left(\tilde{\square}_0\frac{\partial}{\partial\mathbf{A}}\square_0\right)^{(j-2)/2}\mathrm{Tr}\left(\mathbf{A}_0\frac{\partial}{\partial\mathbf{A}}\right)\right\}\times(2\pi)^{(\tau/2\epsilon)+\frac{1}{2}}|\mathbf{A}|^{-\frac{1}{2}}, \tag{32}$$

where Eq. (11) has been used to write the results of the integration over $\mathbf{y}$. Here $\partial/\partial\mathbf{A}$ is a square matrix with components

$$(\partial/\partial\mathbf{A})_{rq}=\partial/\partial A_{rq}. \tag{33}$$

In Eq. (32), $\square_0$'s operate only on $\mathbf{A}_0$ and $\partial/\partial\mathbf{A}$'s operate only on $|\mathbf{A}|$.

The steepest descents approximation is obtained from Eq. (32) by replacing the second exponential by unity. Eliminating l_0 and $\mathbf{k}_0$ and s in Eq. (17) by using Eqs. (20)–(22) we obtain

$$f(l_0,\mathbf{k}_0)=(3N/2)\ln(i)-(3N/2)\ln\beta+\beta\epsilon+(2\beta)^{-1}(\tilde{\phi}-\beta\tilde{b})\mathbf{W}^{-1}(\phi-\beta\mathbf{b})+N\ln(s) \tag{34}$$

with

$$s=\textstyle\sum_r e^{-\phi_r}=(\tau/\epsilon)[1-(\epsilon/\tau)\sum_r(1-e^{-\phi_r})]. \tag{35}$$

The expression, $\tilde{\phi}\mathbf{W}^{-1}\phi$ appearing in Eq. (34) may be rewritten with the aid of Eq. (24) to give

$$\tilde{\phi}\mathbf{W}^{-1}\phi=\beta\tilde{\phi}\mathbf{W}^{-1}\mathbf{b}+(N/s)\beta\tilde{\phi}\mathbf{e}^{-\phi}. \tag{36}$$

Substituting s and $\tilde{\phi}\mathbf{W}^{-1}\phi$ from Eqs. (35) and (36) into f in Eq. (34), f in turn into Eq. (32), and finally J into I with the aid of Eq. (14) gives

$$N^{-N}I=(2\pi)^{-\frac{1}{2}}[(2\pi m/\beta)^{\frac{3}{2}}(\tau/N)]^N\lim_{\epsilon\to 0}(|\mathbf{W}||\mathbf{A}|)^{-\frac{1}{2}}\times\exp\{\beta\mathbf{E}-\tfrac{1}{2}\tilde{\phi}\mathbf{W}^{-1}\mathbf{b}+\tfrac{1}{2}\beta\tilde{b}\mathbf{W}^{-1}\mathbf{b}+N\ln[1-(\epsilon/\tau)\textstyle\sum_r(1-e^{-\phi_r})]+\tfrac{1}{2}(N/s)\tilde{\phi}e^{-\phi}\}. \tag{37}$$

Equation (24) for ϕ may be used to obtain the relation

$$\tfrac{1}{2}\tilde{b}\mathbf{W}^{-1}\phi=\tfrac{1}{2}\beta\tilde{b}\mathbf{W}^{-1}\mathbf{b}+\tfrac{1}{2}(N/s)\beta\tilde{b}\mathbf{e}^{-\phi}. \tag{38}$$

Equation (38) together with Eqs. (6) and (1) for E may

[7] A function of a column or row matrix is here defined to be a column or row matrix whose elements are the function of the corresponding elements of the argument. Thus we have $(\mathbf{e}^{-\phi})_r=e^{-\phi_r}$.

be used to reduce Eq. (37) to

$$N^{-N}I=(2\pi)^{-\frac{1}{2}}[(2\pi m/\beta)^{\frac{1}{2}}(\tau/N)]^N \lim_{\epsilon\to 0}(|\mathbf{W}||\mathbf{A}|)^{-\frac{1}{2}} \times\exp\{\beta E-\beta\sum_{j=1}^{M}(P_j^2/2m)-\beta\sum_{i<j\leqslant M}w(\mathbf{R}_i-\mathbf{R}_j) -\tfrac{1}{2}(N/s)\beta\tilde{b}\mathbf{e}^{-\phi}+N\ln[1-(\epsilon/\tau)\textstyle\sum_{\mathbf{r}}(1-e^{-\phi_{\mathbf{r}}})] +\tfrac{1}{2}(N/s)\tilde{\phi}\mathbf{e}^{-\phi}\}. \quad (39)$$

In Eq. (39), a form like that given by the usual grand canonical ensemble theory is becoming apparent. The third term in the exponential contains the usual pair interaction energy while the fourth term contains the shielding by the particles in the outer region. The last two terms will reduce, in the case of fluids, to surface terms which can be neglected in the limit as τ_i becomes large. The parameter β occurs where we would expect to see $1/kT$, where k is the Boltzmann constant, and T is the temperature. Further evidence that β is $1/kT$ comes from solving Eqs. (6) and (23) to obtain

$$E=(3N/2\beta)+H_1+(N/2\beta s)(\beta\tilde{b}+\tilde{\phi})\mathbf{e}^{-\phi}. \quad (40)$$

If the interaction between particles vanishes, Eq. (24) shows that ϕ will vanish so that Eq. (40) reduces to

$$E\to(3N/2\beta)+\sum_{j=1}^{M}P_j^2/2m. \quad (41)$$

This is just the relation we would expect between E and β for an ideal gas.

V. SHIELDED PAIR POTENTIALS

There are pair potentials which are inconvenient to use because of a long tail on them. The Coulomb potential is a good example of this. For these cases it is possible to introduce a shielded potential similar to that obtained by Debye and Huckel for the Coulomb case. In terms of the matrix notation that we have been using, this shielded potential is given by the equation,

$$\mathbf{U}=\mathbf{W}[\dot{\mathbf{1}}+(N/s)\beta\dot{e}^{-\phi}\mathbf{W}]^{-1} =[\dot{\mathbf{1}}+(N/s)\beta\mathbf{W}\dot{e}^{-\phi}]^{-1}\mathbf{W}, \quad (42)$$

where the dot over a symbol indicates a diagonal matrix. The matrix $\dot{\mathbf{1}}$ is the unit square matrix and $\dot{e}^{-\phi}$ has the same elements as $\mathbf{e}^{-\phi}$ except that they are arranged along the diagonal instead of in a column. A second shielded potential $\mathbf{U}^*$ may be defined by replacing $\dot{e}^{-\phi}$ in Eq. (42) by $\dot{\mathbf{1}}$.

Since the matrices appearing on the right-hand side of Eq. (42) are symmetric, $\mathbf{U}$ and $\mathbf{U}^*$ are symmetric.

The $\mathbf{U}$ defined in Eq. (42) does not reduce to the Debye-Huckel shielded potential for the Coulomb case if τ_i is not zero. It can be found from an equation involving such a matrix $\mathbf{U}_t$ which is related to $\mathbf{W}$ by

$$\mathbf{U}_t=\mathbf{W}-(N/s)\beta(\mathbf{U}_t\mathbf{W})_t, \quad (43)$$

where the subscript t on the parenthesis indicates that the matrix multiplication involves a sum over the total volume of the system. The $\mathbf{U}_t$ may be used to find $\mathbf{U}$ from the equation,

$$\mathbf{U}=\mathbf{U}_t+(N/s)\beta(\mathbf{U}_t\mathbf{U})_t-(N/s)\beta\mathbf{U}_t\dot{e}^{-\phi}\mathbf{U}. \quad (44)$$

The substitution of $\mathbf{U}_t$ from Eq. (43) in place of the first and last $\mathbf{U}_t$'s in Eq. (44) verifies that $\mathbf{U}$ satisfies Eq. (42).

Equation (24) for ϕ may be rewritten in terms of $\mathbf{U}$ instead of $\mathbf{W}$ to give

$$\phi=\beta\mathbf{b}_u+(N/s)\beta\mathbf{U}^*[\mathbf{e}^{-\phi}+\phi], \quad (45)$$

where $\mathbf{b}_u$ is a column matrix given by

$$\mathbf{b}_u=[\dot{\mathbf{1}}+(N/s)\beta\mathbf{W}]^{-1}\mathbf{b}. \quad (46)$$

VI. THE LIMIT AS ε APPROACHES ZERO

At this point it is convenient to examine the above equations in the limit as ϵ approaches zero. In this limit the matrix $\mathbf{W}$ reduces to the pair potential w except for the diagonal elements omitted from $\mathbf{W}$. Thus we may write

$$W_{\mathbf{rq}}\to w(\mathbf{R}-\mathbf{Q})-\delta(\mathbf{R}-\mathbf{Q}) \times\lim\left[\epsilon^{-1}\int_{\text{cell}}\int w(\mathbf{R},\mathbf{Q})d\mathbf{R}d\mathbf{Q}\right], \quad (47)$$

where the integrations are over the cells centered at $\mathbf{r}$ and $\mathbf{q}$ and where the Dirac delta function is recognized to be related to the Kronecker delta by

$$\delta(\mathbf{R}-\mathbf{Q})=\lim_{\epsilon\to 0}\delta_{\mathbf{rq}}/\epsilon. \quad (48)$$

The coefficient of the Dirac delta function in Eq. (47) will vanish as ϵ vanishes so that, for most cases, this term can be ignored.

Also, in this limit Eq. (35) reduces to

$$s=\lim_{\epsilon\to 0}(\tau/\epsilon)\left[1-\tau^{-1}\int\rho(\mathbf{R})d\mathbf{R}\right], \quad (35')$$

where

$$\rho(\mathbf{R})=1-\exp\{-\phi(\mathbf{R})\}. \quad (49)$$

It will become clear later on [see discussion following Eq. (54)] that these terms in $\phi(\mathbf{R})$, whose magnitudes are of the order τ^0, behave like short ranged shielded potentials. There are, in general, terms of order τ^{-1} also present in $\phi(\mathbf{R})$. We shall also see that $\phi(\mathbf{R})$ in the outer region is large near the surface between the two regions but falls off to a negligible value in a distance the order of the average interparticle spacing. Since $\rho(\mathbf{R})$ will behave in a similar manner, the integral in

Eq. (35′) will have a magnitude the order of the area of the inter-region surface measured in units of the average interparticle spacing. Since τ is the volume of the outer region and is very large, we may neglect the second term in the bracket in Eq. (35′) altogether.

Substituting only the first term of Eq. (35′) into Eq. (43) for U_t and taking the limit gives

$$U_t(\mathbf{R}-\mathbf{Q})=w(\mathbf{R}-\mathbf{Q})-(N/\tau)\beta\int_t U_t(\mathbf{R}-\mathbf{R}')\times w(\mathbf{R}'-\mathbf{Q})d\mathbf{R}'. \quad (50)$$

For the Coulomb case, the solution to Eq. (50) is the Debye shielded potential. We expect, in general, that the term in U_t whose magnitude is of order τ^0 will behave like a short range shielded potential.

Equation (44) becomes, in this limit,

$$U(\mathbf{R},\mathbf{Q})=U_t(\mathbf{R}-\mathbf{Q})+(N/\tau)\beta\int_i U_t(\mathbf{R}-\mathbf{R}')\times U(\mathbf{R}',\mathbf{Q})d\mathbf{R}'+(N/\tau)\beta\int U_t(\mathbf{R}-\mathbf{R}')\times\rho(\mathbf{R}')U(\mathbf{R}',\mathbf{Q})d\mathbf{R}'. \quad (51)$$

We have already seen that $U_t(\mathbf{R}-\mathbf{Q})$ is likely to have a short range (the Debye length for the Coulomb case). The second term on the right-hand side (rhs) of Eq. (51) must become small if $\mathbf{R}$ is very far from the inner region because $U_t(\mathbf{R}-\mathbf{R}')$ will vanish. It will become clear in the next paragraph that $\phi(\mathbf{R})$ must become small if $\mathbf{R}$ is very far from the inner region. Since $U_t(\mathbf{R}-\mathbf{R}')$ is small if $\mathbf{R}$ is very far from $\mathbf{R}'$, the last term on the rhs of Eq. (51) must vanish if $\mathbf{R}$ is very far from the inner region. Since $U(\mathbf{R},\mathbf{Q})$ is symmetric in $\mathbf{R}$ and $\mathbf{Q}$, the above statements must also hold for $\mathbf{Q}$. We conclude, then that $U(\mathbf{R},\mathbf{Q})$ becomes equal to $U_t(\mathbf{R}-\mathbf{Q})$ for $\mathbf{R}$ or $\mathbf{Q}$ very far from the inner region. It therefore has a range the order of unity when compared to $\tau^{\frac{1}{3}}$ if it is correct to assume that $\phi(\mathbf{R})$ has a short range. Integrating $\mathbf{R}$ in Eq. (50) over the total volume of the system gives

$$\int_t U_t(\mathbf{R}-\mathbf{Q})d\mathbf{R}=0 \quad (52)$$

because the integral of w has been assumed to vanish. The same integration may be performed in Eq. (51) to show that the corresponding integral for U will also vanish.

Equation (45) for ϕ becomes, in the limit as $\epsilon\to 0$,

$$\phi(\mathbf{R})=\beta\sum_{j=1}^{M} U^*(\mathbf{R},\mathbf{R}_j)+(N/\tau)\beta\int U^*(\mathbf{R},\mathbf{Q})\times[e^{-\phi}+\phi(\mathbf{Q})]d\mathbf{Q}, \quad (53)$$

where the definition of b_τ following Eq. (10) has been used together with Eq. (42). Equation (52) may now be used to rewrite Eq. (53) as

$$\phi(\mathbf{R})=\beta\Big[\sum_{j=1}^{M} U^*(\mathbf{R}-\mathbf{R}_j)-(N/\tau)\int_i U^*(\mathbf{R},\mathbf{Q})d\mathbf{Q}\Big]+(N/\tau)\beta\int U^*(\mathbf{R},\mathbf{Q})[\phi(\mathbf{Q})-\rho(\mathbf{Q})]d\mathbf{Q}. \quad (54)$$

The arguments made above for the short range nature of other functions may now be made for ϕ in Eq. (54), and it is clear that the assumption of the short range nature of ϕ is consistent with this equation. In the outer region, $\phi(\mathbf{R})$ falls off to a very small value in some distance corresponding to a few times the Debye length in a Coulomb system.

We are now in a position to show that ϕ satisfies the Debye-Huckel-Poisson-Boltzmann equation[8] in the Coulomb problem where the particles of the system all have like charges of magnitude z, but move in a uniform background charge of the opposite sign so that the whole system is neutral. This can most easily be done starting from Eq. (24) which becomes

$$\phi(\mathbf{R})=\beta\sum_{j=1}^{M} w(\mathbf{R}-\mathbf{R}_j)-(N/\tau)\beta\int_i w(\mathbf{R}-\mathbf{Q})\rho(\mathbf{Q})d\mathbf{Q}-(N/\tau)\beta\int_i w(\mathbf{R}-\mathbf{Q})d\mathbf{Q}. \quad (55)$$

The pair potential satisfies Poisson's equation

$$\nabla^2 w(\mathbf{R}-\mathbf{Q})=-4\pi z\delta(\mathbf{R}-\mathbf{Q})+4\pi z/\tau. \quad (56)$$

Operating on Eq. (55) with ∇^2 gives

$$\nabla^2\phi(\mathbf{R})=-4\pi z\sum_{j=1}^{M}\delta(\mathbf{R}-\mathbf{R}_j)+4\pi z(N/\tau)\beta\rho(\mathbf{R}) \quad (57)$$

if $\mathbf{R}$ is in the outer region and terms of order τ^{-1} are neglected. In the inner region, the second term on the rhs has unity in place of $\rho(\mathbf{R})$. When M is unity, Eq. (57) reduces to the Debye-Huckel-Poisson-Boltzmann equation. This makes it even more reasonable to believe that ϕ is a shielded potential falling off rapidly as we depart from the surface of the inner region.

Since the integral of U^* over the entire volume of the system vanishes, the same integral of ϕ must also vanish. This can be proved by integrating both sides of Eq. (53) over the entire volume.

VII. THE GRAND CANONICAL DISTRIBUTION FUNCTION

So far the question of evaluating the determinants $|\mathbf{W}||\mathbf{A}|$ appearing in Eq. (39) has been avoided. The

[8] R. H. Fowler and E. A. Gugenheim, *Statistical Thermodynamics* (The Macmillan Company, New York, 1939), 1st ed., p. 390, Eq. (910,6).

procedure for dealing with it is so tedious that it seems wise to relegate it to Appendix II. From Eq. (II-33), we obtain,

$$|\mathbf{W}||\mathbf{A}| = N(5/4\beta^2)\exp\Big\{-(N/\tau)^2\beta^2\int_0^1\lambda \times\int\int U_{t\lambda}(\mathbf{R}-\mathbf{Q})w(\mathbf{Q}-\mathbf{R})d\mathbf{Q}d\mathbf{R}d\lambda\Big\}, \quad (58)$$

where $U_{t\lambda}$ is tiven by Eq. (50) except that (N/τ) is replaced by $\lambda(N/\tau)$.

The most interesting factors in $N^{-N}I$ are those depending upon the positions and momenta of particles in the inner region and the number of particles M in this region. The other factors may be omitted and the normalization determined later by integration. In this sense Eqs. (7) and (39) reduce to

$$P(\mathbf{R}_1,\mathbf{P}_1,\cdots\mathbf{R}_M\mathbf{P}_M)\propto\exp\Big\{-\gamma M-\beta H_1-\beta H_{01}' - (N_t/\tau)\int[1-e^{-\phi}-\tfrac{1}{2}\phi e^{-\phi}]d\mathbf{R}\Big\}, \quad (59)$$

$$\gamma=\ln((2\pi m/\beta)^{\frac{3}{2}}\tau/N_t)+(N_t/\tau)\beta^2 \times\int_0^1\lambda\int U_{t\lambda}(\mathbf{R})w(\mathbf{R})d\mathbf{R}d\lambda, \quad (60)$$

$$H_{01}'=\tfrac{1}{2}(N_t/\tau)\sum_{j=1}^{M}\int w(\mathbf{R}-\mathbf{R}_j)[1-\rho(\mathbf{R})]d\mathbf{R}.$$

Here, use has been made of the fact that the total number of particles in the system N_t is a constant of the order of τ and that $N=N_t-M$. Substitutions have also been made from Eqs. (1) and (58).

Equation (59) gives the desired probability distribution corresponding to that given by the grand canonical ensemble except for additional terms involving integrals over functions of ϕ and w. These terms are important if the inner region is small. In particular, the term H_{01}' adds shielding to the pair potentials due to the particles in the outer region.

It is interesting to see if Eq. (59) reduces to that given by the grand canonical ensemble in the limit of a large inner region. Equation (54) shows that $\phi(\mathbf{R})$ will depend only on the positions of particles in the inner region that lie within the range of U from $\mathbf{R}$. Thus all those integrals in Eq. (59), depending upon particle positions through ϕ alone, can only be affected by particles within a distance, of the order of the range of U, from the surface of the inner region. Thus, if the inner region is increased in size, this number of particles will increase like the surface area while other terms in Eq. (59) will involve all the particles throughout the volume of the inner region. Furthermore, $\phi(\mathbf{R})$ has a short range so that the integrands of these integrals extend into the outer region for only a short distance. Thus the value of the integrals will only increase as the surface area of the inner region, while other terms of Eq. (59) are proportional to the volume of the inner region.

These arguments do not apply so readily to H_{01}' in Eq. (60) unless w is a short ranged function. If, on the other hand, w is a Coulomb potential, the integral in H_{01}' is analogous to the potential acting on a particle due to a charge, proportional to $\rho(\mathbf{R})-1$, distributed through the outer region. If the inner region is spherically symmetric, this potential could be expected to be roughly constant for those particles far from the surface of the inner region. Retaining this shielding term but dropping those discussed in the previous paragraph, we can write for P in the limit of a large inner region.

$$P\propto\exp\{-\gamma M-\beta H_1-\beta H_{01}'\}. \quad (61)$$

This has almost the same form as Hill's[2] equation (6.6) for the grand canonical ensemble if allowance is made for differences in notation.

VIII. THE RADIAL DISTRIBUTION FUNCTION

An expression for the radial (or pair) distribution function may be obtained by choosing the inner region to be a long tube of infinitesimal cross section containing two particles. From Eqs. (59) and (60) it is possible to obtain the relation,

$$P(\mathbf{R}_1,\mathbf{P}_1,\mathbf{R}_2,\mathbf{P}_2) \propto\exp\Big\{-\beta\sum_{j=1}^{2}\Big[(P_j^2/2m) -\tfrac{1}{2}(N_t/\tau)\int w(\mathbf{R}-\mathbf{R}_j)\rho(\mathbf{R})d\mathbf{R}\Big]-\beta w(\mathbf{R}_1-\mathbf{R}_2) - (N_t/\tau)\int[1-e^{-\phi}-\tfrac{1}{2}\phi e^{-\phi}]d\mathbf{R}\Big\}. \quad (62)$$

The integrations in Eq. (62) now extend over almost the entire volume of the system.

In analogy with Eq. (55), we may define ϕ_1 by

$$\phi_1(\mathbf{R}_2-\mathbf{R}_1)=\beta w(\mathbf{R}_2-\mathbf{R}_1) - (N_t/\tau)\int w(\mathbf{R}-\mathbf{R}_2)\rho_1(\mathbf{R})d\mathbf{R}, \quad (63)$$

where

$$\rho_1(\mathbf{R})=1-\exp\{-\phi_1(\mathbf{R})\} \quad (64)$$

which is a shielded potential function at $\mathbf{R}_2$ due to the particle at $\mathbf{R}_1$. Because of symmetry we can write

$$\phi_1(\mathbf{R}_2-\mathbf{R}_1)=\phi_1(\mathbf{R}_1-\mathbf{R}_2). \quad (65)$$

With the aid of this function ϕ_1, Eq. (62) becomes

$$P(\mathbf{R}_1,\mathbf{P}_1,\mathbf{R}_2,\mathbf{P}_2) \propto \exp\left\{-\beta\sum_{j=1,2}(P_j^2/2m)-\phi_1(\mathbf{R}_2-\mathbf{R}_1) - (N/\tau)\int(1-e^{-\phi}-\tfrac{1}{2}\phi e^{-\phi})d\mathbf{R}\right\}. \quad (66)$$

It has already been shown in Eqs. (55)–(57) that ϕ_1 is the solution to the Poisson-Boltzmann equation used by Debye and Huckel.

In the Debye-Huckel theory, the exponential $e^{-\phi}$ is expanded and all terms beyond the first two are neglected. In Eq. (53) for ϕ, this is equivalent to neglecting the integral on the rhs. Under this approximation, ϕ is given by

$$\phi(\mathbf{R})\approx\beta\sum_{j=1,2}U^*(\mathbf{R},\mathbf{R}_j), \quad (67)$$

where U^* is the Debye-Huckel shielded potential. In the case where w is a Coulomb potential,

$$U^*_{\text{Coulomb}}(\mathbf{R})=w_{\text{Coulomb}}e^{-R/\lambda}, \quad (68)$$

where λ is the Debye shielding length.

If the exponentials in the integrand of the last term in the exponent in Eq. (66) are expanded to the second degree in ϕ, the integral vanishes since the integral over ϕ vanishes.

Thus Eq. (66) gives a radial distribution function identical to that obtained by Debye and Huckel.

IX. SUMMARY AND CONCLUSIONS

With the mathematical techniques presented here, we have derived an approximation to the probability distribution of particles in a region of a classical fluid of monatomic particles in thermal equilibrium. This distribution is given by Eq. (59) where ρ, U, ϕ, and γ are given by Eqs. (49)–(51), (55), and (60).

We have seen that this probability distribution reduces, in the limit of a large inner region, to Eq. (61) which is that given by the grand canonical ensemble. The additional term which represents shielding of the inner particles by the particles outside the region surely becomes negligible for short ranged pair potentials, and it seems quite likely that it is negligible when the interactions are Coulomb in the limit of large volume.

From this probability distribution we have derived the radial distribution function obtained by Debye and Huckel with a correction as shown in Eq. (66).

Since the probability distribution given in Eq. (59) is an approximate one, it is worthwhile considering how correction terms for it might be obtained. These correction terms are available from the exponential operator in Eq. (32) by expanding it in a power series where it is understood that terms where $(\partial/\partial\mathbf{A})$ occur to a nonintegral power must be set equal to zero. The first term in this expansion leads to Eq. (59).

Higher terms in Eq. (32) must bear some similarity to those evaluated in reference 5. Figure 5 of that reference presents information that indicates that higher terms increase in magnitude as the temperature and size of the inner region are decreased.

APPENDIX I—DERIVATIVES OF g

The quantity s is defined in Eq. (25) and related to g through Eq. (21). To obtain the derivatives of s entering Eqs. (18), (19), and (32), it is convenient to use the relations

$$\frac{\partial}{\partial l}=\frac{\partial\beta}{\partial l}\frac{\partial}{\partial\beta}+\sum_r\frac{\partial\phi_r}{\partial l}\frac{\partial}{\partial\phi_r} =i\left[\frac{\partial}{\partial\beta}+(2\beta)^{-1}(\tilde{\phi}+\beta\tilde{b})\frac{\partial}{\partial\phi}\right] \quad (I\text{-}1)$$

and

$$\frac{\partial}{\partial k_r}=\frac{\partial\beta}{\partial k_r}\frac{\partial}{\partial\beta}+\sum_q\frac{\partial\phi_q}{\partial k_r}\frac{\partial}{\partial\phi_q}=-i\beta^{\frac{1}{2}}\frac{\partial}{\partial\phi_r}. \quad (I\text{-}2)$$

The derivatives of β and ϕ in the above equations have been obtained from Eqs. (20) and (22). Here $\partial/\partial\phi$ has components $\partial/\partial\phi_r$.

Using these relations, we obtain for the first derivatives of s,

$$\partial s/\partial l=-i(2\beta)^{-1}(\tilde{\phi}+\beta\tilde{b})\mathbf{e}^{-\phi} \quad (I\text{-}3)$$

and

$$\partial s/\partial\mathbf{k}=i\beta^{\frac{1}{2}}\mathbf{e}^{-\phi}, \quad (I\text{-}4)$$

where Eq. (35) has been used for s.

Applying Eqs. (I-1) and (I-2) a second time gives

$$\partial^2 s/\partial l^2=(4\beta^2)^{-1}\{-\tilde{\phi}+\beta\tilde{b}+(\tilde{\phi}+\beta\tilde{b})(\dot{\phi}+\beta\dot{b})\}\mathbf{e}^{-\phi}, \quad (I\text{-}5)$$

$$(\partial/\partial l)(\partial/\partial\mathbf{k})s=(2\beta^{\frac{1}{2}})^{-1}(\dot{\phi}+\beta\dot{b}-\dot{1})\mathbf{e}^{-\phi}, \quad (I\text{-}6)$$

and

$$(\partial/\partial\mathbf{k})\times(\partial/\partial\mathbf{k})s=-\beta\dot{e}^{-\phi}. \quad (I\text{-}7)$$

APPENDIX II—DETERMINANT OF A

The matrix $\mathbf{A}$ is defined in Eq. (28). To compute its elements, we need the second derivatives of f. The first derivatives have already been written as Eqs. (18) and (19). From these we obtain, by differentiation,

$$A_{ll}=-\tfrac{3}{2}Nl_0^{-2}-Ng_0^{-1}\left(\frac{\partial^2 g}{\partial l^2}\right)_0+Ng_0^{-2}\left(\frac{\partial g}{\partial l}\right)_0^2, \quad (II\text{-}1)$$

$$\mathbf{a}=-Ng_0^{-1}\left(\frac{\partial^2 g}{\partial l\partial\mathbf{k}}\right)_0+Ng_0^{-2}\left(\frac{\partial g}{\partial l}\right)_0\left(\frac{\partial g}{\partial\mathbf{k}}\right)_0, \quad (II\text{-}2)$$

and

$$\mathbf{B}=\mathbf{W}^{-1}-Ng_0^{-1}\left(\frac{\partial}{\partial\mathbf{k}}\times\frac{\partial}{\partial\mathbf{k}}g\right)_0 +Ng_0^{-2}\left(\frac{\partial g}{\partial\mathbf{k}}\right)_0\times\left(\frac{\partial g}{\partial\mathbf{k}}\right)_0, \quad (II\text{-}3)$$

where the element $B_{r,q}$ is $A_{r,q}$ and $a_r=A_{lr}$.

We may now use Eqs. (20) and (21) to eliminate l_0 and g_0 and the formulas in Appendix I to evaluate the derivatives. The results may be more conveniently written in terms of the nine quantities,

$$\begin{aligned}
Q_1&=N\beta s^{-2}\tilde{e}^{-\phi}\mathbf{U}\mathbf{e}^{-\phi},\\
Q_2&=s^{-1}(\tilde{\phi}+\beta\tilde{b})\mathbf{e}^{-\phi}-1,\\
Q_3&=s^{-1}(\tilde{\phi}-\beta\tilde{b})\mathbf{e}^{-\phi},\\
Q_4&=s^{-1}(\tilde{\phi}+\beta\tilde{b})(\phi+\beta\dot{b})\mathbf{e}^{-\phi},\\
Q_5&=N\beta s^{-2}\tilde{e}^{-\phi}\mathbf{U}(\phi+\beta\dot{b})\mathbf{e}^{-\phi},\\
Q_6&=N\beta s^{-2}\tilde{e}^{-\phi}(\phi+\beta\dot{b})\mathbf{U}(\phi+\beta\dot{b})\mathbf{e}^{-\phi},\\
Q_7&=(N^2/4\beta s^3)\tilde{e}^{-\phi}\mathbf{U}\dot{e}^{-2\phi}\mathbf{U}\mathbf{e}^{-\phi},\\
Q_8&=(N^2/4\beta s^3)\tilde{e}^{-\phi}\mathbf{U}(\phi+\beta\dot{b})\dot{e}^{-2\phi}\mathbf{U}\mathbf{e}^{-\phi},
\end{aligned}\tag{II-4}$$

and

$$Q_9=(N^2/4\beta s^3)\tilde{e}^{-\phi}\mathbf{U}(\phi+\beta\dot{b})^2\dot{e}^{-2\phi}\mathbf{U}\mathbf{e}^{-\phi},$$

where $\mathbf{U}$ is defined by Eq. (42). We can now write,

$$A_{ll}=\tfrac{3}{2}N\beta^{-2}+(N/4\beta^2)[Q_4-Q_3-(Q_2+1)^2],\tag{II-5}$$

$$\mathbf{a}=-(N/2\beta^{\frac{1}{2}}s)(\phi+\beta\dot{b}+Q_2\dot{\mathbf{i}})\mathbf{e}^{-\phi},\tag{II-6}$$

and

$$\mathbf{B}=\mathbf{W}^{-1}+Ns^{-1}\beta\dot{e}^{-\phi}-N\beta s^{-2}\mathbf{e}^{-\phi}\times\mathbf{e}^{-\phi}.\tag{II-7}$$

By expanding $|\mathbf{A}|$ in terms of the elements of the first column and their cofactors, we have,

$$|\mathbf{A}|=|\mathbf{B}|[A_{ll}-\tilde{a}\mathbf{B}^{-1}\mathbf{a}+Tr(\dot{a}^2\mathbf{B}^{-1})],\tag{II-8}$$

where

$$B_{r,q}^{-1}=|\mathbf{B}|^{-1}\times\text{cofactor of } B_{r,q}\tag{II-9}$$

has been used.

From Eq. (42), it is clear that the first two terms of Eq. (II-7) combine to give

$$\mathbf{B}=\mathbf{U}^{-1}-N\beta s^{-2}\mathbf{e}^{-\phi}\times\mathbf{e}^{-\phi}.\tag{II-10}$$

If we factor $\mathbf{U}^{-1}$ from the expression on the right of Eq. (II-10) and take the reciprocal, we obtain

$$\mathbf{B}^{-1}=[\dot{\mathbf{i}}-N\beta s^{-2}(\mathbf{U}\mathbf{e}^{-\phi})\times\mathbf{e}^{-\phi}]^{-1}\mathbf{U}.\tag{II-11}$$

Expanding the bracket in a power series and multiplying by the $\mathbf{U}$ gives

$$\mathbf{B}^{-1}=\mathbf{U}(\dot{\mathbf{i}}+\mathbf{K}),\tag{II-12}$$

where

$$\mathbf{K}=N\beta s^{-2}(1-Q_1)^{-1}\mathbf{e}^{-\phi}\times(\mathbf{U}\mathbf{e}^{-\phi}).\tag{II-13}$$

To obtain $|\mathbf{B}^{-1}|$, we use the expression

$$\ln|\dot{\mathbf{i}}+\mathbf{K}|=-\sum_{j=1}^{\infty}j^{-1}\mathrm{Tr}(-\mathbf{K})^j\tag{II-14}$$

which can be obtained from the equations on page 121 of reference 9. Raising $\mathbf{K}$ to the jth power gives

$$\begin{aligned}\mathbf{K}^j&=[N\beta s^{-2}(1-Q_1)^{-1}]^j\tilde{e}^{-\phi}\times(\tilde{e}^{-\phi}\mathbf{U}\mathbf{e}^{-\phi})^{j-1}\mathbf{U}\mathbf{e}^{-\phi}\\&=KQ_1{}^{j-1}(1-Q_1)^{-j+1}.\end{aligned}\tag{II-15}$$

[9] W. V. Lovitt, *Linear Integral Equations* (Dover Publications, Inc., New York, 1950).

From Eqs. (II-4) and (II-13), we obtain

$$\mathrm{Tr}\mathbf{K}=Q_1(1-Q_1)^{-1}.\tag{II-16}$$

Thus we have finally, from Eqs. (II-12) and (II-14)–(II-16),

$$|\mathbf{B}^{-1}|=|\mathbf{U}|(1-Q_1)^{-1}.\tag{II-17}$$

To write out $|\mathbf{A}|$, we still must find $\tilde{a}\mathbf{B}\mathbf{a}-Tr(\dot{a}^2\mathbf{B}^{-1})$. The second term subtracts those elements of the first term involving the diagonal elements of $\mathbf{B}^{-1}$. Thus, if we define $\mathbf{U}_x$ to be the same as $\mathbf{U}$ except that its diagonal elements are zero, we may write,

$$\begin{aligned}\tilde{a}\mathbf{B}^{-1}\mathbf{a}-Tr(\dot{a}^2\mathbf{B}^{-1})&=\tilde{a}\mathbf{U}_x\mathbf{a}+N\beta s^{-2}(1-Q_1)^{-1}\\&\times[(\tilde{a}\mathbf{U}\mathbf{e}^{-\phi})^2-\tilde{e}^{-\phi}\mathbf{U}\dot{a}^2\mathbf{U}\mathbf{e}^{-\phi}]\end{aligned}\tag{II-18}$$

where Eqs. (II-12) and (II-13) have been used.

Taking $\mathbf{a}$ from Eq. (II-6), we obtain

$$\tilde{a}\mathbf{U}_x\mathbf{a}=(N/4\beta^2)(Q_{1x}Q_2{}^2+2Q_{5x}Q_2+Q_{6x}),\tag{II-19}$$

where the x subscripts indicate that $\mathbf{U}_x$ replaces $\mathbf{U}$. Similarly we find

$$\tilde{a}\mathbf{U}\mathbf{e}^{-\phi}=(s/2\beta^{\frac{1}{2}})(Q_5+Q_1Q_2)\tag{II-20}$$

and

$$\tilde{e}^{-\phi}\mathbf{U}\dot{a}^2\mathbf{U}\mathbf{e}^{-\phi}=s(Q_7Q_2{}^2+2Q_2Q_8+Q_9).\tag{II-21}$$

Collecting the expressions in Eqs. (II-5), (II-17), (II-18), (II-19), (II-20), and (II-21) and substituting them into Eq. (II-8) we have,

$$\begin{aligned}|\mathbf{A}|=(N/4\beta^2)|\mathbf{U}|^{-1}(1-Q_1)\{&6+[Q_4-Q_3-(Q_2+1)^2]\\&-(Q_{1x}Q_2{}^2+2Q_{5x}Q_2+Q_{6x})-(1-Q_1)^{-1}\\&\times[(Q_5+Q_1Q_2)^2\\&-4s^{-1}\beta^3(Q_7Q_2{}^2+2Q_2Q_8+Q_9)]\}.\end{aligned}\tag{II-22}$$

The expression in Eq. (II-22) is very long. To reduce it, let us consider the order of the terms between the braces under the assumption that U and ϕ are shielded potentials of ranges not far different from the average interparticle spacing. Let us further assume that the interparticle spacing is of order unity in the units we have taken and that the inner region is only a few interparticle spacings across. Thus we shall consider an integral over U, ϕ, or $\rho=1-e^{-\phi}$ to be of order unity.

In the limit as ϵ approaches zero, the expression for Q_1 in Eq. (II-4) becomes

$$\begin{aligned}Q_1=\beta N\tau^{-2}\int\int&[1-\rho(\mathbf{R})]\\&\times U(\mathbf{R},\mathbf{Q})[1-\rho(\mathbf{Q})]d\mathbf{R}d\mathbf{Q}.\end{aligned}\tag{II-23}$$

Since the integral of U over the entire volume of the

system is zero, Eq. (II-23) may be rewritten as

$$Q_1=\beta(N/\tau)\tau^{-1}\Big\{\int_i\int_i U(\mathbf{R},\mathbf{Q})d\mathbf{R}d\mathbf{Q}$$
$$+2\int\rho(\mathbf{R})\int_i U(\mathbf{R},\mathbf{Q})d\mathbf{R}d\mathbf{Q}$$
$$+\int\int\rho(\mathbf{R})U(\mathbf{R},\mathbf{Q})\rho(\mathbf{Q})d\mathbf{R}d\mathbf{Q}\Big\}. \quad \text{(II-24)}$$

Each of these integrals must be of order unity since $\rho(\mathbf{R})$ becomes negligible within a range of order unity of the surface of the inner region. Thus Q_1 is of order τ^{-1}. By this same procedure it is possible to prove that Q_2+1 and Q_3 through Q_9 in Eq. (II-4) are of order τ^{-1}.

In order to prove that some of the above quantities are of order τ^{-1}, it was necessary to assume that w becomes negligible in a distance the order of unity. This is not true for the Coulomb potential. In this case integrals, assumed to be of order unity, may be of order $\tau^{\frac{1}{2}}$ but this will not alter the evaluation of $|\mathbf{A}|$ that will be made below.

Some of the Q's appearing in Eq. (II-22) have the subscript x indicating that U_x replaces U. Since we know

$$U_{xrq}=U_{rq}-U_{rq}\delta_{r,q}, \quad \text{(II-25)}$$

in the limit as ϵ approaches zero, we must have

$$U_x(\mathbf{R},\mathbf{Q})=U(\mathbf{R},\mathbf{Q})-\epsilon U(\mathbf{R},\mathbf{Q})\delta(\mathbf{R}-\mathbf{Q}). \quad \text{(II-26)}$$

Thus, in any integration of $U_x(\mathbf{R}_1,\mathbf{R}_2)$ over $\mathbf{R}_1$ or $\mathbf{R}_2$, we will obtain one integral involving $U(\mathbf{R}_1,\mathbf{R}_2)$ and a second one multiplied by ϵ which vanishes. We can, for this reason, drop the x subscripts in Eq. (II-22).

We are now in a position to eliminate several terms in Eq. (II-22). Since we have been taking s^{-1} to be (ϵ/τ), the term containing it in Eq. (II-22) will vanish as ϵ approaches zero. Keeping only terms of lowest order in τ^{-1}, we may write,

$$|\mathbf{A}|=|\mathbf{U}|^{-1}(5/4)N\beta^{-2}. \quad \text{(II-27)}$$

To determine $|\mathbf{U}|^{-1}$ in Eq. (II-27), we use Eq. (42) to obtain

$$|\mathbf{U}|^{-1}=|\mathbf{W}|^{-1}|\mathbf{i}+(N/s)\beta\mathbf{W}e^{-\phi}|. \quad \text{(II-28)}$$

The second determinant on the rhs may be evaluated with the aid of a matrix $\mathbf{U}_\lambda$ given by

$$\mathbf{U}_\lambda=\mathbf{W}-\lambda(N/s)\beta\mathbf{U}_\lambda e^{-\phi}\mathbf{W}. \quad \text{(II-29)}$$

This determinant may be written as

$$\ln|\mathbf{i}+(N/s)\beta\mathbf{W}e^{-\phi}|=(N/s)\beta\int_0^1(Tr\,\mathbf{U}_\lambda e^{-\phi})d\lambda. \quad \text{(II-30)}$$

Equation (II-30) may be checked by iterating Eq. (II-29) to obtain a series very much like Eq. (II-14).

Substituting $\mathbf{U}_\lambda$ from Eq. (II-29) into the trace in Eq. (II-30) gives

$$\mathrm{Tr}(\mathbf{U}_\lambda e^{-\phi})=-\lambda(N/s)\beta\,\mathrm{Tr}(\mathbf{U}_\lambda e^{-\phi}We^{-\phi}). \quad \text{(II-31)}$$

The first term on the rhs of Eq. (II-29) contributes nothing since the diagonal elements of $\mathbf{W}$ are zero. In the limit as ϵ approaches zero, Eq. (II-30) becomes, substituting from Eq. (II-31),

$$\lim_{\epsilon\to 0}\ln|\mathbf{i}+(N/s)\beta\mathbf{W}e^{-\phi}|$$
$$=(N/\tau)^2\beta^2\int_0^1\lambda\int\int U_\lambda(\mathbf{R},\mathbf{Q})W(\mathbf{Q}-\mathbf{R})d\mathbf{R}d\mathbf{Q}$$
$$-2(N/\tau)^2\beta^2\int_0^1\lambda\int\int U_\lambda(\mathbf{R},\mathbf{Q})\rho(\mathbf{Q})W(\mathbf{Q}-\mathbf{R})d\mathbf{R}d\mathbf{Q}$$
$$-(N/\tau)^2\beta^2\int_0^1\lambda\int\int U_\lambda(\mathbf{R},\mathbf{Q})\rho(\mathbf{Q})$$
$$\times W(\mathbf{Q}-\mathbf{R})\rho(\mathbf{R})d\mathbf{R}d\mathbf{Q}. \quad \text{(II-32)}$$

The last two terms in Eq. (II-32) are smaller by a factor the order of τ^{-1} than the other terms since $\rho(\mathbf{Q})$ becomes negligible if $\mathbf{Q}$ is far from the inner region. We will neglect them for this reason.

Equation (II-29) for $\mathbf{U}_\lambda$ can be obtained from Eq. (42) for $\mathbf{U}$ except for the λ multiplying (N/s). Thus $\mathbf{U}_\lambda$ is identical to $\mathbf{U}$ if the number of particles is reduced by the factor λ. A function $U_{t\lambda}$ can be defined as the solution to Eq. (50) with N reduced by λ and Eq. (51) may be used to relate U_λ and $U_{t\lambda}$. Equation (51) can be used to justify using $U_{t\lambda}$ in Eq. (II-32) in place of U_λ since all but a negligible part of the integration occurs in a region where there is essentially no difference between them.

With the aid of Eqs. (II-28) and (II-32), it is possible to rewrite Eq. (II-27), omitting terms of higher order in τ^{-1}, as

$$|\mathbf{A}|=N^{\frac{5}{4}}\beta^{-2}|\mathbf{W}|^{-1}\exp\Big\{-(N\tau)^2\beta^2\int_0^1\lambda$$
$$\times\int\int U_{t\lambda}(\mathbf{R}-\mathbf{Q})W(\mathbf{Q}-\mathbf{R})d\mathbf{Q}d\mathbf{R}d\lambda\Big\}. \quad \text{(II-33)}$$

Note added in proof: The author is indebted to H. L. Sahlin for pointing out several errors in the original manuscript.

Note on the Albertoni-Bocchieri-Loinger Theorem of Classical Statistical Mechanics

J. S. LOMONT
Department of Mathematics, Polytechnic Institute of Brooklyn, Brooklyn, New York
(Received October 25, 1960)

The A-B-L theorem states that for almost all density functions the time-averaged probability of a system being in a region R of an energy shell in phase space is m_R/m, where m is the volume of the energy shell and m_R is the volume of R. In the present note a stronger form of the A-B-L theorem is proved. It is proved that for almost all density functions, the probability of the system being in R is m_R/m. In particular, it is proved that the time averaging of the original A-B-L theorem is unnecessary.

I. INTRODUCTION

ONE approach to classical equilibrium statistical mechanics is via ergodic theory. The object of ergodic theory is to relate observable time averages of phase functions to calculable phase averages of phase functions. The possibility of making this connection depends on a knowledge of the "metric indecomposability" of phase space, and since this knowledge is usually lacking, the ergodic approach is unsatisfactory.

Albertoni, Bocchieri, and Loinger[1] attempted to remedy this situation by giving a new kind of averaging theorem. They introduced an average over a function space, the space of density functions. The probability $P_R(t)$ of a system being in a certain region R of phase space at a certain time can be expressed in terms of the density function. Let R be a subset of volume m_R of the energy shell of volume m. The A-B-L theorem says that the functional average of the time average of $[P_R(t)-(m_R/m)]^2$ is zero. In other words, for "almost all" density functions the time-averaged probability of the system being in R is m_R/m. This is also the probability when the density function is a constant.

The object of this note is to show that the time averaging of the A-B-L theorem is irrelevant and unnecessary; i.e., that the functional average of $[P_R(t)-(m_R/m)]^2$ is zero. In other words, at all times, and for almost all density functions, the probability of the system being in R is m_R/m. Thus the theorem of this note is stronger than the A-B-L theorem.

Since no time average is involved, the use of this theorem in place of ergodic theory becomes rather unclear.

II. THE A-B-L AVERAGING PROCESS

The A-B-L averaging process will first be described in a slightly more general context than is necessary.[2]

Suppose we are given a space on which integration is defined (e.g., a real finite-dimensional Euclidean space), and that S is a subset of finite, nonzero volume m which can be partitioned into any finite number of subsets of equal volume. Let C be the class of all real nonnegative, integrable, point functions f on S which have

$$\int_S f=1. \tag{2.1}$$

Let Π be a partition of S into N subsets of equal volume, and let Φ_i be the characteristic function[3] of the ith subset. Then the function

$$g=\sum_{i=1}^{N} a_i\Phi_i, \tag{2.2}$$

where

$$\sum_{i=1}^{N} a_i=N/m \quad a_i\geqslant 0 \tag{2.3}$$

is in class C. The functions g belonging to a given partition Π form a class $C(\Pi)$.

If $F[f]$ is any functional defined on C, then

$$F[g]=\mathfrak{F}(a_1,a_2,\cdots,a_N). \tag{2.4}$$

An average of F over $C(\Pi)$ can now be defined. If T is the section of the hyperplane

$$\sum_{i=1}^{N} a_i=N/m \tag{2.5}$$

in N-dimensional a space determined by

$$a_i\geqslant 0 \quad (i=1,2,\cdots,N), \tag{2.6}$$

and if dT is the Euclidean volume element of T, then the A-B-L average of F over $C(\Pi)$ is

$$B^{(\Pi)}[F]=\frac{\int_T d\mathfrak{F}(a_1,a_2,\cdots,a_N)}{\int_T dT} \tag{2.7}$$

if the integral exists. The A-B-L average of F over C is

[1] S. Albertoni, P. Bocchieri, and A. Loinger, J. Math. Phys. 1, 244 (1960).

[2] A similar integration process was discussed by K. O. Friedrichs, "Integration of functionals" (Notes, New York University, 1957).

[3] $\Phi_i=1$ on the ith subset, $=0$ elsewhere.

defined by

$$B[F]=\lim_{N\to\infty} B^{(\Pi)}[F]$$

$$=\lim_{N\to\infty}\frac{\int_T d\mathfrak{F}(a_1,a_2,\cdots,a_N)}{\int_T dT}.$$

if the limit exists.[4]

It may be noted that $B[cF]=cB[F]$ if c is a constant, and $B[F_1+F_2]=B[F_1]+B[F_2]$. Also if $F[f]\equiv 1$ then $B[F]=1$.

III. CALCULATION OF A-B-L AVERAGES

It was shown by A-B-L[1] that

$$\int_T dT=\frac{N^{\frac{1}{2}}}{(N-1)!}\left(\frac{N}{m}\right)^{N-1}, \tag{3.1}$$

$$\frac{\int_T dT a_i}{\int_T dT}=\frac{1}{m}, \tag{3.2}$$

$$\frac{\int_T dT a_i^2}{\int_T dT}=\frac{N^2}{N(N+1)}\frac{2}{m^2}, \tag{3.3}$$

$$\frac{\int_T dT a_i a_j}{\int_T dT}=\frac{N^2}{N(N+1)}\frac{1}{m^2}. \tag{3.4}$$

Let R be any subset of S, let Φ_R be the characteristic function of R, and let h be a bounded ($|h|\leqslant\bar{h}$), complex, function on S with integrable real and imaginary parts. We shall now evaluate the A-B-L average of the functional

$$F[f]=\int_S h\Phi_R f. \tag{3.5}$$

We have

$$\mathfrak{F}(a_1,a_2,\cdots,a_N)=F[g]=\sum_{i=1}^{N} a_i\int_S h\Phi_R\Phi_i \tag{3.6}$$

so

$$B^{(\Pi)}[F]=\frac{1}{m}\sum_{i=1}^{N}\int h\Phi_R\Phi_i$$

$$=\frac{1}{m}\int_S h\Phi_R. \tag{3.7}$$

[4] It should be noted that there may be many partitions belonging to a given N and that $\mathfrak{F}$ depends on Π as well as on N.

Hence,

$$B[F]=\frac{1}{m}\int_S h\Phi_R \tag{3.8}$$

when F has the form (3.5).

Now let

$$F[f]=\left\{\int_S h\Phi_R f\right\}^2. \tag{3.9}$$

We have

$$\mathfrak{F}(a_1,a_2,\cdots,a_N)=F[g]$$

$$=\left\{\sum_{i=1}^{N} a_i\int_S h\Phi_R\Phi_i\right\}^2. \tag{3.10}$$

Let

$$h_{Ri}=\int_S h\Phi_R\Phi_i, \tag{3.11}$$

$$m_{Ri}=\int_S \Phi_R\Phi_i. \tag{3.12}$$

Then

$$\mathfrak{F}(a_1,a_2,\cdots,a_N)=2\sum_{i>j} a_i a_j h_{Ri}h_{Rj}+\sum_i a_i^2 h_{Ri}^2, \tag{3.13}$$

so

$$B^{(\Pi)}[F]=\frac{N^2}{N(N+1)}\frac{2}{m^2}\sum_{i>j} h_{Ri}h_{Rj}$$

$$+\frac{N^2}{N(N+1)}\frac{2}{m^2}\sum_i h_{Ri}^2$$

$$=\frac{N^2}{N(N+1)}\frac{1}{m^2}\Big(\sum_i h_{Ri}\Big)^2$$

$$+\frac{N^2}{N(N+1)}\frac{1}{m^2}\sum_i h_{Ri}^2$$

$$=\frac{N^2}{N(N+1)}\frac{1}{m^2}\left(\int_S h\varphi_R\right)^2$$

$$+\frac{N^2}{N(N+1)}\frac{1}{m^2}\sum_i h_{Ri}^2. \tag{3.14}$$

If $\bar{h}$ is an upper bound of h then

$$|h_{Ri}|\leqslant\bar{h}m_{Ri}\leqslant\bar{h}m/N. \tag{3.15}$$

Hence,

$$\Big|\sum_{i=1}^{N} h_{Ri}^2\Big|\leqslant\frac{\bar{h}^2m^2}{N^2}\sum_{i=1}^{N}=\frac{\bar{h}^2m^2}{N}. \tag{3.16}$$

Thus the last term of (3.14) vanishes in the limit, and we see that when F has the form (3.9)

$$B[F]=\left(\frac{1}{m}\int_S h\Phi_R\right)^2. \tag{3.17}$$

IV. THE A-B-L ENSEMBLE THEOREM

We shall now let the space be a finite-dimensional phase space (of a conservative classical system of point masses), and the subset S be an energy shell in the phase space. The class C will be the class of density functions at $t=0$, and R will be a subset of the energy shell with positive volume. We shall assume the existence of a unitary operator $U(t)$ which describes the evolution of the system with time, and which operates in the Hilbert space of square integrable functions on the phase space.

The probability $P_R(t)$ of a system being in R at time t is

$$P_R(t)=\int_S \Phi_R\rho(t)$$

$$=\int_S \Phi_R U(t)\rho(0)$$

$$=\int_S \rho(0)U^{-1}(t)\Phi_R. \tag{4.1}$$

If we now consider $P_R(t)$ as a functional of $\rho(0)$ and apply the A-B-L averaging process we have by (3.8) [letting $h=U^{-1}(t)\Phi_R$, and replacing Φ_R in (3.8) by Φ, the characteristic function of S]

$$B[P_R(t)]=\frac{1}{m}\int_S \Phi U^{-1}(t)\Phi_R$$

$$=\frac{1}{m}\int_S \Phi_R U(t)\Phi. \tag{4.2}$$

Since[5]

$$U(t)\Phi=\Phi, \tag{4.3}$$

we have

$$B[P_R(t)]=m_R/m, \tag{4.4}$$

where m_R is the volume of R. Similarly[6] (using 3.17),

$$B[P_R{}^2(t)]=B\left[\left(\int_S \rho(0)U^{-1}(t)\Phi_R\right)^2\right]$$

$$=(m_R/m)^2. \tag{4.5}$$

Hence, we have the result

$$B[(P_R(t)-m_R/m)^2]=0. \tag{4.6}$$

Thus $P_R(t)=m_R/m$ for almost all initial density functions at all times. This is the strong form of the A-B-L theorem.

[5] Every function of the energy is an invariant eigenfunction of $U(t)$.

[6] If $h=U^{-1}(t)\Phi_R$ is not bounded the derivation of (3.17) can be altered slightly to take advantage of the properties of Hilbert space. Then

$$|h_{Ri}|=|\langle h|\Phi_R\Phi_i\rangle|\leqslant\|h\|\ \|\Phi_R\Phi_i\|=\|h\|m_{Ri}\leqslant\|h\|m/N.$$

Then $\bar{h}$ is replaced by $\|h\|$ and the proof goes through as before.

A Lattice with an Unusual Frequency Spectrum

ROBERT J. RUBIN AND ROBERT ZWANZIG
National Bureau of Standards, Washington, D. C.

(Received April 11, 1961)

The lattice is a special rooted Cayley tree, generated by N successive m-fold branchings. With each point of the tree are associated a mass M and a position coordinate x_i. All end points are held fixed at $x_i=0$. The potential energy is $V=\frac{1}{2}\sum_{i,j}K_{ij}(x_i-x_j)^2$, where $K_{ij}=K$ if i and j are connected neighbors and neither is an end point, $K_{ij}=\alpha K$ if i and j are connected neighbors and either is a branch tip point, and $K_{ij}=0$ if i and j are not connected neighbors. The allowed frequencies of vibration are obtained for two different cases: In the first case all springs are identical ($\alpha=1$), and in the second case the springs connecting interior points to the branch tips are cut ($\alpha=0$). In the case in which all force constants are the same, the allowed frequencies of vibration, in the limit of infinite N, are given by $\omega(r)=(K/M)^{\frac{1}{2}}[m+1-2m^{\frac{1}{2}}\cos r\pi]^{\frac{1}{2}}$, where r is any rational number between zero and one. The fraction of all normal modes having precisely the value $\omega(r)$ is $\rho[\omega(r)]=(m-1)^2/(m^q-1)$, where r is expressed as the ratio $r=p/q$ of relatively prime integers p and q. The frequency spectrum is dense within the interval $(m^{\frac{1}{2}}-1,\ m^{\frac{1}{2}}+1)$; and $\rho[\omega]$ is discontinuous at every ω for which it does not vanish.

PERHAPS the readers of this journal will be interested in some curious results we have obtained concerning the frequencies of small vibration of a number of mass points connected by springs as in a special rooted Cayley tree.[1] An example is shown in Fig. 1.

The Cayley tree that we are concerned with is made by starting with the trunk connecting the root point 0 to the point 1_1, and adding m branches to the point 1_1 to produce m new branch tip points $1_1 2_1$, $1_1 2_2$, $\cdots$, $1_1 2_m$. Then m branches are added to each of these branch tips to produce the m^2 new points $1_1 2_1 3_1$, $\cdots$, $1_1 2_1 3_m$, $1_1 2_2 3_1$, $\cdots$, $1_1 2_m 3_m$; and so on. This branching process, when repeated N times, results in an Nth-order tree whose m^N branch tip points are labeled $1_1 2_i \cdots (N+1)_j$. We shall refer to m as the branching number. Figure 1 shows a fourth-order tree with a branching number of two. We are concerned especially with an Nth-order tree with a branching number $m\geq 2$ as N becomes indefinitely large.

The root point 0 and the points at the branch tips are called end points, and the others are interior points. Each end point is connected to one neighboring point, and each interior point is connected to $m+1$ neighbors. Thus $m+1$ is the coordination number of the lattice.

It is interesting to note that an Nth-order tree has $(m^N-1)/(m-1)$ interior points and m^N+1 end points. For large N, the fraction of points on the surface of the tree, i.e., branch tip points, is $(m-1)/m$.

With each point of the tree we associate a mass M and a position coordinate x_i. All end points are held fixed to the values $x_i=0$. We suppose the potential energy of interaction of the mass points to be

$$V=\tfrac{1}{2}\sum_{i\neq j}\sum K_{ij}(x_i-x_j)^2 = \tfrac{1}{2}\sum_i\sum_j A(i,j)x_i x_j. \quad (1)$$

The coefficients K_{ij} are

$$K_{ij}=\begin{cases} K, & \text{if } i \text{ and } j \text{ are neighbors, and neither is a branch tip;} \\ \alpha K, & \text{if } i \text{ and } j \text{ are neighbors, and either is a branch tip;} \\ 0, & \text{if } i \text{ and } j \text{ are not neighbors.} \end{cases}$$

The coefficients $A(i,j)$ are determined by these K_{ij}.

By setting α equal to zero, the interior points are entirely disconnected from the branch tips, and the Nth-order tree with fixed tips becomes an $(N-1)$th-order tree with free tips. When α equals one, the interaction of interior points with branch tips is the same as the interaction of interior points with each other.

Our principal results, concerning the frequency spectrum of the tree in the case $\alpha=1$, are the following.

The allowed frequencies of vibration of the tree, in the limit of indefinitely large N, are given by

$$\omega(r)=(K/M)^{\frac{1}{2}}\{m+1-2m^{\frac{1}{2}}\cos r\pi\}^{\frac{1}{2}}, \quad (2)$$

where r is any rational number greater than zero and less than one. The frequency spectrum is dense within the interval

$$(K/M)^{\frac{1}{2}}\{m^{\frac{1}{2}}-1\}<\omega<(K/M)^{\frac{1}{2}}\{m^{\frac{1}{2}}+1\}. \quad (3)$$

The fraction of all normal modes having precisely the frequency $\omega(r)$ is

$$\rho[\omega(r)]=(m-1)^2/(m^q-1), \quad m\geq 2, \quad (4)$$

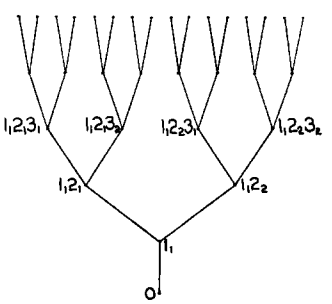

FIG. 1. A fourth-order tree with a branching number of two.

[1] This sort of lattice was brought to the attention of RJR by Dr. Michael Fisher. It is of interest in the study of cooperative phenomena and has been called a Bethe lattice by C. Domb, Phil. Mag. Suppl. 9, 149 (1960).

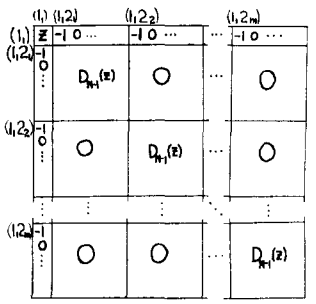

FIG. 2. The general structure of the secular determinant $D_N(z)$.

where we have supposed r to be expressed as the ratio

$$r=p/q \tag{5}$$

of relatively prime integers p and q. Within the stated interval, $\rho[\omega(r)]$ vanishes for irrational r, and is discontinuous at every rational r.

We proceed now to the derivation of the frequency spectrum. The allowed frequencies of vibration of coupled oscillators can be found by solving for the roots of a secular equation. This equation, for an Nth-order tree, is

$$D_N=\det[A(i,j)-M\omega^2\delta_{ij}]=0. \tag{6}$$

The order of the determinant D_N is $(m^N-1)/(m-1)\times(m^N-1)/(m-1)$. The potential energy matrix $A(i,j)$ is positive definite; therefore, the secular equation has $(m^N-1)/(m-1)$ positive roots, here denoted by ω^2.

For convenience we shall suppose that K and M are equal to one; this amounts to measuring frequency in units of $(K/M)^{\frac{1}{2}}$.

The general structure of the secular determinant is shown in Fig. 2. The z in the upper left corner is

$$z=m+1-\omega^2; \tag{7}$$

the entry $A(1_1,1_1)=m+1$ contains the contribution of the $m+1$ springs attached to the first branching point 1_1. This point is connected to 1_12_1, $\cdots$, 1_12_m, (and to no other points), by $A(1_1,1_12_1)=\cdots=A(1_1,1_12_m)=-1$. These elements comprise the nonzero entries in the top row and left column of $D_N(z)$ in Fig. 2.

In the secular determinant, the points 1_12_1, $\cdots$, 1_12_m now may be regarded as the first branching points of m independent trees of order $N-1$. Thus $D_N(z)$ is filled in, on the diagonal, with m smaller determinants $D_{N-1}(z)$. Each D_{N-1} has the same general structure as D_N, including D_2,

$$D_2=\begin{vmatrix} z & -1 & -1 & -1 & \cdots & -1 \\ -1 & D_1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & D_1 & 0 & \cdots & 0 \\ -1 & 0 & 0 & D_1 & \cdots & 0 \\ \vdots & & & & & \\ -1 & 0 & 0 & \cdots & & D_1 \end{vmatrix}. \tag{8}$$

The determinant D_1 consists of the single element

$$D_1=1+m\alpha-\omega^2. \tag{9}$$

If the springs leading to the branch tips are identical with the springs connecting interior points, then $D_1=z$.

By Laplace's expansion, one may easily verify that, for $N>2$,

$$D_N(z)=zD_{N-1}{}^m(z)-mD_{N-1}{}^{m-1}(z)D_{N-2}{}^m(z). \tag{10}$$

The recurrence formula (10) is correct for $N=2$ also, provided we define $D_0(z)$ to be identically one.

It is noteworthy that in the special case $m=1$, i.e., for the familiar one-dimensional chain, the recurrence formula reduces to the well-known linear form, $D_N(z)=zD_{N-1}(z)-D_{N-2}(z)$.

The general recurrence formula or difference equation (10), although nonlinear, turns out to have a remarkably simple solution. We observe first that, by its definition, $D_N(z)$ is a polynomial in z. From the recurrence formula we obtain

$$D_N(z)/D_{N-1}{}^{m-1}(z)=zD_{N-1}(z)-mD_{N-2}{}^m(z); \tag{11}$$

$D_N(z)/D_{N-1}{}^{m-1}(z)$ is also a polynomial in z. But if this is a polynomial, then so is $D_{N-1}(z)/D_{N-2}{}^{m-1}(z)$. Thus,

$$D_N(z)/[D_{N-1}(z)D_{N-2}(z)]^{m-1} = zD_{N-1}(z)/D_{N-2}{}^{m-1}(z)-mD_{N-2}(z) \tag{12}$$

is a polynomial; and so forth.

The preceding observation suggests the substitution

$$D_N(z)=\{D_1(z)D_2(z)\cdots D_{N-1}(z)\}^{m-1}P_N(z), \tag{13}$$

where $P_N(z)$ is a polynomial of the Nth degree in z. Substituting (13) into the recurrence formula (10), once for D_N, once for D_{N-1}, and once for D_{N-2}, we find

$$\begin{aligned}&\{D_1D_2\cdots D_{N-1}\}^{m-1}P_N \\ &\quad=zD_{N-1}{}^{m-1}\{D_1D_2\cdots D_{N-2}\}^{m-1}P_{N-1} \\ &\qquad-mD_{N-1}{}^{m-1}D_{N-2}{}^{m-1}\{D_1D_2\cdots D_{N-3}\}^{m-1}P_{N-2}.\end{aligned} \tag{14}$$

Evidently, $\{D_1D_2\cdots D_{N-1}\}^{m-1}$ can be factored from each term, leaving the *linear* recurrence formula

$$P_N(z)=zP_{N-1}(z)-mP_{N-2}(z). \tag{15}$$

Using the known values of D_0 and D_1, we observe that the P sequence can be started with

$$\begin{aligned}P_0&=D_0=1, \\ P_1&=D_1=z-m(1-\alpha).\end{aligned} \tag{16}$$

The structure of the recurrence formula for P_N suggests that a solution may be found by means of Tchebycheff polynomials. We have done so; the solution, easily verified by substitution, is

$$P_n(z)=m^{n/2}\left\{\frac{\sin(n+1)\theta}{\sin\theta}-m^{\frac{1}{2}}(1-\alpha)\frac{\sin n\theta}{\sin\theta}\right\}, \tag{17}$$

where θ is defined by

$$z=2m^{\frac{1}{2}}\cos\theta=m+1-\omega^2. \tag{18}$$

The relation between the determinant D_N and the

polynomials P_N can be found by repeated application of (13); it is

$$D_N(z)=P_N(z)\prod_{r=1}^{N-1}\{P_r(z)\}^{(m-1)mN-r-1}. \tag{6'}$$

This relation, together with (17) and (18), provides an exact evaluation of the secular determinant. (Even the linear case, $m=1$, in which the product over r in (6') reduces to unity, is included here.)

The problem of finding the distribution of the frequencies of vibration of the tree (for $m\geq 2$) is now reduced to (1) locating the zeros of the polynomials P_n, $n=1, 2, \cdots, N$, and (2) accounting for the multiplicity of these zeros, considered as roots of D_N. We shall first treat the case $\alpha=1$, in which the branch tip springs are identical with the other springs. Then we shall consider what modifications are called for when $\alpha=0$, in which case the interior points are completely detached from the branch tips.

When $\alpha=1$, the expression for P_n is simply

$$P_n=m^{n/2}\sin(n+1)\theta/\sin\theta. \tag{19}$$

The n zeros of P_n are located at

$$\theta_k^{(n)}=k\pi/(n+1), \quad 1\leq k\leq n.$$

It follows that the frequencies of vibration obtained from (18)

$$\omega=\{m+1-2m^{\frac{1}{2}}\cos\theta_k^{(n)}\}^{\frac{1}{2}}, \quad \begin{matrix}1\leq k\leq n,\\ 1\leq n\leq N,\end{matrix} \tag{20}$$

fill the interval $m^{\frac{1}{2}}-1<\omega<m^{\frac{1}{2}}+1$ in the limit of large N.

Now we consider the distribution of frequencies within this interval. In the limit $N\to\infty$, there is a frequency associated with every rational multiple of π,

$$\omega(r)=\{m+1-2m^{\frac{1}{2}}\cos r\pi\}^{\frac{1}{2}}, \tag{20'}$$

where $0<r<1$. By writing $\theta=r\pi$ as $\theta_{q-1}^{(p)}=p\pi/q$, where p and q are relatively prime integers, it is clear that $\theta_{q-1}^{(p)}=r\pi$ is a root of $P_{q-1}, P_{2q-1}, \cdots, P_{nq-1}, \cdots$. Since the total number of roots of D_N is $(m^N-1)/(m-1)$, the fraction of all frequencies having precisely the value $\omega(r)$ is

$$\rho[\omega(r);N]=\left(\frac{m^N-1}{m-1}\right)^{-1}\times\{(m-1)m^{N-q}+(m-1)m^{N-2q}+\cdots\}. \tag{21}$$

In the limit $N\to\infty$, the value of $\rho[\omega(r);N]$ is

$$\rho[\omega(r)]=\lim_{N\to\infty}\rho[\omega(r);N]=(m-1)^2/(m^q-1); \quad m\geq 2. \tag{22}$$

We shall refer to $\rho[\omega(r)]$ as the fractional multiplicity of the frequency $\omega(r)$.

The numerical value of $\rho[\omega(r)]$ is independent of p, except that p must be less than q and relatively prime to q. Then we also have $\rho[\omega(r)]=\rho[\omega(1-r)]$; the two frequencies $\omega(r)=(m+1-2m^{\frac{1}{2}}\cos r\pi)^{\frac{1}{2}}$ and $\omega(1-r)=(m+1+2m^{\frac{1}{2}}\cos r\pi)^{\frac{1}{2}}$ have the same fractional multiplicity. The frequency $\omega(\frac{1}{2})=(m+1)^{\frac{1}{2}}$, which is also the frequency of a tree of order one (and $\alpha=1$) has the largest multiplicity, namely

$$\rho[\omega(\tfrac{1}{2})]=(m-1)/(m+1). \tag{23}$$

Thus there is a band of allowed frequencies of vibration, specified by

$$m^{\frac{1}{2}}-1<\omega<m^{\frac{1}{2}}+1,$$

such that the frequency $\omega(r)=(m+1-2m^{\frac{1}{2}}\cos r\pi)^{\frac{1}{2}}$ within the band has a fractional multiplicity

$$\rho[\omega(r)]=\begin{cases}(m-1)^2/(m^q-1), & \text{if } 0<r<1 \text{ is rational and equal to } p/q \text{ where } p \text{ and } q \text{ are relatively prime integers, } p<q;\\ 0, & \text{if } 0<r<1 \text{ is irrational.}\end{cases}$$

We now consider the case $\alpha=0$, in which the interior points are not attached to the branch tips. The resulting transcendental equation for the roots of P_n is

$$P_n=m^{n/2}\left\{\frac{\sin(n+1)\theta}{\sin\theta}-m^{\frac{1}{2}}\frac{\sin n\theta}{\sin\theta}\right\}=0, \quad m\geq 2. \tag{24}$$

Although we have not been able to find explicit expressions for all the roots of all the P_n's defined by (24), we can nevertheless draw certain qualitative conclusions.

First, according to (16), we observe that $P_1=D_1=1-\omega^2$, or $\omega=1$ is a root of D_N.

Second, with one exception, it can be verified that for $n\geq 2$, P_n in (24) has $n-1$ real roots, θ_{nk}, and one imaginary root $i\phi_n$ (see Appendix for details). The single exception occurs for P_2 when $m=2$, in which case P_2 has two real roots θ_{21} and θ_{22}. The frequencies corresponding to the real roots fill the same open interval as before, $m^{\frac{1}{2}}-1<\omega<m^{\frac{1}{2}}+1$, in the limit of large N. The imaginary root $i\phi_n$ of (24), which is determined conveniently from the equivalent expression

$$e^{2n\phi}=(m^{\frac{1}{2}}-e^{-\phi})/(m^{\frac{1}{2}}-e^{\phi}) \tag{24'}$$

corresponds to a frequency

$$\omega(\phi_n)=(m+1-2m^{\frac{1}{2}}\cosh\phi_n)^{\frac{1}{2}}.$$

This frequency lies in the interval $0<\omega(\phi)<m^{\frac{1}{2}}-1$, $n\geq 2$.

From examination of (24'), it can be seen that the root ϕ_n is such that $1\leq e^{\phi_n}<m^{\frac{1}{2}}$. The ϕ_n's are distinct, with $\phi_n>\phi_m$ for $n>m$; and as $n\to\infty$, $e^{\phi_n}\to m^{\frac{1}{2}}$. The frequency associated with this limiting value is $\omega(\phi_\infty)=0$.

In this way we find, in the case $\alpha=0$, that the frequency spectrum consists of: (1) a band of frequencies in the interval $m^{\frac{1}{2}}-1<\omega<m^{\frac{1}{2}}+1$, and (2) a discrete spectrum in the gap interval $0<\omega<m^{\frac{1}{2}}-1$. The frequency $\omega=0$ is a limit point of the discrete spectrum. For $m\geq 5$, the frequency $\omega=1$ obtained from $P_1=0$ lies in the gap interval, and constitutes an upper bound for the discrete spectrum.

We have not been able to determine the fractional multiplicity of the frequencies in the band (for $\alpha=0$), i.e., to account for possible coincidences of the real roots of different P_n's. However, the fractional multiplicity of the frequencies in the discrete spectrum can be obtained easily. It is

$$\rho[\omega(\phi_n)]=\lim_{N\to\infty}\frac{(m-1)m^{N-1-n}}{(m^N-1)/(m-1)}=\frac{(m-1)^2}{m^{n+1}},\quad n\geq 2.$$

It follows also that the fraction σ of all frequencies lying in the open interval $0<\omega<m^{\frac{1}{2}}-1$ is

$$\sigma=\begin{cases}(m-1)/m^2, & m=2,3,4\\ (m-1)/m, & m\geq 5.\end{cases}$$

APPENDIX

The roots of

$$P_n(\theta)=m^{n/2}\{\sin(n+1)\theta/\sin\theta-m^{\frac{1}{2}}\sin n\theta/\sin\theta\}\text{ for }m\geq 2,$$

where $2m^{\frac{1}{2}}\cos\theta=m+1-\omega^2$.

$P_n(\theta)$ is a polynomial in ω^2 possessing n real roots ω_{nk}^2. The expression for $m^{-n/2}P_n(\theta)$ is the difference of two Tchebycheff polynomials, $\sin(n+1)\theta/\sin\theta$ and $\sin n\theta/\sin\theta$. Because the $n-1$ roots of $\sin n\theta/\sin\theta$ (in the interval $0<\theta<\pi$) interlace the n roots of $\sin(n+1)\theta/\sin\theta$, the difference polynomial,

$$\sin(n+1)\theta/\sin\theta-m^{\frac{1}{2}}\sin n\theta/\sin\theta,$$

has a real zero between every adjacent pair of zeros of $\sin(n+1)\theta/\sin\theta$. Thus there is a total of $n-1$ of these real roots θ_{nk}. With one exception, the last or nth root of $P_n(\theta)$ is obtained from Eq. (24′)

$$e^{2n\phi}=(m^{\frac{1}{2}}-e^{-\phi})/(m^{\frac{1}{2}}-e^{\phi}),$$

where $i\phi=\theta$. The exception can be traced to the fact that for $\epsilon\ll 1/n$, $\sin(n+1)\epsilon/\sin\epsilon<m^{\frac{1}{2}}\sin n\epsilon/\sin\epsilon$ *save* in the case $n=2$, $m=2$. For $m=2$, the roots of P_2 are $\theta_{21}=7\pi/12$ and $\theta_{22}=\pi/12$.

Electrical Conduction in a Noncircular Rod

A. C. PIPKIN AND R. S. RIVLIN
Brown University, Providence, Rhode Island
(Received May 16, 1961)

The general constitutive equations for galvanomagnetic effects in isotropic materials are applied to the study of electrical conduction in rods. It is shown that, in general, rectilinear current flow is not possible, unless the rod has a circular cross section or is an infinite parallel-sided slab.

1. INTRODUCTION

IN a previous paper[1] the constitutive equations for galvanomagnetic effects in isotropic materials were derived from an initial assumption that the electric current density $\mathbf{J}$ and the magnetic field $\mathbf{H}$ are functions of the electric field $\mathbf{E}$ and the magnetic induction $\mathbf{B}$. We now consider the flow of current in an infinitely long cylindrical rod of such material when a constant potential difference is maintained between its ends. It will be shown that rectilinear flow of current is not, in general, possible.

It is natural to assume that the current in a long cylindrical conductor will flow in straight lines parallel to the axis of the cylinder, producing a magnetic field which has no component in the axial direction. However, in materials of the type considered, the assumption of rectilinear flow leads to a system of three partial differential equations for the two nonzero components of the magnetic field. These three equations are independent unless the conducting material is of a certain special type. This special class of materials includes, of course, those which obey the classical Ohm's law.

It is found that if the three equations are independent, the magnetic field inside the conductor must form concentric circles or parallel straight lines. On the other hand, the external field is of the usual solenoidal and irrotational type. Presumably, the internal and external fields cannot be joined properly at the conductor boundary unless the boundary has a special shape, and therefore rectilinear flow is not generally possible. We show, in the case when the material and surrounding dielectric have equal, constant permeabilities, that rectilinear current flow is possible only in circular rods and tubes and in infinite slabs. This result is probably also true in general, when the permeabilities are not equal.

The present problem is somewhat similar to that of the flow of non-Newtonian fluids through tubes. Ericksen[2] pointed out that unless the fluid is of a special type, or the tube has a special cross-sectional shape, rectilinear motion is not possible. For sufficiently small deviations from rectilinear flow, Green and Rivlin[3] have shown that the fluid then flows with a helical motion. In the present case, we expect a similar result; that the electrical current will flow in a helical path when rectilinear flow is not possible.

2. CONSTITUTIVE EQUATIONS

It has been shown[1] that the constitutive equations describing galvanomagnetic effects in holohedral isotropic materials must be expressible in the forms

$$\mathbf{J}=\alpha_1\mathbf{E}+(\alpha_2\mathbf{E}\cdot\mathbf{B})\mathbf{B}+\alpha_3\mathbf{E}\times\mathbf{B} \tag{2.1}$$

and

$$\mathbf{H}=\gamma_1\mathbf{B}+\mathbf{E}\cdot\mathbf{B}(\gamma_2\mathbf{E}+\gamma_3\mathbf{E}\times\mathbf{B}), \tag{2.2}$$

where the coefficients $\alpha_1, \alpha_2, \cdots, \gamma_3$ are functions of

$$\mathbf{E}\cdot\mathbf{E}, \quad \mathbf{B}\cdot\mathbf{B}, \quad \text{and} \quad (\mathbf{E}\cdot\mathbf{B})^2. \tag{2.3}$$

For the purposes of the present paper, it is more convenient to use the inverse equations in which $\mathbf{E}$ and $\mathbf{B}$ are given in terms of $\mathbf{J}$ and $\mathbf{H}$. These may be obtained, by an argument analogous to that employed in obtaining Eqs. (2.1) and (2.2), as

$$\mathbf{E}=a_1\mathbf{J}+(a_2\mathbf{J}\cdot\mathbf{H})\mathbf{H}+a_3\mathbf{J}\times\mathbf{H} \tag{2.4}$$

and

$$\mathbf{B}=c_1\mathbf{H}+\mathbf{J}\cdot\mathbf{H}(c_2\mathbf{J}+c_3\mathbf{J}\times\mathbf{H}), \tag{2.5}$$

where $a_1, a_2, \cdots, c_3$ are functions of

$$\mathbf{J}\cdot\mathbf{J}, \quad \mathbf{H}\cdot\mathbf{H}, \quad \text{and} \quad (\mathbf{J}\cdot\mathbf{H})^2. \tag{2.6}$$

Equations of the forms (2.4) and (2.5) can also be obtained by solving (2.1) and (2.2) for $\mathbf{E}$ and $\mathbf{B}$ in the manner described in the Appendix.

3. RECTILINEAR CURRENT WITH NO AXIAL MAGNETIC FIELD

We consider a straight cylindrical rod of arbitrary, but uniform, cross-section lying parallel to the z axis of a rectangular Cartesian coordinate system x, y, z. The rod is composed of a homogeneous, isotropic conducting material for which the constitutive equations (2.4) and (2.5) are valid. The medium surrounding the rod is a perfect insulator of infinite extent, so that in it

$$\mathbf{J}=0 \quad \text{and} \quad \mathbf{B}=\mu\mathbf{H}, \tag{3.1}$$

where μ is a constant.

If the flow of current in the rod is steady, then Maxwell's equations, which are valid everywhere, take

[1] A. C. Pipkin and R. S. Rivlin, J. Math. Phys. **1**, 542 (1960).
[2] J. L. Ericksen, Quart. J. Appl. Math. **14**, 318 (1956).
[3] A. E. Green and R. S. Rivlin, Quart. J. Appl. Math. **14**, 299 (1956).

the form

$$\nabla\times\mathbf{H}=\mathbf{J}, \tag{3.2}$$

$$\nabla\cdot\mathbf{B}=0, \tag{3.3}$$

$$\nabla\times\mathbf{E}=0. \tag{3.4}$$

The tangential component of $\mathbf{H}$ and the normal component of $\mathbf{B}$ are continuous across the boundary of the conductor. At points in the insulator sufficiently distant from the conductor, $\mathbf{H}=0(1/r)$, where r is the distance of the point considered from the conductor.

We might expect that if a uniform electric field is applied to the rod in the direction of the z axis, the resulting current will be in the same direction and produce a magnetic field for which the z component is zero, so that

$$J_x=J_y=H_z=0. \tag{3.5}$$

We will show that if the validity of (3.5) is assumed, the resulting mathematical problem for the determination of the distribution of current and magnetic field is overdetermined and, in general, does not possess a solution.

From (3.5) and (3.2), it follows that $\mathbf{J}$ and $\mathbf{H}$ are independent of z. We may therefore write

$$J_z=J(x,y). \tag{3.6}$$

From (2.5) and (3.1), it follows that $\mathbf{B}$ is independent of z everywhere and from (2.4) it follows that $\mathbf{E}$ is independent of z inside the rod. With $\partial\mathbf{E}/\partial z=0$, Eq. (3.4) implies that E_z is a constant E (say):

$$E_z=E. \tag{3.7}$$

With (3.5), Eqs. (2.4) yield

$$E_x=-Ja_3H_y, \quad E_y=Ja_3H_x, \quad E_z=Ja_1, \tag{3.8}$$

and Eqs. (2.5) yield

$$B_x=c_1H_x, \quad B_y=c_1H_y, \quad B_z=0, \tag{3.9}$$

where, from (2.6), a_1, a_3, and c_1 are functions of

$$\mathbf{J}\cdot\mathbf{J}=J^2, \quad \mathbf{H}\cdot\mathbf{H}=H_x^2+H_y^2=H^2 \text{ (say)}$$
$$\text{and} \quad (\mathbf{J}\cdot\mathbf{H})^2=0. \tag{3.10}$$

By using Eqs. (3.1) and (3.9) in (3.3), we obtain

$$\nabla\cdot(c_1\mathbf{H})=0 \quad \text{(conductor)} \tag{3.11}$$

and

$$\mu\nabla\cdot\mathbf{H}=0 \quad \text{(insulator)}. \tag{3.12}$$

Since $\mathbf{J}$ and $\mathbf{H}$ are independent of z, Eq. (3.2) yields

$$\frac{\partial H_y}{\partial x}-\frac{\partial H_x}{\partial y}=\begin{cases}J & \text{(conductor)} \quad (3.13)\\ 0 & \text{(insulator)}. \quad (3.14)\end{cases}$$

Equation (3.7), with the last of equations (3.8), yields

$$Ja_1(J^2,H^2,0)=E \quad \text{(constant)}. \tag{3.15}$$

We note that Eqs. (3.11)–(3.15), with $J=0$ in the insulator, provide three equations valid in the conductor and three valid in the insulator. It appears that, together with the boundary conditions on the surface of the conductor and at infinity mentioned earlier, they should be sufficient for the determination of H_x, H_y, and J. However, Eq. (3.15) is not the only consequence of Eq. (3.4). By using (3.8) in (3.4), we also obtain

$$\nabla\cdot(Ja_3\mathbf{H})=0. \tag{3.16}$$

We shall show that solutions of (3.11)–(3.15) cannot also satisfy Eq. (3.16) unless certain restrictions are imposed on the constitutive equations which are valid for the material of the rod or on its cross-sectional shape.

4. DEGENERATE MATERIALS

We may regard Eq. (3.15) as determining J as a function of H^2. With $J=J(H^2)$, the coefficients c_1 and Ja_3 in Eqs. (3.11) and (3.16) can also be regarded as functions of H^2. Equations (3.11) and (3.16) can then be written in the forms

$$c_1\nabla\cdot\mathbf{H}+\frac{dc_1}{d(H^2)}\mathbf{H}\cdot\nabla H^2=0 \tag{4.1}$$

and

$$Ja_3\nabla\cdot\mathbf{H}+\frac{d(Ja_3)}{d(H^2)}\mathbf{H}\cdot\nabla H^2=0. \tag{4.2}$$

Together, these equations imply that either

$$\mathbf{H}\cdot\nabla H^2=0 \tag{4.3}$$

and

$$\nabla\cdot\mathbf{H}=0, \tag{4.4}$$

or

$$c_1\frac{d(Ja_3)}{d(H^2)}=Ja_3\frac{dc_1}{d(H^2)}. \tag{4.5}$$

In the degenerate class of materials for which (4.5) is satisfied, Eqs. (4.1) and (4.2) are compatible and the problem is presumably not overdetermined. For example, if the material obeys Ohm's law, then $a_3=0$ and Eq. (4.5) is satisfied.

5. OVERDETERMINATION

If Eq. (4.5) is not satisfied for the material considered, then the magnetic field inside the conductor must satisfy the independent equations (4.3), (4.4), and (3.13). In (3.13), J is regarded as a known function of H^2, obtained from (3.15). We thus have three equations for the two unknowns H_x and H_y.

Let $\mathbf{t}$ and $\mathbf{n}$ denote unit vectors, tangential and normal, respectively, to the lines of flux and so oriented that $\mathbf{n}$, $\mathbf{t}$, and the direction of the z axis form a right-handed triad at each point. Then,

$$(\mathbf{t}\cdot\nabla)\mathbf{t}=-\mathbf{n}/r \quad \text{and} \quad (\mathbf{t}\cdot\nabla)\mathbf{n}=\mathbf{t}/r, \tag{5.1}$$

and

$$(\mathbf{n}\cdot\nabla)\mathbf{n}=-\mathbf{t}/\rho \quad \text{and} \quad (\mathbf{n}\cdot\nabla)\mathbf{t}=\mathbf{n}/\rho, \tag{5.2}$$

where r and ρ are the radii of curvature to the line of flux and to the orthogonal trajectory of the lines of flux at the point considered. From (5.1) and (5.2), it can be shown, bearing in mind that $\mathbf{t}$ and $\mathbf{n}$ are unit vectors, that

$$(\mathbf{n}\cdot\nabla)(\mathbf{t}\cdot\nabla)+r^{-1}(\mathbf{t}\cdot\nabla)=(\mathbf{t}\cdot\nabla)(\mathbf{n}\cdot\nabla)+\rho^{-1}(\mathbf{n}\cdot\nabla). \quad (5.3)$$

Bearing in mind that $\mathbf{H}=H\mathbf{t}$, we can now rewrite Eqs. (4.3), (4.4), and (3.13) as

$$H\mathbf{t}\cdot\nabla H^2=0, \quad (5.4)$$

$$\nabla\cdot(H\mathbf{t})=\mathbf{t}\cdot\nabla H+H/\rho=0, \quad (5.5)$$

and

$$\nabla\cdot(H\mathbf{n})=\mathbf{n}\cdot\nabla H+H/r=J(H^2). \quad (5.6)$$

Denoting by s distance measured along a line of flux, Eq. (5.4) may be rewritten as

$$H\partial(H^2)/\partial s=0. \quad (5.7)$$

It follows that either $H=0$ or $\partial H/\partial s=0$. If $H=0$, it follows from (5.5) that $\partial H/\partial s=0$. Hence, H is constant along each line of flux.

From (5.5), with $\partial H/\partial s=0$, we see that if $H\neq 0$, then $1/\rho=0$. The orthogonal trajectories of the lines of flux are therefore straight lines and the lines of flux are a family of involutes.

By operating on Eq. (5.6) with $\mathbf{t}\cdot\nabla$, we obtain

$$(\mathbf{t}\cdot\nabla)(\mathbf{n}\cdot\nabla)H+r^{-1}\mathbf{t}\cdot\nabla H-Hr^{-2}\mathbf{t}\cdot\nabla r=\mathbf{t}\cdot\nabla J(H^2). \quad (5.8)$$

By using Eq. (5.3) to interchange the order of the operators $\mathbf{t}\cdot\nabla$ and $\mathbf{n}\cdot\nabla$ in (5.8), we obtain

$$(\mathbf{n}\cdot\nabla)(\mathbf{t}\cdot\nabla)H+2r^{-1}\mathbf{t}\cdot\nabla H-\rho^{-1}\mathbf{n}\cdot\nabla H-Hr^{-2}\mathbf{t}\cdot\nabla r=\mathbf{t}\cdot\nabla J(H^2). \quad (5.9)$$

Since $\mathbf{t}\cdot\nabla H=\partial H/\partial s=0$ and $1/\rho=0$, Eq. (5.9) yields

$$\frac{H}{r^2}\mathbf{t}\cdot\nabla r=\frac{H}{r^2}\frac{\partial r}{\partial s}=0. \quad (5.10)$$

If $H\neq 0$, then either $1/r=0$ or $\partial r/\partial s=0$. A line of flux for which $1/r=0$ is a straight line. A line of flux for which $\partial r/\partial s=0$ is a circle. Thus, the lines of flux are circles of finite or infinite radius.

A family of circular involutes is a family of concentric circles, or degenerately a family of parallel straight lines. In (5.6), $\mathbf{n}\cdot\nabla$ may now be replaced by d/dr and we obtain

$$dH/dr+H/r=J(H^2). \quad (5.11)$$

We conclude that the lines of flux inside the conductor are concentric circular arcs. The magnitude of H is constant along each line of flux and satisfies Eq. (5.11), where r is the radius of the line of flux passing through the point considered.

It is clear that if $\mathbf{H}$ is determined both inside and outside the conductor by using Eqs. (3.11)–(3.15) and the appropriate boundary conditions, the lines of flux inside the conductor will not generally be concentric circles. This contradicts the results of this section, obtained by using Eq. (3.16) as well. In the cases in which this contradiction arises, our initial assumption expressed by Eq. (3.5) must be incorrect.

6. DEGENERATE CROSS-SECTIONAL SHAPES

For materials for which Eq. (4.5) is not valid, we have shown that if there exists a solution of the system of equations (3.11) to (3.16), the lines of flux inside the conductor are concentric circles and the magnitude of $\mathbf{H}$ is constant along each line of flux. It is apparent that these conditions cannot be satisfied in a conducting rod of arbitrary cross-sectional shape. However, if the cross section of the rod is bounded by a circle or by a set of concentric circles, then we may expect the lines of flux to be circular. We wish to show that these are the only possible exceptions. In order to do this, we must show that the solution of Eqs. (3.11)–(3.16) for the interior of the rod, which has already been constructed, cannot be continued into the exterior region, unless the cross-section is circular.

Let the common center of the flux circles in the rod be used as the origin of coordinates. Let the curve C bounding the conductor be given parametrically in terms of the arc length s by $x=x(s)$, $y=y(s)$. Inside C, $\mathbf{H}$ is given by

$$H_x=-H(r)y/r, \quad H_y=H(r)x/r, \quad (6.1)$$

where

$$r^2=x^2+y^2, \quad (6.2)$$

and where $H(r)$ satisfies Eq. (5.11).

By using (6.1) in the condition that the tangential component of $\mathbf{H}$ is continuous across C, we obtain

$$H_x x'(s)+H_y y'(s)=(H/r)[-yx'(s)+xy'(s)], \quad (6.3)$$

where H_x and H_y are now the components of the external field, evaluated on C. Similarly, since $\mathbf{B}=c_1\mathbf{H}$ inside C and $\mathbf{B}=\mu\mathbf{H}$ outside C, continuity of the normal component of $\mathbf{B}$ requires that on C,

$$\mu[-H_x y'(s)+H_y x'(s)]=(c_1H/r)[yy'(s)+xx'(s)]. \quad (6.4)$$

The boundary values on C of H_x and H_y, to be used in determining the field outside the rod are obtained from (6.3) and (6.4) as

$$H_x=(H/r)[x'(xy'-yx')-(c_1/\mu)y'(xx'+yy')], \quad (6.5)$$

and

$$H_y=(H/r)[y'(xy'-yx')+(c_1/\mu)x'(xx'+yy')]. \quad (6.6)$$

Outside C, $\mathbf{H}$ satisfies Eqs. (3.12) and (3.14). From them it follows that H_x-iH_y is an analytic function of $z=x+iy$, so that

$$H_x-iH_y=w(z) \text{ (say)}. \quad (6.7)$$

The condition that H_x and H_y are $O(1/r)$ at infinity implies that $zw(z)$ is regular at infinity. The conditions (6.5) and (6.6) can be combined, by using $z\bar{z}=r^2$ and

$dzd\bar{z}=(ds)^2$, to yield

$$zw(z)=-\tfrac{1}{2}irH\left[1+\frac{c_1}{\mu}-\left(1-\frac{c_1}{\mu}\right)\left(\frac{z}{r}\frac{d\bar{z}}{ds}\right)^2\right]. \quad (6.8)$$

Since (6.8) determines both the real and imaginary parts of $w(z)$ on C, the problem is, in general, overdetermined.

In the particular case when c_1 is a constant and equal to μ, it is easy to show that (6.8) can be satisfied only if the cross-section of the rod is circular. Introducing $c_1=\mu$ into (6.8), we obtain

$$zw(z)=-irH \quad \text{on} \quad C. \quad (6.9)$$

Since $\Re(zw)=0$ on C and is regular outside C and at infinity, $\Re(zw)=0$ outside C. Therefore, $\Im(zw)=$constant outside C. On C, $\Im(zw)=-rH$. Hence, $H=A/r$, where A is a constant, on C. Let r_0 be the radial distance from the origin to the nearest point of the boundary. Then, inside C, for $r\geq r_0$, we have $H=A/r$. By using this result in Eq. (5.11), we obtain $J=0$, for $r>r_0$. This implies that $r=r_0$ is the boundary of the conductor. If the cross section of the conductor is not simply connected, then Eq. (6.9) also holds on each interior boundary and by analogous reasoning it can be seen that each interior boundary is also a circle with center at the origin.

In the more general case when c_1 is a function of H^2, which in turn is a function of r, we have so far not succeeded in proving rigorously that (6.8) can be satisfied only if C is circular. However, if C is the circle $|z|=r_0$, then $d\bar{z}/ds=-ir_0/z$ on C and (6.8) becomes

$$zw(z)=-ir_0H(r_0). \quad (6.10)$$

This boundary condition implies directly its own analytic continuation into the region outside C.

ACKNOWLEDGMENT

The work described in this paper was carried out under a grant from the National Science Foundation. We gratefully acknowledge their support.

APPENDIX. INVERSION OF CONSTITUTIVE EQUATIONS

In this section, we show how Eqs. (2.4) and (2.5) may be obtained directly from (2.1) and (2.2). From (2.1) and (2.2) we obtain

$$\mathbf{J}\times\mathbf{H}=[\beta_1(\mathbf{E}\cdot\mathbf{B})^2+\beta_2\mathbf{B}\cdot\mathbf{B}]\mathbf{E}-(\beta_1\mathbf{E}\cdot\mathbf{E}+\beta_2)(\mathbf{E}\cdot\mathbf{B})\mathbf{B}+\beta_3\mathbf{E}\times\mathbf{B}, \quad (7.1)$$

where

$$\beta_1=\alpha_1\gamma_3-\alpha_3\gamma_2, \quad (7.2)$$

$$\beta_2=\alpha_2\gamma_3(\mathbf{E}\cdot\mathbf{B})^2-\alpha_3\gamma_1, \quad (7.3)$$

and

$$\beta_3=\alpha_1\gamma_1-\alpha_2\gamma_2(\mathbf{E}\cdot\mathbf{B})^2. \quad (7.4)$$

Equations (2.1), (2.2), and (7.1) are linear equations for $\mathbf{E}$, $\mathbf{B}$, and $\mathbf{E}\times\mathbf{B}$ in terms of $\mathbf{J}$, $\mathbf{H}$, and $\mathbf{J}\times\mathbf{H}$, with scalar coefficients. Solving them, we obtain

$$\mathbf{E}=(A_1/\Delta)\mathbf{J}-(A_2/\Delta)(\mathbf{E}\cdot\mathbf{B})\mathbf{H}+(\beta_2/\Delta)\mathbf{J}\times\mathbf{H} \quad (7.5)$$

and

$$\mathbf{B}=(A_3/\Delta)\mathbf{H}-\mathbf{E}\cdot\mathbf{B}[(A_4/\Delta)\mathbf{J}+(\beta_1/\Delta)\mathbf{J}\times\mathbf{H}], \quad (7.6)$$

where

$$A_1=\gamma_1\beta_3+\gamma_3(\beta_1\mathbf{E}\cdot\mathbf{E}+\beta_2)(\mathbf{E}\cdot\mathbf{B})^2, \quad (7.7)$$

$$A_2=\alpha_2\beta_3+\alpha_3(\beta_1\mathbf{E}\cdot\mathbf{E}+\beta_2), \quad (7.8)$$

$$A_3=\alpha_1\beta_3-\alpha_3[\beta_1(\mathbf{E}\cdot\mathbf{B})^2+\beta_2\mathbf{B}\cdot\mathbf{B}], \quad (7.9)$$

$$A_4=\gamma_2\beta_3-\gamma_3[\beta_1(\mathbf{E}\cdot\mathbf{B})^2+\beta_2\mathbf{B}\cdot\mathbf{B}], \quad (7.10)$$

and

$$\Delta=|\beta_1(\mathbf{E}\cdot\mathbf{B})\mathbf{E}+\beta_2\mathbf{B}|^2+\beta_3^2. \quad (7.11)$$

If $\Delta=0$, then $\mathbf{E}$, $\mathbf{B}$, $\mathbf{J}$, and $\mathbf{H}$ are all in the same direction and the inverse equations are nonunique.

From (2.1) and (2.2), we also obtain

$$\mathbf{J}\cdot\mathbf{H}=A_5\mathbf{E}\cdot\mathbf{B},$$

$$A_5=\alpha_1\gamma_1+\alpha_1\gamma_2\mathbf{E}\cdot\mathbf{E}+\alpha_2\gamma_1\mathbf{B}\cdot\mathbf{B}+\alpha_2\gamma_2(\mathbf{E}\cdot\mathbf{B})^2+\alpha_3\gamma_3[(\mathbf{E}\cdot\mathbf{E})(\mathbf{B}\cdot\mathbf{B})-(\mathbf{E}\cdot\mathbf{B})^2]. \quad (7.12)$$

By using (7.12) in (7.5) and (7.6), we obtain

$$\mathbf{E}=(A_1/\Delta)\mathbf{J}-(A_2/A_5\Delta)(\mathbf{J}\cdot\mathbf{H})\mathbf{H}+(\beta_2/\Delta)\mathbf{J}\times\mathbf{H} \quad (7.13)$$

and

$$\mathbf{B}=(A_3/\Delta)\mathbf{H}-\mathbf{J}\cdot\mathbf{H}[(A_4/A_5\Delta)\mathbf{J}+(\beta_1/A_5\Delta)\mathbf{J}\times\mathbf{H}]. \quad (7.14)$$

Equations (7.13) and (7.14) are of the forms (2.4) and (2.5), respectively, except that the scalar coefficients in (7.13) and (7.14) are functions of the invariants (2.3) rather than of the invariants (2.6). It is necessary to find expressions for the quantities (2.6) in terms of the quantities (2.3). The equations to be used are (7.12) and the following, obtained from (2.1) and (2.2):

$$\mathbf{J}\cdot\mathbf{J}=\alpha_1^2\mathbf{E}\cdot\mathbf{E}+\alpha_2^2(\mathbf{E}\cdot\mathbf{B})^2\mathbf{B}\cdot\mathbf{B}+\alpha_3^2[(\mathbf{E}\cdot\mathbf{E})(\mathbf{B}\cdot\mathbf{B})-(\mathbf{E}\cdot\mathbf{B})^2]+2\alpha_1\alpha_2(\mathbf{E}\cdot\mathbf{B})^2, \quad (7.15)$$

$$\mathbf{H}\cdot\mathbf{H}=\gamma_1^2\mathbf{B}\cdot\mathbf{B}+\gamma_2^2(\mathbf{E}\cdot\mathbf{B})^2\mathbf{E}\cdot\mathbf{E}+\gamma_3^2(\mathbf{E}\cdot\mathbf{B})^2\times[(\mathbf{E}\cdot\mathbf{E})(\mathbf{B}\cdot\mathbf{B})-(\mathbf{E}\cdot\mathbf{B})^2]+2\gamma_1\gamma_2(\mathbf{E}\cdot\mathbf{B})^2. \quad (7.16)$$

The inversion cannot be completed until the forms of $\alpha_1, \alpha_2, \cdots, \gamma_3$ have been specified.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

Method for Defining Principal Modes of Nonlinear Systems Utilizing Infinite Determinants

DEMETRIOS G. MAGIROS
General Electric Company, Missile and Space Vehicle Department, Philadelphia, Pennsylvania
(Received April 21, 1961)

A method for calculation of "principal modes" of linear or nonlinear systems is discussed. The physical definition of "principal modes" is formulated mathematically in two ways. The trial solution of the differential equation of the motion of the system is taken in an appropriate structure. The calculation of principal modes leads to infinite determinants of Hill's and von Koch's type, which are analyzed. The above method yields the possibility of getting the "principal modes" in the form of a series, all the coefficients of which can be calculated.

1. INTRODUCTION

THE concept of "principal modes" plays the predominant role in the analysis of the oscillatory systems, no matter what field the systems occur in.

The principal modes of linear systems are, by definition the fundamental set of solutions of which a linear combination gives the general solution of the linear differential equations, which govern the motion of the linear system. This means that any kind of oscillations in linear systems can be discussed in terms of some special modes of oscillation of the system, the "principal modes" of the system.

This definition of "principal modes" is meaningless in nonlinear systems, since the "principle of superposition" does not hold in those systems.

The study of the principal modes of systems, either linear or nonlinear, may be made by using two definitions, namely the "proportionality definition" of principal modes and their definition as solutions of "initial value problems" of special type. Calculations, based on these definitions, are shown for a nonlinear "dual-mode" system. If the solution of the differential equations of this system is taken as an exponential series with complex coefficients, the calculation of the frequency ω and the coefficients of the series leads to a recursion formula, which gives rise to "infinite determinants" of special type. The analysis of the infinite determinants involved, and the solution in its final form is discussed. The nonunit elements of the determinants contain the coefficient of the nonlinearity as a common factor, and, for a weak nonlinearity, we can get an expansion of these determinants in powers of the coefficient, then an appropriate approximation of them. Thus the "frequency equation," in an infinite determinantal form, is reduced to a quartic in ω^2, and the ratio of the determinants of the coefficients of the series to unit.

The solution given is in accordance with both definitions of "principal modes," and imposes a relation between the initial displacements of the masses, and this relation and the condition for the initial velocities to be zero distinguished the special initial value problems appropriate for the "principal modes."

A brief discussion of the present paper has been published in the Proceedings of National Academy of Sciences.[1a,b]

2. THE DEFINITION OF PRINCIPAL MODES, ITS APPLICATION TO A FREE NONLINEAR SYSTEM OF TWO DEGREES OF FREEDOM, AND THE RECURSION FORMULA FOR THE SOLUTION

By using the terminology of mechanical systems with s degrees of freedom, the principal modes of oscillations of the system are defined as those oscillations of the system for which the nonzero amplitudes of the fundamental and the corresponding harmonics of the displacements of any two of the oscillating masses have, separately, a constant ratio. For such motions, the masses all oscillate about their equilibrium positions, where they pass at the same time. Their common frequency is the "principal frequency" of the system.

If x_i, $i=1, 2, \cdots s$ are the displacements of the masses m_i, $i=1, 2, \cdots s$ from their equilibrium positions, and x_{in}, $x_{i'n}$ the corresponding amplitudes of the nth harmonic of the displacements x_i, $x_{i'}$, of the two masses m_i, $m_{i'}$, and if there exist constants c_n such that the conditions

$$x_{in}/x_{i'n}=c_n, \quad i\neq i'; \quad i, i'=1, 2, \cdots s, \quad c_n\neq 0, \infty \qquad (1)$$

are satisfied, then these motions are, by definition, the "principal modes" of the system.

We restrict ourselves, without loss of generality, to a two-degrees-of-freedom nonlinear system; namely, we get—as a mechanical model—the "two-masses-three-springs" system with one of the anchor springs nonlinear and such that the corresponding restoring force is an odd-cubic function of the distance, or—as an electrical model—the "two-inductances-three-capacitances" system with one capacitance variable and the others constant, Fig. 1 (a), (b).

If m_1 and m_2 are the oscillating masses, $\bar{K}_1=K_1+\mu x^2$, K_3 and K_2 the stiffnesses of the first and second anchor springs and of the coupling, μ a constant which characterizes the nonlinearity, and x and y the displacements of m_1 and m_2 from their equilibrium positions, the equa-

[1] (a) D. G. Magiros, Proc. Natl. Acad. Sci. U. S., Dec. (1960). (b) D. G. Magiros, Proc. Natl. Acad. Sci. U. S., June (1961).

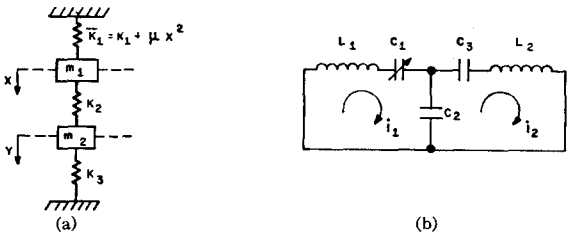

FIG. 1. (a) The mechanical model. (b) The electrical model.

tions of motion of this sytem are:

$$\ddot{x}+\omega_1^2 x+\lambda_1 x^3-\lambda_2 y=0,$$
$$\ddot{y}+\omega_2^2 y-\lambda_3 x=0, \tag{2}$$

where

$$\omega_1^2=\frac{K_1+K_2}{m_1},\quad \omega_2^2=\frac{K_2+K_3}{m_2},$$
$$\lambda_1=\frac{\mu}{m_1},\quad \lambda_2=\frac{K_2}{m_1},\quad \lambda_3=\frac{K_2}{m_2}. \tag{2a}$$

We proceed to find the solutions $x(t)$ and $y(t)$ of (2) with one fundamental frequency ω for both oscillators by imposing a certain organic structure for the functions $x(t)$ and $y(t)$. Assume solutions of (2) of the form

$$x(t)=\sum_{n=-\infty}^{\infty}\alpha_n e^{in\omega t},\quad y(t)=\sum_{n=-\infty}^{\infty}A_n e^{in\omega t}. \tag{3}$$

The coefficients α_n and A_n are complex, then they include the phase angle.

For the reality of the solution, one takes $\bar{\alpha}_n=\alpha_{-n}$, $\bar{A}_n=A_{-n}$, where $\bar{\alpha}_n$ and $\bar{A}_n$ are conjugates of α_n and A_n. Also assume that

$$\sum_{n=-\infty}^{\infty}|\alpha_n|<\infty,\quad \sum_{n=-\infty}^{\infty}|A_n|<\infty,$$
$$\sum_{n=-\infty}^{\infty}n^2|\alpha_n|<\infty,\quad \sum_{n=-\infty}^{\infty}n^2|A_n|<\infty.$$

The first two of these inequalities guarantee the convergence of the series (3), while the second ones the existence of the second derivatives $\ddot{x}$, $\ddot{y}$.

By using the first of (3), x^3 is given by

$$x^3=\sum_{\rho_1=-\infty}^{\infty}\sum_{\rho_2=-\infty}^{\infty}\sum_{\rho_3=-\infty}^{\infty}\alpha_{\rho_1}\alpha_{\rho_2}\alpha_{\rho_3}e^{i(\rho_1+\rho_2+\rho_3)\omega t},$$

which can be written as follows:

$$x^3=\sum_{n}\sum_{\rho_1}\sum_{\rho_2}\alpha_{\rho_1}\alpha_{\rho_2}\alpha_{n-\rho_1-\rho_2}e^{in\omega t}. \tag{3a}$$

Inserting (3) and (3a) into (2), the following system results:

$$\left(\omega_1^2-n^2\omega^2-\frac{\lambda_2\lambda_3}{\omega_2^2-n^2\omega^2}\right)\alpha_n+\lambda_1\sum_{\rho_1}\sum_{\rho_2}\alpha_{\rho_1}\alpha_{\rho_2}\alpha_{n-\rho_1-\rho_2}=0, \tag{4a}$$

$$A_n/\alpha_n=\lambda_3/(\omega_2^2-n^2\omega^2), \tag{4b}$$

where n, ρ_1, ρ_2 are integers. This nonlinear system consists of infinitely many nonlinear ($\lambda_1\neq 0$) homogeneous equations for the infinitely many unknown coefficients α_n and A_n.

The equation (4b) expresses the definition of principal modes applied to the above dual-mode system, and by (4b) the calculation of A_n is deduced from α_n, then the calculation of principal modes of the system is deduced from α_n, by using (4a). If α_1 is the dominant coefficient of the sequence $\{\alpha_n\}$, then the approximate value of the double series of (4a), see Appendix I, is given by

$$3\alpha_1^2\alpha_{n-2}+6|\alpha_1|^2\alpha_n+3\bar{\alpha}_1^2\alpha_{n+2}, \tag{5}$$

with an error $o(\alpha_1^2)$.

Inserting the expression (5) into the place of the double series of (4a), we can get the following recursion formula:

$$\rho_n\alpha_{n-2}+\alpha_n+q_n\alpha_{n+2}=0, \tag{6}$$

where

$$\rho_n=\frac{3\lambda_1\alpha_1^2}{k_n+6\lambda_1|\alpha_1|^2},\quad q_n=\frac{3\lambda_1\bar{\alpha}_1^2}{k_n+6\lambda_1|\alpha_1|^2},$$
$$k_n=\omega_1^2-n^2\omega^2-\frac{\lambda_2\lambda_3}{\omega_2^2-n^2\omega^2}. \tag{6a}$$

It is the recursion formula (6) which will be used for the calculation of the "principal frequency" ω and the coefficients α_n of the solution (3). We notice here that for the convergence of the series (3) it is necessary, according to (4b), that $(\omega_2^2-n^2\omega^2)$ is neither zero nor very small, i.e., ω must not be either a submultiple ω_2 or very close to a submultiple of ω_2.

3. CALCULATION OF THE PRINCIPAL FREQUENCY ω

The recursion formula (6) gives infinitely many homogeneous equations for the infinitely many unknowns α_n. The corresponding infinite matrix of the coefficients of these equations is

$$\begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & 0 & \rho_{-1} & 0 & 1 & 0 & q_{-1} & 0 & \cdots \\
\cdots & 0 & \rho_0 & 0 & 1 & 0 & q_0 & 0 & \cdots \\
\cdots & 0 & \rho_1 & 0 & 1 & 0 & q_1 & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}.$$

For nonzero unknowns α_n, the corresponding infinite determinant must be zero. This doubly infinite determinant, by taking n arbitrary integer, becomes one-sided infinite determinant, and we can write

$$\Delta(n,\infty)=0, \tag{7}$$

where the infinite determinant $\Delta(n,\infty)$ is given by the limit

$$\lim\Delta(n,m)= \begin{array}{l} n=\text{fixed integer} \\ m=0,\,1,\,2,\,3,\,\cdots \\ m\to\infty \end{array} \left\| \begin{array}{ccccccccc} 1 & 0 & q_n & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 1 & 0 & q_{n+1} & 0 & \cdots & \cdots & 0 & 0 \\ \rho_{n+2} & 0 & 1 & 0 & q_{n+2} & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 & q_{n+m-2} \\ 0 & 0 & \cdots & \cdots & \cdots & p_{n+m-1} & 0 & 1 & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 & p_{n+m} & 0 & 1 \end{array} \right\|. \tag{8}$$

Consider a weak nonlinearity, i.e., $\lambda_1|x|^3 \ll \omega_1^2|x|$, or $\lambda_1 x^2 \ll (K_1+K_2)/m_1$. From the first of (3) we get

$$\max x^2=\left[\sum_{n=-\infty}^{\infty}\alpha_n\right]^2=\sum_{\rho_1=-\infty}^{\infty}\sum_{\rho_2=-\infty}^{\infty}\alpha_{\rho_1}\alpha_{\rho_2},$$

and since α_1 is considered as dominant element of that of the sequence $\{\alpha_n\}$, when $\max x^2=2\alpha_1^2$, then

$$\lambda_1 \ll (K_1+K_2)/(2m_1\alpha_1^2). \tag{9a}$$

For a weak nonlinearity we can write

$$\rho_n=\frac{3\lambda_1\alpha_1^2}{k_n+6\lambda_1|\alpha_1|^2}=\frac{3\lambda_1\alpha_1^2}{k_n} \times\left\{1-\lambda_1\frac{6|\alpha_1|^2}{k_n}+\lambda_1^2\left(\frac{6|\alpha_1|^2}{k_n}\right)^2-\cdots\right\} =\lambda_1\frac{3\alpha_1^2}{k_n}-\lambda_1^2\frac{18\alpha_1^2|\alpha_1|^2}{k_n^2}+\cdots, \tag{9}$$

$$q_n=\lambda_1\frac{3\bar{\alpha}_1^2}{k_n}-\lambda_1^2\frac{18\bar{\alpha}_1^2|\alpha_1|^2}{k_n^2}+\cdots.$$

All the elements not in the main diagonal of the determinant (8) have the small coefficient λ_1 as a common factor. Then, by applying formula (B) of Appendix II, and taking the first terms of ρ_n and q_n from formulas (9), the determinant (8) can be written approximately as

$$\Delta(n,\infty)=1-9\lambda_1^2|\alpha_1|^4\sum_{m=0}^{\infty}\frac{1}{k_{n+m}k_{n+m+2}}+0(\lambda_1^3). \tag{10}$$

$k_{\bar{n}}$, given by (6a), is an even function of $\bar{n}$. To examine the convergence of the series in (10) we confine ourselves to non-negative integers $\bar{n}$. Since $\{|k_{\bar{n}}|\}$ is a sequence with positive terms monotonically increasing with $\bar{n}$, and $|k_{\bar{n}}|\to\infty$, as $\bar{n}\to\infty$, the series in (10) converges. Convergence requirements of the series in (10) necessitates the ω^2 must not be a zero of $k_{\bar{n}}(\omega^2)$; hence

$$\omega^2\mp\omega_{0\pm}^2=\frac{1}{2\bar{n}^2}\{\omega_1^2+\omega_2^2\pm[(\omega_1^2-\omega_2^2)^2+4\lambda_2\lambda_3]^{\frac{1}{2}}\}. \tag{11}$$

The principal frequency ω can be determined from (7) by using (10).

Then, if we confine ourselves to the first term of the series in (10) and put $n=1$, the principal frequency ω is approximately a root of

$$k_1k_3=9\lambda_1^2|\alpha_1|^4,$$

or root of

$$\left(\omega_1^2-\omega^2-\frac{\lambda_2\lambda_3}{\omega_2^2-\omega^2}\right)\left(\omega_1^2-9\omega^2-\frac{\lambda_2\lambda_3}{\omega_2^2-9\omega^2}\right) -9\lambda_1^2|\alpha_1|^4=0, \tag{12}$$

which is quartic in ω^2. Formula (12) gives the principal frequency ω of order $O(\lambda_1^2)$. For values of the principal frequency of order higher than $O(\lambda_1^2)$, we find in the expansion of $\Delta(n,\infty)$ terms of order higher than $O(\lambda_1^2)$,[2] and continue in the same way to find the corresponding algebraic equation.

4. CALCULATION OF THE COEFFICIENTS

To calculate the coefficients α_n, we use in the recursion formula (6) the notation

$$u_n=\alpha_n/\alpha_{n-2}, \tag{13}$$

when we can get

$$u_n=-\frac{\rho_n}{1+q_nu_{n+2}},$$

which leads to the infinite continued fraction

$$\frac{\alpha_n}{\alpha_{n-2}}=-\cfrac{\rho_n}{1-\cfrac{\rho_{n+2}q_n}{1-\cfrac{\rho_{n+4}q_{n+2}}{1-\cdots}}}, \tag{14}$$

where the ρ's and q's are given by (9). The formula (14) is written as

$$\alpha_n/\alpha_{n-2}=-\rho_nZ_{n,\infty}, \tag{15}$$

[2] W. Magnus, "Infinite determinants in the theory of Mathieu's and Hill's equations," Research Report No. BR-1, Mathematical Research Group, Washington Square College of Arts and Science, New York University, 1953.

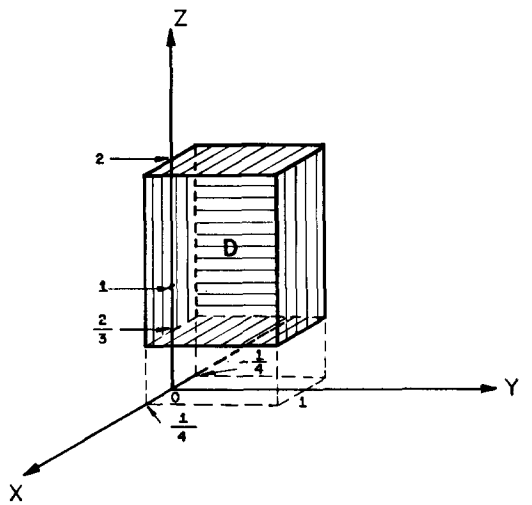

FIG. 2. The domain D in the X, Y, Z space is the appropriate one for the convergence of the continued fraction (16).

where

$$Z_{n,\infty}=\cfrac{1}{1+\cfrac{X_{n,1}}{1+\cfrac{X_{n,2}}{1+\cdots}}}. \tag{16}$$

$$X_{n,m}=-\rho_{n+2m}q_{n+2(m-1)}, \tag{16a}$$

n=fixed integer, $m=1,2,3,\cdots$.

Since the elements $X_{n,m}$ of (16) are functions of the real variable ω, the regions E, Y, V, defined in Appendix III, are segments of lines, and according to von Koch's[3] and Worpitzky's theorems, as stated in Appendix III,

$$\begin{aligned} E:&\quad -\tfrac{1}{4}\leq X_{n,m}\leq\tfrac{1}{4},\\ Y:&\quad 0<Y_{n,m}=\sum_{m=2}|X_{n,m}|<1,\\ V:&\quad \tfrac{2}{3}\leq Z_{n,m}\leq 2. \end{aligned} \tag{17}$$

Consider a Cartesian coordinate system in space, and let us take the region element on the X axis, the series region on the Y axis, and the value region on the Z axis. The inequalities (17) correspond then to the interior of the orthogonal parallelepiped D, Fig. 2, which is an open domain.
The domain

$$D:\quad -\tfrac{1}{4}\leq X\leq\tfrac{1}{4},\quad 0<Y<1,\quad \tfrac{2}{3}\leq Z\leq 2$$

is the appropriate one for the continued fraction (16).

To evaluate the continued fraction (16) we apply the theory of Appendix III. The denominator of the ratio, which gives the mth approximant Z_m of the continued fraction (16) is

$$B_{n,m-1}=\begin{vmatrix} 1 & -1 & 0 & \cdots & 0 & 0 & 0\\ -\rho_{n+2}q_n & 1 & -1 & \cdots & 0 & 0 & 0\\ 0 & -\rho_{n+4}q_{n+2} & 1 & \cdots & 0 & 0 & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0\cdots & & \cdots & -\rho_{n+2(m-1)}q_{n+2(m-2)} & 1 & -1\\ 0 & 0\cdots & & \cdots & 0 & -\rho_{n+2m}q_{n+2(m-1)} & 1 \end{vmatrix}. \tag{18}$$

The numerator $A_{n,m-1}$ of the ratio can be obtained from (18) if we omit its first row and first column. Taking the limit as $m\to\infty$, we obtain

$$\frac{\alpha_n}{\alpha_{n-2}}=-\rho_n\frac{\bar{\Delta}(n+2,\infty)}{\bar{\Delta}(n,\infty)}, \tag{19}$$

where

$$\bar{\Delta}(n,\infty)=\lim_{m\to\infty}B_{n,m-1},$$

$$\bar{\Delta}(n+2,\infty)=\lim_{m\to\infty}A_{n,m-1}. \tag{19a}$$

The determinants of (19) are of von Koch's type and they converge by von Koch's rule, that is, when the series $\sum_n|\rho_nq_{n-2}|$ converges, which happens here, as was pointed out at the discussion of the convergence of the series of (10).

Since $\alpha_n=\bar{\alpha}_n$, we may restrict ourselves to non-negative integers for the calculation of the coefficients α_n. Formula (19) suggests starting with $n=2$; then α_0 and α_1 are arbitrary. The induction procedure applied to (19) for the coefficients with even index gives

$$\alpha_{2n}=(-1)^n\alpha_0\rho_2\rho_4\cdots\rho_{2n}\frac{\bar{\Delta}(2n+2,\infty)}{\bar{\Delta}(2,\infty)},$$

$$n=1,2,3\cdots. \tag{20}$$

For the determination of the coefficients with odd index, we first calculate the coefficient α_3. This, according to Appendix I, is an exception.

If, according to Appendix I, we take α_1^3 instead of the double series of the formula (4a), and we apply this formula for $n=3$, there results

$$\alpha_3=-(\lambda_1\alpha_1^3/k_3), \tag{21}$$

where

$$k_3=\omega_1^2-9\omega^2-[\lambda_2\lambda_3/(\omega_2^2-9\omega^2)]. \tag{21a}$$

[3] H. von Koch, Compt. rend., **120**, 144 (1895).

Now, by applying the induction procedure to (19) starting from the coefficient α_5 and using the value of α_3 given by (21), one can get

$$\alpha_{2n+1}=(-1)^{n+1}\frac{\lambda_1\alpha_1^3}{k_3}\rho_5\rho_7\cdots\rho_{2n+1}\frac{\bar{\Delta}(2n+3,\infty)}{\bar{\Delta}(5,\infty)},\qquad n=2,3,4\cdots. \quad (22)$$

By applying formula (B) of Appendix II to the determinants of (20) and (22), there is obtained

$$\frac{\bar{\Delta}(2n+2,\infty)}{\bar{\Delta}(2,\infty)}=1-9\lambda_1^2|\alpha_1|^4\sum_{m=0}^{\infty}\left(\frac{1}{k_{2n+2+m}k_{2n+4+m}}-\frac{1}{k_{2+m}k_{4+m}}\right)+O(\lambda_1^3),$$

$$\frac{\bar{\Delta}(2n+5,\infty)}{\bar{\Delta}(5,\infty)}=1-9\lambda_1^2|\alpha_1|^4\sum_{m=0}^{\infty}\left(\frac{1}{k_{2n+5+m}k_{2n+7+m}}-\frac{1}{k_{5+m}k_{7+m}}\right)+O(\lambda_1^3). \quad (23)$$

The formulas (21), (20), and (22) give the coefficients of the first of the series (3) for any positive n, with arbitrary α_0 and α_1. For the determination of the coefficients α_{-n} we use the property $\alpha_{-n}=\bar{\alpha}_n$. The coefficient α_0 is real, and α_1 in general complex, $\alpha_1=|\alpha_1|e_i^{\varphi_1}$. If we take α_1 real, the solution (3) can be written as follows:

$$x(t)=\alpha_0+2\alpha_1\cos\omega t-2\alpha_3\cos3\omega t+2\sum_n\alpha_N\cos N\omega t,$$

$$y(t)=\frac{\lambda_3}{\omega_2^2}\alpha_0+\frac{2\lambda_3}{\omega_2^2-\omega^2}\alpha_1\cos\omega t-\frac{2\lambda_3}{\omega_2^2-3^2\omega^2}\alpha_3\cos3\omega t+2\sum_n\frac{\lambda_3}{\omega_2^2-N^2\omega^2}\alpha_N\cos N\omega t, \quad (24)$$

where the coefficients α_3, $\alpha_{N=2n}$, $\alpha_{N=2n+1}$ are given by the formulas (21), (20), (22), respectively. The formulas (24) give the solution in its final form, and the formulas (20)–(23) can be used for the calculation of as many coefficients of the solution (24) as we want in terms of powers of λ_1.

We can easily see that for the calculation of the coefficients, in terms of powers of λ_1, does not need but only the unit as value of the ratios of the determinants (23). The first two terms of the solution (24) are independent of λ_1. For terms of order $O(\lambda_1)$, we take the first term of ρ_n of (9) and combining it with (20) for $n=1$ we get the 2nd harmonic, which, with (21), gives

$$x(t)=\alpha_0+2\alpha_1\cos\omega t-2\lambda_1\left\{\frac{3\alpha_0\alpha_1^2}{k_2}\cos2\omega t+\frac{\alpha_1^3}{k_3}\cos3\omega t\right\}+0(\lambda_1^2),$$

$$y(t)=\frac{\lambda_3}{\omega_2^2}\alpha_0+\frac{2\lambda_3}{\omega_2^2-\omega^2}\alpha_1\cos\omega t-2\lambda_1\left\{\frac{3\lambda_3\alpha_0\alpha_1^2}{k_2(\omega_2^2-2^2\omega^2)}\cos2\omega t+\frac{\lambda_3\alpha_1^3}{k_3(\omega_2^2-3^2\omega^2)}\cos3\omega t\right\}+0(\lambda_1^2). \quad (25)$$

If we take the terms of ρ_n of (9) up to the order $0(\lambda_1^2)$, and combine them with (22) for $n=2$ we obtain the 4th harmonic terms of order $0(\lambda_1^2)$:

$$18\lambda_1^2\frac{\alpha_0\alpha_1^4}{k_2k_4}\cos4\omega t,\quad 18\lambda_1^2\frac{\alpha_0\alpha_1^4\lambda_3}{k_2k_4(\omega_2^2-4^2\omega^2)}\cos4\omega t, \quad (25a)$$

of $x(t)$ and $y(t)$, respectively. The above procedure indicates how we can get higher harmonics in terms of higher powers of λ_1. The solution (24), constructed as indicated above, must be convergent and its coefficient of the fundamental term must be much larger than any other coefficient. These requirements imply that the following conditions are satisfied:

$$\alpha_0\ll\alpha_1,\quad \lambda_1\ll\min\left\{\frac{K_1+K_2}{2m_1\alpha_1^2},\frac{k_2}{2\alpha_0\alpha_1},\frac{k_3}{\alpha_1^2}\right\}. \quad (26)$$

For the second condition, the inequality (6a) was taken into account. The 4th, 5th, $\cdots$ harmonic terms are of order $O(\lambda_1^2)$, $O(\lambda_1^3)$, $\cdots$, and the convergence is guaranteed.

Since α_1 is much larger compared to α_0, the solution in the linear case is approximately

$$x(t)=2\alpha_1\cos\omega t,\qquad x(t)=2\frac{\lambda_3}{\omega_2^2-\omega^2}\alpha_1\cos\omega t, \quad (25b)$$

and the motions of the oscillators are "in phase" for $\omega_2<\omega$, and "180° out of phase" for $\omega_2<\omega$.

5. THE PRINCIPAL MODES AS SOLUTIONS OF INITIAL VALUE PROBLEMS

Another approach for the determination of principal modes may be based on the manner in which the system is set into motion. This is equivalent to considering the principal modes as solutions of special initial value problems.

The differential equations of the "dual-mode" system are considered subject to the restriction that the

masses are displaced from their equilibrium positions either both up, or one up and the other down, by amounts x_0 and y_0, respectively, and released without velocity; i.e.,

$$x(0)=x_0, \quad y(0)=y_0, \quad \dot{x}(0)=\dot{y}(0)=0, \tag{27a}$$

$$x(0)=x_0, \quad y(0)=-y_0, \quad \dot{x}(0)=\dot{y}(0)=0. \tag{27b}$$

If the initial displacements x_0 and y_0 are appropriately related, then each one of these initial conditions gave rise to special vibration modes, which are, by definition, the "principal modes" of the system. To calculate the principal modes of our system utilizing infinite determinants and using the above definition, assume a solution in the form of complex exponential series (3), as in the previous case. The calculation of the principal frequency and the coefficients of the series has been completed throughout the preceding sections and the solution is found to be in the form given by (25). An approximation of this solution is given by (26), associated with the initial conditions

$$x_0=2\alpha_1, \quad y_0=\frac{\lambda_3 x_0}{\omega_2^2-\omega^2}=\frac{2\alpha_1\lambda_3}{\omega_2^2-\omega^2}. \tag{28}$$

Formulas (28) give the relations of the initial displacements required for the solution to be of "principal modes" type. These sinusoidal motions are "in phase" for the initial conditions (27a) if $\omega<\omega_2$.

Both definitions of principal modes lead to the same solution; they have the same physical interpretation and they are equivalent.

The discussion here is based on two definitions of principal modes and the final solution found by analyzing the infinite determinants involved. It may be mentioned that G. W. Hill, in his *Lunar Theory* brought into notice the infinite determinants, and H. Poincaré first gave conditions for their convergence.

ACKNOWLEDGMENT

The author is greatly indebted to Professor L. Brillouin for his unflagging interest and advice.

APPENDIX I. THE DOMINANT SUM OF A DOUBLE SERIES

Suppose in the sequence $\{\alpha_n\}$, $n=0, \pm1, \pm2, \cdots$ the complex elements have the property $\alpha_{-n}=\bar{\alpha}_n$. If α_1 and α_{-1} are dominant elements in the sequence, then it is easily seen that the dominant sum of the double series of (4a) is given by

$$3\alpha_1^2\alpha_{n-2}+6|\alpha_1|^2\alpha_n+3\bar{\alpha}_1^2\alpha_{n+2}, \tag{A1}$$

where n is any integer except $n=\pm1, \pm3$. For these exceptions the dominant terms of the double series are

$$3\bar{\alpha}_1\alpha_1^2 \text{ for } n=1, \quad 3\alpha_1\bar{\alpha}_1^2 \text{ for } n=-1,$$
$$\alpha_1^3 \text{ for } n=3, \quad \bar{\alpha}_1^3 \text{ for } n=-3. \tag{A2}$$

APPENDIX II. APPROXIMATE VALUE OF AN INFINITE DETERMINANT

If in an infinite convergent determinant,

$$\Delta=\|B_{m,n}\|_{-\infty}^{+\infty},$$

the elements in the main diagonal are equal to unity, and all the elements not in the main diagonal have a small common factor,[4] say ϵ, i.e., if $B_{m,m}=1$, $B_{m,n}=\epsilon\bar{B}_{m,n}$, $m\neq n$, we may get an expression of the determinant in powers of ϵ. The first term in this expression is independent of ϵ; it is the product of all the elements in the main diagonal, that is 1. The next terms in the expansion are in ϵ^2 and are obtained by replacing the product $\prod_m B_{m,m}$ the elements in the main diagonal (m,m) and (n,n) by the elements not in the main diagonal (m,n) and (n,m). These terms have a minus sign, according to the laws of determinants, and they are $-\epsilon^2\sum_m\sum_n \bar{B}_{m,n}\bar{B}_{n,m}$; the determinant Δ can be written in the form

$$\Delta=1-\epsilon^2\sum_m\sum_n \bar{B}_{m,n}\bar{B}_{n,m}+0(\epsilon^3). \tag{B}$$

APPENDIX III. A CONTINUED FRACTION AS A RATIO OF TWO INFINITE DETERMINANTS

Given the continued fraction

$$Z_\infty=\cfrac{1}{1+\cfrac{X_2}{1+\cfrac{X_3}{1+\cdots}}}, \tag{C1}$$

where the complex elements X are subject to specified conditions, its mth approximant Z_m, obtained by stopping with the mth partial quotient, can be estimated If the elements X of the sequence $\{X_m\}$ of (C1) have arbitrary values in a region, the "region element" E, then the correspondent series $\sum_{m=2}^{\rho}|X_m|$, $\rho=2, 3, 4, \cdots$ has its values in the "series region" Y, and the approximants Z_m have all their values in the value region V. The following theorems give relationships between the above regions.[5]

H. von Koch's Theorem

"If

$$Y: \quad \sum_{m=2}^{\rho}|X_m|<1, \quad \rho=2, 3, 4, \cdots \tag{C2}$$

then: the continued fraction (C1) converges."

Worpitzky's Theorem

"If

$$E: \quad |X_m|\leq\tfrac{1}{4}, \quad m=2, 3, 4\cdots \tag{C3}$$

then

$$V: \quad |Z_m-\tfrac{4}{3}|\leq\tfrac{2}{3}." \tag{C4}$$

[4] L. Brillouin, *Wave Propagation in Periodic Structures* (Dover Publications, New York, 1953), 2nd ed., pp. 34, 35.

[5] H. Wall, *Analytic Theory of Continued Fractions* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1948), pp. 26, 42, and 51.

The mth approximant of (C1), Z_m, can have the form of a ratio of two determinants. To show that, one associates the continued fraction (C1) a sequence of linear transformations

$$X_1(v)=1, \quad X_m=1/(1+X_m v), \quad m=2, 3, 4, \cdots.$$

If:

$$X_\rho \neq 0, \quad \rho=2, 3, \cdots, m; \quad X_\rho=0, \quad \rho>m,$$

then the product of m of the above transformations is

$$Z_m=X_1X_2\cdots X_m(v)=\cfrac{1}{1+\cfrac{X_2}{1+\cdot\cdot\cdot\cfrac{X_{m-1}}{1+X_m v}}}=\frac{X_m v A_{m-2}+A_{m-1}}{X_m v B_{m-2}+B_{m-1}} \tag{C5}$$

where the A's and B's may be calculated by means of the recursion formulas

$$A_\rho=A_{\rho-1}+X_\rho A_{\rho-2}, \quad B_\rho=B_{\rho-1}+X_\rho B_{\rho-2}, \quad \rho=1, 2, 3, \cdots. \tag{C6}$$

For the above we require the initial values

$$A_{-1}=1, \quad A_0=0, \quad B_{-1}=0, \quad B_0=1, \quad X_1=1.$$

The mth approximant of (C1), Z_m, is given by (C5) if $v=0$, then it is equal to the ratio A_{m-1}/B_{m-1}.

The recursion formulas (C6) give two systems of homogeneous linear equations, one in the variables A, the other in the variables B. These systems give rise to two determinants, which give the values of the A's and B's. The B's are given by the determinant

$$B_{m-1}=\begin{vmatrix} 0 & -1 & 0 & 0 & \cdots & & \cdots & 0 \\ X_2 & 1 & -1 & 0 & \cdots & & \cdots & 0 \\ 0 & X_3 & 1 & -1 & \cdots & & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & & & & \cdots & X_{m-1} & 1 & -1 \\ 0 & & & & \cdots & 0 & X_m & 1 \end{vmatrix}, \tag{C7}$$

$$m=2, 3, \cdots.$$

The determinant for the A's can be obtained from the above determinant by omitting its first row and its first column. These determinants are different from zero.

The value of the continued fraction (C1) is given, by definition, by the limit $\lim_{\rho\to\infty}(A_\rho/B_\rho)$.

Note on the Algebraic Aspect of the Integration of a System of Ordinary Linear Differential Equations

EYVIND H. WICHMANN
*Department of Physics, University of California, Berkeley, California**
(Received March 16, 1961)

In this note the Lie algebra generated by the coefficient matrix of a system of ordinary, linear, first-order differential equations is considered. A systematic discussion, based on some well-known results in the theory of Lie albegras, is given for the reduction of the problem of integration of such a system. For the purposes of this note the integration of a system of equations for which the coefficient matrix does not depend on the independent variable is regarded as "elementary." It will be shown that the problem of integrating any system of linear ordinary differential equations can be reduced to the problem of integrating a set of such systems, each one of which has the property that the corresponding Lie algebra is *simple*, and in such a way that the sum of the dimensionalities of the Lie algebras of the reduced systems in the set does not exceed the dimensionality of the Lie algebra of the original system.

The application of the reduction principle to the equations of motion in classical mechanics and in quantum mechanics is considered. It is shown that the principle in question applies to a class of Hamiltonian equations of motion not customarily regarded as describing linear systems.

GROUP THEORETICAL INTERPRETATION OF THE GENERAL SOLUTION OF A SYSTEM OF ORDINARY LINEAR DIFFERENTIAL EQUATIONS

1

IN this part we shall define the class of ordinary linear differential equations which shall be the primary object of our study, and review briefly some well-known facts about such equations.

We consider a system E of linear ordinary differential equations. Imagine the system in first-order form; the problem of complete integration then consists in finding an N-by-N matrix $U(t)$ such that

$$dU(t)/dt=M(t)U(t), \quad U(0)=I. \tag{E}$$

Here, t is the real independent variable, restricted to an open interval S, which we take to contain the point $t=0$. The matrix $M(t)$, "the coefficient matrix," is defined on S; for simplicity we shall furthermore assume $M(t)$ to be continuous on S. This means no essential loss of generality in the discussion to be given.

It is well-known that, under the stated assumptions, a unique differentiable $U(t)$ exists on S, satisfying the equation and boundary conditions given in (E).

2

Let V be the *real* vector space of matrices, spanned by the matrices $M(t)$ as t varies on S. (The matrices $M(t)$ need *not*, of course, be real). Let $L(\mathrm{E})$ be the real Lie algebra of matrices generated by the elements of V, the Lie product being defined as the commutator. Let $G(\mathrm{E})$ be the group of matrices generated by the exponentials of the elements of $L(\mathrm{E})$.

As is well-known, the solution $U(t)$ of the system (E) may be interpreted as a parametrization of a continuous curve on the group manifold of $G(\mathrm{E})$. To make the matter clear, let $L(\mathrm{E})$ be of dimensionality d, and let $B_1, B_2, \cdots, B_d$ be a basis of $L(\mathrm{E})$. We may then write

$$M(t)=\sum_{k=1}^{d} B_k m_k(t), \tag{1}$$

where the $m_k(t)$ are real continuous functions of t on S.

Let t_0 be any point of S. We define

$$W(t;t_0)=U(t)U^{-1}(t_0).$$

It follows that W satisfies the equation

$$dW(t;t_0)/dt=M(t)W(t;t_0); \quad W(t_0;t_0)=I \tag{2}$$

for all t in S.

In some neighborhood N_0 of t_0, contained in S, W will be sufficiently close to the identity so that

$$W(t;t_0)=\exp(F(t;t_0)), \tag{3}$$

where $F(t;t_0)$ is an element of $L(\mathrm{E})$, and a continuous (matrix valued) function of t in N_0, such that $F(t_0;t_0)=0$.

Let us write

$$F(t;t_0)=\sum_{k=1}^{d} B_k f_k(t;t_0). \tag{4}$$

As shown by Magnus,[1] the functions $f_k(t;t_0)$ are uniquely determined in the neighborhood N_0 as particular solutions of a set of first-order ordinary differential equations, which in general are nonlinear. These differential equations are determined solely by the structure of the Lie algebra $L(\mathrm{E})$, and by the expansion coefficients $m_k(t)$, through which $M(t)$ is expressed with respect to some basis, as in (1).

We may write $U(t)=\exp(F(t;t_0))U(t_0)$ when t is in N_0. When t varies on N_0, the point $U(t)$ in the group manifold $G(\mathrm{E})$ traces out a segment of a continuous curve. This segment, by (4), is entirely determined by the differential equations of Magnus, and hence by the

* Research supported in part by the National Science Foundation.

[1] W. Magnus, Communs. Pure and Appl. Math. **VII**, 649 (1954).

structure of $L(\mathrm{E})$ and the expansion coefficients $m_k(t)$. No other properties of the matrix $M(t)$ enter into consideration.

3

In view of what has been said, the essence of the system (E) is thus the structure of the Lie algebra $L(\mathrm{E})$, and the way in which, as in (1), the coefficient matrix $M(t)$ is expressed as an element of $L(\mathrm{E})$.

Consider the Lie algebra of matrices $L(\mathrm{E})$ as a faithful representation of an abstract Lie algebra L. Let $R(L)$ be any faithful representation of L, such that the group G_R generated by the exponentials of $R(L)$ covers the group $G(\mathrm{E})$. Let

$$M_R(t)=\sum_{k=1}^{d} R(B_k)m_k(t).$$

Then the solution $U_R(t)$ of the system, (t in S)

$$dU_R(t)/dt=M_R(t)U_R(t), \quad U_R(0)=I \tag{5}$$

uniquely determines $U(t)$ through the mapping of G_R onto $G(\mathrm{E})$. The system (5) is thus equivalent to the system (E), and this procedure may be used in practice to "simplify" the integration of (E). Actually, it is not necessary that G_R covers $G(\mathrm{E})$ since the *curve* $U(t)$ on $G(\mathrm{E})$ is anyway uniquely determined by the *curve* $U_R(t)$ on G_R and vice versa, by the conditions that both curves be continuous. We may thus solve the system (E) by solving the system (5) where $R(L)$ is any faithful representation of L.

REDUCTION PRINCIPLES FOR THE SYSTEM (E)

4

In this note the term "reduction" is used in the following sense: As before let $L(\mathrm{E})$, or L, be the Lie algebra generated by the coefficient matrix $M(t)$ of the system (E), and let L be of dimensionality d. Suppose that two other systems of differential equations (E′) and (E″) with corresponding Lie algebras L' and L'' of dimensionalities $d'>0$ and $d''>0$, respectively, where $d'+d''=d$, can be explicitly constructed, such that the solution of (E) can be obtained by quadrature from the solutions of (E′) and (E″); then we say that the problem of integrating the system (E) has been reduced to the problem of integrating the systems (E′) and (E″). The systems (E′) and (E″) possibly may be further reduced. We will, in fact, show that the problem of integrating any system of linear ordinary differential equations can, in the above sense, be reduced to the problem of integrating a set of systems of equations, each one of which has the property that the corresponding Lie algebra is simple.

5

Consider the system (E) and the associated Lie algebra $L(\mathrm{E})$. Suppose that $L(\mathrm{E})$ is the direct sum of two proper ideals L_1 and L_2 of $L(\mathrm{E})$. Let L_1 and L_2 be of dimensionality d_1 and d_2, respectively; hence, $d=d_1+d_2$. Let $M(t)=M_1(t)+M_2(t)$ such that $M_1(t)\in L_1$ and $M_2(t)\in L_2$. M_1 and M_2 are clearly continuous on S. Furthermore, for any t', t'' in S

$$[M_1(t'),M_2(t'')]=0. \tag{6}$$

Consider the two systems of differential equations

$$\begin{aligned} dU_1(t)/dt&=M_1(t)U_1(t), \quad U_1(0)=I,\\ dU_2(t)/dt&=M_2(t)U_2(t), \quad U_2(0)=I. \end{aligned} \tag{7}$$

From (6) it follows that $[U_1(t'),M_2(t'')]=0$ for all t' and t'' on S. Hence, $U(t)=U_1(t)U_2(t)$ is the solution of (E).

Let G_1 and G_2 be the matrix groups generated by the exponentials of the elements of the algebras L_1 and L_2. The group $G(\mathrm{E})$ is then the direct product of the two normal subgroups G_1 and G_2. $U_1(t)$ and $U_2(t)$ are continuous curves on the group manifolds of G_1 and G_2, respectively; these curves are the images of $U(t)$ under the homomorphisms $G(\mathrm{E})\rightarrow G_1$ and $G(\mathrm{E})\rightarrow G_2$, respectively.

6

Let us now consider the case when $L(\mathrm{E})$ is the semidirect sum of the two Lie algebras L_1 and L_2. Let thus L_2 be a proper ideal of $L(\mathrm{E})$ such that the quotient algebra $L(\mathrm{E})/L_2$ is isomorphic to the Lie subalgebra L_1 of $L(\mathrm{E})$, that is,

$$\begin{aligned} [L_2,L(\mathrm{E})]&\subseteq L_2,\\ [L_1,L_1]&\subseteq L_1. \end{aligned} \tag{8}$$

Again, let $M(t)=M_1(t)+M_2(t)$, where $M_1(t)\in L_1$ and $M_2(t)\in L_2$. Let $U_1(t)$ be defined by

$$dU_1(t)/dt=M_1(t)U_1(t), \quad U_1(0)=I, \tag{9}$$

and let $U_2(t)$ be defined by $U(t)=U_1(t)U_2(t)$.

It follows that

$$dU_2(t)/dt=\bar{M}_2(t)U_2(t), \quad U_2(0)=I, \tag{10}$$

where

$$\bar{M}_2(t)=U_1^{-1}(t)M_2(t)U_1(t). \tag{11}$$

Since L_2 is an ideal of $L(\mathrm{E})$, it follows that $\bar{M}_2(t)\in L_2$ for all t in S.

Let G_1 and G_2 be the two matrix groups generated by the exponentials of L_1 and L_2. Then $U_1(t)\in G_1$ and $U_2(t)\in G_2$. The group $G(\mathrm{E})$ is the semidirect product of G_1 and G_2, such that G_2 is a normal subgroup of $G(\mathrm{E})$, and G_1 is a subgroup of $G(\mathrm{E})$ isomorphic to $G(\mathrm{E})/G_2$.

We thus have the result that, if the solution of (9)

is known, we may find $\bar{M}_2(t)$; if we then solve (10), we have in effect obtained the solution of (E). The two equations (9) and (10) have associated with them Lie algebras L_1 and L_2, which, by our assumption, are of lower dimensionalities than $L(\mathrm{E})$. We therefore consider the replacement of the original equation (E) by Eqs. (9)-(11) a reduction of the system (E).

7

We next consider two special cases of the reduction principles of Secs. 5 and 6.

(a) Suppose $L(\mathrm{E})$ is commutative. Then $L(\mathrm{E})$ is the direct sum of d one-dimensional Lie algebras. We have the well-known solution

$$U(t)=\exp\left(\int_0^t ds M(s)\right). \tag{12}$$

(b) Suppose $L(\mathrm{E})$ is solvable. As is well-known, the system (E) is soluble by quadrature in this case too. To see this we may employ the procedure of Sec. 6, or, much more simply, we may rely on the fact that every representation of a solvable Lie algebra is similar to a representation by triangular matrices. Hence, without loss of generality we may assume that $M(t)$ is triangular for all t in S. Such an equation is, however, directly soluble by quadrature.

8

Consider now the general case of Eq. (E). Let L_2 be the maximum solvable ideal of $L(\mathrm{E})$. Then, by Levi's theorem,[2] there exists a Lie subalgebra L_1 of $L(\mathrm{E})$ which is isomorphic to the quotient algebra $L(\mathrm{E})/L_2$, i.e., $L(\mathrm{E})$ is the semidirect sum of L_1 and L_2. Furthermore, L_1 is semi-simple.

If $L_2=L(\mathrm{E})$, then $L_1=0$, and the discussion of Sec. 7(b) applies and the system (E) is soluble by quadrature.

Suppose now that $L_1\neq 0$ and $L_2\neq 0$. The reduction principle of Sec. 6 applies. Notice that once the system (9) has been solved, the system (10) can be solved by quadrature, since the corresponding Lie algebra L_2, by the assumption that L_2 is the maximum solvable ideal of $L(\mathrm{E})$, is solvable. In solving the system (E), the essential problem is thus the solution of (9). The system (9) is distinguished by the property that the corresponding Lie algebra L_1, generated by the coefficient matrix, is semi-simple.

It may, of course, happen that $L_2=0$, in which case $L(\mathrm{E})$ is semi-simple. Summarizing, we have the result that, given an arbitrary system (E), the integration problem is either trivial (if $L(\mathrm{E})$ is solvable) or else equivalent to the problem of integrating an analogous system E' with the property that $L(\mathrm{E}')$ is semi-simple, and such that $\dim(L(\mathrm{E}'))\leqq\dim(L(\mathrm{E}))$.

[2] J. H. C. Whitehead, Proc. Cambridge Phil. Soc. 32, 229 (1936).

It may, however, be possible to carry the reduction a step further. Every semi-simple Lie algebra is the direct sum of simple Lie algebras; therefore, if $L(\mathrm{E}')$ is not simple, we may reduce the system (E') by the principles of Sec. 5. We thus have the general result:

Given an arbitrary system (E) of linear ordinary first-order differential equations, unless the Lie algebra $L(\mathrm{E})$ is simple, then the problem of integrating (E) always can be reduced as follows: (a) If $L(\mathrm{E})$ is solvable, then the system (E) is soluble by quadrature. (b) If $L(\mathrm{E})$ is not solvable the problem of integrating (E) can be reduced to the problem of integrating a set $\mathrm{E}^{(1)}, \mathrm{E}^{(2)}, \cdots, \mathrm{E}^{(p)}$ of systems of differential equations, each one of which has the property that the corresponding Lie algebra $L(\mathrm{E}^{(k)})$ is simple, and such that $d_1+d_2+\cdots+d_p\leqq d$, where $d_k=\dim(L(\mathrm{E}^{(k)}))$, and $d=\dim(L(\mathrm{E}))$. In the particular case that $p=1$ we have $d_1<d$. This means that once the solutions of the reduced systems $\mathrm{E}^{(k)}$ are all known, the solution of (E) may be obtained by quadrature.

DISCUSSION OF THE EQUATIONS OF MOTION IN CLASSICAL MECHANICS AND IN QUANTUM MECHANICS

9

In this part we shall consider the equations of motion of physical systems in the light of the theory presented in the preceding parts. Our aim is to emphasize a certain unifying point of view which is the following: Consider a quantum mechanical system, or a classical system within the framework of Hamiltonian mechanics. The motion, or time-development, of the quantum mechanical system is described by a one-parameter family $U(t)$ of unitary transformations defined on the Hilbert space of all state vectors of the system, where $U(0)=I$. The transformations $U(t)$ are determined, heuristically, by an equation of motion of the form (E), with the matrix $M(t)$ replaced by $(-iH)$, H being the Hamiltonian. If H does not depend on the time t explicitly, we have the case regarded as "trivial" in this paper, namely $U(t)$ forms a one-dimensional continuous (Abelian) group such that $U(t)=\exp(-itH(0))$. If, however, $H=H(t)$ does depend on the time t explicitly, the group generated by the transformations $U(t)$ is, in general, not one dimensional. It *may* happen, however, that the group generated by the transformations $U(t)$ is a finite dimensional Lie group G_q, in which case the motion of the quantum mechanical system is described by a continuous curve on G_q; to every point of G_q corresponds a unique unitary transformation U on the Hilbert space of states. To determine the curve on G_q we may employ the procedure outlined in paragraph 3, according to which we reformulate the original equations of motion into Eq. (5), where the *finite* dimensional matrices $M_R(t)$ are associated with any faithful finite dimensional representation of the Lie algebra L of G_q, i.e., with the

Lie algebra generated by the operators $(-iH(t))$ as t varies on some interval S. In this manner, we may determine the curve on G_q and thus determine $U(t)$ as a group element of G_q. The problem of determining how this group element acts *explicitly* on the Hilbert space still remains and is "trivial" in about the same sense as the solution of the equations of motion is "trivial" when H does not depend on the time. Nevertheless, something has been gained as we shall explain later.

In the case of a classical Hamiltonian system, we have an analogous situation. The motion of the system may be thought of as a motion of points in phase-space, or as a one-parameter family of contact transformations. Let us denote the contact transformations by $U(t)$; it is again of interest to consider the case when the transformations $U(t)$ generate a finite dimensional Lie group G_c. If this is the case, we may, by solving an equation like Eq. (5), determine $U(t)$ as a group element of G_c after which it remains to determine the explicit action of $U(t)$ on the canonical variables, i.e., to find the explicit realization of G_c as a group of contact transformations. It should be noted that the above discussion is *not* restricted to a linear system, i.e., a dynamical system for which the canonical variables satisfy linear equations of motion.

We may express the matter as follows: The equations of motion, in classical as well as in quantum mechanics, define a certain Lie group. If this group is finite dimensional, we may solve the equations of motion in the sense that we determine a curve on the group manifold by solving an equation like Eq. (5). If the problem of determining the explicit realization of the group elements as transformations on all dynamical variables is regarded as "trivial," the essence of the motion is thus the curve on the group manifold. This is an old and well-established principle, both in classical and quantum mechanics. Since it is a connecting link between the two disciplines, we felt it worthwhile to state it once more.

In view of what has been said, it is clear that the reduction principles formulated in paragraphs 4–8 are applicable to the equations of motion of a dynamical system, provided the Lie algebra associated with the equations of motion is finite dimensional.

10

Let us elaborate further on the equations of motion of a system in classical dynamics.

The equations (E) arise naturally as the actual equations of motion when we consider a linear dynamical system with a finite number of degrees of freedom; whether the system is described by a Hamiltonian or not is immaterial. The equations of motion may, of course, be inhomogeneous linear equations; if we then adjoin a constant to the dynamical variables we may achieve the homogeneous form (E). The well-known procedure whereby the complete solution to a system of inhomogeneous equations is obtained by quadrature in terms of the general solution of the corresponding homogeneous system and the "driving forces" is a special case of the reduction principle stated in paragraph 6. In the particular case that the system is described by a Hamiltonian function, the group $G(\mathrm{E})$ is either a subgroup of a real symplectic group (if H is a homogeneous quadratic function of the canonical coordinates and momenta), or else a subgroup of the semidirect product of the symplectic group with the translation group (if H is quadratic but not homogeneous).

Consider next a dynamical system, described by a Hamiltonian, and nonlinear in the sense that the canonical variables (q_k,p_k) do not satisfy linear equations of motion. This does not exclude the possibility that we may be able to find a set of dynamical variables, X_k, $k=1, 2, \cdots, N$, which do satisfy linear equations of motion. We are, of course, only interested in the case when this set can be determined *before* the Hamiltonian equations of motion have been integrated; otherwise nothing is gained by studying the variables X_k.

Let us thus assume that a set C_X of linearly independent dynamical variables $X_k=X_k(q,p)$ can be found such that the variables X_k do not depend explicitly on the time t, but only depend on t through their dependence on the canonical variables (q_k,p_k), and such that the linear vector space spanned by these variables is closed under the Poisson-bracket operation with the Hamiltonian $H(q,p;t)$. In actual motion of the system, the variables $X_k=X_k(t)$ then satisfy the equations of motion

$$\frac{d}{dt}X_k(t)=[-H(t),\ X_k(t)]'=\sum_{n=1}^{N}\bar{M}_{kn}(t)X_n(t), \quad (13)$$

where $[\ ,\ ]'$ denotes the Poisson bracket. These equations have the solution

$$X_k(t)=\sum_{n=1}^{N}\bar{U}_{kn}(t)X_n(0), \quad (14)$$

where the matrix $\bar{U}(t)$ satisfies the equation

$$d\bar{U}(t)/dt=\bar{M}(t)\bar{U}(t);\quad \bar{U}(0)=I. \quad (15)$$

11

Let us consider an arbitrary classical dynamical system described by the Hamiltonian $H(q,p;t)=H(t)$. Suppose the system is such that the Lie algebra L generated by $H(t)$ as t varies on some interval S, and where the Lie product is realized by the Poisson bracket, is finite dimensional of dimensionality d. Let $U(t)$ denote the contact transformation which relates the canonical variables (q_k,p_k) at time t to the canonical variables at time $t=0$. $U(t)$ is then a curve on the group manifold of the finite dimensional Lie group G_c generated by the exponentials of L, and the differential equation of the curve is given by Eq. (E), interpreted

abstractly, provided we write $M(t)=-H(t)$. To find this curve, we may solve the matrix differential equation (5) associated with a faithful representation of L, exploiting the fact that such a representation always exists.

Let us select a basis $B_k=B_k(q,p)$, $k=1, 2, \cdots, d$, of L where the dynamical variables B_k do not depend on the time t explicitly. We then may write $M(t)=-H(t)$ in the form given in Eq. (1).

A particular representation of L is the adjoint representation on L; the dynamical variables B_k form a set of variables with the properties of the set C_X discussed in paragraph 10. By solving the corresponding equations (13), we can thus obtain the time-dependence in actual motion of the d variables B_k; all of these need however not be independent and the number of constants of motion obtained in this manner cannot exceed d but may be smaller.

Consider again any set C_X of linearly independent dynamical variables as in paragraph 10. The Lie algebra of matrices $\bar{L}$ generated by the matrices $\bar{M}(t)$ is a representation of the Lie algebra L; precisely stated $\bar{L}$ is isomorphic with the quotient algebra of L with respect to the ideal J consisting of all dynamical variables in L which have vanishing Poisson brackets with all the X_k. If J is nonempty, the solution $\bar{U}(t)$ of (15) thus, under no circumstances, can give us the complete solution $U(t)$ of the Hamiltonian equations. It should be noted, however, that in many applications the complete solution is not necessarily desired.

We may express the matter as follows: We cannot, if we wish to obtain the complete solution, avoid integrating the system (E). A contact transformation which does not involve the time t cannot, of course, change the structure of L. Integrating equations like (15) for a set of variables C_X, in effect, amounts to integrating the equations which arise in the reduction procedure described in paragraphs 4–8. A time-dependent contact transformation on the other hand naturally changes the structure of L.

12

The situation described in paragraph 10 has its exact analog in quantum mechanics. The unitary transformations $U(t)$ mentioned in paragraph 9 have an action by conjugation on all dynamical variables (operators); in the Heisenberg picture, $U(t)$ describes the time development of the observables. If it happens that a finite dimensional real vector space spanned by the Hermitean operators $X_k(t)$, $k=1, 2, \cdots, N$, is stable under $U(t)$, we have, in analogy with (13)

$$\frac{d}{dt}X_k(t)=[iH(t), X_k(t)]=\sum_{n=1}^{N}\bar{M}_{kn}(t)X_n(t), \quad (16)$$

which equations are solved by $\bar{U}(t)$ given by Eq. (15). The matrix group generated by $\bar{U}(t)$ is a representation of the group G_q generated by $U(t)$.

In case G_q is finite dimensional, the discussion of paragraph 11 applies with minor changes. In particular, the method in which the adjoint representation is exploited should be noted. This method is employed, for instance, when one studies the precession of a spinning particle in a time-dependent magnetic field; i.e., the solutions are obtained from the solutions of the "classical" equations.[3]

ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. Friedman, and Professor B. Kostant, of the University of California, Berkeley, and Dr. A. Novikov of Stanford Research Institute, as well as Professor W. Magnus of New York University, New York, for advice and helpful discussion.

[3] F. Bloch, Phys. Rev. **70**, 460 (1946).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

General Perturbational Solution of the Harmonically Forced van der Pol Equation*

RAIMOND A. STRUBLE AND JOHN E. FLETCHER
North Carolina State College, Raleigh, North Carolina

(Received June 20, 1961)

Some formal techniques for the study of nonlinear oscillations are illustrated through a development of a general perturbational solution of the harmonically forced van der Pol equation. These techniques provide for the study of perturbations of nearly linear oscillations in almost complete generality. The intricate resonance problems associated with small divisors, and in this case leading to the entrainment phenomena, are treated with unusual ease and serve to illustrate both the versatility and the generality of the techniques.

I. INTRODUCTION

IN this paper, a general perturbational solution of the van der Pol equation

$$\frac{d^2x}{dt^2}-\mu(1-x^2)\frac{dx}{dt}+x=F\cos\lambda t \quad (1)$$

is obtained in the form of an asymptotic series. Typically the series takes the form

$$x=A\cos(t-\omega)+\mu x_1+\mu^2x_2+\cdots+\mu^Nx_N, \quad (2)$$

where μ is treated as a small positive constant. The formal solution (2) reveals the well-known characteristics of the nonlinear oscillator for small μ including the entrainment of harmonic, subharmonic, and

* Work sponsored by the U. S. Army Research Office, Durham, North Carolina.

abstractly, provided we write $M(t)=-H(t)$. To find this curve, we may solve the matrix differential equation (5) associated with a faithful representation of L, exploiting the fact that such a representation always exists.

Let us select a basis $B_k=B_k(q,p)$, $k=1, 2, \cdots, d$, of L where the dynamical variables B_k do not depend on the time t explicitly. We then may write $M(t)=-H(t)$ in the form given in Eq. (1).

A particular representation of L is the adjoint representation on L; the dynamical variables B_k form a set of variables with the properties of the set C_X discussed in paragraph 10. By solving the corresponding equations (13), we can thus obtain the time-dependence in actual motion of the d variables B_k; all of these need however not be independent and the number of constants of motion obtained in this manner cannot exceed d but may be smaller.

Consider again any set C_X of linearly independent dynamical variables as in paragraph 10. The Lie algebra of matrices $\bar{L}$ generated by the matrices $\bar{M}(t)$ is a representation of the Lie algebra L; precisely stated $\bar{L}$ is isomorphic with the quotient algebra of L with respect to the ideal J consisting of all dynamical variables in L which have vanishing Poisson brackets with all the X_k. If J is nonempty, the solution $\bar{U}(t)$ of (15) thus, under no circumstances, can give us the complete solution $U(t)$ of the Hamiltonian equations. It should be noted, however, that in many applications the complete solution is not necessarily desired.

We may express the matter as follows: We cannot, if we wish to obtain the complete solution, avoid integrating the system (E). A contact transformation which does not involve the time t cannot, of course, change the structure of L. Integrating equations like (15) for a set of variables C_X, in effect, amounts to integrating the equations which arise in the reduction procedure described in paragraphs 4–8. A time-dependent contact transformation on the other hand naturally changes the structure of L.

12

The situation described in paragraph 10 has its exact analog in quantum mechanics. The unitary transformations $U(t)$ mentioned in paragraph 9 have an action by conjugation on all dynamical variables (operators); in the Heisenberg picture, $U(t)$ describes the time development of the observables. If it happens that a finite dimensional real vector space spanned by the Hermitean operators $X_k(t)$, $k=1, 2, \cdots, N$, is stable under $U(t)$, we have, in analogy with (13)

$$\frac{d}{dt}X_k(t)=[iH(t), X_k(t)]=\sum_{n=1}^{N}\bar{M}_{kn}(t)X_n(t), \quad (16)$$

which equations are solved by $\bar{U}(t)$ given by Eq. (15). The matrix group generated by $\bar{U}(t)$ is a representation of the group G_q generated by $U(t)$.

In case G_q is finite dimensional, the discussion of paragraph 11 applies with minor changes. In particular, the method in which the adjoint representation is exploited should be noted. This method is employed, for instance, when one studies the precession of a spinning particle in a time-dependent magnetic field; i.e., the solutions are obtained from the solutions of the "classical" equations.[3]

ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. Friedman, and Professor B. Kostant, of the University of California, Berkeley, and Dr. A. Novikov of Stanford Research Institute, as well as Professor W. Magnus of New York University, New York, for advice and helpful discussion.

[3] F. Bloch, Phys. Rev. **70**, 460 (1946).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 2, NUMBER 6 NOVEMBER-DECEMBER, 1961

General Perturbational Solution of the Harmonically Forced van der Pol Equation*

RAIMOND A. STRUBLE AND JOHN E. FLETCHER
North Carolina State College, Raleigh, North Carolina

(Received June 20, 1961)

Some formal techniques for the study of nonlinear oscillations are illustrated through a development of a general perturbational solution of the harmonically forced van der Pol equation. These techniques provide for the study of perturbations of nearly linear oscillations in almost complete generality. The intricate resonance problems associated with small divisors, and in this case leading to the entrainment phenomena, are treated with unusual ease and serve to illustrate both the versatility and the generality of the techniques.

I. INTRODUCTION

IN this paper, a general perturbational solution of the van der Pol equation

$$\frac{d^2x}{dt^2}-\mu(1-x^2)\frac{dx}{dt}+x=F\cos\lambda t \quad (1)$$

is obtained in the form of an asymptotic series. Typically the series takes the form

$$x=A\cos(t-\omega)+\mu x_1+\mu^2x_2+\cdots+\mu^Nx_N, \quad (2)$$

where μ is treated as a small positive constant. The formal solution (2) reveals the well-known characteristics of the nonlinear oscillator for small μ including the entrainment of harmonic, subharmonic, and

* Work sponsored by the U. S. Army Research Office, Durham, North Carolina.

superharmonic responses. In addition, it reveals the global operation of the system and the intricate interplay of the system parameters.

The generality of the approach represents a departure from previous studies[1] which have been devoted mainly to the analysis of periodic or almost-periodic solutions and their stability characteristics. These special solutions are of first importance in applications but reveal only a fraction of the total operational behavior of a nonlinear system. Aside from their prime importance in applications, they appear to have received such overwhelming attention mainly due to the absence of any really sound and workable techniques for treating other cases. As a result, various formal procedures[2] have been introduced in an effort to exhibit some of the intermediate or transient behavior. These formal procedures often appear to indicate genuine characteristics[3] but are limited to what one might call "favorable cases". In contrast, the asymptotic method illustrated here provides for the study of the perturbational characteristics of nearly linear oscillations in almost complete generality. Thus, for example, one may now understand the special role played by periodic or almost-periodic responses, for these become imbedded within the general solutions. The stability characteristics, so important in applications, are revealed and all transient and intermediate behavior is exhibited. The basic limitation imposed by the method reflects the perturbational character of the results which are otherwise quite general.

The techniques employed in this paper were originally developed for the study of the motion of artificial satellites.[4] The application considered here serves to illustrate these basic techniques in another important area of nonlinear oscillation problems. Since our principal purpose is to illustrate these basic techniques, we have limited the discussion to those resonance aspects already well established. In future papers we shall discuss some of the more intricate resonance problems associated with almost-periodic forcing, multi-dimensional systems and related small divisor problems. The study here hinges mainly upon an appropriate treatment of the small divisor problem, but the situation is relatively simple owing to the restriction to a single degree of freedom and to a single input frequency. The procedure followed is related to a number of well-known approximate methods. However, it does not appear to be appropriate nor desirable in this preliminary, illustrative work to emphasize these relationships. In keeping with this view, we have discussed a number of results which may be found elsewhere and have made no special effort to single out new results.

II. FREE OSCILLATIONS

It is instructive to first illustrate some of the features of the method as applied to the free oscillations of (1). We determine the asymptotic solution (2) by the familiar process of successive approximations, with certain innovations. For $\mu=0$, (2) yields the elementary general solution $x=A\cos(t-\omega)$, where the amplitude A and phase ω are arbitrary constants. For $\mu\neq 0$, we shall permit variations in each of A and ω and proceed to determine these variations successively to increasing orders in powers of μ together with the additive corrections $x_1, x_2, \cdots, x_N$. Thus we employ both the variation of parameters technique and a power series expansion procedure in one and the same process. There appears to be more degrees of freedom in our solution than the system (1) warrants, but this turns out to be at the very root of the success of the method. The extra degrees of freedom are needed to appropriately distribute the perturbations. The two constants of integration for the general solution are conveniently assigned to the variable parameters A and ω and each of the additive corrections $x_1, x_2, \cdots, x_N$ are to be expressed uniquely in terms of A, ω, and the independent variable t. The term $A\cos(t-\omega)$ of the expansion (2) will be referred to as the *principal* term.

With $F=0$, we write (1) in the form

$$\frac{d^2x}{dt^2}+x=\mu(1-x^2)\frac{dx}{dt}. \tag{3}$$

Then if we substitute (2) in (3), there results

$$\left[2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2\right]\cos(t-\omega)+\left[A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt}\right]\sin(t-\omega)+\mu\left(\frac{d^2x_1}{dt^2}+x_1\right)$$

$$+\mu^2\left(\frac{d^2x_2}{dt^2}+x_2\right)+\cdots+\mu^N\left(\frac{d^2x_N}{dt^2}+x_N\right)=\mu\{1-[A\cos(t-\omega)+\mu x_1+\cdots+\mu^N x_N]^2\}$$

$$\times\left\{A\sin(t-\omega)\frac{d\omega}{dt}-A\sin(t-\omega)+\frac{dA}{dt}\cos(t-\omega)+\mu\frac{dx_1}{dt}+\cdots+\mu^N\frac{dx_N}{dt}\right\}. \tag{4}$$

[1] For recent examples, see J. Hale, Ann. Math. **73**, 496 (1961); L. Cesari, Ann. Math. Studies **45**, 115 (1960); N. Bogoliubov and I. Mitropolsky, *Asymptotic Methods in the Theory of Nonlinear Oscillations*, Gos. Izd. Fiz. Mat., Moscow (1958); I. Malkin, *Some Problems in the Theory of Nonlinear Oscillations*, AEC-Translation 3766 (1956); S. Diliberto and G. Hufford, Ann. Math. Studies **36**, 207 (1956).

[2] N. Minorsky, *Introduction to Nonlinear Mechanics* (Edwards Brothers, Inc., Ann Arbor, Michigan, 1947).

[3] J. Ford, J. Math. Phys. **2**, 387 (1961).

[4] R. Struble, Arch. Rational Mech. Anal. **7**, 87 (1961).

If we consider only the terms through the first power in μ, (4) reduces to

$$\left[2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2\right]\cos(t-\omega)+\left[A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt}\right]\sin(t-\omega)+\mu\left(\frac{d^2x_1}{dt^2}+x_1\right)$$

$$=-\frac{\mu}{4}A(4-A^2)\sin(t-\omega)+\mu\frac{A^3}{4}\sin 3(t-\omega)+\mu A\sin(t-\omega)\frac{d\omega}{dt}$$

$$+\frac{\mu}{4}(4-3A^2)\cos(t-\omega)\frac{dA}{dt}-\frac{\mu}{4}A^2\cos 3(t-\omega)\frac{dA}{dt}+\frac{\mu}{4}A^3\sin 3(t-\omega)\frac{d\omega}{dt}. \quad (5)$$

An examination of the explicit terms in (5) suggests the following distribution

$$\begin{aligned} 2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2&=\frac{\mu}{4}(4-3A^2)\frac{dA}{dt},\\ A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt}&=\frac{-\mu A}{4}(4-A^2)+\mu A\frac{d\omega}{dt}, \end{aligned} \quad (6)$$

and

$$\frac{d^2x_1}{dt^2}+x_1=\frac{A^3}{4}\sin 3(t-\omega)\left(1-\frac{d\omega}{dt}\right)-\frac{A^2}{4}\cos 3(t-\omega)\frac{dA}{dt}. \quad (7)$$

Equations such as in (6) will be called *variational* equations while an equation such as (7) will be called a *perturbational* equation. At any step of the process the variational equations are always to be associated with the fundamental harmonic terms $\cos(t-\omega)$ and $\sin(t-\omega)$ and the perturbational equation with the remaining nonresonant terms.

Since we seek a solution of Eqs. (6) which is correct to first order in μ, they may be reduced to

$$\begin{aligned} 2A(d\omega/dt)&=0\\ 2(dA/dt)&=(\mu A/4)(4-A^2). \end{aligned} \quad (8)$$

Equations (8) yield the well-known first approximations

$$\omega=\omega_0, \text{ a constant},$$

$$A^2=4\Big/\left[1-\left(1-\frac{4}{A_0^2}\right)e^{-\mu t}\right] \quad (9)$$

with arbitrary ω_0 and $A_0\neq 0$. The perturbational equation (7) possesses the (approximate) particular integral

$$x_1=-(A^3/32)\sin 3(t-\omega) \quad (10)$$

which is correct to first order in μ, inasmuch as the derivatives of A and ω are each of at least first order in μ. Thus the solution for the free oscillation to first order in μ is given by the expression

$$x=A\cos(t-\omega)-\mu(A^3/32)\sin 3(t-\omega),$$

where A and ω are given by (9). Nontrivial amplitudes A always tend to the value 2 as $t\to\infty$ while $A=2$ corresponds to a stationary solution of the amplitude equation in (8). This is the familiar first approximation to the amplitude of the unique asymptotically, orbitally stable periodic solution of (3).

Using (10) in (4) and retaining all terms through second order in μ leads to the variational system

$$\begin{aligned} 2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2&=\frac{\mu^2A^5}{128}+\frac{\mu^2A(4-3A^2)(4-A^2)}{32}\\ A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt}&=\frac{-\mu A}{4}(4-A^2), \end{aligned} \quad (11)$$

and the perturbational equation

$$\mu^2\left(\frac{d^2x_2}{dt^2}+x_2\right)=\frac{9}{16}\mu A^2\frac{dA}{dt}\cos 3(t-\omega)+\frac{5\mu^2A^5}{128}\cos 5(t-\omega)-\frac{\mu^2A^3}{64}(14-5A^2)\cos 3(t-\omega). \quad (12)$$

Using (8), we may reduce the variational system to

$$\begin{aligned} \frac{d\omega}{dt}&=\frac{\mu^2}{16}+\frac{\mu^2}{256}(A^2-4)(A^2+4)+\frac{\mu^2}{128}(4-3A^2)(4-A^2)\\ \frac{dA}{dt}&=\frac{\mu}{8}A(4-A^2) \end{aligned} \quad (13)$$

which retains accuracy to second order in μ. The second of these is as before while the first may be expressed in the form

$$\frac{d\omega}{dt}=\frac{\mu^2}{16}+\frac{\mu}{32A}(4-7A^2)\frac{dA}{dt}.$$

Hence

$$\omega=\omega_0+\frac{\mu^2}{16}t+\frac{\mu}{8}\ln A-\mu\frac{7A^2}{64}. \quad (14)$$

Equation (14) introduces a second-order correction in the fundamental frequency of the periodic solution in addition to a slowly varying "phase" for the nonperiodic solutions. The amplitude A to second order in μ is

given by (9). Using (8) on the right in (12) yields

$$\frac{d^2x_2}{dt^2}+x_2=\frac{5A^5}{128}\cos5(t-\omega)+\frac{A^3}{128}(A^2+8)\cos3(t-\omega) \quad (15)$$

which in turn leads to the expression

$$x_2=-\frac{5A^5}{3072}\cos5(t-\omega)-\frac{A^3}{1024}(A^2+8)\cos3(t-\omega) \quad (16)$$

for the second-order additive correction. The complete second-order solution is

$$x=A\cos(t-\omega)-\mu\frac{A^3}{32}\sin3(t-\omega)$$
$$-\frac{\mu^2A^3}{1024}(A^2+8)\cos3(t-\omega)-\frac{\mu^25A^5}{3072}\cos5(t-\omega), \quad (17)$$

where A and ω are given by (9) and (14), respectively.

Using (16) in (4) and retaining all terms through third order in μ leads to the variational system

$$\frac{dA}{dt}=-\frac{\mu A}{8}\left[\mu^2\frac{37}{1024}A^6-\mu^2\frac{35}{128}A^4+\left(1+\frac{\mu^2}{2}\right)A^2-4\right]$$
$$\frac{d\omega}{dt}=\mu^2\frac{A^4}{256}+\frac{\mu^2}{128}(4-3A^2)(4-A^2). \quad (18)$$

The amplitude of the periodic solution now appears as a root of the algebraic equation

$$\mu^2\frac{37}{1024}A^6-\mu^2\frac{35}{128}A^4+\left(1+\frac{\mu^2}{2}\right)A^2-4=0. \quad (19)$$

Of course, its absolute value is very nearly equal to 2. In fact it can be shown[5] that for $0\leq\mu\leq0.5$, the pertinent root A_0 of (19) satisfies the inequalities

$$4+\frac{4\mu^2}{3\mu^2+64.064}\leq A_0^2\leq4+\frac{4\mu^2}{3\mu^2+64}.$$

The third-order perturbational equation reduces to

$$\frac{d^2x_3}{dt^2}+x_3=\frac{A^3}{1025}[7A^4-42A^2+16]\sin3(t-\omega)$$
$$-\frac{A^5}{(12)(1024)}[15A^2+280]\sin5(t-\omega)$$
$$-\frac{7A^7}{1536}\sin7(t-\omega) \quad (20)$$

and yields the third-order additive correction

$$x_3=\frac{-A^3}{8192}[7A^2-42A^2+16]\sin3(t-\omega)$$
$$+\frac{A^5}{294\,912}[15A^2+280]\sin5(t-\omega)$$
$$+\frac{7A^7}{73\,728}\sin7(t-\omega). \quad (21)$$

Solutions of arbitrary order in μ can be obtained in a similar manner; however, the labor involved soon becomes enormous. Clearly we view here an extension, to general solutions, of the classical Lindstedt-Poincaré,[6] procedure for the expansion of periodic solutions. In this instance, the procedure is equivalent to the method of higher approximations of Krylov and Bogoliubov.[7] The certainty with which one may recognize the resonant terms is a noteworthy feature of this type of application.

III. FORCED OSCILLATIONS

We now turn our attention to the forced oscillations. The versatility and generality of the asymptotic method becomes particularly evident in this type of application. We examine first the "soft" forced case, where the magnitude of the forcing term is small with μ. Equation (1) is written in the form

$$\frac{d^2x}{dt^2}+x=\mu(1-x^2)\frac{dx}{dt}+\mu k\cos\lambda t, \quad (22)$$

where μ, k, and λ are positive constants. Substituting the expansion (2) in (22) yields

$$\left[2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2\right]\cos(t-\omega)+\left[A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\right.$$
$$\left.\times\frac{d\omega}{dt}-2\frac{dA}{dt}\right]\sin(t-\omega)+\mu\left(\frac{d^2x_1}{dt^2}+x_1\right)+\cdots$$
$$+\mu^N\left(\frac{d^2x_N}{dt^2}+x_N\right)=\mu k\cos\lambda t-\frac{\mu A}{4}(4-A^2)$$
$$\times\sin(t-\omega)+\frac{\mu A^3}{4}\sin3(t-\omega)+0(\mu^2). \quad (23)$$

On the right we have anticipated that the derivatives dA/dt and $d\omega/dt$ will be first order in μ. The distribution of the terms in (23) into variational equations and a perturbational equation is no longer a straightforward

[5] R. Struble and J. Fletcher, ERD-106 Tech. Memo. 7, North Carolina State College, Raleigh, North Carolina.

[6] H. Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste* (Gauthier-Villars, Paris, 1892); A. Lindstedt, Mem. Sci. St. Petersbourg 31, 4 (1883).

[7] N. Krylov and N. Bogoliubov, *Introduction to Nonlinear Mechanics* (Princeton University Press, Princeton, New Jersey, 1947).

matter. For if λ is nearly equal to 1, the forcing term $\mu k \cos\lambda t$ is nearly resonant and would produce either a small divisor or a secular term if incorporated into the perturbational equation. In this case we write

$$\mu k \cos\lambda t = \mu k\{\cos(t-\omega)\cos[(\lambda-1)t+\omega] - \sin(t-\omega)\sin[(\lambda-1)t+\omega]\}$$

and in this form we may associate these terms with the variational equations in an obvious manner. The potentially small divisor or secular term is thus avoided. The first-order variational system for (22) becomes

$$\begin{aligned} &2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2=\mu k\cos[(\lambda-1)t+\omega], \\ &A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt} \\ &\quad=-\frac{\mu A}{4}(4-A^2)-\mu k\sin[(\lambda-1)t+\omega], \end{aligned} \tag{24}$$

and the perturbational equation becomes, as in the unforced case,

$$\frac{d^2x_1}{dt^2}+x_1=\frac{A^3}{4}\sin 3(t-\omega). \tag{25}$$

Each solution of the following reduced system

$$\begin{aligned} &\frac{dA}{dt}=\mu\frac{A}{8}(4-A^2)+\mu\frac{k}{\lambda+1}\sin[(\lambda-1)t+\omega], \\ &A\frac{d\omega}{dt}=\mu\frac{k}{\lambda+1}\cos[(\lambda-1)t+\omega] \end{aligned} \tag{26}$$

will satisfy (24) to first order in μ. The system (26), therefore, is sufficient to depict the first-order variations in amplitude A and phase ω for the resonant or nearly resonant case. We refer to this as the *harmonic* resonance case since resonance occurs with the forcing frequency near the frequency of the fundamental harmonic of the free oscillation. The perturbational equation (25) possesses the approximate particular integral (10). Thus the resonant or nearly resonant solution to first order in μ is

$$x=A\cos(t-\omega)-\mu\frac{A^3}{32}\sin 3(t-\omega), \tag{27}$$

as in the unforced case, where now, however, A and ω are determined to within integration constants by (26). In Sec. IV we discuss the implications of the system (26). In particular, we find the well-known phenomenon of fundamental harmonic frequency entrainment displayed.

When the forcing frequency λ is sufficiently different from 1, we need not shift the forcing term to the variational equations. The perturbational equation is then written

$$\frac{d^2x_1}{dt^2}+x_1=k\cos\lambda t+\frac{A^3}{4}\sin 3(t-\omega) \tag{28}$$

and the variational equations, as for the free oscillation, reduce to (8). The additive correction

$$x_1=\frac{k}{1-\lambda^2}\cos\lambda t-\frac{A^3}{32}\sin 3(t-\omega) \tag{29}$$

will satisfy (28) to first order in μ and consequently we may accept, as a first-order nonresonant solution,

$$x=A\cos(t-\omega)-\mu\frac{A^3}{32}\sin 3(t-\omega)+\mu\frac{k}{1-\lambda^2}\cos\lambda t, \tag{30}$$

where (nontrivial) A and ω are given by (9). The trivial solution $A=0$ of (8) here corresponds to an unstable periodic solution of frequency λ. The solution (30) contains both the forced and natural frequencies superimposed as though the system were linear. Since nontrivial $A \to 2$, the free response generally dominates.

Using (27) and (8) in (22) and retaining terms through second order in μ leads again to the variational system (13) and to the perturbational equation

$$\begin{aligned} \frac{d^2x_2}{dt^2}+x_2=&\frac{5A^5}{128}\cos 5(t-\omega)+\frac{A^3}{128}(A^2+8)\cos 3(t-\omega) \\ &-\frac{\lambda k}{2(1-\lambda^2)}(2-A^2)\sin\lambda t \\ &+\frac{A^2k(2+\lambda)}{4(1-\lambda^2)}\sin[(2+\lambda)t-2\omega] \\ &+\frac{A^2k(2-\lambda)}{4(1-\lambda^2)}\sin[(2-\lambda)t-2\omega], \end{aligned} \tag{31}$$

provided there are no resonant or nearly resonant terms in the latter. We observe that if λ is nearly equal to 3, however, the last term

$$\frac{A^2\mu k(2-\lambda)}{4(1-\lambda^2)}\sin[(2-\lambda)t-2\omega]$$

is nearly resonant and would introduce either a small divisor or a secular term in the additive correction x_2. In Sec. V it is shown that under favorable circumstances this leads to the phenomenon of entrainment of the free response at the subharmonic frequency $\lambda/3$. For λ

sufficiently different than 3, the second-order correction

$$x_2=-\frac{5A^5}{3072}\cos 5(t-\omega)-\frac{A^3}{1024}(A^2+8)\cos 3(t-\omega)$$
$$-\frac{\lambda k}{2(1-\lambda^2)^2}(2-A^2)\sin\lambda t-\frac{A^2k(2+\lambda)}{4(1-\lambda)(1+\lambda)^2(\lambda+3)}$$
$$\times\sin[(2+\lambda)t-2\omega]-\frac{A^2k(2-\lambda)}{4(1-\lambda)^2(1+\lambda)(\lambda-3)}$$
$$\times\sin[(2-\lambda)t-2\omega]$$

may be accepted since it satisfies (31) to zero order in μ. Repeated iterations which determine higher-order approximate solutions will reveal, at each step, new resonant or nearly resonant possibilities. However, increasing orders of the coefficients tend to decrease the practical importance of these phenomena inasmuch as the resonant ranges become correspondingly, increasingly small. The magnitudes of the effects, however, at or sufficiently near resonance may be significant. If the magnitude of the forcing term is large, some of the resonance phenomena appear in earlier iterations. If we write (22) in the form

$$\frac{d^2x}{dt^2}+x=\mu(1-x^2)\frac{dx}{dt}+F\cos\lambda t, \tag{32}$$

where the coefficient of the forcing term may not be small, the asymptotic series takes the form

$$x=A\cos(t-\omega)+\frac{F}{1-\lambda^2}\cos\lambda t+\mu x_1+\cdots+\mu^N x_N. \tag{33}$$

Of course we must except values of λ near 1 here.

Using (33) in (32) results in the equation

$$\left[2A\frac{d\omega}{dt}+\frac{d^2A}{dt^2}-A\left(\frac{d\omega}{dt}\right)^2\right]\cos(t-\omega)+\left[2A\frac{d^2\omega}{dt^2}+2\frac{dA}{dt}\frac{d\omega}{dt}-2\frac{dA}{dt}\right]\sin(t-\omega)+\mu\left(\frac{d^2x_1}{dt^2}+x_1\right)+\cdots+\mu^N\left(\frac{d^2x_N}{dt^2}+x_N\right)$$
$$=-\frac{\mu}{4}A[2(2-F_0^2)-A^2]\sin(t-\omega)+\mu\frac{A^3}{4}\sin 3(t-\omega)+\mu\frac{\lambda}{4}F_0[F_0^2-2(2-A^2)]\sin\lambda t+\mu F_0^3\frac{\lambda}{4}\sin 3\lambda t$$
$$+\mu\frac{A}{4}F_0^2(1+2\lambda)\sin[(1+2\lambda)t-\omega]+\mu\frac{A}{4}F_0^2(1-2\lambda)\sin[(1-2\lambda)t-\omega]+\mu F_0\frac{A^2}{4}(2+\lambda)\sin[(2+\lambda)t-2\omega]$$
$$+\mu F_0\frac{A^2}{4}(2-\lambda)\sin[(2-\lambda)t-2\omega]+0(\mu^2), \tag{34}$$

where $F_0=F/(1-\lambda^2)$. In this expansion, we observe three potentially resonant possibilities: λ near 3, λ near $\frac{1}{3}$ and λ near 0. The first of these possibilities occurred in (31) while the other two did not. (They are encountered at a later stage of the development in the soft forced case.) We discuss these resonant possibilities in Secs. V, VI, and VII.

If resonance is not a problem in (34), one obtains the variational system

$$\frac{dA}{dt}=\frac{\mu A}{8}[2(2-F_0^2)-A^2],\quad d\omega/dt=0$$

which possess the nontrivial solution

$$A^2=\frac{2(2-F_0^2)}{1-\left[1-\frac{2(2-F_0^2)}{A_0^2}\right]\exp\left[-\frac{\mu}{2}(2-F_0^2)t\right]},\quad \text{for } F_0^2\neq 2$$
or
$$A^2=\left(\frac{\mu t}{4}+\frac{1}{A_0^2}\right)^{-1},\quad \text{for } F_0^2=2 \tag{35}$$

with $\omega=$constant. It also possesses the trivial solution $A=0$. Thus if $F_0^2>2$, A^2 tends to zero as $t\to\infty$, while if $F_0^2<2$, A^2 tends to $2(2-F_0^2)$ as $t\to\infty$. In the former case the impressed force drives out the free response portion and the steady state is always periodic with frequency λ, while in the latter case, the free response portion persists into the steady state along with the forced response. The steady state is then termed a *combination* oscillation.

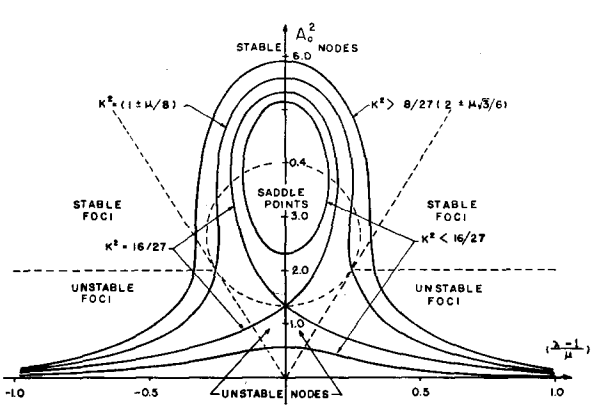

FIG. 1. Response curves.

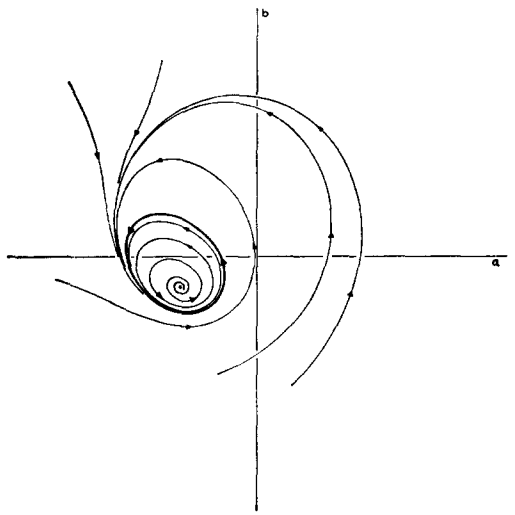

FIG. 2. Unstable focus.

IV. HARMONIC RESONANCE

The variational system (26) may be re-expressed in the autonomous form

$$\begin{aligned}\frac{da}{dt}&=\frac{\mu}{8}(4-A^2)a-(\lambda-1)b\\ \frac{db}{dt}&=\frac{\mu}{8}(4-A^2)b+(\lambda-1)a+\frac{\mu k}{\lambda+1}\end{aligned}\tag{36}$$

with

$$a=A\cos[(\lambda-1)t+\omega]$$

and

$$b=A\sin[(\lambda-1)t+\omega].$$

This system was originally derived by van der Pol[8] and has been discussed in detail by Andronow and Witt.[9]

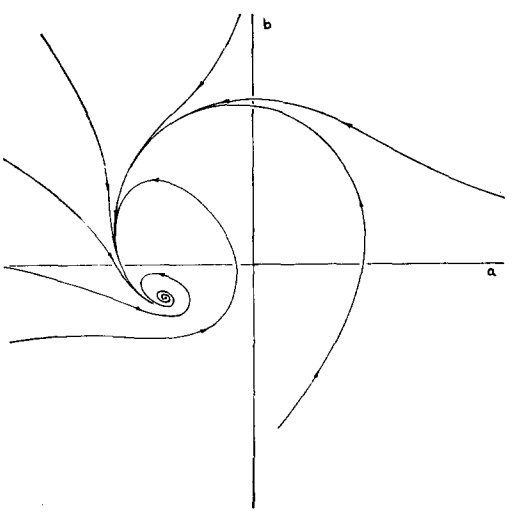

FIG. 3. Stable focus.

[8] B. van der Pol, Phil. Mag. 3, 65 (1927).
[9] A. Andronow and A. Witt, Arch. Electrotech. 24, 99 (1930).

However, the forms treated by these investigators limited the application to small values of the detuning $|(\lambda-1)/\mu|$. Here Eqs. (36) apply for both large and small values of the detuning, though for large values, it is best to re-express the results in the nonresonant form (30). Critical points of the system (36) depict stationary oscillations of (22). These are periodic solutions of frequency λ. In fact, with "$(\lambda-1)t+\omega$" constant, the frequency of the fundamental term $A\cos(t-\omega)$ becomes $1-(d\omega/dt)=\lambda$. A critical point (a_0,b_0) satisfies the algebraic system

$$\begin{aligned}\tfrac{1}{8}(4-A_0^2)a_0-[(\lambda-1)/\mu]b_0&=0\\ \tfrac{1}{8}(4-A_0^2)b_0+[(\lambda-1)/\mu]a_0+k/(\lambda+1)&=0\end{aligned}\tag{37}$$

and is conveniently expressed in the form

$$\begin{aligned}a_0&=[(1-\lambda^2)A_0^2]/\mu k\\ b_0&=[(1-\lambda^2)(4-A_0^2)A_0^2]/8\mu k,\end{aligned}\tag{38}$$

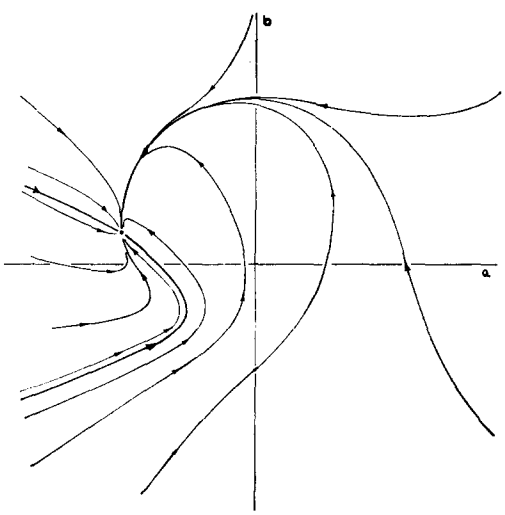

FIG. 4. Stable node.

where A_0^2 is a positive root of the equation

$$A_0^2\left[\frac{1}{64}(4-A_0^2)^2+\left(\frac{\lambda-1}{\mu}\right)^2\right]=\left(\frac{k}{\lambda+1}\right)^2.\tag{39}$$

The nature of a critical point is generally characterized by the Jacobian matrix

$$\begin{bmatrix}\tfrac{1}{8}(4-A_0^2-2a_0^2) & \dfrac{1-\lambda}{\mu}-\dfrac{a_0b_0}{4}\\ -\dfrac{1-\lambda}{4}-\dfrac{a_0b_0}{4} & \tfrac{1}{8}(4-A_0^2-2b_0^2)\end{bmatrix}$$

with determinant

$$\Delta=\frac{1}{64}(4-A_0^2)(4-3A_0^2)+\left(\frac{\lambda-1}{\mu}\right)^2$$

and trace

$$\Omega=\tfrac{1}{2}(2-A_0^2).\tag{40}$$

The locus $\Delta=0$ is an ellipse in the $[A_0^2, (\lambda-1)/\mu]$ plane, interior to which, $\Delta<0$ and exterior to which, $\Delta>0$. Thus interior to the ellipse $\Delta=0$, a critical point is always a saddle point while exterior to the ellipse $\Delta=0$, a critical point is a stable node or focus if $\Omega<0$ or an unstable node or focus if $\Omega>0$. The stability characteristics of a critical point of (36) reflect the orbital stability characteristics of the corresponding periodic solution of (22). Figure 1 illustrates some of the well-known response curves corresponding to harmonic resonance. Figures 2–10 depict a variety of the possibilities.[10] In Fig. 2, there is a single critical point which is an unstable focus. In such a case, a unique, asymptotically stable limit cycle forms about the critical point and the corresponding steady-state response of (22) is a combination oscillation. In Fig. 3, the single critical point is a stable focus and the corresponding steady-state response depicts harmonic entrainment. In Fig.

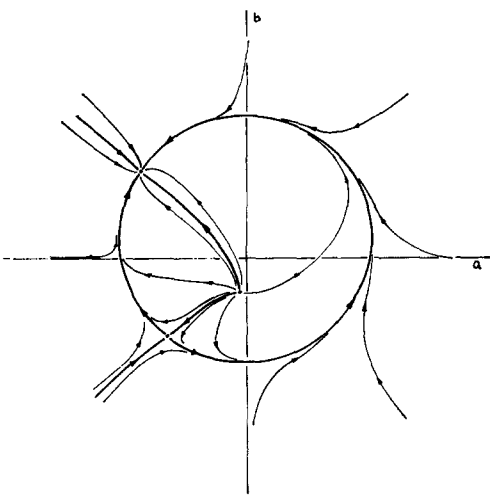

FIG. 5. Saddle point—stable node—unstable node.

4, harmonic entrainment is reflected in a stable node. Figure 5 depicts the circumstance wherein Eq. (39) possesses three real roots. One critical point is a saddle point, one is a stable node and one is an unstable node. In Fig. 6, the saddle point and stable node have coalesced to form an interesting unstable, asymptotically stable critical point. In Fig. 7, the saddle point and unstable node have coalesced to form an unstable critical point. In Fig. 8, one sees a saddle point, a stable node and an unstable focus. In Fig. 9, the focus has become stable. In the latter case, there are two possible steady-state responses, both depicting harmonic entrainment. In Fig. 10, there are two stable nodes and a saddle point. Again there are two possible steady-state responses depicting harmonic entrainment.

The second-order (harmonic) resonant equations are obtained by substituting the first-order solution (27) in (22) and using (26) to evaluate the second-order terms involving the derivatives dA/dt, $d\omega/dt$, d^2A/dt^2

[10] See also reference 9.

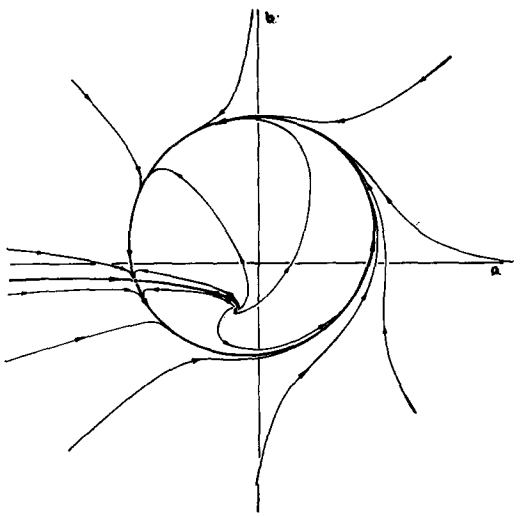

FIG. 6. Unstable, asymptotically stable point—unstable node.

and $d^2\omega/dt^2$. Here λ is assumed to be very nearly unity so that $\lambda-1$ is treated as a first order term in μ. When the terms are appropriately distributed, one obtains the second-order variational system

$$\begin{aligned}\frac{dA}{dt}&=\frac{\mu A}{8}(4-A^2)+\frac{\mu k}{2}\sin\Phi-\frac{\mu k(\lambda-1)}{4}\sin\Phi\\&\quad+\frac{\mu^2 k}{16}(4-A^2)\cos\Phi\\A\frac{d\Phi}{dt}&=A(\lambda-1)+\frac{\mu k\cos\Phi}{2}+\frac{\mu^2A^5}{256}+\frac{\mu^2A}{128}(4-3A^2)(4-A^2)\\&\quad+\frac{\mu^2k^2}{16A}+\frac{\mu^2k}{16}(4-3A^2)\sin\Phi-\frac{\mu k}{4}(\lambda-1)\cos\Phi\\&\quad-\frac{\mu^2k^2}{16A}\sin2\Phi+\frac{\mu^2k^2}{16A}\cos2\Phi,\end{aligned}\tag{41}$$

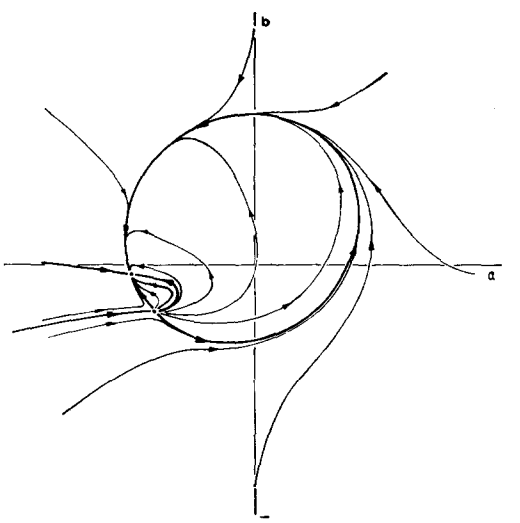

FIG. 7. Stable node—unstable critical point.

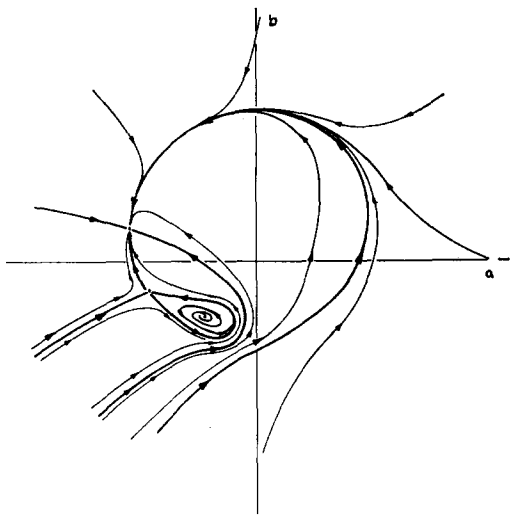

FIG. 8. Saddle point—stable node—unstable focus.

where $\Phi=(\lambda-1)t+\omega$. Of course for μ small, the near solutions of (41) are qualitatively similar to those of the first-order system. The second-order perturbational equation already contains a multitude of terms and so we choose not to exhibit it here. The higher-order resonant systems may be obtained though the computational work soon becomes enormous.

It is, perhaps, of some interest to consider the transition from nearly resonant to nonresonant operation. As the detuning $|(\lambda-1)/\mu|$ becomes large, the operation of the system is reflected in the response curves of Fig. 1 which are far removed from the A_0^2 axis. Unless the forcing amplitude factor k becomes correspondingly large, Eq. (39) possesses a single root which depicts an unstable focus, very near the origin in the ab plane. In such a case, the stable limit cycle which forms about this root (see Fig. 2) expands out to, and becomes almost coincident with, the circle $a^2+b^2=4$. The period of this limit cycle is no longer large. In fact, the fundamental frequency is approximately $\lambda-1$. The resulting steady-state operation given by (27) is a combination oscillation which is equivalent, to first order in μ, to that given by (30). Of course, the phase variable ω is markedly different in these two formulations. It is readily shown that the stability demarcation given by (40), which is the determining factor in producing a steady-state combination oscillation, leads, as $(\lambda-1)/\mu$ becomes large, to that obtained previously for the hard forced case at the end of Sec. III.

V. SUBHARMONIC RESONANCE

Subharmonic resonance occurs for λ near 3. In such a case, the term

$$\mu F_0(A^2/4)(2-\lambda)\sin[(2-\lambda)t-2\omega]$$

on the right in (34) is nearly resonant. It may be expressed in the form

$$\mu F_0(A^2/4)(2-\lambda)\{\cos[(\lambda-3)t+3\omega]\sin(t-\omega)+\sin[(\lambda-3)t+3\omega]\cos(t-\omega)\}$$

and so introduced into the variational system. To first order in μ, the latter may be reduced to

$$\begin{aligned} d\Phi/dt&=\lambda-3+\tfrac{3}{4}\mu\beta F_0 A\sin\Phi\\ dA/dt&=\mu/8[2(2-F_0^2)-A^2]A-\tfrac{1}{4}\mu\beta F_0 A^2\sin\Phi, \end{aligned} \tag{42}$$

where $\beta=(\lambda-2)/(\lambda-1)$ and $\Phi=(\lambda-3)t+3\omega$. With $a=A\cos\Phi$ and $b=A\sin\Phi$, the system (42) may be

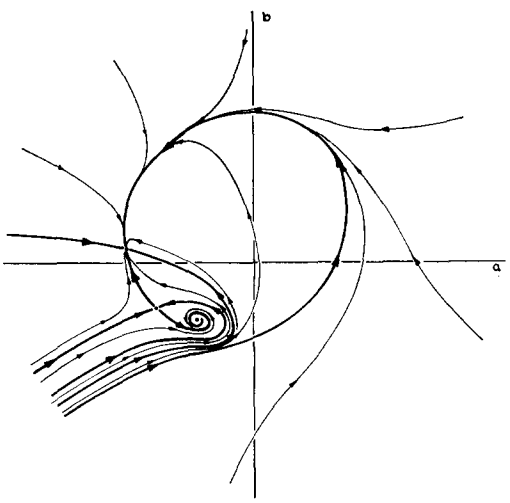

FIG. 9. Saddle point—stable node—stable focus.

expressed in the form

$$\begin{aligned} \frac{da}{dt}&=-\frac{\mu}{8}[2(2-F_0^2)-A^2]a-(\lambda-3)b-\frac{\mu}{4}\beta F_0(3b^2+a^2)\\ \frac{db}{dt}&=-\frac{\mu}{8}[2(2-F_0^2)-A^2]b+(\lambda-3)a+\frac{\mu}{2}\beta F_0 ab. \end{aligned} \tag{43}$$

Singular solutions of (42) depict stationary oscillations of (32). We find them to be $\frac{1}{3}$ subharmonic responses since $\Phi=(\lambda-3)t+3\omega$ is constant and hence $1-(d\omega/dt)=\lambda/3$. These correspond to the nontrivial critical points of (43). In addition (43) possesses a critical point at the origin. This also depicts a stationary oscillation of (32) but one which is a harmonic response. The singular solutions of (42) are given by

$$\begin{aligned} a_0&=(2\beta F_0)^{-1}[2(2-F_0^2)-A_0^2]\\ b_0&=-(4/3\beta F_0)[(\lambda-3)/\mu], \end{aligned} \tag{44}$$

where

$$A_0^2=2(2-F_0^2)+2\beta^2F_0^2\pm2\beta F_0\times\left[\beta^2F_0^2-\frac{16}{9\beta^2F_0^2}\left(\frac{\lambda-3}{\mu}\right)^2+2(2-F_0^2)\right]^{\frac{1}{2}}, \tag{45}$$

provided

$$\beta^2F_0^2-\frac{16}{9\beta^2F_0^2}\left(\frac{\lambda-3}{\mu}\right)^2+2(2-F_0^2)$$

is positive. Thus, subharmonic entrainment may occur for λ near 3 only if

$$F_0^2\leq 4/(2-\beta^2) \tag{46}$$

and

$$\left(\frac{\lambda-3}{\mu}\right)^2\leq\frac{9\beta^2F_0^2}{16}[\beta^2F_0^2+2(2-F_0^2)]. \tag{47}$$

The nature of a critical point (a_0,b_0) of the system (43), is generally characterized by the Jacobian matrix

$$\begin{bmatrix} \frac{\mu}{8}[2(2-F_0^2)-A_0^2-2b_0^2]+\frac{\mu}{2}\beta F_0a_0 & -\frac{\mu}{4}a_0b_0+\lambda-3+\frac{\mu}{2}\beta F_0b_0 \\ -\frac{\mu}{4}a_0b_0-(\lambda-3)-\frac{3\mu}{2}\beta F_0b_0 & \frac{\mu}{8}[2(2-F_0^2)-A_0^2-2a_0^2]-\frac{\mu}{2}\beta F_0a_0 \end{bmatrix}$$

with determinant

$$\Delta=\frac{\mu^2}{16}(2-F_0^2-A_0)^2-\frac{\mu^2}{64}A_0^2+\mu^2\beta F_0\left[2\left(\frac{\lambda-3}{\mu}\right)b_0^2-\frac{1}{8}a_0A_0^2\right]+\mu^2\beta^2F_0^2[b_0^2-\tfrac{1}{4}A_0^2]+(\lambda-3)^2 \tag{49}$$

and trace

$$\Omega=(\mu/2)(2-F_0^2-A_0^2). \tag{50}$$

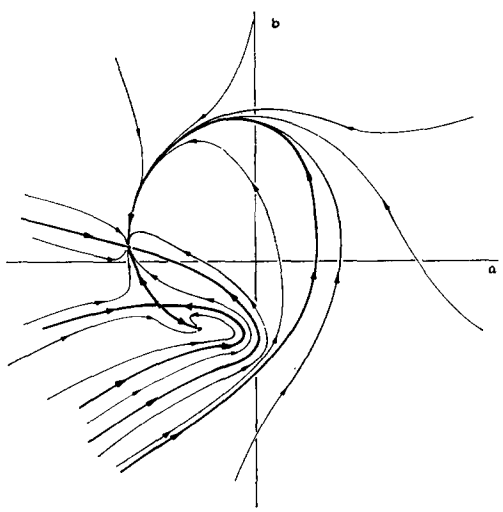

FIG. 10. Saddle point—stable node—stable node.

In addition, one readily obtains the equation

$$d(A^2)/dt=-(\mu A^2/4)[A^2-2(2-F_0^2)+2\beta F_0a] \tag{51}$$

which guarantees that for A^2 large, the trajectories are necessarily directed inward. Thus the index of a large circle about the origin is necessarily $+1$. Now the trivial solution $a=b=0$ of (43) corresponds to the circumstances

$\Delta=\mu^2[\frac{1}{16}(2-F_0^2)^2+(\{\lambda-3\}/\mu)^2]$ and $\Omega=(\mu/2)(2-F_0^2)$,

where $\Omega^2-4\Delta=-(\lambda-3)^2/4\leq 0$. Thus, except for degenerate cases, the origin is always a focus which is stable for $F_0^2>2$ and unstable for $F_0^2<2$. This is in agreement with the conclusions drawn at the end of Sec. III in regard to the harmonic periodic solutions. The sum of the indices of the remaining critical points of (43) necessarily vanishes. Thus if (44) and (45) define two nontrivial critical points, one is a saddle point and one is a node or a focus. The latter turns out to be a stable node. Of course, the origin will be the only critical point if either of the inequalities (46) or (47) is violated. Figures 11–14 depict the various possibilities. In Fig. 11, the origin is unstable and there is a saddle point and the stable node. The latter depicts subharmonic entrainment at the subharmonic $\lambda/3$. In Fig. 12, the saddle point and node have coalesced to form an unstable critical point. Note that every nontrivial trajectory has for its positive limit set this unstable, asymptotically stable critical point. In this case the system would not exhibit a stable subharmonic but the subharmonic would appear to depart and recur sporadically. In Fig. 13, the saddle point has coalesced with the origin to form an unstable critical point. Subharmonic entrainment is depicted by the remaining stable node. In Fig. 14, the saddle point has moved to an intermediate position between the other two singu-

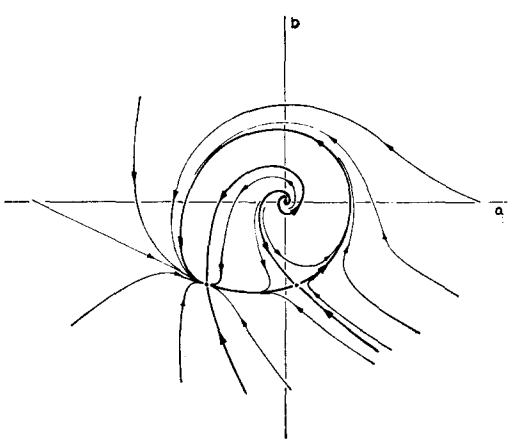

FIG. 11. Saddle point—stable node—unstable origin.

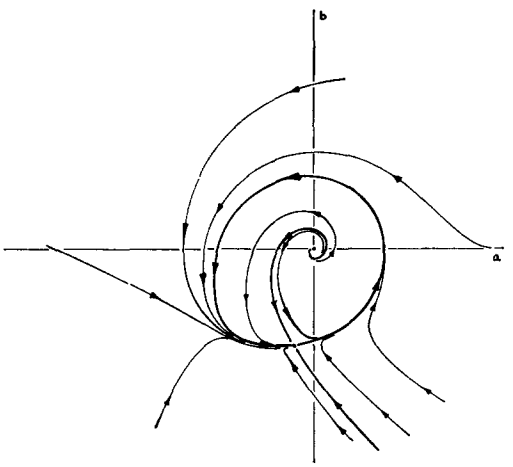

FIG. 12. Unstable, asymptotically stable point—unstable origin.

larities with the latter then stable. This depicts the very interesting circumstances wherein both stable subharmonic and stable harmonic entrainment takes place. Which steady state results, of course, depends upon the initial values.

Higher-order subharmonic resonance equations may be derived much as in the harmonic resonance case, but once again the computational work becomes enormous.

VI. SUPERHARMONIC RESONANCE

Superharmonic resonance occurs for λ near $\frac{1}{3}$. In such a case, the term

$$\mu F_0^3(\lambda/4)\sin 3\lambda t$$

on the right in (34) is nearly resonant. It may be expressed in the form

$$\mu F_0^3(\lambda/4)\{\cos[(3\lambda-1)t+\omega]\sin(t-\omega)+\sin[(3\lambda-1)t+\omega]\cos(t-\omega)\}$$

and so introduced into the variational system. To first order in μ, the latter may be reduced to

$$\begin{aligned}A\frac{d\omega}{dt}&=\mu\frac{F_0^3\lambda}{4(3\lambda+1)}\sin[(3\lambda-1)t+\omega]\\ \frac{dA}{dt}&=\frac{\mu A}{8}[2(2-F_0^2)-A^2]-\mu\frac{F_0^3\lambda}{2(3\lambda+1)}\cos[(3\lambda-1)t+\omega].\end{aligned}\tag{52}$$

With

$$a=A\cos[(3\lambda-1)t+\omega]$$

and

$$b=A\sin[(3\lambda-1)t+\omega],$$

(52) may be expressed in the form

$$\begin{aligned}\frac{da}{dt}&=\frac{\mu}{8}[2(2-F_0^2)-A^2]a-(3\lambda-1)b-\mu\frac{F_0^3\lambda}{4(3\lambda+1)}\\ \frac{db}{dt}&=\frac{\mu}{8}[2(2-F_0^2)-A^2]b+(3\lambda-1)a.\end{aligned}\tag{53}$$

For $F_0^2<2$, this system possesses exactly the same structure as the harmonic resonant system (36). In fact, a critical point (a_0,b_0) of (53) satisfies the algebraic equations

$$\begin{aligned}\frac{\mu}{8}[2(2-F_0^2)-A^2]a_0-(3\lambda-1)b_0-\frac{\mu}{4}\frac{F_0^3\lambda}{4(3\lambda+1)}&=0\\ \frac{\mu}{8}[2(2-F_0^2)-A^2]b_0+(3\lambda-1)a_0&=0,\end{aligned}\tag{54}$$

and from these we obtain the response equation

$$A_0^2\left\{\frac{1}{64}[2(2-F_0^2)^2-A_0^2]+\left(\frac{3\lambda-1}{\mu}\right)^2\right\}=\left[\frac{F_0^3\lambda}{4(3\lambda+1)}\right]^2.\tag{55}$$

To each positive root A_0^2 of (55) there corresponds a critical point of the system (53) given by the expressions

$$\begin{aligned}b_0&=-[4A_0^2(9\lambda^2-1)]/\mu\lambda F_0^3\\ \text{and}\qquad a_0&=\{[2(2-F_0^2)-A_0^2]A_0^2(3\lambda+1)\}/2\lambda F_0^3.\end{aligned}\tag{56}$$

These formulas parallel those of the harmonic resonance case. A critical point of (53) corresponds to a stationary oscillation of (32). However, the frequency in this case is readily found to be the superharmonic 3λ. The stability of the superharmonic response is generally

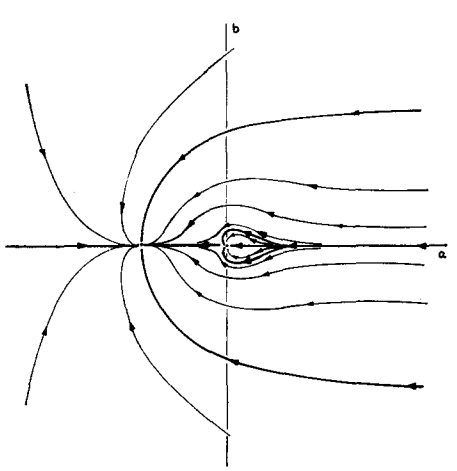

FIG. 13. Stable node—unstable origin.

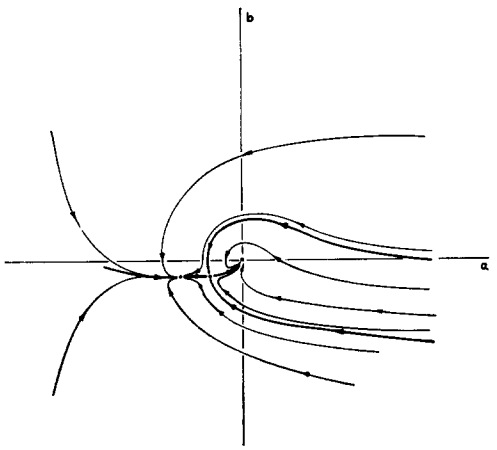

FIG. 14. Saddle point—stable node—stable origin.

reflected in the Jacobian matrix

$$\begin{bmatrix} \frac{1}{8}[2(2-F_0^2)^2-A_0^2-2b_0^2] & -\frac{3\lambda-1}{\mu}-\frac{a_0b_0}{4} \\ \frac{3\lambda-1}{\mu}-\frac{a_0b_0}{4} & \frac{1}{8}[2(2-F_0^2)-A_0^2-2a_0^2] \end{bmatrix}$$

with determinant

$$\Delta=\frac{1}{64}[2(2-F_0^2)-A_0^2][2(2-F_0^2)-3A_0^2]+\left(\frac{3\lambda-1}{\mu}\right)^2$$

and trace

$$\Omega=\tfrac{1}{2}[(2-F_0^2)-A_0^2].$$

The locus of saddle points is the interior of the ellipse

$$\frac{9}{4}\left[\frac{A_0^2-\frac{4}{3}(2-F_0^2)}{2-F_0^2}\right]^2+48\left[\frac{(3\lambda-1)/\mu}{2-F_0^2}\right]^2=1,$$

and if $2-F_0^2<A_0^2$ for any other critical point, it is certain to be stable.

For $F_0^2>2$, the system (53) possesses exactly one stable critical point. In this circumstance, the free response portion is always entrained at the frequency 3λ.

VII. OTHER RESONANCE PHENOMENA

When the input frequency λ is small, the two terms

$$(\mu AF_0^2/4)(1+2\lambda)\sin[(1+2\lambda)t-\omega]$$

and

$$(\mu AF_0^2/4)(1-2\lambda)\sin[(1-2\lambda)t-\omega]$$

on the right in (34) become nearly resonant. If their sum is expressed in the form

$$(\mu AF_0^2/2)[\cos 2\lambda t\sin(t-\omega)+2\lambda\sin 2\lambda t\cos(t-\omega)]$$

and introduced into the variational equations, the reduced system becomes

$$\begin{aligned} dA/dt&=\mu/8A[2(2-F_0^2)-A^2]-(\mu AF_0^2/4\sigma)\cos 2\lambda t \\ d\omega/dt&=(\mu F_0^2\lambda/2\delta)\sin 2\lambda t, \end{aligned} \tag{57}$$

where $\sigma=1-2\lambda^2$ and $\delta=(1-2\lambda^2)/(1-\lambda^2)$. It is readily verified that the formulas

$$A^2=A_*^2[1-(\mu F_0^2/4\lambda\sigma)\sin 2\lambda t]$$

and

$$\omega=\omega_0-(\mu F_0^2/4\delta)\cos 2\lambda t,$$

where A_*^2 is the amplitude variation given in (35), yield a solution of the system (57) which is accurate to first order in μ.

In seeking higher order solutions, either in the hard forced case or the soft forced case, other subharmonic and superharmonic resonance cases are disclosed. These may be investigated along similar lines. In addition, higher-order solutions for each of the resonant cases encountered may be obtained. The explicit calculations considered in this paper cover those cases of practical interest and are, perhaps, a sufficient indication of the general procedure to be followed.

The phase-plane solutions illustrated in the figures have been obtained with the aid of an analog computer. Except for Figs. 7–10, where the separation of the critical points is exaggerated, these illustrations are tracings of the machine drawn curves.

Transient Response of a Dipole Antenna*

TAI TSUN WU
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts
(Received May 11, 1961)

The current distributed for a dipole antenna driven by a step-function voltage is found shortly after the switch-on of the voltage.

ALTHOUGH many approximate methods have been devised to deal with the dipole antenna of finite length, no case has been solved exactly in a form appropriate for computation insofar as the author is aware. In this paper, the initial behavior of the transient response of a model of the dipole antenna is considered. If h is the half-length of the antenna, then for $x_0=ct<h$ the dipole behaves as though it were infinitely long and the problem can be solved exactly. The model of the dipole antenna which we shall adopt is a perfectly conducting, infinitely thin tube located at $r=a$, in our cylindrical coordinate system r, θ, x. The external voltage is applied at $r=a$, $x=0$:

$$E_x^{\text{ext}}=\begin{cases}\delta(x) & t>0,\\ 0 & t<0.\end{cases} \tag{1}$$

The problem is rotationally symmetrical with respect to θ.

Let G be the retarded Green's function for the scalar wave equation

$$(\nabla^2-\partial^2/\partial x_0^2)G(\mathbf{r},x_0)=-c\delta(\mathbf{r})\delta(x_0), \tag{2}$$

then the total current $\hat{x}I(x,x_0)$ on the dipole induces the vector potential

$$\mathbf{A}(\mathbf{r},x_0)=\mu_0\hat{x}\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dt'I(x',x_0')(2\pi)^{-1}\times\int_0^{2\pi}d\theta'(\mathbf{r}-\mathbf{r}',x_0-x_0'). \tag{3}$$

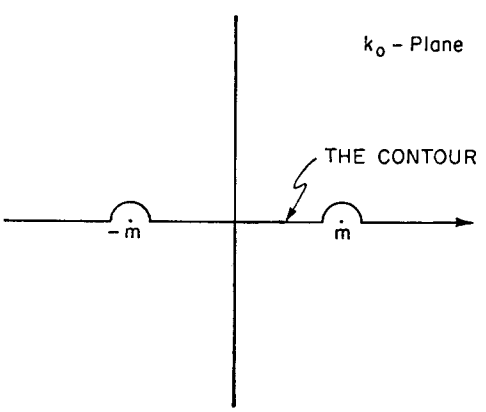

FIG. 1. Projection of C_0 on k_0 plane.

* This work was supported in part by contract.

On the other hand, it follows from (1) that for $r=a$, $\mathbf{A}$ satisfies

$$(\partial^2/\partial x^2-\partial^2/\partial x_0^2)\mathbf{A}\cdot\hat{x}=-c^{-1}\delta(x)\delta(x_0). \tag{4}$$

Since G is explicitly given by

$$G(\mathbf{r},x_0)=(4\pi x_0)^{-1}c\delta(r-x_0), \tag{5}$$

(3) and (4) lead to the integral equation for I

$$(\partial^2/\partial x^2-\partial^2/\partial x_0^2)\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dx_0'I(x',x_0')\times K(x-x',x_0-x_0')=-4\pi\zeta_0^{-1}\delta(x)\delta(x_0), \tag{6}$$

where ζ_0 is the characteristic impedance of free space, and the kernel K is

$$K(x,x_0)=(2\pi x_0)^{-1}\int_{-\pi}^{\pi}d\theta\delta([x^2+(2a\sin\theta/2)^2]^{\frac{1}{2}}-x_0). \tag{7}$$

Equation (6) may be solved exactly by a Fourier transformation with respect to both variables. In general, it is very hard to compute numerically the value of a double Fourier transform. However, in the present case, the double Fourier transform may be reduced to a single integral due to the invariance of (6) under the formal one-dimensional orthochronous Lorentz transformation

$$\begin{aligned}\mathfrak{L}x&=x\cosh\phi+x_0\sinh\phi,\\ \mathfrak{L}x_0&=x\sinh\phi+x_0\cosh\phi.\end{aligned} \tag{8}$$

To make use of this invariance, introduce a formal "photon mass" $m>0$, and consider (6) to be the limit as $m\to 0+$ of

$$(\partial^2/\partial x^2-\partial^2/\partial x_0^2-m^2)\int_{-\infty}^{\infty}dx'\int_{-\infty}^{\infty}dx_0'I(x',x_0')\times K(x-x',x_0-x_0')=-4\pi\zeta_0^{-1}\delta(x)\delta(x_0). \tag{9}$$

The Green's function g for the one-dimensional Klein-Gordon equation

$$(\partial^2/\partial x^2-\partial^2/\partial x_0^2-m^2)g(x,x_0)=-\delta(x)\delta(x_0) \tag{10}$$

has the representation

$$g(x,x_0)=-(2\pi)^{-2}\int_{S_0}dkdk_0(k_0^2-k^2-m^2)^{-1}\times\exp[i(kx-k_0x_0)]. \tag{11}$$

It is desired to choose the complex surface of integration S_0 so that g of (11) is the retarded Green's function and also that S_0 is invarient under the formal Lorentz transformation in the momentum space.

$$\mathfrak{L}k = k\cosh\phi + k_0\sinh\phi,$$
$$\mathfrak{L}k_0 = k\sinh\phi + k_0\cosh\phi. \tag{12}$$

Let C_0 be a contour where $k=0$ and k_0 ranges from $-\infty$ to ∞ with detours near $k_0=\pm m$ as shown in Fig. 1. Note that C_0 is required to go through the origin in k_0. Then a possible choice for S_0 is:

$$S_0: k, k_0 \text{ both real if } |k|>|k_0|$$
$$S_0=\mathfrak{L}C_0 \quad \text{if} \quad |k_0|>|k|.$$

The kernel K has the presentation

$$K(x,x_0)=(2\pi)^{-2}\int_{-\infty}^{\infty}dk\int_{-\infty}^{\infty}dk_0 L(k,k_0) \times\exp[i(kx-k_0x_0)], \tag{13}$$

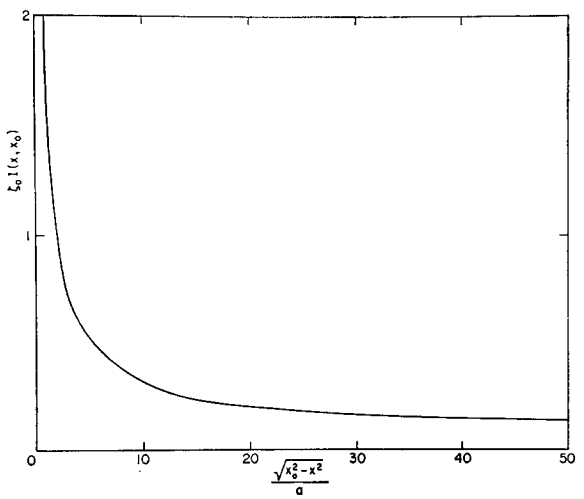

FIG. 2. Current distribution as a function of $(x_0^2-x^2)^{\frac{1}{2}}/a$.

where

$$L(k,k_0)=\begin{cases}\pi i H_0^{(1)}[a(k_0^2-k^2)^{\frac{1}{2}}]J_0[a(k_0^2-k^2)^{\frac{1}{2}}], \\ \qquad\qquad \text{for} \quad k_0>|k|, \\ -\pi i H_0^{(2)}[a(k_0^2-k^2)^{\frac{1}{2}}]J_0[a(k_0^2-k^2)^{\frac{1}{2}}], \\ \qquad\qquad \text{for} \quad k_0<-|k|, \\ 2K_0[a(k^2-k_0^2)^{\frac{1}{2}}]I_0[a(k^2-k_0^2)^{\frac{1}{2}}], \\ \qquad\qquad \text{for} \quad |k_0|<|k|.\end{cases} \tag{14}$$

Thus the solution of (9) is

$$I(x,x_0)=(2\pi)^{-1}\int_{S_0}dk\,dk_0 J(k,k_0)\exp[i(kx-k_0x_0)], \tag{15}$$

where

$$J(k,k_0)=-4\pi\zeta_0^{-1}(k_0^2-k^2-m^2)^{-1}[L(k,k_0)]^{-1}. \tag{16}$$

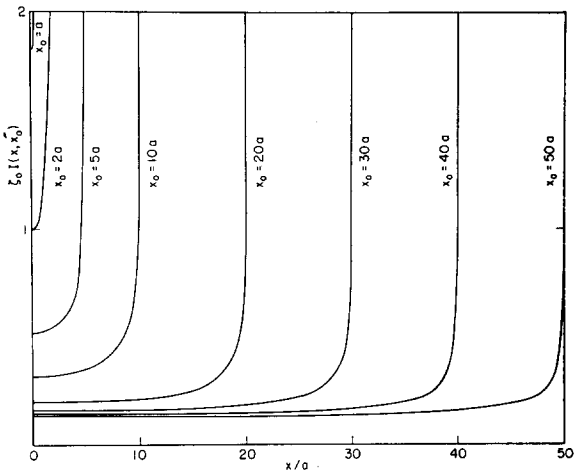

FIG. 3. Current distribution as a function of x for fixed x_0.

As shown in the Appendix, (15) may be reduced to, in the limit $m\to 0+$,

$$I(x,x_0)=4(\pi\zeta_0)^{-1}\int_0^{\infty}\frac{d\zeta}{\zeta}\frac{J_0[\zeta(x_0^2-x^2)^{\frac{1}{2}}/a]}{[J_0(\zeta)]^2+[Y_0(\zeta)]^2} \tag{17}$$

for $x_0>x$ and is zero otherwise. For the purpose of numerical computation it may be advantageous to use the alternative form

$$I(x,x_0) = 2\pi\zeta_0^{-1}\int_0^{\infty}\frac{d\zeta}{\zeta}\frac{K_0[\zeta(x_0^2-x^2)^{\frac{1}{2}}/a]I_0(\zeta)}{K_0(\zeta)\{[K_0(\zeta)]^2+\pi^2[I_0(\zeta)]^2\}}. \tag{18}$$

It is seen from (17) that, as $x_0^2-x^2\to 0+$,

$$I(x,x_0)\sim 2a\zeta_0^{-1}(x_0^2-x^2)^{-\frac{1}{2}}. \tag{19}$$

For $x=0$, the singularity at $x_0=0$ is x_0^{-1}. This is not integrable and is responsible for the logarithmic singularity previously found for harmonic time dependence at $x=0$.[1] On the other hand, in the limit $x_0^2-x^2\to\infty$, various approximate techniques may be tried.

The behavior of the integral on the right-hand side of (18) is shown in Fig. 2. This curve contains all the information about the current distribution. In particular, it is straightforward to read off the dependence of I on x for various fixed values of x_0. This is shown in Fig. 3.

Two remarks may be added. First, the transient response of the dipole as found is very different from that of a lossy transmission line excited in the same manner. Secondly, the response to a rectangular pulse may be obtained by superimposing two oppositely directed step functions separated by a finite interval of time.

[1] T. T. Wu and R. W. P. King, J. Appl. Phys. **30**, 74 (1959).

ACKNOWLEDGMENTS

For helpful discussions, I am indebted to Professor Ronold W. P. King, Professor George F. Carrier, and Professor Hans J. Schmitt.

APPENDIX

In order to derive (17) from (15), let the surface S_0 be divided into four pieces:

(1) $\mathrm{Re}k_0 > |k|$; (2) $k > |k_0|$;
(3) $-\mathrm{Re}k_0 > |k|$; and (4) $-k > |k_0|$.

Let I_i, $i=1\cdots4$, be the contribution to I of (15) from these four regions, respectively. With the variable $(k_0^2-k^2)^{\frac{1}{2}}$, it may be verified that

$$I_i(x,x_0) = -(\pi\zeta_0)^{-1}\int_{C_i}\zeta d\zeta[(\zeta^2-m^2) \times \pi i H_0^{(1)}(a\zeta)J_0(a\zeta)]^{-1}F_i(x,x_0;\zeta). \quad \text{(A1)}$$

In (A1), the four contours are as follows: C_1 is from 0 to ∞ along the positive real axis except for an upward detour near $\zeta=m$; C_2 and C_4 are both from $i\infty$ to 0 along the positive imaginary axis; and C_3 is from 0 to $-\infty$ along the negative real axis except for an upward detour near $\zeta=-m$. The functions F_i are explicitly

$$F_1(x,x_0;\zeta)=\begin{cases} -\pi i H_0^{(2)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], \ \text{for} \quad x>|x_0|, & \\ \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], \ \text{for} \quad -x>|x_0|, & \end{cases} \quad \text{(A2)}$$

$$F_2(x,x_0;\zeta)=\begin{cases} \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}]+2\pi i I_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x>|x_0|, & \\ \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], \ \text{for} \quad -x>|x_0|, & \end{cases} \quad \text{(A3)}$$

$$F_3(x,x_0;\zeta)=\begin{cases} \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}]+2\pi i J_0[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}]+2\pi i I_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x>|x_0|, & \\ \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}]+2\pi i I_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x>|x_0|, & \end{cases} \quad \text{(A4)}$$

and

$$F_4(x,x_0;\zeta)=\begin{cases} \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], \ \text{for} \quad x>|x_0|, & \\ \pi i H_0^{(1)}[\zeta(x_0^2-x^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x_0>|x|, & \\ 2K_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}]+2\pi i I_0[\zeta(x^2-x_0^2)^{\frac{1}{2}}], & \\ \qquad\qquad \text{for} \quad -x>|x_0|. & \end{cases} \quad \text{(A5)}$$

From (A2)–(A5) it is seen that $I_1+I_4=I_2+I_3=0$ for $x_0<x$ and $I_1+I_2=I_3+I_4=0$ for $x_0<-x$. Thus $I=0$ unless $x_0>|x|$. For $x_0>|x|$, (A2)–(A5) give

$$I(x,x_0)=2(\pi\zeta_0)^{-1}\int\zeta d\zeta(\zeta^2-m^2)^{-1}J_0[\zeta(x_0^2-x^2)^{\frac{1}{2}}] \times[H_0^{(1)}(a\zeta)J_0(a\zeta)]^{-1} \quad \text{(A6)}$$

where the contour of integration is that of Fig. 1. Equation (17) now follows readily.