JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 3, NUMBER 5 SEPTEMBER–OCTOBER 1962

On Feynman Quantization*

S. S. SCHWEBER†
Department of Physics and Laboratory for Nuclear Science,
Massachusetts Institute of Technology,
Cambridge, Massachusetts
(Received January 4, 1962)

The observation by Klauder that in the space of the $a = (1/\sqrt{2})(p + iq)$ variables, the Feynman integral can be defined in terms of a Gaussian measure, forms the basis of a presentation of the Feynman formulation of nonrelativistic quantum mechanics. The extension of this formulation to the case of a Bose field is sketched.

INTRODUCTION

THE alternate approach to quantum mechanics given by Feynman[1,2] has a great deal of intuitive appeal as it makes clear, *ab initio*, the logical structure of quantum mechanics.[3] It also makes clearer the nature of the limiting situation encompassed by classical mechanics.[4] Moreover, it may be the case that this particular formulation is somewhat more general than the "historical" one based on the correspondence between observables and linear self-adjoint operators and states to vectors in Hilbert space.

In Feynman's formulation of the quantum dynamics for a one-particle system, a prescription is given for the calculation of the transformation function $\langle q''t'' \mid q't' \rangle$ (the absolute value squared of which gives the probability density for the particle, whose dynamics is described by a Hamiltonian operator H, to be found at the position q'' at time t'' if its initial position at time t' was q') by assigning a complex probability amplitude to each space-time path starting at q' at t' and ending at q'' at t''. To the space-time path $q(t)$ [with $q(t') = q'$ and $q(t'') = q''$] Feynman associates the complex amplitude Φ

$$\Phi\,[\text{path } q(t)] = \exp\left[\frac{i}{\hbar}\int_{t'}^{t''} L(q(t), \dot{q}(t))\,dt\right], \quad \text{(1a)}$$

where $L(q(t), \dot{q}(t))$ is the classical Lagrangian for the particle. In the right-hand side of Eq. (1a), the action

$$I_{[q(t)]}(t'', t') = \int_{t'}^{t''} L(q(t), \dot{q}(t))\,dt \quad \text{(1b)}$$

is evaluated for the particular path $q(t)$ under consideration. The amplitude $\langle q''t'' \mid q't' \rangle$ is then obtained as the sum of the amplitudes Φ over all paths joining $q't'$ to $q''t''$ and is usually written in the form

* This work is supported in part through AEC Contract AT(30-1)-2098, by funds provided by the U. S. Atomic Energy Commission, the Office of Naval Research, and the Air Force Office of Scientific Research.

† Permanent address: Department of Physics, Brandeis University, Waltham, Massachusetts.

[1] R. P. Feynman, Ph.D. dissertation, Princeton University (1942).

[2] R. P. Feynman, Revs. Modern Phys. **20**, 367 (1948).

[3] R. P. Feynman, "The Concept of Probability in Quantum Mechanics" in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, California, 1954). See also A. R. Hibbs in M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Inc., New York, 1959).

[4] See in this connection P. A. M. Dirac, Physik Z. Sowjetunion **3**, 64 (1933) and Revs. Modern Phys. **17**, 195 (1945), as well as the interesting article by E. T. Whittaker, Proc. Roy. Soc. (Edinburgh) **A, 61**, 1 (1941).

$$\langle q''t'' \mid q't' \rangle = \int \cdots \int \mathfrak{D}[q(t)] \times \exp\left[\frac{i}{\hbar}\int_{t'}^{t''} L(q(t), \dot{q}(t))\, dt\right]. \quad (2)$$

In order for Eq. (2) to provide a practical alternative formulation of quantum mechanics, it is necessary to make more precise the concept of integration over paths involved therein. Thus Feynman[1] interprets Eq. (2) as follows: Divide the time interval $t'' - t'$ into N equal parts of duration $\epsilon = (t'' - t')/N$ and denote the successive times by $t_0 = t', t_1, \cdots, t_N = t''$, with $t_{i+1} - t_i = \epsilon$. A path $q(t)$ is specified by the sequence of points $q_0 = q(t_0), \cdots q_k = q(t_k), \cdots$. In the limit as $N \to \infty$, $\epsilon \to 0$ but $N\epsilon = t'' - t'$, this sequence is expected to approximate the path $q(t)$. It is then assumed that the motion in the time interval t_k to t_{k+1} is described by the classical path joining q_k to q_{k+1}. The action associated with this segment is

$$S(q_{k+1}, q_k) = \int_{t_k, q_k}^{t_{k+1}, q_{k+1}} L(q(t), \dot{q}(t))\, dt. \quad (3)$$

Since in (3) the integration is carried out over the classical path joining q_k to q_{k+1}, for ϵ small enough S will only depend on q_{k+1} and q_k. The amplitude $\Phi[q(t)]$ associated with the path $q(t)$ is then written as $\Phi(q_0, \cdots q_N)$ with

$$\Phi(q_0, \cdots q_N) = \frac{1}{\mathfrak{N}} \exp\left[\frac{i}{\hbar}\sum_{k=0}^{N-1}\int_{t_k, q_k}^{t_{k+1}, q_{k+1}} L\, dt\right] = \prod_{k=0}^{N-1} \frac{\exp\left[i/\hbar \cdot S(q_{k+1}, q_k)\right.}{A_k}, \quad (4)$$

where $\mathfrak{N}$ and A_k are normalization factors. The A_k are actually independent of the particular path q_k to q_{k+1} and depend only on the mass of the particle and the time inteval. The Feynman principle [Eq. (2)] is then interpreted as stating that

$$\langle q''t'' \mid q't' \rangle = \lim_{\substack{\epsilon \to 0 \\ N \to \infty \\ N\epsilon = t'' - t'}} \int \cdots \int \prod_{k=0}^{N} \exp\left[iS(q_{k+1}, q_k)\right] \frac{dq_k}{A_k}. \quad (5)$$

Feynman[2] has shown that for any Lagrangian of the form $L = \frac{1}{2} m\dot{q}^2 - V_1(q)\dot{q} - V_2(q)$, the transformation function defined by the right-hand side of Eq. (5) satisfies for $t'' \neq t'$ the Schrödinger equation

$$\{i\hbar\, \partial t'' - H(q'', -i\hbar\, \partial q'')\}\langle q''t'' \mid q't' \rangle = 0 \quad (6a)$$

$$H(q, p) = p\dot{q} - L; \qquad p = \partial L/\partial \dot{q}, \quad (6b)$$

and the initial condition

$$\lim_{t'' \to t''} \langle q''t'' \mid q't' \rangle = \delta(q'' - q') \quad (6c)$$

as well as the composition law

$$\langle q''t'' \mid q't' \rangle = \int \langle q''t'' \mid q'''t''' \rangle \times \langle q'''t''' \mid q't' \rangle\, dq'''. \quad (7)$$

These results are identical to those obtained by computing

$$\langle q''t'' \mid q't' \rangle = \langle q'' \mid \exp\left[-(i/\hbar)H(t'' - t')\right] \mid q' \rangle, \quad (8)$$

where H is the Hamiltonian operator for the system.

If one does not wish to use this "Riemann" approach to the Feynman integral but instead tries to adopt Eq. (2) as the defining equation for quantum mechanics, one is then faced with two problems. The first consists in giving a precise specification of the class of paths over which to integrate. The "ansatz" of Feynman[5] essentially specifies this set to be the class of all continuous functions connecting $(q't')$ to $(q''t'')$. Although in the case of a single particle this prescription is evidentally sufficient, if one wishes to incorporate the Bose or Fermi statistics within the Feynman formalism when describing a system of identical particles, classes of "unruly" histories must also be considered.[6] The second difficulty is connected with the fact that in evaluating the right-hand side of Eq. (5) one encounters nonconvergent integrals of the form

$$\int_{-\infty}^{+\infty} \exp\left(ix^2\right) dx.$$

To give meaning to such integrals, Feynman gives the mass m of the particle a small negative imaginary part $-i\epsilon$. The rule is then that one is to pass to the limit $\epsilon \to 0+$ after all the integrations have been carried out. More generally, one can try to replace the factor i in the exponent by a parameter $\lambda = -\epsilon + i$ and take the limit $\epsilon \to 0+$ after the computations or, better still, note the analyticity properties of the Feynman integral as a function of λ and attempt an analytic continuation to $\lambda = i$. It is, however, not clear whether either of these procedures works in general.[7] This lack of absolute convergence also implies that mathematical difficulties are encountered in trying to give a rigorous meaning to the Feynman integral in terms of a measure over a suitably defined function space (i.e., in the space of paths).

[5] See in this connection K. O. Friedrichs, *Integration of Functionals* (New York University Institute of Mathematical Sciences, New York, 1957).

[6] J. R. Klauder, Ann. Phys. (New York) 11, 123 (1960).

[7] See in this connection the review article by I. M. Gel'fand and A. M. Yaglom, J. Math. Phys. 1, 48 (1960).

It has been noted by Klauder,[6] that this last difficulty is not encountered if, instead of dealing with the observable q of the particle, one deals with the non-Hermitian variables $a = (1/\sqrt{2})(q + ip)$ and $a^* = (1/\sqrt{2})(q - ip)$. These variables define a realization of Hilbert space in terms of entire analytic functions. This has been investigated in detail by Bargmann.[8] We shall, in the following exposition, establish the Feynman formalism in terms of the a variables, using the work of Bargmann as a basis. Thus, in Sec. II we review the necessary mathematical background. In III we derive the transformation function for the a variables. In IV we give a statement of the Feynman principle in terms of these variables. The extension of this method to the case of Bose fields is then established in Sec. V. In Secs. I–IV we shall limit ourselves to the case of a single nonrelativistic particle with one degree of freedom. These results can be easily generalized to the case of a finite number of degrees of freedom and to any finite number of particles.

II. THE HILBERT SPACE OF ENTIRE FUNCTIONS

In the usual formulation of quantum mechanics in the Heisenberg picture, a particle (with one degree of freedom) is described by a time independent state vector $|\Psi\rangle$ and by the Hermitian operators $q(t)$, $p(t)$ corresponding to the position and momentum observables. These operators obey the commutation rules

$$[q(t), p(t)] = i \qquad (\hbar = 1), \tag{9a}$$

$$[q(t), q(t)] = [p(t), p(t)] = 0. \tag{9b}$$

The dynamical behavior of the system is determined by the equations of motion of these operators

$$i\,\partial_t q(t) = -[q(t), H], \tag{10a}$$

$$i\,\partial_t p(t) = -[p(t), H], \tag{11}$$

where H is the Hamiltonian operator. The Schrödinger wave function $\psi(q', t)$ is given by the projection of $|\Psi\rangle$ on the eigenvectors $|q', t\rangle$ of the position operator $q(t)$:

$$\psi(q', t) = \langle q', t \mid \Psi\rangle, \tag{12a}$$

with

$$q(t) \mid q', t\rangle = q' \mid q', t\rangle, \tag{12b}$$

$$\int |q', t\rangle\, dq'\, \langle q', t| = 1. \tag{12c}$$

The vectors $|q''t'\rangle$ and $|q''t''\rangle$ at times t' and t'' are related by a unitary operator $U(t', t'')$

$$\langle q'', t''| = \langle q'', t'|\, U(t'', t'). \tag{13a}$$

For a conservative system, as the one under consideration, H does not depend on the time explicitly and is, in fact, constant in time. In this case, the time translation operator U is given by

$$U(t'', t') = e^{-iH(t''-t')}. \tag{13b}$$

We are interested in computing the transformation function

$$\begin{aligned}\langle q''t'' \mid q't'\rangle &= \langle q'', t'' \,|U(t'', t')|\, q', t'\rangle \\ &= \langle q'' \,|e^{-iH(t''-t')}|\, q'\rangle,\end{aligned} \tag{14}$$

which corresponds to the probability amplitude for the system to undergo a transition from the initial state $|q'\rangle$ at time t' to the state $|q''\rangle$ at time t''. The time evolution of the system arising from an initial configuration $\psi(q', t')$ can then be calculated as follows:

$$\begin{aligned}\langle q''t'' \mid \Psi\rangle &= \psi(q'', t'') \\ &= \int \langle q''t'' \mid q't'\rangle\, dq'\, \langle q't' \mid \Psi\rangle.\end{aligned} \tag{15}$$

To indicate the correspondence of the usual formulation of quantum mechanics with the Feynman prescription for computing the propagator $\langle q''t'' \mid q't'\rangle$, subdivide the time interval $t'' - t'$ into $N + 1$ equal time intervals of duration ϵ and repeatedly insert the completeness relation (12b) to obtain

$$\begin{aligned}\langle q''t'' \mid q't'\rangle = \int \cdots \int dq_1 \cdots dq_N\, \langle q'' \,|e^{-iH\epsilon}|\, q_N\rangle \\ \langle q_N \,|e^{-iH\epsilon}|\, q_{N-1}\rangle \cdots \langle q_1 \,|e^{-iH\epsilon}|\, q'\rangle.\end{aligned} \tag{16}$$

In the limit as $N \to \infty$, it can be shown[9] that for a suitable class of Hamiltonian Eq. (16) reduces to Eq. (5), the Feynman definition of the transformation function. This approach, however, suffers from the difficulties mentioned above, namely, that one en-

[8] V. Bargmann, Commun. Pure and Appl. Math. **14**, 187 (1961). Such Hilbert spaces of analytic functions have also been considered by I. E. Segal in his lectures at the University of Colorado during the summer of 1960. J. Schwinger has made extensive use of such non-Hermitian variables see, e.g., "Differential Equations of Quantum Field Theory" lectures by J. Schwinger, given at Stanford during the summer of 1957. (unpublished). The first use of such non-Hermitian variables is due to V. A. Fock, Z. Physik. **49**, 339 (1928) and Physik. Z. Sowjetunion **6**, 425 (1934). For the applications of related methods to problems in functional integration in field theory see also the review article by J. V. Novozilov and A. V. Tolub, Fortschr. Physik **6**, 50 (1958). See also J. Schwinger, Proc. Natl. Acad. Sci. U. S. **46**, 1401 (1960). I am indebted to Professor Schwinger for a personal communication regarding this work which contains material related to the present investigation.

[9] See, for example, W. Tobocman, Nuovo cimento (10), **3**, 1213 (1956).

counters integrals of rapidly oscillating functions, which in general will not be convergent, and it is necessary to resort to certain devices such as considering the mass m of the particle to have an imaginary part and to carry out the limiting procedure of allowing this imaginary part to go to zero only after all the integrations have been performed.

In the above, the resolution of the identity corresponding to the position operator has been made use of and some of the convergence difficulties encountered with the Feynman integral can be traced to this decomposition. It is, of course, true that it is in terms of such a decomposition that the integral over paths has its most intuitive meaning. It turns out however that some of the convergence difficulties are alleviated if instead of the complete set $|q', t'\rangle$, one considers the set of eigenvectors $|a', t'\rangle$ corresponding to the operator

$$a(t) = (1/\sqrt{2})\{q(t) + ip(t)\}. \tag{17a}$$

The properties of this operator and of its adjoint

$$a^*(t) = (1/\sqrt{2})\{q(t) - ip(t)\} \tag{17b}$$

which will be needed are the following: First that $a(t)$ is not a Hermitian operator; it is not even a normal operator since from the commutation rules (9) for $p(t)$ and $q(t)$ we deduce that

$$[a(t), a^*(t)] = 1, \tag{18a}$$

$$[a(t), a(t)] = [a^*(t), a^*(t)] = 0. \tag{18b}$$

From these commutation rules one verifies in a well-known fashion that the positive semidefinite operator

$$N(t) = a^*(t)a(t) \tag{19}$$

has as its eigenvalues the positive integers and zero:

$$N(t) \mid n, t\rangle = n \mid n, t\rangle, \qquad n = 0, 1, 2, \cdots, \tag{20a}$$

with

$$|n, t\rangle = (n!)^{-1/2}\{a^*(t)\}^n \mid 0\rangle, \tag{20b}$$

and where $|0\rangle$ is the "no-particle" state characterized by

$$a(t) \mid 0\rangle = 0. \tag{21}$$

An explicit representation of the operators $a(t)$ and $a^*(t)$ is given by

$$a(t) \mid n, t\rangle = n^{1/2} \mid n - 1, t\rangle, \tag{22a}$$

$$a^*(t) \mid n, t\rangle = (n + 1)^{1/2} \mid n + 1, t\rangle. \tag{22b}$$

The eigenfunctions of the operator $a(t)$, $|a', t\rangle$, can be expressed in terms of the basis vectors $|n, t\rangle$ as follows:

$$|a', t\rangle = \sum_{n=0}^{\infty} (n!)^{-1/2}a'^n \mid n, t\rangle$$

$$= e^{a'a^*(t)} \mid 0\rangle. \tag{23}$$

In the right-hand side of Eq. (23) a' can take on all *complex* values. One verifies, using Eq. (22a), that the so-defined vector $|a', t\rangle$ has the property that

$$a(t) \mid z, t\rangle = z \mid z, t\rangle. \tag{24}$$

In view of the fact that the eigenvalue z can take on all complex values we must be careful in defining the adjoint vector to $|z\rangle$. This question does not arise in the case of eigenfunctions of Hermitian operators since their eigenvalues are real. We define the adjoint of the vector $|z\rangle$ by using the defining equation, Eq. (23), i.e., we define[10]

$$(|z', t\rangle)^* = \sum_{n=0}^{\infty} (n!)^{-1/2}\bar{z}'^n \langle n, t| = \langle 0| \exp [\bar{z}'a(t)] \tag{25a}$$

$$\equiv \langle z', t|. \tag{25b}$$

Note that we write $\langle z', t|$ even though it is $\bar{z}'$ which appears in the right-hand side of Eq. (25a).[10] By the vector $\langle \bar{z}', t|$ we mean

$$\langle \bar{z}', t| = \sum_{n=0}^{\infty} (n!)^{-1/2}z'^n \langle n, t|. \tag{26}$$

The vector $\langle z', t|$ is a left eigenvector of the operator $a^*(t)$ with eigenvalue $\bar{z}'$. These vectors are not normalized, nor are they orthogonal for different values of z' and z'' since

$$\langle z' \mid z''\rangle = \sum_{n,m=0}^{\infty} (n!\, m!)^{-1/2}\bar{z}'^n z''^m \langle n \mid m\rangle$$

$$= \sum_{n=0}^{\infty} \frac{\bar{z}'^n z''^n}{n!} = \exp [\bar{z}'z''], \tag{27a}$$

or alternatively

$$\langle z' \mid z''\rangle = \langle 0| \exp [\bar{z}'a(t)] \exp [z''a^*(t)] \mid 0\rangle$$

$$= \exp [\bar{z}'z''] \langle 0| \exp [z''a^*(t)] \exp [\bar{z}'a(t)] \mid 0\rangle$$

$$= \exp [\bar{z}'z'']. \tag{27b}$$

The specification of an arbitrary normalizable vector $|f\rangle$ by its components along the "axes" specified by the vectors $|z\rangle$ is then given by the quantities

$$f(z) = \langle \bar{z} \mid f\rangle = \sum_{n=0}^{\infty} (n!)^{-1/2}z^n \langle n \mid f\rangle. \tag{28}$$

The expansion (28) defines an analytic function of the complex variable z. Since Eq. (28) is defined for all values of the complex variable z, it in fact defines an *entire* analytic function. The vector space $\{f(z)\}$,

[10] $\bar{z}$ denotes the complex conjugate of z, i.e., $\bar{z} = x - iy$.

whose elements are entire analytic functions, can be made into a Hilbert space, $\mathfrak{H}_B$, by defining on it the following Hermitian scalar product

$$\langle f \mid g\rangle = \int d\mu(z)\, \overline{f(z)}\, g(z) \equiv (f, g), \tag{29}$$

where $d\mu(z)$ is the following real measure ($z = x + iy$)

$$d\mu(z) = (1/\pi) \exp(-\bar{z}z)\, dx\, dy. \tag{30}$$

In (29) the integration is carried out over all values of x and y; explicitly

$$\langle f \mid g\rangle = \frac{1}{\pi} \iint_{-\infty}^{+\infty} dx\, dy \exp(-\bar{z}z)\, \overline{f(z)}\, g(z). \tag{31}$$

Bargmann[8] has made a detailed study of this (separable) Hilbert space of entire functions, as well as some of its applications to quantum mechanics and group theory. We here collect the formulas that will be of relevance for our exposition.

If $f(z)$ and $g(z)$ are two entire functions with the following power-series expansions

$$f(z) = \sum_{n=0}^{\infty} \frac{\alpha_n}{(n!)^{1/2}} z^n, \tag{32a}$$

$$g(z) = \sum_{n=0}^{\infty} \frac{\beta_n}{(n!)^{1/2}} z^n, \tag{32b}$$

then

$$\langle f \mid g\rangle = \frac{1}{\pi} \iint_{-\infty}^{+\infty} dx\, dy\, e^{-|z|^2} \sum_{n,m=0}^{\infty} \frac{\bar{\alpha}_n \beta_m}{(n!\, m!)^{1/2}} \bar{z}^n z^m. \tag{33}$$

The integrations can be carried out by introducing polar coordinates, $z = |z|\, e^{i\phi}$, $\bar{z} = |z|\, e^{-i\phi}$, in which case

$$\iint dx\, dy \longrightarrow \tfrac{1}{2} \int_0^{\infty} d\,|z|^2 \int_0^{2\pi} d\phi.$$

In (33), the angular integration vanishes unless $n = m$, so that

$$\langle f \mid g\rangle = \int_0^{\infty} d\,|z|^2 \sum_{n=0}^{\infty} \frac{\bar{\alpha}_n \beta_n}{n!} |z|^{2n} e^{-|z|^2},$$

$$= \sum_{n=0}^{\infty} \bar{\alpha}_n \beta_n. \tag{34}$$

The square of the norm of the vector $|f\rangle$ is thus given by

$$||f||^2 = \langle f \mid f\rangle = \sum_{n=0}^{\infty} |\alpha_n|^2 < \infty. \tag{35}$$

From the above, one readily deduces that z^m (m integer) is orthogonal to z^n (n, integer), and that the vectors $|m\rangle$ with components $\langle z \mid m\rangle = u_m(z) = z^m/m!$, $m = 0, 1, 2, \cdots$ form an orthonormal basis in $\mathfrak{H}_B$. Their completeness is verified by noting that if $f(z)$ is given by Eq. (32a) then

$$\langle m \mid f\rangle = (u_m, f) = \alpha_m, \tag{36a}$$

so that

$$(f, f) = \sum_m \bar{\alpha}_m \alpha_m = \sum_m (f, u_m)(u_m, f). \tag{36b}$$

These facts are essentially a reflection of the completeness of the eigenfunctions of the number operator a^*a. Within $\mathfrak{H}_B$ the operators a and a^* have the following representation

$$(f, a^* g) = \int d\mu(z)\, \overline{f(z)}\, z g(z), \tag{37a}$$

$$(f, ag) = \int d\mu(z)\, \overline{f(z)}\, \frac{dg(z)}{dz}, \tag{37b}$$

i.e., the operator a^* corresponds to multiplication by z, and a to differentiation by d/dz. The latter operation is of course well defined, since the elements of $\mathfrak{H}_B$ are entire analytic functions.

The following useful bounds on $|f(z)|$ can be obtained by applying Cauchy's inequality to the right-hand side of the expression

$$|f(z)| < \sum_{m=0}^{\infty} |\alpha_m| \left| \frac{z^m}{(m!)^{1/2}} \right|$$

and yields

$$|f(z)|^2 < \left(\sum_{m=0}^{\infty} |\alpha_m|^2 \right) \left(\sum_{m=0}^{\infty} \frac{|z|^{2m}}{m!} \right)$$

$$< ||f||^2 e^{\bar{z}z}$$

so that

$$|f(z)| < ||f||\, \exp[(1/2)\bar{z}z].$$

By applying this inequality to the difference of two functions

$$|f(z) - g(z)| < \exp[(1/2)\bar{z}z]\, ||f - g||$$

one infers that the strong convergence in $\mathfrak{H}_B$ implies ordinary point-wise convergence on bounded sets. However the condition

$$|f(z)| < ||f||\, e^{(1/2)|z|^2}$$

is not sufficient to guarantee that the vector $|f\rangle$ is in $\mathfrak{H}_B$. The condition

$$|f(z)| < C \exp[(\lambda^2/2)\bar{z}z]$$

turns out to be sufficient.[8] Calling $f_\lambda(z) = f(\lambda z)$, it can be shown[8] that if $||f_\lambda|| < C$ for all λ such that $0 < \lambda < 1$, then $|f\rangle$ is an element of $\mathfrak{H}_B$ and

$$\lim_{\lambda \to 1} ||f - f_\lambda|| \to 0.$$

Consider next the mapping of the vectors $|f\rangle$ on the complex-valued functions $f(z)$ such that $\lambda_1 \mid f_1\rangle + \lambda_2 \mid f_2\rangle$ is mapped into $\lambda_1 f_1(z) + \lambda_2 f_2(z)$

for every $|f\rangle$, $|f_1\rangle$, $|f_2\rangle$ in $\mathfrak{H}$ and every complex number λ. If we denote this mapping by T_z, i.e.,

$$T_z\,|f\rangle = f(z), \tag{38a}$$

then T_z is a linear functional on $\mathfrak{H}$. By Riesz' theorem, it must be of the form

$$T_z\,|f\rangle = f(z) = (e_z, f). \tag{38b}$$

Using Eq. (28) and the definition of the scalar product Eq. (29), we infer that $|e_z\rangle = |\bar{z}\rangle$ and

$$\langle \bar{z}\mid e_a\rangle = e_a(z) = \exp(\bar{a}z), \tag{39a}$$

or equivalently

$$(e_a, e_b) = \exp(a\bar{b}). \tag{39b}$$

Equation (39a) implies that

$$\begin{aligned} f(a) = (e_a, f) &= \int d\mu(z)\,\exp\overline{(\bar{a}z)}f(z) \\ &= \int d\mu(z)\,\exp(a\bar{z})f(z) \end{aligned} \tag{40}$$

In other words $\exp(a\bar{z})$ acts like a delta function in the Hilbert space of entire analytic functions. Although neither orthogonal nor normalized [see Eq. (39b)], the vectors $|e_a\rangle$ are complete in the sense that

$$\langle f\mid g\rangle = \int \langle f\mid e_a\rangle\, d\mu(a)\langle e_a\mid g\rangle. \tag{41}$$

The proof is immediate since

$$\langle f\mid e_a\rangle = \overline{f(a)}, \tag{42}$$

and

$$\langle e_a\mid g\rangle = g(a), \tag{42b}$$

so that Eq. (41) reduces to the definition of the scalar product. The completeness relation is exhibited in another fashion by the assertion that the operation of taking the trace of an operator A can be accomplished by the operation $\int d\mu(a')\langle a'\,|A|\,a'\rangle$, i.e.,

$$\mathrm{Tr}\,A = \int d\mu(a')\langle a'\,|A|\,a'\rangle. \tag{43}$$

For example

$$\begin{aligned} \mathrm{Tr}\,\{e^{i\alpha p+i\beta q}\} &= \int d\mu(a')\langle a'\,|e^{i\alpha p+i\beta q}|\,a'\rangle \\ &= \int d\mu(a')e^{-(1/4)(\alpha^2+\beta^2)} \\ &\quad\times\langle a'\,|e^{(1/\sqrt{2})(i\beta-\alpha)a^*} \\ &\quad\times e^{(1/\sqrt{2})(i\beta+\alpha)a}|\,a'\rangle \qquad (44a) \\ &= \frac{1}{\pi}\iint da_r'\,da_i'\,e^{\sqrt{2}i\beta a_r} \\ &\quad\times e^{\sqrt{2}i\alpha a_i}\,e^{-(1/4)(\alpha^2+\beta^2)} \qquad (44b) \\ &= 2\pi\,\delta(\alpha)\,\delta(\beta), \qquad (44c) \end{aligned}$$

the familiar result. In (44b) the factor $\exp(-|a'|^2)$ from the measure element $d\mu(a)$ canceled the factor $\exp|a'|^2 = \langle a'\mid a'\rangle$.

We conclude this section by exhibiting the unitary operator $\langle \bar{z}\mid q\rangle$ which maps $\mathfrak{H}_B$ on the Hilbert space of configuration space wave functions [i.e., $L^{(2)}(-\infty, +\infty)$]. If we call $\langle q\mid \Psi\rangle = \psi(q)$, and to avoid confusion, $\langle \bar{z}\mid \Psi\rangle = f(z)$ we then wish to exhibit the kernel

$$\langle \bar{z}\mid q\rangle = A(z, q), \tag{45}$$

having the property that[11]

$$f(z) = \int A(z, q)\psi(q)\,dq, \tag{46a}$$

$$\psi(q) = \int \overline{A(z, q)}f(z)\,d\mu(z), \tag{46b}$$

and guarantees that $||f|| = ||\psi||$ and that the map is one-to-one. The conditions on the kernel $A(z, q)$ which ensure that these requirements are satisfied are that

$$\int dq\,\overline{A(z, q)}\,A(z, q) = \exp(\bar{z}'z) = e_{z'}(z), \tag{47a}$$

$$\int d\mu(z)\,\overline{A(z, q)}\,A(z, q) = \delta(q - q'). \tag{47b}$$

The unitarity of the kernel can be inferred from the completeness of the set $e_{z'}(z)$ occurring on the right-hand side of (46a). An explicit representation of $A(z, q)$ is obtained by noting that

$$\begin{aligned} \langle z'\,|a^*|\,q'\rangle &= (1/\sqrt{2})\langle z'\,|q - ip|\,q'\rangle \\ &= (1/\sqrt{2})\left(q' + \frac{\partial}{\partial q'}\right)\langle \bar{z}'\mid q'\rangle \\ &= z'\langle \bar{z}'\mid q'\rangle, \end{aligned} \tag{48a}$$

and similarly that

$$\begin{aligned} \langle \bar{z}'\,|a|\,q'\rangle &= (d/dz)'\langle \bar{z}'\mid q'\rangle, \\ &= (1/\sqrt{2})\langle \bar{z}'\,|q + ip|\,q'\rangle \\ &= (1/\sqrt{2})\left(q' - \frac{\partial}{\partial q'}\right)\langle \bar{z}'\mid q'\rangle, \end{aligned} \tag{48b}$$

[11] The inverse relation (46b should be interpreted as

$$\psi(q) = \underset{\lambda\to 1}{\text{l.i.m.}} \int \overline{A(z, q)}f_\lambda(z)\,d\mu(z)$$

since (46b) does not always converge, whereas this last equation does. See the explicit representation for the kernel $A(z,q)$ given by Eq. (48).

which equations imply that

$$A(z, q) = C \exp\left[-\tfrac{1}{2}(z^2 + q^2) + \sqrt{2}\, zq\right] \tag{49}$$

The conditions (47a, b) further assert that $C = \pi^{-1/4}$. In an analogous manner one verifies that

$$B(z, p) = \langle \bar{z} \mid p \rangle = (1/\sqrt[4]{\pi}) \exp\left[-\tfrac{1}{2}p^2 + \tfrac{1}{2}z^2 + i\sqrt{2}\, zp\right] \tag{50}$$

We are now ready to derive the Feynman integral representation for the transformation function $\langle a''t'' \mid a't' \rangle$.

III. THE FEYNMAN INTEGRAL OVER PATHS IN PHASE SPACE

We are interested in deriving a Feynman integral over paths representation for $\langle q''t'' \mid q't' \rangle$ for a system with one degree of freedom described by the Hamiltonian H. It will be assumed in the following that H is written in normal form with all a^* operators standing to the left of all a operators so that $\langle a'' \mid H \mid a' \rangle = H(\bar{a}'', a')\langle a'' \mid a' \rangle = c$ number. The procedure that we will employ will again be to break up the time interval $t'' - t'$ into $N + 1$ equal intervals of duration ϵ and to repeatedly insert the unit operator expressed as

$$\int |e_a(t)\rangle\, d\mu(a)\, \langle e_a(t)| = 1. \tag{51}$$

It is clearly sufficient to consider the transformation function

$$\langle a''t'' \mid a't' \rangle = \langle a'' \mid e^{-iH(t''-t')} \mid a' \rangle = \langle a'' \mid (e^{-iH\epsilon})^{N+1} \mid a' \rangle, \tag{52}$$

since

$$\langle q''t'' \mid q't' \rangle = \iint d\mu(a'')\, d\mu(a')\, \langle q''t'' \mid a''t'' \rangle \langle a''t'' \mid a't' \rangle \langle a't' \mid q't' \rangle, \tag{53}$$

and the transformation function $\langle a't \mid q't \rangle$ is known. Upon inserting the resolution of the identity (51) into (52) we obtain

$$\langle a''t'' \mid a't' \rangle \int \cdots \int d\mu(a_1) \cdots d\mu(a_n) \times \langle a'' \mid e^{-iH\epsilon} \mid a_N \rangle \langle a_N \mid e^{-iH\epsilon} \mid a_{N-1} \rangle \cdots \langle a_1 \mid e^{-iH\epsilon} \mid a' \rangle. \tag{54}$$

In the limit as $N \to \infty$, i.e., for infinitesimal ϵ, we can evaluate each factor to order ϵ as follows:

$$\langle a_{j+1} \mid e^{-i\epsilon H} \mid a_j \rangle \simeq \langle a_{j+1} \mid 1 - i\epsilon H \mid a_j \rangle = \langle a_{j+1} \mid a_j \rangle - i\epsilon H(\bar{a}_{j+1}, a_j)\langle a_{j+1} \mid a_j \rangle \simeq e^{-i\epsilon H(a_{j+1}, a_j)} \langle a_{j+1} \mid a_j \rangle. \tag{55}$$

The expression (54) for the propagator $\langle a''t'' \mid a't' \rangle$ valid to order ϵ thus takes the form

$$\langle a''t'' \mid a't' \rangle = \int \cdots \int \prod_{j=0}^{N} [\exp[-iH(\bar{a}_{j+1}, a_j)\epsilon] \times \langle a_{j+1} \mid a_j \rangle] \prod_{j=1}^{N} d\mu(a_j), \tag{56}$$

with $a_0 \equiv a'$ and $a_{N+1} \equiv a''$. We can further simplify Eq. (56) by using the normalization condition $\langle a_{j+1} \mid a_j \rangle = \exp(\overline{a_{j+1}} a_j)$. We next explicitly separate from the measure element $d\mu(a_j)$ the Gaussian factor and write

$$d\mu(a_j) = e^{-|a_j|^2}\, d^B a_j, \tag{57a}$$

$$d^B a = (1/\pi)\, da_r\, da_i \qquad (a = a_r + ia_i), \tag{57b}$$

so that the expression for the propagator becomes

$$\langle a''t'' \mid a't' \rangle = \int \cdots \int \prod_{m=1}^{N} d^B a_m\, e^{(1/2)|a'|^2 + (1/2)|a''|^2} \times \exp\left[\sum_{l=0}^{N} (-\tfrac{1}{2}\,|a_{l+1}|^2 + \overline{a_{l+1}} a_l - \tfrac{1}{2}\,|a_l|^2 - i\epsilon H(\overline{a_{l+1}}, a_l)\right]. \tag{58}$$

The limit $N \to \infty$, $(N + 1)\epsilon = t'' - t'$ defines the Feynman sum. As $\epsilon \to 0$, the exponent in Eq. (58) can also be written as

$$\lim_{\epsilon \to 0} \sum_{l=0}^{\infty} \left\{ i\epsilon \frac{i}{2} \left[\overline{a_{l+1}} \frac{a_{l+1} - a_l}{\epsilon} - \overline{\frac{a_{l+1} - a_l}{\epsilon}}\, a_l \right] - i\epsilon H(\bar{a}_{l+1}, a_l) \right\} \to i \int_{a't'}^{a''t''} dt \left\{ \frac{i}{2} \left(\overline{a(t)} \frac{da(t)}{dt} - \overline{\frac{da(t)}{dt}}\, a(t) \right) - H(\bar{a}(t), a(t)) \right\}. \tag{59}$$

If we symbolize

$$\lim_{N \to \infty} \prod_{m=1}^{N} d^B a_m = \mathfrak{D}[a(t)]. \tag{60}$$

Equation (58) is then the desired expression for the Feynman integral over paths representation of the propagator $\langle a''t'' \mid a't' \rangle$

$$e^{[-(1/2)|a''|^2]}\langle a''t'' \mid a't'\rangle e^{[-(1/2)|a'|^2]} = \int \cdots \int \mathfrak{D}[a(t)] \exp\left[i \int_{a't'}^{a''t''} \times L(a(t), \bar{a}(t), \dot{a}(t), d\bar{a}/dt)\, dt\right], \quad (61)$$

since

$$\frac{i}{2}\left(\overline{a(t)}\,\frac{da(t)}{dt} - \frac{\overline{da(t)}}{dt}\,a(t)\right) - H(a(t), \overline{a(t)}) \equiv L \quad (62)$$

is the Lagrangian for the system when the dynamical variables of the system are taken to be the a variables. The factor $\exp(-\frac{1}{2}|a''|^2)$ and $\exp(-\frac{1}{2}|a'|^2)$ occurring on the left-hand side of (61) are normalization factors arising from the fact that our states $|a'\rangle$ are not normalized but satisfy the normalization condition $\langle a' \mid a'\rangle = \exp(|a'|^2)$. The state $\exp(-\frac{1}{2}|a'|^2)\cdot|a'\rangle$ is a normalized state.

The advantage of defining the Feynman integral in terms of the a variables lies in the following circumstances: In configuration space the expression $\langle q''t'' \mid q't'\rangle$ is not necessarily an ordinary function but rather a distribution. (Recall that for $t'' = t'$ it is the delta function.) Questions of convergence are therefore "delicate." In "phase space" on the other hand $\langle a''t'' \mid a't'\rangle$ is not only a function, it even is analytic. For example, in the case of a harmonic oscillator whose Hamiltonian is $H = \omega a^* a$, so that $H(\bar{a}'', a') = \omega\bar{a}''a'$, the finite number of Gaussian integrals involved can readily be evaluated to yield

$$\langle a''t'' \mid a't'\rangle = \exp(\bar{a}''a'\, e^{-i\epsilon(N+1)\omega}) = \exp(\bar{a}''a'\, e^{-i\omega(t''-t')}), \quad (64)$$

and in fact only absolutely convergent integrals are encountered. It is of course true that some of the intuitive appeal of the original Feynman approach has been lost in that the integration is now over phase space [since $d^B a = (1/\pi)\, dp\, dq$]. However this very fact suggests that this particular approach ought to be particularly useful in the problems encountered in the determination of the classical limit of both quantum and quantum statistical mechanics. Our motivation for the study of the Feynman formulation in terms of the a variables is that in such a formulation some of the mathematical ambiguities found in other approaches are not encountered. In particular, this approach permits a study of the absolute convergence (or lack thereof) of integrals such as those encountered in Eq. (58) for certain classes of Hamiltonians. It thus paves the way for a rigorous formulation in terms of a measure over paths in phase space as well as for the study differential version of the dynamical principle and thus make contact with the Schwinger action principle.[12] Here we shall only study the "derivation" of the usual rules of quantum mechanics if the form (61) is adopted as the basic dynamical principle.

IV. THE FEYNMAN PRINCIPLE

Let us adopt as the Feynman principle[13] the assertion that the transformation function $\langle a''t'' \mid a't'\rangle$ is given Eq. (61) with the integration over paths defined by Eqs. (58)–(60). We shall explicitly assume that our Lagrangian is of the form given by Eq. (62), i.e., that it is of first order in $\partial_t a(t)$ and $\partial_t(\overline{a(t)})$. We shall then show that the usual rules of quantum mechanics follow. If the Lagrangian is of first order in $\partial_t a$ and $\partial_t \bar{a}$, we can approximate the exponent in (62) by

$$\int_{t'a'}^{t''a''} L\, dt = \sum_{l=0}^{N} L\left(a_l, \bar{a}_l, \frac{a_{l+1} - a_l}{\epsilon}, \frac{\bar{a}_{l+1} - \bar{a}_l}{\epsilon}\right). \quad (65)$$

Hence upon breaking up this sum

$$\sum_{l=0}^{N} L = \left(\sum_{l=1}^{j-1} + \sum_{l=j}^{N}\right) \times L\left(a_l, \bar{a}_l, \frac{a_{l+1} - a_l}{\epsilon}, \frac{\bar{a}_{l+1} - \bar{a}_l}{\epsilon}\right) \quad (66)$$

(with $a_{N+1} = a''$ and $a_0 = a'$), we deduce that

$$\begin{aligned}\langle a''t'' \mid a't'\rangle &= e^{+(1/2)|a''|^2} \times \int \cdots \int \exp\left[\frac{i}{\hbar}\sum_{l=j}^{N} L(a_l, a_{l+1})\right] \prod_{l=j+1}^{N-1} d^B a_l \\ &\times \int \cdots \int \exp\left[\frac{i}{\hbar}\sum_{l=0}^{j-1} L(a_l, a_{l+1})\right] \\ &\times \prod_{l=1}^{j-1} d^B a_l \cdot d^B a_j e^{+(1/2)|a'|^2} \\ &= \int \langle a''t'' \mid a_j t_j\rangle\langle a_j t_j \mid a't'\rangle e^{-|a_j|^2}\, d^B a_j \\ &= \int \langle a''t'' \mid a_j t_j\rangle\, d\mu(a_j)\langle a_j t_j \mid a't'\rangle. \end{aligned} \quad (67)$$

If we now identify the left-hand side of Eq. (62) as the Hermitian scalar product of two vectors in

[12] J. Schwinger, Phys. Rev. **82**, 914 (1951); **91**, 713 (1953).

[13] This section is patterned after the work of J. C. Polkinghorne, Proc. Roy. Soc. (London) **A 230**, 272 (1955). In the present section we shall differentiate between operators and eigenvalues by denoting the former by boldface letters.

Hilbert space [which is consistent with the right-hand side of Eq. (62)], Eq. (67) then allows us to infer that the set $|a_i t_i\rangle$ is complete:

$$\int |a_i t_i\rangle \, d\mu(a_i) \, \langle a_i t_i| = 1. \tag{68}$$

We next define the operator $\mathbf{a}(t)$ in terms of its matrix elements, by the equation

$$e^{-(1/2)|a''|^2}\langle a''t'' \,|\mathbf{a}(t)|\, a't'\rangle \, e^{-(1/2)|a'|^2}$$
$$= \int \cdots \int \mathfrak{D}[a(t)] \cdot a(t) \exp\left[-\frac{i}{\hbar}\int_{a't'}^{a''t''} L \, dt\right]. \tag{69}$$

If we set $t = t'$, then since $a(t)$ is not integrated over in the right-hand side, we deduce that

$$\langle a''t'' \,|\mathbf{a}(t)|\, a't'\rangle = a'\langle a''t'' \mid a't'\rangle. \tag{70a}$$

and using the completeness relation (70), we obtain

$$\mathbf{a}(t) \, |a't\rangle = a' \, |a', t\rangle. \tag{70b}$$

The definition (69) of the operator $\mathbf{a}(t)$ therefore implies the interpretation of the kets $|a', t'\rangle$ as eigenvectors of the operator $\mathbf{a}(t')$. Similarly one readily verifies, using the decomposition property (66), that

$$e^{-(1/2)|a''|^2}\langle a''t'' \,|P(\mathbf{a}(t_1) \cdots \mathbf{a}(t_n))|\, a't'\rangle e^{-(1/2)|a'|^2}$$
$$= \int \cdots \int \exp\left[\frac{i}{\hbar}\int_{a't'}^{a''t''} L \, dt\right]$$
$$\times a(t_1) \cdots a(t_n) \, \mathfrak{D}[a(t)], \tag{71}$$

where P denotes the time-ordering operator.

The statement that the operator $\mathbf{a}(t)$ satisfies the same equations of motion as the classical c-number variables $a(t)$ follows from the fact that

$$\langle a''t'' \,|\delta\mathbf{I}/\delta\mathbf{a}(t)|\, a't'\rangle = e^{(1/2)(|a'|^2+|a''|^2)}$$
$$\times \int \cdots \int \mathfrak{D}[a(t)] e^{(i/\hbar)I[a(t)]} \frac{\delta I[a(t)]}{\delta a(t)} \tag{72a}$$

$$I[a(t)] = \int_{a't'}^{a''t''} L[a(t)] \, dt. \tag{72b}$$

The integral on the right-hand side of Eq. (72a) vanishes since the integrand is a total derivative, namely,

$$(\hbar/i)[\delta/\delta a(t)] \exp(iI[a(t)]/\hbar),$$

and the particular variable of $\mathfrak{D}a(t)$ referring to the time t in question can be directly integrated. For both limits of $a(t)$ the integrand, as interpreted by Eq. (58), will vanish, so that

$$\delta\mathbf{I}/\delta\mathbf{a}(t) = \delta\mathbf{I}/\delta\mathbf{a}^*(t) = 0. \tag{72c}$$

It should be stressed that in (72a) the symbol $\delta\mathbf{I}/\delta\mathbf{a}(t)$ is only given a meaning by the right-hand side, where $\delta I/\delta a(t)$ assumes its usual variational meaning in terms of the classical variables a and $\bar{a}$, namely: Consider new functions $a(t, \epsilon)$ such that $a(t, 0) = a(t)$ and such that at $t = t'$ and $t = t''$, $\partial a(t, \epsilon)/\partial\epsilon = 0$. By $\delta I/\delta a$ we then mean

$$\delta I = (d/d\epsilon) I[a(t, \epsilon)]|_{\epsilon=0}. \tag{73}$$

We next define the operator $\mathbf{b}(t)$ by[14]

$$\langle a''t'' \,|\mathbf{b}(t')|\, a't'\rangle = \bar{a}''\langle a''t'' \mid a't'\rangle. \tag{74}$$

From the fact that we have interpreted $\langle a''t'' \mid a't'\rangle$ as the scalar product of two vectors in Hilbert space, upon taking the complex conjugate of Eq. (74) and comparing the resulting equation with Eq. (70a) we deduce that

$$\mathbf{b}(t) = \mathbf{a}^*(t). \tag{75}$$

One next computes, using the above enumerated properties that

$$\langle a''t'' \,|[\mathbf{a}(t''), \mathbf{b}(t'')]|\, a't'\rangle$$
$$= \int \langle a''t'' \,|\mathbf{a}(t'')|\, a'''t''\rangle$$
$$\times \, d\mu(a''')\langle a'''t'' \,|\mathbf{b}(t'')|\, a't'\rangle$$
$$- \int \langle a''t'' \,|\mathbf{b}(t'')\mathbf{a}(t'')|\, a'''t''\rangle$$
$$\times \, d\mu(a''')\langle a'''t'' \mid a't'\rangle$$
$$= \int (|a'''|^2 - \bar{a}''a''')\langle a''t'' \mid a'''t''\rangle$$
$$\times \langle a'''t'' \mid a't'\rangle \, d\mu(a''')$$
$$= \int (|a'''|^2 - \bar{a}''a''')e^{\bar{a}''a'''}$$
$$\times \langle a'''t'' \mid a't'\rangle \, d\mu(a'''). \tag{76}$$

The right-hand side of Eq. (76) can be evaluated by substituting for $\langle a'''t'' \mid a't'\rangle$ its explicit form as given by Eq. (62). The result is the expected one, namely,[15]

$$\langle a''t'' \,|[\mathbf{a}(t''), \mathbf{b}(t'')]|\, a't'\rangle = \langle a''t'' \mid a't'\rangle, \tag{77}$$

whence

$$[\mathbf{a}(t), \mathbf{a}^*(t)] = 1. \tag{78}$$

---

[14] Alternatively we can define $\mathbf{b}(t)$ by

$$\langle a''t'' \,|\mathbf{b}(t')\,|\, a't'\rangle = [\partial/\partial a'(t')]\langle a''t'' \mid a't'\rangle.$$

[15] The correctness of the resulting formula is checked by noting that

$$\int d\mu(a''')a'''(\overline{a'''} - \overline{a''})e^{\bar{a}''a'''}(a''')^n$$
$$= n\overline{(a'')^n} - (n-1)\overline{(a'')^n}$$
$$= \overline{(a'')^n}.$$

V. THE FIELD THEORETIC CASE

The treatment of a quantized field can be patterned after the example of a system with a finite degree of freedom. We shall here consider only the simplest situation, namely that of a real scalar field. To develop the formalism by analogy to the case of a system of particles, we expand the field $\phi(x)$ $[x = (x^0 = t, \mathbf{x})]$ and its canonically conjugate variable $\pi(x)$ into a complete set of *real* orthonormal functions $\varphi_n(\mathbf{x})$ which have the properties that

$$\int d^3x \varphi_m(\mathbf{x})\varphi_n(\mathbf{x}) = \delta_{mn}, \tag{79}$$

$$\sum_n \varphi_m(\mathbf{x})\varphi_m(\mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'). \tag{80}$$

The expansions are

$$\phi(x) = \sum_n q_n(t)\varphi_n(\mathbf{x}), \tag{81a}$$

$$q_n(t) = \int d^3x \varphi_n(\mathbf{x})\phi(\mathbf{x}), \tag{81b}$$

and

$$\pi(x) = \sum_n p_n(t)\varphi_n(\mathbf{x}) = \frac{\partial \mathfrak{L}}{\partial \dot{\phi}_0(x)}, \tag{82}$$

$$p_n(t) = \int d^3x \varphi_n(\mathbf{x})\pi(x), \tag{82b}$$

where $\mathfrak{L}$ is the Lagrangian density of the field system. In the quantized theory the expansion coefficients $q_n(t)$ and $p_n(t)$ are Hermitian operators which satisfy the commutation rules

$$[q_n(t), p_m(t)] = i\hbar\,\delta_{nm}, \tag{83a}$$

$$[q_n(t), q_m(t)] = [p_n(t), p_m(t)] = 0, \tag{83b}$$

which in turn imply the usual equal-time canonical commutation rules for the field operators $\phi(x)$ and $\pi(x)$:

$$[\phi(x), \pi(x')]_{x_0=x_0'} = i\hbar\,\delta(\mathbf{x} - \mathbf{x}'). \tag{84}$$

We next introduce the operators

$$a_n(t) = (2\hbar)^{-1/2}(q_n(t) + ip_n(t)), \tag{85a}$$

$$a_n^*(t) = (2\hbar)^{-1/2}(q_n(t) - ip_n(t)), \tag{85b}$$

which satisfy the commutation rules

$$[a_n(t), a_m^*(t)] = \delta_{nm}, \tag{86a}$$

$$[a_n(t), a_m(t)] = [a_n^*(t), a_m^*(t)] = 0. \tag{86b}$$

Simultaneous eigenfunctions of the operators $a_n(t)$ $(n = 0, 1, 2 \cdots)$ can be written in the form $|z_1, \cdots z_n, \cdots ; t\rangle$ with

$$a_n(t)\,|z_1, \cdots z_n, \cdots; t\rangle = z_n\,|z_1, \cdots z_n, \cdots; t\rangle$$
$$n = 0, 1, 2, \cdots. \tag{87}$$

These eigenfunctions are vectors in a separable Hilbert space which is the direct product of the Hilbert spaces $\{|z_1\rangle\}$, $\{|z_2\rangle\}$, $\cdots$. The mathematical characterization of such direct product spaces has been given by Von Neumann.[16-18] We can express the vectors $|z_1, \cdots z_n, \cdots ; t\rangle$ in terms of the eigenfunctions of the number operators

$$N_i(t) = a_i^*(t)a_i(t), \tag{88a}$$

$$N_i(t)\,|n_1, \cdots n_i, \cdots; t\rangle = n_i\,|n_1, \cdots n_i, \cdots; t\rangle, \tag{88b}$$

as follows:

$$|z_1, \cdots z_n, \cdots; t\rangle = \sum_{n_1, \cdots n_i, \cdots}^{\infty} \frac{(z_1)^{n_1} \cdots (z_i)^{n_i} \cdots}{(n_1!\, n_2! \cdots n_i! \cdots)^{1/2}} \times |n_1, n_2, \cdots n_i \cdots; t\rangle = \exp\left[\sum_i z_i a_i^*(t)\right]|0\rangle. \tag{89}$$

An arbitrary normalizable state vector of the field system $|\Psi\rangle$ can be expanded in terms of these basis vectors, the expansion coefficients

$$\Psi(z_1, \cdots z_n, \cdots) = \langle \bar{z}_1, \cdots \bar{z}_n, \cdots | \Psi\rangle \tag{90}$$

now being entire analytic functions in each of the variables $z_1, z_2, \cdots$. The Hermitian scalar product of two vectors $|\Psi\rangle$, $|\Phi\rangle$ is defined as

$$\langle \Phi | \Psi\rangle = \int \cdots \int \prod_{i=1}^{\infty} d\mu(z_i) \times \overline{\Phi(z_1, \cdots z_n, \cdots)}\Psi(z_1, \cdots z_n, \cdots), \tag{91a}$$

$$d\mu(z_i) = \pi^{-1} \exp(-\bar{z}_i z_i)\, dx_i\, dy_i \tag{91b}$$

the integration over each set of variables x_i, y_i being carried out over all values of x_i, y_i: $-\infty < (x_i, y_i) < \infty$. We here note that

$$\langle z_1'', z_2'', \cdots | z_1', z_2', \cdots\rangle = \prod_i \exp(\bar{z}_i z_i) = \exp\left(\sum \bar{z}_i z_i\right), \tag{92}$$

and also that in terms of the fields $U(\mathbf{x})$ and $\bar{U}(\mathbf{x})$ defined by

---

[16] J. von Neumann, Compositio Math. **6**, 1 (1939).

[17] It can be shown that in terms of von Neumann's theory, one is dealing here with the *incomplete* direct product for the equivalence class $\mathfrak{C}$ of $\{f^0{}_i\}$ where all $f^0{}_i(z) = 1$. This is as should be since the Hilbert space is isomorphic to Fock space (see Bargmann[18]).

[18] V. Bargmann, Proc. Nat'l. Acad. Sci. U. S. **48**, 199 (1962).

$$U(\mathbf{x}) = \sum_j z_j \varphi_j(\mathbf{x}), \tag{93a}$$

$$\bar{U}(\mathbf{x}) = \sum_j \bar{z}_j \varphi_j(\mathbf{x}), \tag{93b}$$

the expression $\sum_i \bar{z}_i z_i$ can be written as

$$\sum_j \bar{z}_j z_j = \int d^3x\, \bar{U}(\mathbf{x}) U(x). \tag{94}$$

Formulas which are the generalization of Eqs. (45) to (50) can also readily be written down for the present situation.

We next turn to the computation of the transformation function $\langle \bar{z}_1, \cdots \bar{z}_n, \cdots | \chi \rangle$ where the state $|\chi\rangle$ is an eigenstate of the field operator $\phi(\mathbf{x})$, i.e.,

$$\phi(x)\ |\chi\rangle = \chi(x)\ |\chi\rangle. \tag{95}$$

$\langle \chi\ |\Psi\rangle$ is thus the amplitude for finding the field system described by the Heisenberg state vector $|\Psi\rangle$ in the field configuration $\chi(\mathbf{x}, x_0)$ at time $t = x_0$.[19] An explicit representation of the basis vectors $|\chi\rangle$ can be obtained in terms of the eigenstates of the operators $q_n(t)$, $n = 0, 1, 2, \cdots$. If we denote these eigenstates by $|q_1', q_1', \cdots ; t\rangle$ with

$$q_i(t)\ |q_1', q_2', \cdots q_i' \cdots; t\rangle = q_i'\ |q_1', q_2', \cdots q_i', \cdots; t\rangle, \tag{96}$$

then the vector $|\chi\rangle$ has the following representation

$$|\chi\rangle = |\chi_1, \chi_2, \cdots \chi_n, \cdots; t\rangle, \tag{97}$$

where the χ_n's are the expansion coefficients of $\chi(x)$ in terms of the orthonormal set φ_n, i.e.,

$$\chi(x) = \sum_n \chi_n(t)\varphi_n(\mathbf{x}). \tag{98}$$

The kernel of interest $\langle \bar{z}_1, \cdots \bar{z}_n \cdots | \chi_1, \cdots \chi_n, \cdots \rangle$ is then readily computed and the result is

$$\langle e_{z_1 z_2 \cdots z_n}\ |\chi_1, \cdots \chi_n, \cdots\rangle = \prod_i \pi^{-1/4} \exp\left[-(1/2)(z_i^2 + \chi_i^2) + \sqrt{2}\, z_i \chi_i\right]. \tag{99}$$

This transformation function allows one to compute the transformation function $\langle \chi'' t'' \mid \chi' t' \rangle = \langle \chi_1'', \chi_2'' \cdots ; t'' \mid \chi_1', \chi_2' \cdots ; t' \rangle$ from the knowledge of $\langle z_1'', z_2'' \cdots ; t'' \mid z_1', z_2' \cdots ; t' \rangle$. Let us therefore compute the latter. The steps in this derivation are analogous to those involved in obtaining Eqs. (58) to (62). One again breaks up the time interval $t'' - t'$ into $N + 1$ equal intervals of duration ϵ and repeatedly inserts the resolution of the identity

$$\int \cdots \int |z_1, z_2 \cdots\rangle \prod_i d\mu(z_i) \langle z_1, z_2, \cdots| = 1. \tag{100}$$

One thus obtains the following expression for the propagator

$$\begin{aligned}\langle z_1'', z_2'', \cdots; t'' \mid z_1', z_2', \cdots; t'\rangle &= \int \cdots \int \prod_{i_1} d\mu(z_{i_1}^{(1)}) \prod_{i_2} d\mu(z_{i_2}^{(2)}) \cdots \prod_{i_N} d\mu(z_{i_N}^{(N)}) \\ &\times \langle z_1'', z_2'', \cdots |e^{-i\epsilon H}| z_1^{(N)}, z_2^{(N)}, \cdots\rangle \cdots \\ &\times \langle z_1^{(1)}, z_2^{(1)}, \cdots |e^{-i\epsilon H}| z_1', z_2', \cdots\rangle,\end{aligned} \tag{101}$$

where H is the Hamiltonian of the field system which will be assumed to have been written in normal form (i.e., all creation operators standing to the left of all annihilation operators) so that

$$\langle z_1^{(i+1)}, z_2^{(i+1)}, \cdots |H| z_1^{(i)}, z_2^{(i)} \cdots\rangle = H(\overline{Z^{(i+1)}}, Z^{(i)}) \langle Z^{(i+1)} \mid Z^{(i)}\rangle. \tag{102}$$

We have denoted by $Z^{(i)}$ the set $z_1^{(i)}, z_2^{(i)}, \cdots$. If we again separate from the measure element (916) the Gaussian factor and write

$$\prod_l d\mu(z_l^{(j)}) = \prod_l e^{-|z_l^{(j)}|^2} \frac{1}{\pi} dx_l^{(j)}\, dy_l^{(j)}, \tag{103}$$

the expression for the propagator becomes, with $Z^{(0)} = Z'$ and $Z^{(N+1)} = Z''$,

$$\begin{aligned}\langle Z'' t'' \mid Z' t'\rangle &= \int \cdots \int \prod_{l_1} d\mu(z_{l_1}^{(1)}) \cdots \prod_{l_N} d\mu(z_{l_N}^{(N)}) \\ &\times \prod_{j=0}^{N} \exp\left[\sum_l \overline{z_l^{(j+1)}} z_l^{(j)} - i\epsilon H(\bar{Z}^{(j+1)}, Z^{(j)})\right] \\ &= \int \cdots \int \prod_{l_1} \frac{dx_{l_1}\, dy_{l_1}}{\pi} \cdots \prod_{l_N} \frac{dx_{l_N}\, dy_{l_N}}{\pi} \\ &\times \prod_{j=0}^{N} \exp\Big[\sum_i \{-\tfrac{1}{2} |z_i^{(j+1)}|^2 \\ &+ \overline{z_i^{(j+1)}} z_i^{(j)} - \tfrac{1}{2} |z_i^{(j)}|^2\}\Big] \\ &\times \exp\left[-i\epsilon H(\overline{Z^{(j+1)}}, Z^{(j)})\right] \\ &\times \exp\Big[\tfrac{1}{2} \sum_i \{|z_i''|^2 + |z_i'|^2\}\Big],\end{aligned} \tag{104}$$

which in the limit $\epsilon \to 0$ becomes equal to

[19] In the case of a noninteracting real scalar field, the vectors $|\chi\rangle$ will in general only exist only for $\chi(x)$ which are normalizable solutions of the Klein-Gordon equation, i.e., if we denote by $\chi(k)$ the Fourier transform of $\chi(x)$

$$\chi(\mathbf{x}) = \int d^4k\ \delta(k^2 - \mu^2) \tilde{\chi}(k) e^{-ik\cdot x},$$

$$\chi(-k) = \overline{\chi(k)},$$

$|\chi\rangle$ will exist only for χ's such that $d^4k\delta(k^2-\mu^2)\,|\chi(k)|^2 < \infty$ In the case of interacting fields the exact conditions have not been investigated.

$$\langle Z''t'' \mid Z't' \rangle$$

$$= \exp\left[\tfrac{1}{2}\sum_i \{|z_i''|^2 + |z_i'|^2\}\right] \int \cdots \int \mathfrak{D}[Z(t)]$$

$$\times \exp\left[i \int_{Z't'}^{Z''t''} dt \{ \tfrac{1}{2} \sum (\overline{z_i(t)}\dot{z}_i(t) - \overline{\dot{z}_i(t)} z_i(t)) - iH(\bar{Z}(t), Z(t)) \} \right], \tag{105}$$

where we have written $\mathfrak{D}[Z(t)]$ for

$$\mathfrak{D}[Z(t)] = \lim_{N\to\infty} \prod_{l_1}^{N} \frac{dx_{l_1}\, dy_{l_1}}{\pi} \cdots \prod_{l_N}^{N} \frac{dx_{l_N}\, dy_{l_N}}{\pi}. \tag{106}$$

When expressed in terms of the fields $U(x) = \sum_i z_i(t)\varphi_i(\mathbf{x})$, Eq. (105) can be written as

$$\exp\left[\tfrac{1}{2}\sum_l |z_l''|^2\right]\langle Z''t'' \mid Z't'\rangle \exp\left[\tfrac{1}{2}\sum_l |z_l'|^2\right]$$

$$= \int \cdots \int \mathfrak{D}[U(x)]$$

$$\times \exp\left[i \int_{Z''t''}^{Z''t''} dt \int d^3x \{ \tfrac{1}{2}(\overline{U(x)} \partial_t U(x) - \overline{\partial_t U(x)} U(x)) - \mathfrak{IC}(\bar{U}, U) \} \right] \tag{107a}$$

$$= \int \cdots \int \mathfrak{D}[U(x)] \exp\left[i \int d^4x \mathfrak{L}(U) \right] \tag{107b}$$

where $\mathfrak{IC}(\bar{U}, U)$ is the Hamiltonian density for the classical field, expressed in terms of U and $\bar{U}$, with $H = \int d^3x \mathfrak{IC}$, and $\mathfrak{L}$ is the Lagrangian density. The first-order form for the Lagrangian is again to be noted. In Eqs. (107a) and (107b) the expression $\mathfrak{D}[U(x)]$ is given meaning by Eqs. (106) and (105).

ACKNOWLEDGMENTS

I should like to thank Professor V. Bargmann for introducing me to Hilbert spaces of analytic functions and for stimulating discussions on the subject. I should also like to acknowledge a communication from Dr. Klauder pointing out an error in the original manuscript in the derivation of Eq. (73) as well as raising an interesting question concerning the possible need of an "action option" if one adopts Eq. (61) as defining the Feynman integral.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

On Covariant Formulations of the Maxwell-Lorentz Theory of Electromagnetism*

MENDEL SACHS†
McGill University, Montreal, Canada
AND
SOLOMON L. SCHWEBEL
The University of Cincinnati, Cincinnati, Ohio
(Received February 27, 1962)

Maxwell's equations are formulated in a number of different representations: (a) As a single four-component spinor equation whose transformation properties are almost identical with those of the Dirac equation. (b) As a pair of uncoupled two-component spinor equations, in two different representations. One of these is similar to the Weyl equation for the neutrino field and the other to the two-component spinor form of the Dirac equation. (c) As a single equation in which the field variables are 2×2 matrices. In terms of these new field variables corresponding conservation laws are derived. Identification of these conservation laws with the standard ones is made. The transformations that are appropriate to the maintainance of covariance of the field equations are discussed.

I. INTRODUCTION

MAXWELL'S equations have an important place in the development of both the classical and the quantum theory of fields. As a field theory describing macroscopic classical electrodynamics it is eminently successful. The application of the theory to microscopic phenomena by Lorentz, however, gave rise to a number of difficulties such as the infinite self-energy of the point electron, its self-acceleration, and the concomitant radiation by the accelerated point electron.

The classical electromagnetic field served as a model for Yukawa's meson theory. It has been taken as the prototype for the construction of quantum field theories because of a greater understanding of electromagnetic forces as compared with other types of forces. However, some difficulties of the present-day quantum field theories are traceable to those present in classical theory.

A theory, developed by us,[1] attempts to resolve these difficulties by considering the interaction between particles rather than the free particle fields as the elementary entity. The field variables are then reinterpreted in terms of the interactions between particles instead of the intrinsic properties of the isolated particle. Contained in this theory is a reinterpretation of the Maxwell field equations. According to this interpretation, Maxwell's equations are nothing else but a covariant means of obtaining force field variables through which to represent a source field or vice-versa. Thus, within the framework of the theory, we can only accept the particular solutions of Maxwell's equations as physically meaningful.

Within this interpretation, it follows that there is no special reason for maintaining the vector form of the Maxwell field equations if it is possible to express them in other covariant forms. It is our object in this paper to present the reader with several new mathematical forms of Maxwell's equations. The new forms are particularly interesting because of their transformation properties. These lead to generalizations because of the appearance of generalized conservation laws. One of these forms [Eq. (7)] is used in our formulation of quantum electrodynamics.[1] However, it should be emphasized that, independent of microscopic phenomena, these equations can also be useful because solutions of Maxwell's equations applied to macroscopic phenomena may be more readily determined in some cases from the forms of the equations that we present here than from the conventional form.

The mathematical forms of the Maxwell field equations that we present in Sec. II require only that the principle of covariance be satisfied in order to obtain the transformations of the field variables defined by these equations. In particular, two of these forms of Maxwell's equations [Eqs. (7) and (8)] are so similar to the corresponding forms of the Dirac equation that both sets of variables, one representing the Maxwell field and the other the Dirac field, have the same transformation properties under the elements of the proper Lorentz group. In Sec. III we discuss the conservation laws which are intimately connected with the transformation

* The work of one of the authors (MS) was supported in part by a grant from the National Research Council of Canada.
† Present address: Physics Department, Boston University, Boston, Massachusetts.
[1] M. Sachs and S. L. Schwebel, Nuovo cimento suppl. **21**, 197 (1961).

properties that prescribe the covariance of these new forms of the field equations.

II. FORMS OF MAXWELL'S EQUATIONS AND THEIR TRANSFORMATION PROPERTIES

II.1 Review of the Conventional Forms

The well-known form of Maxwell's equations is[2]

$$\begin{aligned} \nabla \times \mathbf{E} + \partial_t \mathbf{H} &= 0 \qquad & \nabla \cdot \mathbf{H} &= 0 \\ \nabla \times \mathbf{H} - \partial_t \mathbf{E} &= 4\pi \mathbf{j} \qquad & \nabla \cdot \mathbf{E} &= 4\pi\rho. \end{aligned} \tag{1}$$

In this form, the electric field variable $\mathbf{E}(\mathbf{x})$ transforms like a polar vector and the magnetic field variable $\mathbf{H}(\mathbf{x})$ transforms like an axial vector when the charge density $\rho(\mathbf{x})$ is assumed to transform like a scalar. Further, the imposed requirement of covariance of (1) requires that proper Lorentz transformations on $\mathbf{E}$ and $\mathbf{H}$ will yield field variables in the new inertial frame that are particular combinations,[3] respectively, in terms of $\mathbf{E}$, $\mathbf{v} \times \mathbf{H}$ and $\mathbf{H}$, $\mathbf{v} \times \mathbf{E}$.

Next, a form of Maxwell's equations that makes its covariance more obvious is in terms of two field equations in the field variable $F_{\mu\nu}$—an antisymmetric matrix. The field equations are[3]

$$\begin{aligned} \partial_\nu \mathrm{F}_{\mu\nu} &= 4\pi j_\mu \\ \partial_\rho \mathrm{F}_{\mu\nu} + \partial_\nu \mathrm{F}_{\rho\mu} + \partial_\mu \mathrm{F}_{\nu\rho} &= 0. \end{aligned} \tag{2}$$

Here again, it is only when we specify that the four variables $[\rho(\mathbf{x}), j_1(\mathbf{x}), j_2(\mathbf{x}), j_3(\mathbf{x})]$ must transform as a four-vector that the principle of covariance ensures that $F_{\mu\nu}$ transforms as a second-rank tensor.

In the covariant forms of Maxwell's equations discussed below, no additional statement has to be made about the *a priori* transformation properties of the source terms. The principle of covariance alone determines these properties. Also, in view of the ensuing discussion, it should be emphasized that (2) was constructed by identifying the elements of $F_{\mu\nu}$ with components of $\mathbf{E}$ and $\mathbf{H}$ in a particular Lorentz frame. (This is the procedure that will be followed below.) However, Lorentz transformations preserve the covariance of (1) if the field variables transform as three-vectors while the covariance of (2) is maintained when the solutions transform as a tensor field. In the forms of Maxwell's equations demonstrated below, the field variables transform (isomorphic with the proper Lorentz group) neither as vectors nor as tensors.

The conserved quantities follow from the equations of continuity. These are the equations that relate to the four-divergence of a function whose time component (or components) are the conserved quantities. For example, in the case of equations of the type (2), we have the continuity equation

$$\partial_\mu T_{\mu\nu} = k_\nu \equiv j_\mu F_{\mu\nu} \tag{3}$$

where k_ν are the components of the four-Lorentz force density,

$$T_{\mu\nu} = (1/4\pi)(F_{\mu\rho}F_{\nu\lambda} - \tfrac{1}{4}F^2_{\rho\lambda}\,\delta_{\mu\nu})$$

is the energy-momentum tensor and $\int T_{\mu 0}\, d^3x$ is the conserved energy-momentum vector of the field. [This, of course, follows from the integral form of (3) combined with Gauss' theorem.]

In the discussion that follows, we will demonstrate new covariant forms of the Maxwell field equations and corresponding conservation laws. It will be seen how the new conservation laws contain the old ones and therefore represent a generalization. Throughout this development it will be assumed that the underlying symmetry that governs the structure of the field equations is described by the *proper Lorentz group* (i.e., discontinuous transformations such as spatial and temporal reflections are not contained in the group). Also, we do not restrict ourselves to a spin-one field, rather we allow a reduction from the vector field to a spinor field representation. The justification for this follows from our interpretation of the Maxwell field equations.

II.2 A Four-Component Spinor Form

Spinor formulations of Maxwell's equations have been proposed by several authors.[4] It is our aim here to present different forms of spinor equations than those cited in reference 4. In particular, some of these equations have the same transformation properties as the Dirac equation (in its two-component form). We start with a discussion of the four-component spinor form.

Defining the complex vector

$$\mathbf{G} = \mathbf{H} - i\mathbf{E}$$

the two-component functions

---

[2] Units are chosen with $c = 1$.

[3] See, for example, L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1951).

[4] O. Laporte and G. E. Uhlenbeck, Phys. Rev. **37**, 1380 (1931); J. R. Oppenheimer, *ibid.* **38**, 725 (1931); G. Moliere, Ann. Physik **6**, 146 (1949); T. Ohmura, Progr. Theoret. Phys. (Kyoto) **16**, 684 (1956); R. H. Good, Jr., Phys. Rev. **105**, 1914 (1957); H. E. Moses, *ibid.* **113**, 1670 (1959).

$$\phi_1 = \begin{bmatrix} G_3 \\ G_1 + iG_2 \end{bmatrix} \qquad \Upsilon_1 = 4\pi i \begin{bmatrix} -\rho + j_3 \\ j_1 + ij_2 \end{bmatrix}$$
$$\phi_2 = \begin{bmatrix} -G_1 + iG_2 \\ G_3 \end{bmatrix} \qquad \Upsilon_2 = 4\pi i \begin{bmatrix} -j_1 + ij_2 \\ \rho + j_3 \end{bmatrix} \tag{4}$$

and the four-component functions

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \qquad \Upsilon = \begin{bmatrix} -i\Upsilon_2 \\ i\Upsilon_1 \end{bmatrix} \tag{5}$$

and using the notation

$$\mathbf{x} = \{x_0 = it; x_k\} \qquad \partial(\equiv \partial/\partial\mathbf{x}) = \{\partial_0 = -i\,\partial_t; \partial_k\}$$
$$\tilde{\gamma} = \{\gamma_0\gamma_5; \gamma_k\} \qquad \gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3 \qquad \gamma_0 = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}$$
$$\gamma_k = -i \begin{bmatrix} 0 & \sigma_k \\ -\sigma_k & 0 \end{bmatrix} \qquad \bar{\phi} = \phi^\dagger\gamma_0 \qquad \phi^\dagger_{ij} \equiv \phi^*_{ji}$$
$$(k = 1, 2, 3)$$

where σ_k are the usual Pauli matrices and I is the unit 2×2 matrix, it is a straightforward matter to verify that the bispinor field equation

$$\tilde{\gamma}_\nu\, \partial_\nu \phi(\mathbf{x}) = \Upsilon(\mathbf{x}) \tag{6}$$

is a representation of Maxwell's equations in the form given by Eq. (1).

Since $\tilde{\gamma}_\nu$ differs from γ_ν only in the time component $\tilde{\gamma}_0 = \gamma_0\gamma_5$, the similarity between the form of Maxwell's equations (6) and the Dirac equation

$$\gamma_\nu\, \partial_\nu \psi(\mathbf{x}) = -m\psi(\mathbf{x}) \tag{6'}$$

should be noted.

It should be emphasized here that as far as the field equation (6) is concerned, the transformed spinors $\phi'(\mathbf{x}') = A\phi(\mathbf{x})$ and $\Upsilon'(\mathbf{x}') = B\Upsilon(\mathbf{x})$ are not form-invariant with respect to the conventional field variables $\mathbf{E}$, $\mathbf{H}$, $\mathbf{j}$ and ρ; e.g., $\phi(\mathbf{E}, \mathbf{H}) \nrightarrow \phi(\mathbf{E}', \mathbf{H}')$. Thus, once a connection between the new field equation and (1) is established in one Lorentz frame, then we choose to consider the new form of the field equations thereby abandoning the old vector field variables for the new spinor field variables. The conserved quantities do, however, match those of the conventional formalism, along with the predictions of new conserved quantities.

II.3 Two-Component Spinor Forms

The Weyl Form

If we operate on Eq. (6) from the left with γ_5, this field equation splits into two uncoupled two-component spinor equations

$$\sigma_\mu\, \partial_\mu \phi_\alpha(\mathbf{x}) = \Upsilon_\alpha(\mathbf{x}) \qquad (\alpha = 1, 2) \tag{7}$$

where $\sigma_0 = iI$. With $\Upsilon_\alpha = 0$, Eq. (7) is the Weyl equation for the neutrino field.[5] The covariance of this form of the field equations (7) is demonstrated in the Appendix. It is shown there that when Lorentz transformations take $\phi_\alpha(\mathbf{x})$ into $S\phi_\alpha(\mathbf{x})$, the source term $\Upsilon_\alpha(\mathbf{x})$ must transform as $(S^\dagger)^{-1}\Upsilon_\alpha(\mathbf{x})$ in order to maintain the covariance of these equations. Thus, in terms of these variables, $\phi^\dagger_\alpha\Upsilon_\beta$ ($\alpha, \beta = 1, 2$) are invariant forms.

The invariant metrics which define the two-component spinor algebra are[6]

$$\phi_2^{tr} i\sigma_2 \phi_1 = I_1 \qquad (\phi^{tr}_{ij} \equiv \phi_{ji})$$
$$\Upsilon_2^{tr} i\sigma_2 \Upsilon_1 = I_2$$

If we express ϕ_α and Υ_α in terms of the conventional variables (in a particular Lorentz frame) according to Eq. (4), the invariants take the form

$$I_1 = E^2 - H^2 + 2i\mathbf{E}\cdot\mathbf{H} = F^2_{\mu\nu} + 2i\epsilon_{\mu\nu\lambda\rho}F_{\mu\nu}F_{\lambda\rho}$$
$$I_2 = j^2 - \rho^2 \qquad\qquad = j^2_\mu$$

which are the invariants of the conventional formulations.

The Dirac Form

Operating on the left-hand side of one of Eqs. (7) (say, the equation with $\alpha = 2$) with the Wigner time-reversal operator $\hat{W}$, Eqs. (7) become

$$\sigma_\mu\, \partial_\mu \phi_1 = \Upsilon_1$$
$$\tilde{\sigma}_\mu\, \partial_\mu \hat{\phi}_2 = \hat{\Upsilon}_2 \tag{8}$$

where

$$\tilde{\sigma} = \{\sigma_0; -\sigma_k\}$$
$$\hat{\phi}_2 = \hat{W}\phi_2 \qquad \hat{\Upsilon}_2 = \hat{W}\Upsilon_2$$

and

$$\hat{W} = (i\sigma_2)K,$$

where K denotes complex conjugation.

The covariance of the two-component spinor equations (8) are determined by the same transformation properties as those associated with the Dirac equation in its two-component form

$$\sigma_\mu\, \partial_\mu \varphi = -im\chi$$

---

[5] See P. Roman, *Theory of Elementary Particles* (North-Holland Publishing Company, Amsterdam, 1960), p. 107.

[6] For a more comprehensive discussion of spinor algebra, see the article by Laporte and Uhlenbeck.[4]

$$\bar{\sigma}_\mu \partial_\mu \chi = -im\varphi$$

This is derived from Eq. (6′) by calling

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \quad \text{and} \quad \varphi = \psi_1 + \psi_2, \qquad \chi = \psi_1 - \psi_2.$$

II.4 A Matrix Form of Maxwell's Equations

Still another form of the field equations is obtained by combining the uncoupled set of equations (7) to form a single equation in which the field variable and source term are themselves 2×2 matrices. When this is done, the result is

$$\sigma_\mu \partial_\mu \Phi(\mathbf{x}) = \Sigma(\mathbf{x}) \tag{9}$$

where (9) was constructed by allowing

$$\Phi = (\phi_1 \; -\phi_2) \qquad \Sigma = (\Upsilon_1 \; -\Upsilon_2)$$

to be the 2×2 matrices involved in the field equation. In specifying the covariance of the field equation (9) however, just as before, we do not maintain form invariance in terms of the variables ϕ_1, ϕ_2. Thus, the variables $\Phi(\mathbf{x})$ and $\Sigma(\mathbf{x})$ are to be regarded as 2×2 matrices and covariance of (9) is maintained with

$$\Phi'(\mathbf{x}') \rightarrow \Phi(\mathbf{x}) = S^{-1}\Phi'(\mathbf{x}')S$$
$$\Sigma'(\mathbf{x}') \rightarrow \Sigma(\mathbf{x}) = S^\dagger \; \Sigma'(\mathbf{x}')S$$

where S is defined in the Appendix [Eq. (A11)]. This representation of Maxwell's equations is closely related to the conventional treatment in terms of second-rank spinors with mixed dotted and undotted indices. Since this formalism has an appreciable literature,[4] we will not develop it here anew.

It is also noted that (9) may also be expressed in terms of quaternions by choosing as units the identity I and the Pauli matrices (multiplied by i). Equation (9) then takes the form

$$\sigma_\mu \partial_\mu \{\sigma_k(H_k - iE_k)\} = 4\pi i \sigma_\nu j_\nu$$

where $H_0 - iE_0 \equiv 0$.

III. CONSERVATION LAWS

The procedure for deriving relations that are interpretable as conservation laws is very similar for all forms of Maxwell's equations. To eliminate such repetitive manipulations we shall undertake only the calculations for Eqs. (6) and (7) and leave to the interested reader the task of obtaining the counterparts to these results for the remaining equations.

Multiply Eq. (6) on the left by $\phi^\dagger \hat{O}_i$, where $\hat{O}_i$ is one of the operators of the set

$$\{\hat{O}_i\} = \{I, \gamma_0, \gamma_5, \gamma_0\gamma_5\}.$$

Take the Hermitian adjoint of Eq. (6) and multiply it on the right with $O_i\phi$. Adding or subtracting the equations obtained with the same selection for O_i, we find that

$$\partial_\nu(\bar{\phi}\tilde{\gamma}_\nu\phi) = \bar{\phi}\Upsilon - \bar{\Upsilon}\phi \tag{10a}$$
$$\partial_\nu(\bar{\phi}\gamma_0\tilde{\gamma}_\nu\phi) = \bar{\phi}\gamma_0\Upsilon + \bar{\Upsilon}\gamma_0\phi \tag{10b}$$
$$\partial_\nu(\bar{\phi}\gamma_0\gamma_5\tilde{\gamma}_\nu\phi) = \bar{\phi}\gamma_0\gamma_5\Upsilon - \bar{\Upsilon}\gamma_0\gamma_5\phi \tag{10c}$$
$$\partial_\nu(\bar{\phi}\gamma_5\tilde{\gamma}_\nu\phi) = \bar{\phi}\gamma_5\Upsilon + \bar{\Upsilon}\gamma_5\phi. \tag{10d}$$

The procedure with regard to Eqs. (7) is simpler. Multiply one of these equations (which we denote by the subscript α) by $\phi^\dagger_\beta$ and the Hermitian adjoint of the other of these equations (denoted by the subscript β) by ϕ_α and then add. We thus obtain

$$\partial_\mu(\phi^\dagger_\beta\sigma_\mu\phi_\alpha) = \phi^\dagger_\beta\Upsilon_\alpha + \Upsilon^\dagger_\beta\phi_\alpha \qquad (\alpha, \beta = 1, 2). \tag{11}$$

Equations (10) and (11) are the conservation laws that we sought. Expressed in the field variables ϕ and Υ these equations are unfamiliar ones indeed. However, if we undertake to express them in terms of the usual field variables $\mathbf{E}$, $\mathbf{H}$, ρ, and $\mathbf{j}$, we find that Eqs. (10) and (11) give the familiar forms which describe the conservation of energy and momentum of the electromagnetic field. For example, Eq. (10d) becomes

$$(1/8\pi)\,\partial_t(E^2 + H^2) + (1/4\pi)\nabla\cdot(\mathbf{E}\times\mathbf{H}) = -\mathbf{E}\cdot\mathbf{j} \tag{12}$$

which is the usual energy-momentum conservation law.

Equation (11), which is a more compact form of Maxwell's equations, yields the above conservation equation when the two equations obtained for $\alpha = \beta = 1$ and $\alpha = \beta = 2$ are added.

A rather interesting distinction between Eqs. (6) and (7) arises from their invariance under transformations belonging to the proper Lorentz group. This is, that the transformation properties for the field variables and the source terms of Eq. (7) are such that all four equations that appear in (11) are Lorentz invariants, while such is not the case for the conservation laws (10) that are associated with the bispinor form (6) of Maxwell's equations. This additional *symmetry* along with the fact that Eq. (7) represents Maxwell's original formulation as a set of two uncoupled spinor equations serves as a strong stimulus to explore further. In addition to the application of the form (7) of Maxwell's

equations to our formulation of quantum electrodynamics,[1] the implications in the formulation of the classical electrodynamical interaction have also been discussed[7] and will be published in the near future.

APPENDIX

A proof of the covariance of Eqs. (7),

$$\sigma_\mu \partial_\mu \phi_\alpha = \Upsilon_\alpha \qquad (\alpha = 1, 2) \tag{A1}$$

under a proper Lorentz transformation follows below. The Lorentz transformation is given by

$$x'_\mu = a_{\mu\nu} x_\nu, \tag{A2}$$

where $a_{\mu\nu}$ is real for $\mu, \nu = 1, 2, 3$, a_{00} is real, and $a_{\mu 0}$ and $a_{0\mu}$ are pure imaginary for $\mu = 1, 2, 3$. From the invariance of x_μ^2, we obtain the relations

$$\begin{aligned} a_{\mu\nu} a_{\mu\sigma} &= \delta_{\nu\sigma} \\ a_{\mu\sigma} a_{\nu\sigma} &= \delta_{\mu\nu}. \end{aligned} \tag{A3}$$

The principle of covariance requires that in the coordinate system $\{x'_\mu\}$ Eq. (A1) becomes

$$\sigma_\mu \partial'_\mu \phi'_\alpha = \Upsilon'_\alpha \qquad (\alpha = 1, 2). \tag{A4}$$

If we assume that

$$\phi'_\alpha = S\phi_\alpha \qquad \Upsilon'_\alpha = T\Upsilon_\alpha \tag{A5}$$

where S and T are 2×2 matrices independent of $\{x'_\mu\}$, and that $\{x'_\mu\}$ is related to $\{x_\mu\}$ by Eq. (A2), then (A4) takes the form

$$T^{-1}\sigma_\mu a_{\mu\nu} S \, \partial_\nu \phi_\alpha = \Upsilon_\alpha \qquad (\alpha = 1, 2)$$

Consequently, the principle of covariance can be satisfied if

$$T^{-1}\sigma_\mu a_{\mu\nu} S = \sigma_\nu \tag{A6}$$

or

$$\sigma_\mu a_{\mu\nu} = T\sigma_\nu S^{-1} \qquad (\nu = 0, 1, 2, 3). \tag{A7}$$

For $\nu = 0$

$$\sigma_\mu a_{\mu 0} = iTS^{-1}. \tag{A7'}$$

We now take the Hermitian adjoint of (A7'). Because of the Hermiticity of the Pauli matrices σ_k, $k = 1, 2, 3$, the reality of a_{00} and the pure imaginary character of σ_0 and $a_{\mu 0}$ for $\mu = 1, 2, 3$, we find that

$$(TS^{-1})^\dagger = TS^{-1}. \tag{A8}$$

For $\nu \neq 0$, we also find that

$$(T\sigma_\nu S^{-1})^\dagger = (T\sigma_\nu S^{-1}). \tag{A9}$$

[7] M. Sachs and S. L. Schwebel, Bull. Am. Phys. Soc. 5, 505 (1960).

Writing this equation out, we find

$$(S^{-1})^\dagger \sigma_\nu T^\dagger = T\sigma_\nu S^{-1}$$

or

$$\sigma_\nu T^\dagger S = S^\dagger T \sigma_\nu \qquad (\nu = 1, 2, 3).$$

But from (A8), $T^\dagger S = S^\dagger T$. Thus, $T^\dagger S$ commutes with the three Pauli matrices σ_1, σ_2, and σ_3. It follows that

$$T^\dagger S = bI \tag{A10}$$

with b a real constant. That b is real follows from (A8) or from the fact that $T^\dagger S = S^\dagger T$. We can normalize T by requiring its determinant to be unity. Then from (A7') it follows that the determinant of S is unity and hence the constant b in (A10) must be 1. Thus,

$$T = (S^{-1})^\dagger = (S^\dagger)^{-1}. \tag{A10'}$$

Equation (A6) may now be rewritten as

$$S^\dagger \sigma_\mu a_{\mu\nu} S = \sigma_\nu$$

or

$$S^\dagger \sigma_\mu S = a_{\mu\nu} \sigma_\nu. \tag{A11}$$

To finish the proof, we must establish the existence of S. To do this we consider an infinitesimal proper Lorentz transformation

$$x'_\mu = (\delta_{\mu\nu} + \epsilon_{\mu\nu}) x_\nu \tag{A12}$$

where $\epsilon_{\mu\nu}$ are small. The invariance of x_μ^2 requires that $\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}$. Under this transformation we have

$$S = I + \epsilon_{\mu\nu} \lambda_{\mu\nu} \tag{A13}$$

and

$$S^\dagger = I + \bar{\epsilon}_{\mu\nu} \lambda_{\mu\nu}^\dagger$$

where $\lambda_{\mu\nu}$ are the matrices which are to be determined and $\bar{\epsilon}_{\mu\nu}$ are the complex conjugates of $\epsilon_{\mu\nu}$. Since $\epsilon_{\mu\nu} = -\epsilon_{\nu\mu}$ these parameters are not linearly independent. However, if we specify that $\lambda_{\nu\mu} = -\lambda_{\mu\nu}$, we may then treat the parameters $\epsilon_{\mu\nu}$ as linearly independent for we then obtain redundant equations. The alternative is to define the double summation $\epsilon_{\mu\nu}\lambda_{\mu\nu}$ as $\sum_{\mu>\nu} \epsilon_{\mu\nu}\lambda_{\mu\nu}$. We will use the first procedure.

Applying Eqs. (A12) and (A13) to Eq. (A11), we find

$$\bar{\epsilon}_{\alpha\beta} \lambda_{\alpha\beta}^\dagger \sigma_\mu + \epsilon_{\alpha\beta} \sigma_\mu \lambda_{\alpha\beta} = \epsilon_{\mu\nu} \sigma_\nu. \tag{A14}$$

If $\mu = 0$ ($\sigma_0 = iI$), then

$$i(\bar{\epsilon}_{\alpha\beta} \lambda_{\alpha\beta}^\dagger + \epsilon_{\alpha\beta} \lambda_{\alpha\beta}) = \epsilon_{0\nu} \sigma_\nu.$$

The Hermitian adjoint of this equation is

$$-i(\epsilon_{\alpha\beta}\lambda_{\alpha\beta} + \bar{\epsilon}_{\alpha\beta}\lambda^{\dagger}_{\alpha\beta}) = \bar{\epsilon}_{0\nu}\sigma_\nu.$$

Thus,

$$\epsilon_{0\nu} = -\bar{\epsilon}_{0\nu} \qquad (\nu = 1, 2, 3). \tag{A15}$$

When $\mu \neq 0$, the Hermitian adjoint of (A14) is

$$\epsilon_{\alpha\beta}\sigma_\mu\lambda_{\alpha\beta} + \bar{\epsilon}_{\alpha\beta}\lambda^{\dagger}_{\alpha\beta}\sigma_\mu = \bar{\epsilon}_{\mu k}\sigma_k - \bar{\epsilon}_{\mu 0}\sigma_0 = \epsilon_{\mu\nu}\sigma_\nu.$$

The last equality follows from (A14). The summation over k is for $k = 1, 2, 3$. Since the σ_k's are linearly independent it follows that

$$\begin{aligned} &\epsilon_{\mu k} = \bar{\epsilon}_{\mu k} \qquad k = 1, 2, 3; \qquad \mu \neq 0 \\ &\epsilon_{\mu 0} = -\bar{\epsilon}_{\mu 0}. \end{aligned} \tag{A16}$$

Returning to Eq. (A14) and making use of Eqs. (A15) and (A16), we find that for $\mu \neq 0$, $\nu \neq 0$

$$\sigma_\nu = 2(\lambda^{\dagger}_{\mu\nu}\sigma_\mu + \sigma_\mu\lambda_{\mu\nu}). \tag{A17}$$

The right-hand side is not to be summed over μ. For $\mu = 0$, $\nu \neq 0$

$$\sigma_\nu = -2(\lambda^{\dagger}_{0\nu}\sigma_0 - \sigma_0\lambda_{0\nu}) = -2i(\lambda^{\dagger}_{0\nu} - \lambda_{0\nu}). \tag{A18}$$

For $\mu \neq 0$, $\nu = 0$

$$\sigma_0 = -2(\lambda^{\dagger}_{\mu 0}\sigma_\mu - \sigma_\mu\lambda_{\mu 0}). \tag{A19}$$

Equation (A17) has the solution

$$\lambda_{\mu\nu} = \tfrac{1}{4}\sigma_\mu\sigma_\nu \qquad (\mu \neq 0, \nu \neq 0)$$

with

$$\lambda_{\mu\nu} = -\lambda_{\nu\mu}.$$

Equation (A18) has the solution

$$\lambda_{0\nu} = -\tfrac{1}{4}\sigma_0\sigma_\nu.$$

Equation (A19) is solved with

$$\lambda_{\mu 0} = \tfrac{1}{4}\sigma_0\sigma_\mu$$

Note that $\lambda_{0\nu} = -\lambda_{\nu 0}$ as required.

Thus, we have evaluated all $\lambda_{\mu\nu}$ and determined the matrices S of Eq. (A13).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER–OCTOBER 1962

Dynamical Mappings of Density Operators in Quantum Mechanics. II. Time Dependent Mappings*

THOMAS F. JORDAN, MARK A. PINSKY,† AND E. C. G. SUDARSHAN
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received March 14, 1962)

The most general continuous time-dependent evolution of a physical system is represented by a continuous one-parameter semi-group of linear mappings of density operators to density operators. It is shown that if these dynamical mappings form a group they can be represented by a group of unitary operators on the Hilbert space of state vectors. This proof does not assume that the absolute values of inner products of state vectors or "transition probabilities" are preserved but *deduces* this fact from the requirement that density operators are mapped linearly to density operators. An example is given of a continuous one-parameter semi-group of dynamical mappings which is not a group.

I. INTRODUCTION

THE most general dynamical transformation of a physical system can be represented by a linear mapping of density operators to density operators. It was pointed out in an earlier paper[1] that there are many such dynamical mappings which are not Hamiltonian mappings, that is, there are linear mappings of the set of density operators into itself which can not be represented by unitary transformations on the Hilbert space of state vectors. The present paper is a continuation of the investigation begun in reference 1, and answers some questions which were left open there. In particular, we consider whether there can be non-Hamiltonian dynamical mappings which represent a continuous time dependent evolution of a physical system.

In Sec. II the property that a family of dynamical mappings represent a continuous time dependent evolution of a system is formulated in the requirement that it forms a continuous one-parameter semi-group. The requirement that it form a con-

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

† Research assistant visiting under the cooperative program from Antioch College, Yellow Springs, Ohio.

[1] T. F. Jordan and E. C. G. Sudarshan, J. Math. Phys. 2, 772 (1961).

$$-i(\epsilon_{\alpha\beta}\lambda_{\alpha\beta} + \bar{\epsilon}_{\alpha\beta}\lambda^{\dagger}_{\alpha\beta}) = \bar{\epsilon}_{0\nu}\sigma_\nu.$$

Thus,

$$\epsilon_{0\nu} = -\bar{\epsilon}_{0\nu} \qquad (\nu = 1, 2, 3). \tag{A15}$$

When $\mu \neq 0$, the Hermitian adjoint of (A14) is

$$\epsilon_{\alpha\beta}\sigma_\mu\lambda_{\alpha\beta} + \bar{\epsilon}_{\alpha\beta}\lambda^{\dagger}_{\alpha\beta}\sigma_\mu = \bar{\epsilon}_{\mu k}\sigma_k - \bar{\epsilon}_{\mu 0}\sigma_0 = \epsilon_{\mu\nu}\sigma_\nu.$$

The last equality follows from (A14). The summation over k is for $k = 1, 2, 3$. Since the σ_k's are linearly independent it follows that

$$\begin{aligned} &\epsilon_{\mu k} = \bar{\epsilon}_{\mu k} \qquad k = 1, 2, 3; \qquad \mu \neq 0 \\ &\epsilon_{\mu 0} = -\bar{\epsilon}_{\mu 0}. \end{aligned} \tag{A16}$$

Returning to Eq. (A14) and making use of Eqs. (A15) and (A16), we find that for $\mu \neq 0$, $\nu \neq 0$

$$\sigma_\nu = 2(\lambda^{\dagger}_{\mu\nu}\sigma_\mu + \sigma_\mu\lambda_{\mu\nu}). \tag{A17}$$

The right-hand side is not to be summed over μ. For $\mu = 0$, $\nu \neq 0$

$$\sigma_\nu = -2(\lambda^{\dagger}_{0\nu}\sigma_0 - \sigma_0\lambda_{0\nu}) = -2i(\lambda^{\dagger}_{0\nu} - \lambda_{0\nu}). \tag{A18}$$

For $\mu \neq 0$, $\nu = 0$

$$\sigma_0 = -2(\lambda^{\dagger}_{\mu 0}\sigma_\mu - \sigma_\mu\lambda_{\mu 0}). \tag{A19}$$

Equation (A17) has the solution

$$\lambda_{\mu\nu} = \tfrac{1}{4}\sigma_\mu\sigma_\nu \qquad (\mu \neq 0, \nu \neq 0)$$

with

$$\lambda_{\mu\nu} = -\lambda_{\nu\mu}.$$

Equation (A18) has the solution

$$\lambda_{0\nu} = -\tfrac{1}{4}\sigma_0\sigma_\nu.$$

Equation (A19) is solved with

$$\lambda_{\mu 0} = \tfrac{1}{4}\sigma_0\sigma_\mu$$

Note that $\lambda_{0\nu} = -\lambda_{\nu 0}$ as required.

Thus, we have evaluated all $\lambda_{\mu\nu}$ and determined the matrices S of Eq. (A13).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER–OCTOBER 1962

Dynamical Mappings of Density Operators in Quantum Mechanics. II. Time Dependent Mappings*

THOMAS F. JORDAN, MARK A. PINSKY,† AND E. C. G. SUDARSHAN
Department of Physics and Astronomy, University of Rochester, Rochester, New York
(Received March 14, 1962)

The most general continuous time-dependent evolution of a physical system is represented by a continuous one-parameter semi-group of linear mappings of density operators to density operators. It is shown that if these dynamical mappings form a group they can be represented by a group of unitary operators on the Hilbert space of state vectors. This proof does not assume that the absolute values of inner products of state vectors or "transition probabilities" are preserved but *deduces* this fact from the requirement that density operators are mapped linearly to density operators. An example is given of a continuous one-parameter semi-group of dynamical mappings which is not a group.

I. INTRODUCTION

THE most general dynamical transformation of a physical system can be represented by a linear mapping of density operators to density operators. It was pointed out in an earlier paper[1] that there are many such dynamical mappings which are not Hamiltonian mappings, that is, there are linear mappings of the set of density operators into itself which can not be represented by unitary transformations on the Hilbert space of state vectors. The present paper is a continuation of the investigation begun in reference 1, and answers some questions which were left open there. In particular, we consider whether there can be non-Hamiltonian dynamical mappings which represent a continuous time dependent evolution of a physical system.

In Sec. II the property that a family of dynamical mappings represent a continuous time dependent evolution of a system is formulated in the requirement that it forms a continuous one-parameter semi-group. The requirement that it form a con-

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

† Research assistant visiting under the cooperative program from Antioch College, Yellow Springs, Ohio.

[1] T. F. Jordan and E. C. G. Sudarshan, J. Math. Phys. 2, 772 (1961).

tinuous one-parameter group and the requirement that it is a continuous family of Hamiltonian mappings, i.e., that it is representable by a continuous group of unitary operators on the Hilbert space of state vectors, are apparently more restrictive conditions. In Sec. III it is shown that the latter two requirements are equivalent. Every continuous group of dynamical mappings is a continuous family of Hamiltonian mappings. But in Sec. IV an example is given of a continuous semi-group of dynamical mappings which is not a group and, therefore, is not a family of Hamiltonian mappings.

The one-parameter groups of dynamical transformations which we call "time dependent" can be used to represent other symmetry transformations on the system besides transformations in time. In fact, it is no problem to generalize our results from one-parameter groups to representations by dynamical mappings of many continuous groups such as the Lorentz group. That such representations must be generated by continuous unitary representations on the Hilbert space of state vectors[2] then follows in analogy to our proof in Sec. III that the group property implies that a time-dependent family of dynamical mappings is a family of Hamiltonian mappings. This proof does not require the assumption that the absolute values of inner products of state vectors or "transition probabilities" are preserved but proves this fact from the requirement that density operators are mapped linearly to density operators.

II. TIME-DEPENDENT DYNAMICAL MAPPINGS

In the quantum mechanical description of a physical system by operators on a seperable Hilbert space $\mathfrak{H}$, the state of the system can be represented by a density operator on $\mathfrak{H}$. A density operator is a positive semi-definite, self-adjoint operator which has unit trace. These operators form a convex set. The extremal elements of this convex set are the pure state density operators; they are projection operators onto one-dimensional subspaces of $\mathfrak{H}$. The properties of density operators were outlined in some detail in reference 1. It was shown that density operators belong to the Hilbert space $\mathfrak{L}$ of operators ρ on $\mathfrak{H}$ for which[3] $\mathrm{Tr}\,(\rho^\dagger\rho)$ is finite, the inner product in $\mathfrak{L}$ being defined by

$$(\rho, \sigma) = \mathrm{Tr}\,(\rho^\dagger\sigma). \tag{2.1}$$

The pure state density operators span the space $\mathfrak{L}$, so that a linear mapping on the density operators uniquely defines a linear mapping on $\mathfrak{L}$. A dynamical transformation of the physical system may be represented by a linear transformation of $\mathfrak{L}$ which maps the convex set of density operators into itself. If A is a linear operator on $\mathfrak{L}$ such that, if ρ is a density operator, then

$$\rho' = A\rho \tag{2.2}$$

is also a density operator, we will call A a dynamical mapping.[4]

In order to represent dynamics in the usual sense, that is as a continuous time-dependent evolution of the state of the system, we must have a family of dynamical mappings $A(t)$,

$$\rho \rightarrow \rho(t) = A(t)\rho \tag{2.3}$$

depending on a real parameter t, such that

$$A(t)A(s) = A(t + s) \tag{2.4}$$

for non-negative values of t and s, and

$$A(0) = I. \tag{2.5}$$

In other words we must require that the dynamical mappings $A(t)$ form a one-parameter semi-group. In addition we must require that the expectation value

$$\langle\sigma\rangle_t = \mathrm{Tr}\,(\sigma\rho(t)) = (\sigma, A(t)\rho) \tag{2.6}$$

of the self-adjoint operator σ belonging to $\mathfrak{L}$, for the time dependent state $\rho(t)$, be a continuous function of the parameter t. Since the trace of the product is the inner product in $\mathfrak{L}$, as is indicated in Eq. (2.6), this means that $A(t)$ must be weakly continuous as a function of t. The mathematical condition for a time dependent evolution of density operators is then that we have a family of linear transformations of the form (2.3) on $\mathfrak{L}$, and that:

(I) $A(t)$, $0 \leq t < \infty$, is a weakly continuous one-parameter *semi-group* of dynamical mappings (linear transformations of $\mathfrak{L}$ that map the convex subset of density operators into itself).

If we want the dynamics to be reversible, that is if we require that every dynamical mapping have an inverse which is a dynamical mapping, then we need the stronger condition that:

(II) $A(t)$, $-\infty < t < +\infty$, is a weakly con-

---

[2] E. P. Wigner, Ann. Math **40**, 149 (1939).

[3] It is sufficient for our purposes to consider only bounded operators which form a linear space without causing any problems of domains.

[4] We will use the same notational convention as in reference 1. Capital letters A will represent operators on $\mathfrak{L}$ and Greek letters ρ, ω, σ, operators on $\mathfrak{H}$ (elements of $\mathfrak{L}$). Greek letters ϕ, ψ will denote vectors in $\mathfrak{H}$, and small letters a, c will denote scalars.

tinuous one-parameter *group* of dynamical mappings $(A(t)^{-1} = A(-t))$.

Finally we are interested in time dependent Hamiltonian dynamical mappings which are defined by the condition that:

(III) There exists a (strongly or weakly) continuous one-parameter group of unitary operators $\omega(t)$ on $\mathfrak{H}$, such that $A(t)\rho = \omega(t)\rho\ \omega^\dagger(t)$ for each ρ belonging to $\mathfrak{L}$.

Clearly (III) implies (II) implies (I). We will see that conversely (II) implies (III), but that (I) does not imply (II).

III. HAMILTONIAN MAPPINGS

In this section we will show that a time dependent family of dynamical mappings is a family of Hamiltonian mappings whenever it is a one-parameter group. In reference 1 conditions were given which are necessary and sufficient for a dynamical mapping of the general form (2.2) to be a Hamiltonian mapping. However, no consideration was given to the time dependence of these mappings, or to the implications of time dependence which could possibly restrict the allowed dynamical mappings to Hamiltonian mappings. It is the purpose of this paper to consider these questions and thus complete the study of the relation of Hamiltonian quantum dynamics to the more general dynamics of density operators.

If a dynamical mapping takes pure state density operators to pure state density operators, then it defines a mapping of normalized vectors in $\mathfrak{H}$ to normalized vectors in $\mathfrak{H}$. For each vector this mapping is defined up to a phase factor. If these phase factors can be chosen so as to yield a linear mapping on $\mathfrak{H}$, we say that the dynamical mapping induces a linear mapping on $\mathfrak{H}$. In the earlier paper[1] it was stated that if a dynamical mapping maps pure state density operators to pure state density operators and induces a linear mapping on $\mathfrak{H}$, then it is a Hamiltonian mapping. This statement is true only for those dynamical mappings which map the set of pure state density operators *onto* itself. In Theorem 2 of reference 1, the possibility for a dynamical mapping to map the set of pure state density operators one-to-one onto a proper subset of itself was not given proper consideration. Before moving on to the new questions, we will give a corrected statement of this theorem, giving explicit attention to this particular feature:

Theorem. Equivalent necessary and sufficient conditions for a dynamical mapping to be a Hamiltonian dynamical mapping are:

(i) There exists a linear unitary operator ω on $\mathfrak{H}$ such that the dynamical mapping maps each operator ρ in $\mathfrak{L}$ to $\omega\rho\omega^\dagger$. (This can be taken as the definition of a Hamiltonian dynamical mapping.)

(ii) The dynamical mapping maps the set of pure state density operators *onto* itself and induces a linear mapping on $\mathfrak{H}$.

(iii) For each member $\phi^{(i)}$ of any set of basis vectors in $\mathfrak{H}$, there exists a normalized vector $\psi^{(i)}$, such that the set of these vectors spans $\mathfrak{H}$, and the dynamical mapping maps $\phi^{(i)}\phi^{(j)\dagger}$ to $\psi^{(i)}\psi^{(j)\dagger}$.[5]

(iv) There exist linear operators ω and σ on $\mathfrak{H}$, which have inverses, such that the dynamical mapping maps each operator ρ on $\mathfrak{H}$ to $\omega\rho\sigma^\dagger$.

Now we can proceed to the consideration of conditions under which time dependent dynamical mappings represent Hamiltonian dynamics.

Theorem. A time dependent family of dynamical mappings $A(t)$ is a family of Hamiltonian mappings (satisfying condition III) if it is a weakly continuous one-parameter group (satisfying condition II).

Proof. If the dynamical mappings $A(t)$ form a group, then for any value of t the dynamical mapping $A(t)$ has an inverse dynamical mapping $A(-t)$. Now $A(-t)$ can not map a density operator ρ which is not a pure state density operator to a pure state density operator. For let $\rho = a\rho_1 + (1 - a)\rho_2$, where $0 < a < 1$, and ρ_1 and ρ_2 are distinct density operators. Then $A(-t)\rho = aA(-t)\rho_1 + (1 - a)\,A(-t)\rho_2$ is not a pure state density operator unless $A(-t)\rho_1 = A(-t)\rho_2$, which can not be true, since $A(-t)$ must be one-to-one if it is to have an inverse. Hence, only pure state density operators can be mapped to pure state density operators by $A(-t)$. From this we can conclude that $A(t)$ must map all pure state density operators to pure state density operators, and must, in fact, map the set of pure state density operators one-to-one onto itself, since it has an inverse dynamical mapping. The group property, therefore, implies that we have an induced mapping of $\mathfrak{H}$ one-to-one onto itself. We need to determine that this induced mapping is linear.

Let ρ_ϕ be the projection operator whose range is the one-dimensional subspace of $\mathfrak{H}$ spanned by the normalized vector ϕ. Then,

$$\mathrm{Tr}\,(\rho_\psi\rho_\phi) = |(\psi, \phi)|^2. \tag{3.1}$$

The density operators are the operators ρ on $\mathfrak{H}$ of

[5] The notation is the same as in reference 1. $\psi\phi^\dagger$ is a linear operator defined on $\mathfrak{H}$ by its matrix elements with respect to any set of basis vectors $\phi^{(i)}$ in $\mathfrak{H}$ as $(\phi^{(i)}, \psi\phi^\dagger\phi^{(j)}) = (\phi^{(i)}, \psi)(\phi, \phi^{(j)})$.

the form

$$\rho = \sum_i a_i \rho_{\phi_i},$$

where $0 \leq a_i \leq 1$, $\sum_i a_i = 1$, and ϕ_i are a set of orthonormal vectors in $\mathfrak{H}$. Since $A(t)$ maps the set of pure state density operators one-to-one onto itself, we can let

$$A(t)\rho_{\phi_i} = \rho_{\phi_i'},$$

where ϕ_i' form a set of distinct normalized vectors in $\mathfrak{H}$. Then,

$$\rho' = A(t)\rho = \sum_i a_i \rho_{\phi_i'}.$$

Now

$$\mathrm{Tr}\,(\rho^2) = \sum_i a_i^2$$

and

$$\mathrm{Tr}\,(\rho'^2) = \sum_i a_i^2 + \sum_{i \neq j} a_i a_j \,|(\phi_i', \phi_j')|^2 \geq \mathrm{Tr}\,(\rho^2).$$

By applying the same argument to $A(-t)$, which maps ρ' to ρ, we get that

$$\mathrm{Tr}\,(\rho^2) \geq \mathrm{Tr}\,(\rho'^2)$$

which, together with the previous result, implies that

$$\mathrm{Tr}\,(\rho'^2) = \mathrm{Tr}\,(\rho^2),$$

which can be true only if

$$\mathrm{Tr}\,(\rho_{\phi_i'}\rho_{\phi_j'}) = |(\phi_i', \phi_j')|^2 = 0$$

for $i \neq j$. Hence, we can conclude that orthogonal projections are mapped to orthogonal projections by $A(t)$, or in other words that sets of orthonormal vectors in $\mathfrak{H}$ are mapped to sets of orthonormal vectors by the mapping induced by $A(t)$. From this it follows that, if ρ is a completely continuous symmetric operator belonging to $\mathfrak{L}$ which is mapped to ρ' by $A(t)$, then

$$\mathrm{Tr}\,(\rho'^2) = \mathrm{Tr}\,(\rho^2)$$

since a completely continuous symmetric operator has a pure point spectrum. In particular, since $\rho_\psi - \rho_\phi$ is a completely continuous symmetric operator, we have that

$$\mathrm{Tr}\,([\rho_\psi - \rho_\phi]^2) = 2 - \mathrm{Tr}\,(\rho_\psi\rho_\phi + \rho_\phi\rho_\psi)$$

is equal to

$$\mathrm{Tr}\,([\rho_{\psi'} - \rho_{\phi'}]^2) = 2 - \mathrm{Tr}\,(\rho_{\psi'}\rho_{\phi'} + \rho_{\phi'}\rho_{\psi'}),$$

where $\rho_{\psi'}$ and $\rho_{\phi'}$ are the images under the mapping $A(t)$ of ρ_ψ and ρ_ϕ, respectively, which implies, by Eq. (3.1), that

$$|(\psi, \phi)|^2 = |(\psi', \phi')|^2.$$

The mapping induced on $\mathfrak{H}$ by $A(t)$ therefore preserves the absolute value of inner products.

Wigner[6] has shown that if a mapping of a Hilbert space preserves the absolute value of inner products and is defined up to phase factors, then these phase factors can be chosen to make the mapping either linear and unitary or antilinear and antiunitary. The latter possibiliby is eliminated by the requirement that the dynamical mappings $A(t)$ form a weakly continuous one-parameter group with $A(0) = 1$.[7] Hence, for each value of t we have that, for any operator ρ belonging to $\mathfrak{L}$,

$$A(t)\rho = \omega(t)\rho\omega^\dagger(t),$$

where $\omega(t)$ is a unitary linear operator on $\mathfrak{H}$ which is defined up to a phase factor. The operators $\omega(t)$ form a one-parameter group up to a phase factor, that is

$$\omega(t)\omega(s) = c(t, s)\omega(t + s),$$

where $c(t, s)$ is a complex number of absolute value one.

From the weak continuity of $A(t)$ it follows that, for any vectors ψ and ϕ in $\mathfrak{H}$,

$$\mathrm{Tr}\,(\omega(t)\rho_\psi\omega^\dagger(t)\rho_\phi) = |(\omega(t)\psi, \phi)|^2$$

is a continuous function of t. The operators $\omega(t)$, therefore, give a continuous unitary ray representation of the additive group of real numbers. It has been shown by Bargmann[8] that in such a case the phase factors of the $\omega(t)$ can be chosen so that the $\omega(t)$ form a continuous one-parameter group of unitary operators on $\mathfrak{H}$. This completes the proof of the theorem.

We notice that the proof of this theorem remains valid if, instead of a representation of the additive group of real numbers, we are interested in a representation by dynamical mappings of any locally compact topological group for which we can use the theorems of Bargmann[8] to substitute a continuous unitary representation for a ray representation (representation up to a phase factor). For any symmetry group of this type, such as the Lorentz

[6] E. P. Wigner, *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press Inc., New York, 1959).

[7] The continuity condition is not needed here. That $\omega(t)$ is unitary is implied by the group property. For $\omega(t)$ is equal to within a phase factor to the square of $\omega(t/2)$, which is unitary whether $\omega(t/2)$ is unitary or antiunitary.

[8] V. Bargmann, Ann. Math. **59**, 1 (1954).

group, we can deduce the necessity of a representation by unitary transformations on the Hilbert space of state vectors.[2] The requirement that the induced mappings preserve the absolute values of inner products, or that "transition probabilities" be preserved by the symmetry transformations, need not be assumed; it can be *proved* from the requirement that each symmetry transformation maps density operators linearly to density operators.

IV. NON-HAMILTONIAN MAPPINGS

In this section we will show that the group property is necessary if a time dependent family of dynamical mappings is to be a family of Hamiltonian mappings. We present an example of a family of dynamical mappings $A(t)$ which have almost every property one might ask for except the group property. Let $\omega(t)$ be a continuous one-parameter semi-group of operators on $\mathfrak{H}$ which are isometric but not unitary. Examples of such semi-groups are well known.[9] For each operator ρ belonging to $\mathfrak{L}$, let

$$A(t)\rho = \omega(t)\rho\omega^{\dagger}(t). \tag{4.1}$$

It is easy to see that the $A(t)$ form a continuous one-parameter semi-group, and that, since pure state density operators are mapped to pure state density operators, each $A(t)$ is a dynamical mapping. Therefore, the $A(t)$ satisfy condition I. But each $A(t)$ does not have an inverse dynamical mapping, so they do not satisfy condition II.

In addition to the property of time dependence, we have that for each t the $A(t)$ of this example is an extremal element of the convex set of dynamical mappings which maps the set of pure state density operators one-to-one onto a subset of itself, induces a linear mapping on $\mathfrak{H}$, preserves the entropy or the trace of the square of density operators, and preserves the multiplication of operators on $\mathfrak{H}$.

Every continuous one-parameter semi-group of isometric operators is generated by a maximal symmetric operator,[10] so the transformation (4.1) can be thought of as the solution of a "Schrödinger equation" for the pure states with a maximal symmetric Hamiltonian operator. It is also known that such a semi-group can be made into a unitary group by extending the Hilbert space. The extension of the maximal symmetric generator of the semi-group is the self-adjoint generator of the unitary group.[9] Hence, the dynamical mappings defined by the isometric semi-group can be thought of as the restriction of a family of Hamiltonian dynamical mappings to a subspace of the pure states.

ACKNOWLEDGMENT

We would like to acknowledge the pleasure of discussing some of these topics with Professor J. E. Moyal of the Institute of Advanced Studies, Australian National University.

[9] F. Riesz and B. Sz. Nagy, *Functional Analysis*, translated from the second French edition by L. F. Boron; (F. Ungar Publishing Company, New York, 1955), p. 396.

[10] J. L. B. Cooper, Ann. Math **48**, 827 (1947); Proc. London Math Soc. **50**, 11 (1948).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

The Primitive Domains of Holomorphy for the 4 and 5 Point Wightman Functions*

A. C. MANOHARAN†
Department of Physics, Brandeis University, Waltham, Massachusetts
(Received April 2, 1962)

A method is given for estimating the dimensions of different types of boundary surfaces which are relevant for the domains of holomorphy, without using local commutativity, of the 4 and 5 point Wightman functions in scalar product space. The procedure is a straightforward application of the explicit parametrization given by Källén and Wightman for vectors on the boundary. The *DANAD* and other hypersurfaces for the 4 point function are deduced. For the 5 point function, in addition to generalizations of hypersurfaces corresponding to lower point functions, a new type of hypersurface appears which can be denoted in a matrix form $Z = DUMUD$ but not in the form $Z = DANAD$.

1. INTRODUCTION

MUCH effort and interest has centered in recent years on the study of relativistic quantum field theory based on the properties of vacuum expectation values of products of Heisenberg field operators, or in terms of Wightman functions.[1] Beginning with very general physical requirements on the theory, such as Lorentz invariance, absence of negative energy states and local commutativity, one attempts to obtain the mathematical and physical consequences of these assumptions. In the deduction of the consequences of the general requirements, the theory of analytic functions of several complex variables has played an important role since, by virtue of the spectral conditions and relativistic invariance, the Wightman functions are boundary values of analytic functions of several complex variables analytic in a region called the future tube. By virtue of their analyticity and invariance properties these functions are also analytic functions of the scalar products of their 4-vector variables, or of their invariant variables.[2] Local commutativity allows an extension of the domain of analyticity of these functions.

A complete determination of the domain of holomorphy,[3] using local commutativity, has been carried out[4] for the 3 point but not for the higher point functions. A possible first step in the procedure of determining the domain of holomorphy for these functions is to find the image in scalar product space of the mapping of the tube. We call this image the *primitive domain*. Källén and Wightman[5] have given an explicit parametrization of vectors corresponding to the boundary of the primitive domain, for all the functions. This enables one to investigate in detail the types of boundary surfaces that could arise. In this note we make a straightforward application of this parametrization to derive the boundary surfaces for the 4 and 5 point functions in scalar product space.[6] A detailed study of the hypersurfaces for the 4 point function has been hitherto carried out by several investigators.[7-12] The results we derive for the 4 point function have also been obtained by Källén using different techniques.

The contents of the present note are as follows: In Sec. 2 we outline the arguments leading to the definition of $B_n^{(m)}(N)$ surfaces ($n \leq 4$) which make up the boundary. In Sec. 3 we estimate the dimensions of these sets and pick only those which can possibly be hypersurfaces. It is shown that there are only five basic types of hypersurfaces. The first three (the Cut, S, and F) have already been studied[5] by the methods of this note. In Sec. 4 we show the equivalence of the fourth hypersurface to the *DANAD*. (It appears for the 4 and 5 point functions). Finally in Sec. 5 we prove that the fifth

* Supported in part by the National Science Foundation and the Office of Naval Research.
† Present address: Palmer Physical Laboratory, Princeton University, Princeton, New Jersey.
[1] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[2] D. Hall and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **31**, No. 5 (1957).
[3] For the definition of this term, see A. S. Wightman, in *Dispersion Relations and Elementary Particles* (John Wiley & Sons, Inc., New York, 1960), p. 229.
[4] G. Källén and A. S. Wightman, Kgl. Danske Videnskab. Selskab, Mat.-fys. Skrifter **1**, No. 6 (1958).
[5] Appendix II of reference 4.
[6] Professor A. S. Wightman informs me that he has obtained a similar derivation for the 4 point function boundary [J. Indian Math. Soc. **24**, 625 (1960)].
[7] D. J. Kleitman, 1959 (unpublished), see reference 9.
[8] R. Jost, 1959, see reference 10.
[9] G. Källén, in *Dispersion Relations and Elementary Particles* (John Wiley & Sons, Inc., New York, 1960), p. 389.
[10] G. Källén, Nuclear Phys. **25**, 568 (1961).
[11] C. Fronsdal, J. Math. Phys. **2**, 748 (1961).
[12] G. Eriksson (to be published).

type of hypersurface, which appears for the 5 point function, cannot be expressed in the *DANAD* form, but can be expressed in a modified form $Z = DUMUD$. In Appendix A a list is made of the equations in terms of invariant variables for the basic surfaces, and in Appendix B an alternative proof that points of *DUMUD* lie on the boundary of the 5 point is given. The arguments there indicate that the primitive domain is a domain of holomorphy.

2. DEFINITIONS AND PRELIMINARIES

As usual we restrict ourselves to the case of scalar fields $A_i(x)$. Our study is then concerned with the properties of the Wightman function

$$W^{(n+1)}(\xi_1, \xi_2, \cdots \xi_n)$$
$$= W^{(n+1)}(x_1 - x_2, x_2 - x_3, \cdots x_n - x_{n+1}) \quad (2.1)$$
$$= \langle 0 \,|A_1(x_1)A_2(x_2) \cdots A_{n+1}(x_{n+1})|\, 0\rangle \quad (2.2)$$

which, due to the absence of states of negative energy and spacelike momenta, are boundary values of a function

$$W^{(n+1)}(\zeta_1, \zeta_2, \cdots \zeta_n), \qquad \zeta_i = \xi_i - i\eta_i, \qquad i = 1, \cdots n \quad (2.3)$$

analytic[1] in the open set[13]

$$-\infty < \xi_i^{(\mu)} < \infty, \qquad \eta_i^2 > 0, \qquad \eta_i^{(0)} > 0, \qquad i = 1, \cdots n \quad (2.4)$$

called the future tube T_n. The Wightman functions are invariant under transformations of the proper orthochronous homogeneous Lorentz group $L_+^\uparrow$. The Bargmann-Hall-Wightman theorem[2] states that $W^{(n+1)}$ is also an analytic function of the scalar products $\zeta_j\zeta_k = z_{jk}$, $j, k = 1, \cdots n$. As ζ_i varies over the tube T_n, z_{jk} varies over a domain $\mathfrak{M}_n$ in scalar product space (*$\mathfrak{M}$ space*).

Our object is to determine the boundary (denoted $\partial\mathfrak{M}_n$) of $\mathfrak{M}_n$ for $n = 3, 4$. To achieve this, it is useful to consider the extended tube T_n' which is defined as the set of all points $\Lambda\zeta_i$, $i = 1, \cdots n$, where $\zeta_i \in T_n$ and $\Lambda \in \mathfrak{L}_+$ the group of all complex Lorentz transformations of determinant $+1$. The extended tube T_n' contains the tube T_n as a subset. Since a complex Lorentz transformation leaves scalar products of space-time vectors invariant, the extended tube also maps onto $\mathfrak{M}_n$.

To determine $\partial\mathfrak{M}_n$, one uses certain general properties[2,14] of the mapping of vectors ζ_i onto their scalar products $\zeta_i\zeta_k = z_{jk}$. It can be shown that the image of $\partial T_n'$ is $\partial\mathfrak{M}_n$ and that $\partial\mathfrak{M}_n$ is contained in the image of ∂T_n. Hence it is not necessary to consider all points of $\partial T_n'$. Let $B_n = (\partial T_n) \cap (\partial T_n')$, i.e., the intersection of the boundary of the tube and the boundary of the extended tube. In fact the image of B_n itself is $\partial\mathfrak{M}_n$. On the other hand, there exists *no* complex Lorentz transformation $\Lambda \in \mathfrak{L}_+$ which carries points of B_n into T_n. It is thus possible to partition the points of B_n into disjoint subsets $B_n^{(m)}$, $m = 1, \cdots n$: $B_n^{(m)}$ is the set of points ζ_i, $i = 1, \cdots n$, of B_n such that there exists a $\Lambda \in \mathfrak{L}_+$ carrying *any* $(m-1)$ element subset of ζ_i into T_{m-1}, but there exists no $\Lambda \in \mathfrak{L}_+$ carrying *some* m element subset of ζ_i into T_m.

Let those vectors of $B_n^{(m)}$ which cannot be transformed into T_m be relabeled ζ_i, $i = 1, \cdots m$. They are characterized by[15]

$$\eta_i^2 = 0, \qquad \eta_i^{(0)} > 0, \qquad i = 1, \cdots m \quad (2.5)$$

$$\sum_1^m \lambda_i(\xi_i \wedge \eta_i) = 0, \qquad \sum \lambda_i = 1, \qquad \lambda_i > 0, \qquad i = 1, \cdots m, \quad (2.6)$$

where $\wedge$ denotes the antisymmetric tensor product.[16]

Suppose there are N linearly independent vectors among the η_i, $i = 1, \cdots m$. We relabel them η_i, $i = 1, \cdots N$. We can then express the other vectors in terms of them as

$$\eta_k = \sum \beta_{kj}\eta_j, \qquad k = N+1, \cdots m, \qquad j = 1, \cdots N \quad (2.7)$$

$$\eta_l^2 = 0, \qquad \eta_l^{(0)} > 0, \qquad l = 1, \cdots m \quad (2.8)$$

and Eq. (2.6) may be written[5] in the more explicit form

$$\left.\begin{aligned} &\lambda_i\xi_i + \sum \lambda_k\beta_{ki}\xi_k = \sum \alpha_{ij}\eta_j, \\ &\tilde{\alpha} = \alpha \quad \text{(real)}[17] \\ &k = N+1, \cdots m, \quad i, j = 1, \cdots N \end{aligned}\right\} \quad (2.9)$$

$$\sum \lambda_l = 1, \qquad \lambda_l > 0, \qquad l = 1, \cdots m. \quad (2.10)$$

We see from Eqs. (2.7) and (2.9) that of the $2m$

[13] The scalar product of ζ is defined here as $\zeta^2 = \zeta^{(0)2} - \zeta^2$.

[14] A good part of what follows in Sec. 2 is a summary of the results of reference 5. We omit a discussion of the limit points mentioned there.

[15] We have normalized the λ_i, and also set them strictly positive, since if one is zero, say λ_j, then ζ_i, $i = 1, \cdots (j-1), (j+1), \cdots m$ belongs to $B_n^{(m-1)}$ too, which contradicts the definition of $B_n^{(m)}$.

[16] $(\xi_i \wedge \eta_i)^{(\mu\nu)} = \xi_i^{(\mu)}\,\eta_i^{(\nu)} - \xi_i^{(\nu)}\,\eta_i^{(\mu)}$, where $\mu < \nu$ are space-time indices.

[17] $\tilde{\alpha}$ denotes the transpose of the matrix α.

real vectors (ξ_i, η_i, $i = 1, \cdots m$), the maximum number that are linearly independent[18] is m.

We shall denote by $B_n^{(m)}(N)$ the subset of $B_n^{(m)}$ defined by Eqs. (2.7) to (2.10) where the indices of the vectors in these equations are permutations of the original ones in Eq. (2.3). The subset of $\partial\mathfrak{M}_n$ arising by mapping the points of $B_n^{(m)}(N)$ will be referred to as *the surface* $B_n^{(m)}(N)$.

3. THE SURFACES $B_n^{(m)}(N)$ $n = 2, 3, 4$

From the definitions of N and m we have $1 \leq N \leq m \leq n$, and since we are dealing with 4-space, we also have $N \leq 4$. We shall restrict our discussion, in the following, to the 3, 4, and 5 point functions, so that $n = 2, 3, 4$, respectively. The number of complex scalar products is $(n/2)(n+1)$ and the vectors ζ_i giving rise to them are, in general, linearly independent since $n \leq 4$. A surface of real dimension $n(n+1) - 1 = D$ (say) in the space $\mathfrak{C}_{(n/2)(n+1)}$ of $(n/2)(n+1)$ complex variables [or $n(n+1)$ real variables] will be called a *hypersurface*. For a given n we wish to know which of the $B_n^{(m)}(N)$ sets could give rise to hypersurfaces. We expect the boundary $\partial\mathfrak{M}_n$ to be covered by all such sets together with their limit points.

We shall first consider the surfaces $B_n^{(n)}(N)$ and obtain upper bounds for their real dimension. This upper bound will be denoted by P. We take, as our point of departure, the explicit parametrization given by Eqs. (2.7) to (2.10) for vectors of $B_n^{(m)}(N)$. Since $m = n$ we have $1 \leq N \leq \min(n, 4)$.

The real parameters λ_i, α_{ij}, β_{ki}, will be allowed to range over all values subject to the restrictions (2.10), $\tilde{\alpha} = \alpha$, and (2.7) to (2.8), respectively. The η_k, $k = N+1, \cdots n$, depend on the η_j, $j = 1, \cdots N$, through Eq. (2.7) and the ξ_i, $i = 1, \cdots N$, depend on the η_j through Eq. (2.9). From Eq. (2.9) we have

$$\lambda_i\zeta_i = \sum \alpha_{ij}\eta_j - i\lambda_i\eta_i - \sum \lambda_k\beta_{ki}\xi_k,$$
$$i, j = 1, \cdots N, \qquad k = N+1, \cdots n. \tag{3.1}$$

Also, from Eq. (2.7) we obtain

$$\zeta_k = \xi_k - i\sum \beta_{ki}\eta_j,$$
$$j = 1, \cdots N, \qquad k = N+1, \cdots n. \tag{3.2}$$

The scalar products $z_{ij} = \zeta_i\zeta_j$, $i, j = 1, \cdots n$, would then contain the real parameters λ_i, α_{ij}, β_{kj}, $\eta_i\eta_j$ $(i < j)$, $\eta_i\xi_k$, $\xi_k\xi_l$ $(k \leq l)$, $i, j = 1, \cdots N$, $k, l = N+1, \cdots n$. The total number of these parameters, the restrictions in (2.7) to (2.10) having been taken into account, will be our estimated upper bound for the (real) dimension of the surface $B_n^{(n)}(N)$ and is denoted by P.

The number of independent λ_i is $(n-1)$, of α_{ij} is $\frac{1}{2}N(N+1)$, of $\eta_i\eta_j$ is $\frac{1}{2}N(N-1)$, of $\eta_i\xi_k$ is $N(n-N)$, and of $\xi_k\xi_l$ is $\frac{1}{2}(n-N)(n-N+1)$. If $N = 1$ the number of β_{ki} is $(n-1)$. But if $N > 1$, the condition (2.8) that all the η_i lie on the forward cone (none being zero) gives a relation on the β's and they are not all independent. For example, if $N = 2$, $n = 3$, then $\eta_3 = \beta_1\eta_1 + \beta_2\eta_2$ and Eq. (2.8) implies $\beta_1\beta_2\eta_1\eta_2 = 0$. Now $\eta_1\eta_2 > 0$ since η_1, η_2 are linearly independent and by our stipulation $\eta_i^{(0)} > 0$ in Eq. (2.8) we cannot have $\beta_1 = 0 = \beta_2$. Thus $\beta_1 = 0$ or $\beta_2 = 0$. If $N = 3$, $n = 4$, then $\eta_3 = \beta_1\eta_1 + \beta_2\eta_2 + \beta_3\eta_3$, giving $\sum_{i<j} \beta_i\beta_j\eta_i\eta_j = 0$ and at least one $\beta_i \neq 0$, which equation may, in principle, be solved for one of the β_i in terms of the others and the $\eta_i\eta_j$ $(i < j)$. Hence, if $N > 1$, the number of independent parameters β_{ki} is $(n-N)(N-1)$. Adding up these numbers, we obtain

$$P = \begin{cases} \frac{1}{2}n(n+5) - 2 & \text{if } N = 1 \\ \frac{1}{2}n(n+1) - \frac{1}{2}N(N-1) + nN - 1 & \text{if } N > 1. \end{cases}$$

Thus, letting $R = D - P$,

$$R = n(n+1) - 1 - P$$
$$= \begin{cases} \frac{1}{2}(n-2)(n-1), & N = 1 \\ \frac{1}{2}(n-N)(n-N+1), & N > 1. \end{cases} \tag{3.3}$$

No hypersurface is possible if $R > 0$. On the other hand if $R \leq 0$ then we have suitable candidates for hypersurfaces. In fact inspection of Eq. (3.3) shows that the upper bound P has been chosen judiciously enough to give $R \geq 0$ for all the $B_n^{(n)}(N)$ surfaces $(n \leq 4)$. We see that (since $n \geq N$)

$$R = 0, \text{ if and only if } \begin{cases} n = 1, 2 & \text{for } N = 1 \\ n = N & \text{for } N > 1. \end{cases} \tag{3.4}$$

The values in Eq. (3.4) $(1 \leq n \leq 4)$ are

$$(n, N) = (1, 1), (2, 1), (2, 2), (3, 3), (4, 4). \tag{3.5}$$

The surfaces are,[19] respectively,

$$B_1^{(1)}(1), \quad B_2^{(2)}(1), \quad B_2^{(2)}(2), \quad B_3^{(3)}(3), \quad B_4^{(4)}(4)$$

and denoted by the respective names

[18] This result has been proved also by other means in reference 11.

[19] The 2 point function cut is only obtained by taking a suitable limit for the points of B_1. This is because we have excluded the tip of the light cone for the vector η in our discussion. This is also a boundary point of the tube. So, to be quite accurate, we could possibly have $N = 0$ when Eqs. (2.5) and (2.6) no longer apply. However we notice that $\zeta^2 = \xi^2$ in this case, and that if $\xi^2 \geq 0$ we are also on the boundary of the extended tube. This boundary $(\zeta^2 \geq 0)$ is in fact the cut.

$$\text{Cut} \quad S \quad F \quad DANAD \; DUMUD. \tag{3.5a}$$

They are, except for the last, already well known[5,10] and one readily satisfies oneself that they are in fact hypersurfaces and not of lower dimension. The *DUMUD* hypersurface for the 5 point function is discussed below in Sec. 5. We notice that the 3 point function is exceptional in the sense that it gives a hypersurface for $N < n$ as well as for $N = n$. The numbers P are given in Table I. For brevity we shall sometimes refer to the *DANAD* and *DUMUD* as the A and U surfaces, and denote the cut by C.

The surfaces $B_n^{(m)}(N)$, $m < n$ remain to be dealt with. In these cases Eqs. (2.7) to (2.10) parametrize *some* m-element (proper) subset of the vectors ζ_i, $i = 1, \cdots n$, each remaining ζ_i lying within the tube T_1. The totality of points of $B_n^{(m)}(N)$ is given by vectors $\zeta_{\sigma_1}, \cdots \zeta_{\sigma_m}$ satisfying Eqs. (2.7) to (2.10), where σ denotes *any* permutation of $1, \cdots n$. Of the $\frac{1}{2}n(n+1)$ scalar products $z_{\sigma_i\sigma_j}$, $i, j = 1, \cdots n$, those with $i, j = 1, \cdots m$ are parametrized in precisely the same form as $z_{\sigma_i\sigma_j}$, $i, j = 1, \cdots m$ in $B_m^{(m)}(N)$. Thus $B_n^{(m)}(N)$, $m < n$, can give rise to hypersurfaces[20] for no values of (m, N) other than

$$(m, N) = (1, 1), (2, 1), (2, 2), (3, 3). \tag{3.6}$$

We introduce generalizations of the surfaces C, S, F, A of (3.5a). The generalized surfaces are denoted by

$$C_i, S_{ij}, F_{ij}, A_{ijk} \quad \text{where} \quad i < j < k,$$
$$i, j, k = 1, \cdots n. \tag{3.6a}$$

The z_{rs}, $r, s = i, j, k$, in A_{ijk} satisfy the same equations as z_{rs}, $r, s = 1, 2, 3$, in A, the other z_{rs}, $r, s = 1, 2, \cdots 4$, $r, s \neq i, j$ or k, being free. The F_{ij}, S_{ij}, C_i, are defined in similar fashion. Then all points on the surface $B_n^{(m)}(N)$, $m < n$ (apart from limit points) must lie[21] on the appropriate surfaces of (3.6a). For example, any point of $B_4^{(3)}(3)$ lies on A_{123}, A_{124}, A_{134}, or A_{234}.

To summarize, for points on $\partial\mathfrak{M}_n$ it is sufficient to consider the hypersurfaces $B_n^{(n)}(N)$ with (n, N) given by Eq. (3.5), together with the surfaces $B_n^{(m)}(N)$, $m < n$, with (m, N) given by Eq. (3.6). The former set of surfaces $B_n^{(n)}(N)$ are precisely C, S, F, A, and U, whereas each of the latter set $B_n^{(m)}(N)$, $m < n$ is contained in the generalizations of C, S, F, and A. There are, in all, only the Cut, and four other basic types of surfaces. For example, in the case of $\partial\mathfrak{M}_4$ we need only consider $B_4^{(4)}(4)$, $B_4^{(3)}(3)$, $B_4^{(2)}(2)$, $B_4^{(2)}(1)$, and $B_4^{(1)}(1)$. Points of $B_4^{(4)}(4)$ form the U (or *DUMUD*) surface (Sec. 5).

TABLE I. Upper bounds for dimensions of $B_n^{(n)}(N)$ surfaces.

$(n+1)$ point function n	Dimension of boundary $\partial\mathfrak{M}_n$ $D = n(n+1) - 1$	No. of linearly independent η_i N	Upper bound for dimension of $B_n^{(n)}(N)$ P
2	5	2	5
		1	5
3	11	3	11
		2	10
		1	10
4	19	4	19
		3	18
		2	16
		1	16

[20] Roughly speaking, each $W^{(n+1)}$, $n > 1$, can have "lower point function boundaries." This is already clear for the 3 point. For the 4 point, Källén has numerically plotted plane sections of $\partial\mathfrak{M}_3$ exhibiting the intersection of *DANAD* with S_{ij} or F_{ij} (reference 10).

4. THE *DANAD* BOUNDARY FOR THE 4 POINT FUNCTION

We demonstrate now that the *DANAD* boundary for the 4 point function is precisely the surface $B_3^{(3)}(3)$. The equations for the latter, from Eqs. (2.7) to (2.10) are

$$\lambda_i\xi_i = \sum \alpha_{ij}\eta_j, \qquad \eta_i^2 = 0, \qquad \tilde{\alpha} = \alpha \quad \text{(real)},$$
$$\lambda_i > 0, \qquad \sum_1^3 \lambda_i = 1, \qquad i, j = 1, 2, 3, \tag{4.1}$$

where the η_i are 3 (variable) linearly independent vectors on the future light cone. Equation (4.1) gives

$$\zeta_i = \sum_j (\alpha_{ij}/\lambda_i - i\,\delta_{ij})\eta_j \tag{4.2}$$

and

$$z_{ik} = \zeta_i\zeta_k = \sum_{j,l} (\alpha_{ij}/\lambda_i - i\,\delta_{ij})\eta_j\eta_l \times (\alpha_{lk}/\lambda_k - i\,\delta_{lk}). \tag{4.3}$$

Let $\eta_j\eta_l = M_{jl}$, $j, l = 1, 2, 3$. Then $\tilde{M} = M$. $M_{jj} = 0$ and $M_{jl} > 0$, $j \neq l$. It is possible to express the scalar products of the η_i in the form

$$\eta_j\eta_l = M_{jl} = \mu_j N_{jl} \mu_l$$
$$\text{where} \quad \mu_j > 0, \quad j = 1, 2, 3,$$
$$\text{and} \quad N_{jl} = \begin{cases} 0 & j = l \\ 1 & j \neq l; \quad j, l = 1, 2, 3. \end{cases} \tag{4.4}$$

[21] On the other hand, parts of the surfaces (3.6a) certainly lie in the interior of the *singular* region for the function.

In fact, the μ_j are given uniquely by

$$\mu_1 = (M_{12}M_{13}/M_{23})^{1/2}, \quad \mu_2 = (M_{12}M_{23}/M_{13})^{1/2}, \quad \mu_3 = (M_{13}M_{23}/M_{12})^{1/2}. \tag{4.5}$$

The use of Eq. (4.4) in (4.3), with the substitutions

$$A_{ij} = \mu_i(\alpha_{ij} - i\lambda_i\,\delta_{ij})\mu_j \qquad D_{ij} = d_i\,\delta_{ij}, \quad d_i = (\mu_i\lambda_i)^{-1}, \tag{4.6}$$

gives

$$z_{ik} = \sum_{j,l} d_i A_{ij} N_{jl} A_{lk}\, d_k,$$

or in matrix form[6]

$$Z = DANAD\ (3 \times 3), \qquad (\tilde{A} = A, d_i > 0).$$

This expression is identical to that of Jost[8] and Källén[9,10]; the real part of A is a symmetric matrix and the imaginary part is diagonal with strictly negative elements.

The other hypersurfaces for the 4 point are parts of the 3 point type boundaries F_{ij}, S_{ij} (three of each) and the three cuts C_i, $1 \leq i < j \leq 3$. Equations for these surfaces (including the *DANAD*) in terms of the 6 complex variables z_{ij} in $\mathfrak{M}$ space are given in Appendix A. The *DANAD* boundary is denoted by A_{123}.

5. THE *DUMUD* BOUNDARY FOR THE 5 POINT FUNCTION

In this section it is proved that a new type of surface appears when we consider the 5 point function. This is the surface $B_4^{(4)}(4)$ which we show below can be written, in matrix notation, as $Z = DUMUD\ (4 \times 4)$ but *not* in the form $Z = DANAD\ (4 \times 4)$.

The equations for vectors on the boundary satisfy the same Eqs. (4.1) to (4.3) except that now the subscripts range from 1 to 4. There are 4 linearly independent vectors η_i on the future light cone, giving six independent scalar products $\eta_i\eta_j (i < j)$. It is not possible, in general, to find four μ_j, $j = 1, \cdots 4$ satisfying Eq. (4.4). However those points of $B_4^{(4)}(4)$ where Eq. (4.4) is satisfied give us a 17-dimensional subset forming the (4×4) *DANAD*. This surface certainly lies on the boundary $\partial\mathfrak{M}_4$ but is of 2 dimensions too low to be a hypersurface.[22]

Let

$$\eta_i\eta_l = M_{il}, \qquad j, l = 1, \cdots 4,$$
$$(\tilde{M} = M, M_{jj} = 0, M_{jl} > 0 \ \text{ for } \ j \neq l), \tag{5.1}$$

and then Eq. (4.3) can be written in the form

$$Z = DUMUD, \tag{5.2}$$

where

$$D_{ij} = d_i\,\delta_{ij}, \qquad d_i = (\lambda_i)^{-1} > 0 \tag{5.3}$$

and

$$U = \alpha - i\,D^{-1} = \tilde{U}. \tag{5.4}$$

This is 19-dimensional hypersurface.[23] The other possible hypersurfaces for the 5 point are parts of the four 4 point type boundaries A_{ijk}, the 3 point type boundaries F_{ij}, S_{ij} (six of each) and the four cuts C_i, $1 \leq i < j < k \leq 4$. Equations for these surfaces, in terms of the 10 complex variables z_{ij} in $\mathfrak{M}$ space, are given in Appendix A.

Another proof that *DUMUD* gives points on the boundary is given in Appendix B. We note here that the points of the surfaces $B_2^{(2)}(2)$, $B_3^{(3)}(3)$, as well as $B_4^{(4)}(4)$, are all expressible in the form (5.2) with 2-, 3-, and 4-dimensional matrices, respectively. The results of Appendix B give specific functions $(\Delta^{(+)})$ having singularities at all points of these surfaces. Those surfaces not covered by the formalism of Appendix B are analytic hypersurfaces (C_i, S_{ij}). (So is F_{ij}.) It follows that the primitive domain is a domain of holomorphy.

ACKNOWLEDGMENTS

I am indebted to Professor S. S. Schweber for several suggestions, and the keen interest with which he has directed this investigation. I am also grateful to Professor A. S. Wightman for informing me of his derivation of similar results and for an illuminating conversation. Many stimulating discussions with Professor D. J. Kleitman are also gratefully acknowledged.

APPENDIX A. $\mathfrak{M}$ SPACE EQUATIONS FOR THE BASIC TYPES OF SURFACES

We use the notation $z_i = z_{ii}$, $y_i = \text{Im}\ z_i$; α is a real symmetric matrix, M is a symmetric matrix with strictly positive nondiagonal elements, zero diagonal elements, and $d_i > 0$. Then we have the following equations:

The cut $C_i : z_i = \rho$, $0 \leq \rho < \infty$.

$S_{ij}(i < j)$: $2z_{ij} + kz_i + z_j/k = 0$,

$$0 < k < \infty, \qquad y_iy_j < 0.$$

[22] An alternative proof that all points $Z = DANAD\ (4 \times 4)$ lie on the 5 point function boundary has been given in reference 9.

[23] The *DANAD* (4×4) surface is that part of the *DUMUD* (4×4) satisfying the 2 restricting real conditions $M_{12}M_{34} = M_{13}M_{24} = M_{14}M_{23}$. It is of dimension 17.

$F_{ij}(i < j)$: $2z_{ij} = z_i z_j / r + r$,

$$0 < r < \infty, \qquad y_i y_j > 0.$$

A_{ijk} $(i < j < k)$: After elimination of only the diagonal elements of α, we have

$$y_p y_q z_{pq} = M_{pq} z_p z_q + a_{pq} z_p + b_{pq} z_q + c_{pq},$$

where

$$y_p = -2\, d_p(\alpha_{pq} M_{pq} + \alpha_{pr} M_{pr})$$

$$a_{pq} = 2\, d_q^2 \alpha_{qr}^2 M_{pr} M_{qr}, \qquad b_{pq} = 2\, d_p^2 \alpha_{pr}^2 M_{pr} M_{qr}$$

$$c_{pq} = 4\, d_p^2\, d_q^2 \alpha_{pq}^2 M_{pq} (\alpha_{pq} M_{pq} + \alpha_{pr} M_{pr} + \alpha_{qr} M_{qr})^2,$$

and $p < q$ are two of the numbers i, j, k, r being the third.

U: The 10 invariant scalar products are

$$z_{ij} = d_i\, d_j \sum \alpha_{ik} M_{kl} \alpha_{lj} - M_{ij} - i[d_i \sum \alpha_{ik} M_{kj} + d_j \sum \alpha_{ik} M_{ki}], \qquad 1 \le i \le j \le 4.$$

APPENDIX B. ALTERNATIVE PROOF THAT $Z = DUMUD$ LIES ON BOUNDARY

We give, in this Appendix, an alternative proof that all points $Z = DUMUD$ $(n \times n)$ lie on the boundary $\partial\mathfrak{M}_n$ for $n = 2$, 3, or 4.

Let $A = \alpha - i\Lambda$, $(n \times n)$, $n = 2$, 3, or 4, where $\tilde{\alpha} = \alpha$ (real),

$$\Lambda_{ij} = \lambda_i\, \delta_{ij}, \qquad \lambda_i > 0, \tag{B1}$$

and let η_i, $i = 1, \cdots n$, be linearly independent vectors on the forward light cone, i.e., $\eta_i^2 = 0$, $\eta_i^{(0)} > 0$. Then the points

$$\zeta_i = d_i \sum A_{ij} \eta_j \tag{B2}$$

lie within the tube T_n or on its boundary. Writing $D_{ij} = d_i\, \delta_{ij}$, $d_i > 0$, $\eta_i \eta_j = M_{ij}$ (which is > 0 or $= 0$ according as $i \neq j$ or $i = j$, $\tilde{M} = M$), and $z_{ij} = \zeta_i \zeta_j$, we obtain

$$Z = DAMAD \tag{B3}$$

and that these scalar products must lie inside or on $\partial\mathfrak{M}_n$.

On the other hand, the Fourier transform G of the boundary value $W^{(n+1)}(\xi_i)$ of $W^{(n+1)}(\zeta_i)$ is given[10,24] by

$$W^{(n+1)}(\xi_1, \cdots \xi_n) = \int dp_1 \cdots dp_n \times \exp(-i \sum p_i \xi_i) G(p_i p_k) \quad \theta(p_1) \cdots \theta(p_n). \tag{B4}$$

This enables one to write

$$W^{(n+1)}(z_{lm}) = \int \prod_{i \le k} da_{jk} G(a_{jk})\, \Delta_{n+1}^{(+)}(z_{lm}; a_{jk}) \tag{B5}$$

where

$$\Delta_{n+1}^{(+)}(z_{lm}; a_{jk}) = \int dp_1 \cdots dp_n \exp(-i \sum p_i \zeta_i) \times \theta(p_1) \cdots \theta(p_n) \prod_{i \le k} \delta(p_i p_k - a_{jk}) \tag{B6}$$

and the integration range over the masses in Eq. (B5) being defined by

$$a_{jk} \ge 0, \qquad -\begin{vmatrix} a_{jj} & a_{jk} \\ a_{jk} & a_{kk} \end{vmatrix} \ge 0 \text{ all } j, k,$$

$$\cdots (-)^{n+1} \det(a_{jk}) \ge 0. \tag{B7}$$

The z dependence of $W(z)$ is contained in the $\Delta^{(+)}$ function which has been explicitly evaluated.[24] Writing $Y = Za$ and denoting the eigenvalues of Y by σ_i, the singular manifold for the $\Delta_{n+1}^{(+)}$ function is given by[25]

$$\sum_{i=1}^{n} \pm (\sigma_i)^{1/2} = t \quad \text{(real number)}.$$

Let us choose masses $a = D^{-1} M D^{-1}$, that is, $a_{ij} = d_i^{-1} M_{ij} d_j^{-1}$. These masses satisfy the condition (B7) and are thus within the range of integration in Eq. (B5). We also have, for this choice of a,

$$Y = DAMAMD^{-1} = D(AM)^2 D^{-1}.$$

The eigenvalues of $(AM)^2$ are those of Y, namely, σ_i. Thus

$$\sum \pm (\sigma_i)^{1/2} = \operatorname{Tr} AM = \sum A_{ij} M_{ij} = \sum (\alpha_{ij} - i\Lambda_{ij}) M_{ij} = \sum \alpha_{ij} M_{ij} = t \quad \text{(say)}, \qquad \text{(real number)}.$$

This means that the points Z in Eq. (B3) must lie on or outside $\partial\mathfrak{M}_n$. Combining with the previous result, $Z = DAMAD$ lies *on* $\partial\mathfrak{M}_n$.

The above argument is valid for any choice of Λ provided its elements are strictly positive. The extra parameters may, in fact, be dropped by setting $\Lambda = D^{-1}$, when we get the $DUMUD$ form considered in the body of the text.

Notes added in proof.

(a) N. H. Möller, using techniques different from ours, has independently obtained (preprint) hypersurfaces for the 5 point function. His method consisted of investigating the singularity domain of the $\Delta_5^{(+)}$ function. A transformation $d_i \alpha_{ij} d_j \to \alpha_{ij}$,

[24] G. Källén and H. Wilhelmsson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Skrifter 1, No. 9 (1959).

[25] There is no correlation of $\pm$ signs in the sum.

$M_{ij} \to d_i M_{ij} d_j$ in our equation $Z = DUMUD$ (cf. the last equation of Appendix A) takes it to the form $Z = UMU$ and preserves the restrictions on the parameters α_{ij}, M_{ij}, and U_{ij}. The latter form is identical with his.

(b) The article cited in footnote 6 contains proofs of results we have summarized in Sec. 2 regarding the parametrization of vectors which give rise to $\partial\mathfrak{M}_n$. Using the new results contained in that article it is possible to give a direct treatment of the main limit points we have not discussed in the text. Special points of $B_n^{(m)}$ arise when $\xi_i \wedge \eta_i = 0$ for some of the vectors ζ_i, $i = 1, \cdots m$. In the case $\eta_i \neq 0$ we have $\xi_i = \alpha\eta_i$ which means that these points belong to $B_n^{(1)}$, i.e., the cut C_i. On the other hand, in the case $\eta_i = 0$, that is to say some of the vectors ζ_i are real, it has been shown that the parametric Eqs. (2.6), (2.7), (2.9), and (2.10) still hold with the following provisions: When Im $\zeta_i = 0$ for some values of i, the η_i with the same values of i no longer stand for Im ζ_i but denote auxiliary vectors satisfying $\xi_i\eta_i = 0$, $\eta_i^2 \geq 0$, $\eta_i^{(0)} > 0$. Parameters may again be counted almost exactly as done in Sec. 3. The only difference now is that on account of the additional restriction Im $\zeta_i = 0$, which is seen to be independent of the parametric equations, these special points of $B_n^{(m)}$ cannot give rise to hypersurfaces by themselves.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER–OCTOBER 1962

Heisenberg Fields which Vanish on Domains of Momentum Space*

O. W. GREENBERG
University of Maryland, College Park, Maryland
(Received April 18, 1962)

If a local Heisenberg field vanishes, or where appropriate has an infinite zero, on one of the momentum space domains $\mathfrak{A}$, $p^2 = -a^2$; $\mathfrak{B}$, $0 \leq p^2 < m^2$, and $p = 0$; or $\mathfrak{C}$, $p^2 > M^2$, then the field is a generalized free field. Counter examples show that this conclusion cannot be drawn if the field vanishes on the momentum space domains $\mathfrak{D}$, $0 \leq M_1^2 < p^2 < M_2^2$, $p \neq 0$; or $\mathfrak{E}$, $p = 0$. It follows that if two fields in the same Borchers class are equal on one of the domains $\mathfrak{A}$, $\mathfrak{B}$, or $\mathfrak{C}$, then the fields differ at most by a generalized free field in their Borchers class.

I. INTRODUCTION

LOCAL quantum field theories, at present, seem to separate into two types: one type which can neither be solved exactly nor be shown to exist mathematically, and another which can be constructed in closed form. The first kind, hopefully, allows the description of a wide range of physical phenomena, such as the interactions of elementary particles. The second kind, typically, describes trivial systems, such as collections of noninteracting particles. Some middle ground would be very valuable—cases of field theory which are soluble, in some sense, and have physical interest.

Of less importance, but still of interest, is the question: What restrictions, in addition to the general requirements of field theory, force a theory to be physically empty? Answering this question will help narrow our search for this middle ground, and may increase our understanding of the structure of quantum field theory.

In this paper we discuss the effect of the requirement that the Fourier transform of the field $\tilde{A}(p)$, should vanish or have an infinite zero on some domain in momentum space.[1-4] For Lorentz in-

* This research was in part supported by the U.S. Air Force through the Air Force Office of Scientific Research and Development Command under Contract AF 49(638)-24.

[1] G. F. Dell'Antonio, J. Math. Phys. 2, 759 (1961) gave a weaker result of the kind discussed here. Dell'Antonio showed that $A(x)$ is a generalized free field if $A(x)$ satisfies the usual repuirements of local quantum field theory and $\tilde{A}(p)$ vanishes for both $p^2 < 0$ and $p^2 > M^2$. We thank Dr. Dell'Antonio for communicating his results prior to publication.

[2] H. J. Borchers (private communication) has found results equivalent to our cases $\mathfrak{a}$ and $\mathfrak{c}$ below (and also equivalent counter examples for cases $\mathfrak{D}$ and $\mathfrak{E}$) using methods of the theory of analytic functions of several complex variables rather than the wave equation technique we use. Dr. Borchers informed us of his result for case $\mathfrak{C}$ before we obtained a result for this case. Our knowledge of Dr. Borchers' proof using the kantensatz was helpful in constructing our proof via the wave equation. In addition, Dr. Borchers has pointed out that since $\rho(k^2) = 0$, $k^2 > M^2$, where ρ is the Källén-Lehmann weight of the two point function, implies $\tilde{A}(k)\Psi_0 = 0$, $k^2 > M^2$, which in turn implies the same region of vanishing for $f_p(q)$ as $\tilde{A}(k) = 0$, $k^2 > M^2$ [see Sec. 3

$M_{ij} \to d_i M_{ij} d_j$ in our equation $Z = DUMUD$ (cf. the last equation of Appendix A) takes it to the form $Z = UMU$ and preserves the restrictions on the parameters α_{ij}, M_{ij}, and U_{ij}. The latter form is identical with his.

(b) The article cited in footnote 6 contains proofs of results we have summarized in Sec. 2 regarding the parametrization of vectors which give rise to $\partial\mathfrak{M}_n$. Using the new results contained in that article it is possible to give a direct treatment of the main limit points we have not discussed in the text. Special points of $B_n^{(m)}$ arise when $\xi_i \wedge \eta_i = 0$ for some of the vectors ζ_i, $i = 1, \cdots m$. In the case $\eta_i \neq 0$ we have $\xi_i = \alpha\eta_i$ which means that these points belong to $B_n^{(1)}$, i.e., the cut C_i. On the other hand, in the case $\eta_i = 0$, that is to say some of the vectors ζ_i are real, it has been shown that the parametric Eqs. (2.6), (2.7), (2.9), and (2.10) still hold with the following provisions: When Im $\zeta_i = 0$ for some values of i, the η_i with the same values of i no longer stand for Im ζ_i but denote auxiliary vectors satisfying $\xi_i\eta_i = 0$, $\eta_i^2 \geq 0$, $\eta_i^{(0)} > 0$. Parameters may again be counted almost exactly as done in Sec. 3. The only difference now is that on account of the additional restriction Im $\zeta_i = 0$, which is seen to be independent of the parametric equations, these special points of $B_n^{(m)}$ cannot give rise to hypersurfaces by themselves.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER–OCTOBER 1962

Heisenberg Fields which Vanish on Domains of Momentum Space*

O. W. GREENBERG
University of Maryland, College Park, Maryland
(Received April 18, 1962)

If a local Heisenberg field vanishes, or where appropriate has an infinite zero, on one of the momentum space domains $\mathfrak{A}$, $p^2 = -a^2$; $\mathfrak{B}$, $0 \leq p^2 < m^2$, and $p = 0$; or $\mathfrak{C}$, $p^2 > M^2$, then the field is a generalized free field. Counter examples show that this conclusion cannot be drawn if the field vanishes on the momentum space domains $\mathfrak{D}$, $0 \leq M_1^2 < p^2 < M_2^2$, $p \neq 0$; or $\mathfrak{E}$, $p = 0$. It follows that if two fields in the same Borchers class are equal on one of the domains $\mathfrak{A}$, $\mathfrak{B}$, or $\mathfrak{C}$, then the fields differ at most by a generalized free field in their Borchers class.

I. INTRODUCTION

LOCAL quantum field theories, at present, seem to separate into two types: one type which can neither be solved exactly nor be shown to exist mathematically, and another which can be constructed in closed form. The first kind, hopefully, allows the description of a wide range of physical phenomena, such as the interactions of elementary particles. The second kind, typically, describes trivial systems, such as collections of noninteracting particles. Some middle ground would be very valuable—cases of field theory which are soluble, in some sense, and have physical interest.

Of less importance, but still of interest, is the question: What restrictions, in addition to the general requirements of field theory, force a theory to be physically empty? Answering this question will help narrow our search for this middle ground, and may increase our understanding of the structure of quantum field theory.

In this paper we discuss the effect of the requirement that the Fourier transform of the field $\tilde{A}(p)$, should vanish or have an infinite zero on some domain in momentum space.[1-4] For Lorentz in-

* This research was in part supported by the U.S. Air Force through the Air Force Office of Scientific Research and Development Command under Contract AF 49(638)-24.

[1] G. F. Dell'Antonio, J. Math. Phys. 2, 759 (1961) gave a weaker result of the kind discussed here. Dell'Antonio showed that $A(x)$ is a generalized free field if $A(x)$ satisfies the usual repuirements of local quantum field theory and $\tilde{A}(p)$ vanishes for both $p^2 < 0$ and $p^2 > M^2$. We thank Dr. Dell'Antonio for communicating his results prior to publication.

[2] H. J. Borchers (private communication) has found results equivalent to our cases $\mathfrak{a}$ and $\mathfrak{c}$ below (and also equivalent counter examples for cases $\mathfrak{D}$ and $\mathfrak{E}$) using methods of the theory of analytic functions of several complex variables rather than the wave equation technique we use. Dr. Borchers informed us of his result for case $\mathfrak{C}$ before we obtained a result for this case. Our knowledge of Dr. Borchers' proof using the kantensatz was helpful in constructing our proof via the wave equation. In addition, Dr. Borchers has pointed out that since $\rho(k^2) = 0$, $k^2 > M^2$, where ρ is the Källén-Lehmann weight of the two point function, implies $\tilde{A}(k)\Psi_0 = 0$, $k^2 > M^2$, which in turn implies the same region of vanishing for $f_p(q)$ as $\tilde{A}(k) = 0$, $k^2 > M^2$ [see Sec. 3

variant theories, it suffices to consider invariant regions in momentum space corresponding to intervals in the variable p^2, together with the origin, $p = 0$. We will treat the cases:

$$\mathfrak{A}, p^2 = -a^2; \quad \mathfrak{B}, 0 \leq p^2 < m^2, \text{ and } p = 0;$$

$$\mathfrak{C}, p^2 > M^2; \quad \mathfrak{D}, 0 \leq M_1^2 < p^2 < M_2^2, p \neq 0;$$

$$\text{and } \mathfrak{E}, p = 0.$$

We use the same script, capital letter to label the case and the corresponding region in momentum space. With one rather singular exception,[5] the cases we consider exhaust the possible connected invariant regions in momentum space. For each of cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$, we show that vanishing of $\tilde{A}(p)$ on the corresponding domain implies that $A(x)$ is a generalized free field.[1,6,7] For cases $\mathfrak{D}$ and $\mathfrak{E}$ we exhibit a counter example to show that in general, the above conclusion cannot be drawn for $\tilde{A}(p)$ vanishing in $\mathfrak{D}$ or $\mathfrak{E}$.

As corollaries of these results, we show that if fields which commute with each other at space-like separation in configuration space (i.e., are in the same Borchers class[8]) are equal in either of the regions $\mathfrak{A}$, $\mathfrak{B}$, or $\mathfrak{C}$, then the fields differ at most by a generalized free field which commutes with both of them at space-like separation in configuration space (i.e., is in their Borchers class).

In Sec. 2, we discuss the main mathematical tool which we use in this work, Gårding's analysis of distributions on a four-vector (in a Lorentz metric space) which vanish for space-like value of their argument. Section 3 contains the proof of our results including the corollary for cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$ and Sec. 4 contains counter examples for cases $\mathfrak{D}$ and $\mathfrak{E}$.

2. GÅRDING'S ANALYSIS OF LOCAL DISTRIBUTIONS

Our main mathematical tool in this work will be Gårding's analysis[9,10] of temperate local distributions $\tilde{f}(x)$, where by "local" we mean that $\tilde{f}(x) = 0$, $x^2 < 0$. Since this work is not well known, we give a brief summary of Gårding's analysis following Wightman's exposition[9] and refer the reader to that exposition for a more detailed discussion.

Gårding showed that for every local temperate distribution $\tilde{f}(x)$ there corresponds a unique temperate distribution $G(q, \sigma)$ defined by

$$G(q, \sigma) = \frac{1}{(2\pi)^4} \int \cos \sigma(x^2)^{1/2} e^{iq\cdot x} \tilde{f}(x)\, d^4x,$$

which satisfies the five-dimensional wave equation,

$$\left(\frac{\partial^2}{\partial q^{02}} - \sum_1^3 \frac{\partial^2}{\partial q^{i2}} - \frac{\partial^2}{\partial \sigma^2}\right) G(q, \sigma) = 0, \tag{1}$$

is even in σ

$$G(q, \sigma) = G(q, -\sigma), \tag{2}$$

and whose restriction to the plane $\sigma = 0$ satisfies

$$G(q, 0) = f(q) = \frac{1}{(2\pi)^4} \int e^{iq\cdot x} \tilde{f}(x)\, d^4x. \tag{3}$$

Conversely, for every temperate G satisfying Eqs. (1) and (2) there is a unique local temperate $\tilde{f}$ whose Fourier transform f is given by Eq. (3); and in particular the null functions $G = 0$ and $\tilde{f} = 0$ correspond uniquely to each other.

According to Eq. (3), statements concerning the support of $f(q)$ in momentum space lead directly to statements about $G(q, \sigma)$ in the plane $\sigma = 0$.[11] If G vanishes in some region of (q, σ) space there are three principles which allow this region to be enlarged:

1. If G and all its normal derivatives vanish on a time-like curve, then G vanishes in the double cone region subtended by that curve.[9]

(d)], it suffices to assume that ρ vanishes above some mass for case $\mathfrak{C}$. We thank Dr. Borchers for communicating his results prior to publication.

[3] After the present paper was completed, we received a preprint "Support of a field in momentum space," by Derek W. Robinson, Institut für Theoretische Physik, ETH, Zurich, which contains a result equivalent to our case $\mathfrak{A}$. We thank Dr. Robinson for communicating his results prior to publication. *Note added in proof.* Dr. Robinson has informed us later (private communication) that he has also obtained independently results equivalent to our cases $\mathfrak{B}$ and $\mathfrak{C}$.

[4] There is an earlier result of a similar kind, namely, that if the Källén-Lehmann weight of a field is concentrated on a single mass, then this field is a free field of that mass. This result was proved independently by P. G. Federbush and K. A. Johnson, Phys. Rev. **120**, 1926 (1960); R. Jost, *Lectures on Field Theory and the Many Body Problem*, edited by E. R. Caianiello (Academic Press Inc., New York, 1961), pp. 127–145; and B. Schroer (unpublished).

[5] The exception is $\mathfrak{F}$, $p^2 = 0$, and $p = 0$.

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

[7] A. L. Licht and J. S. Toll. Nuovo cimento **21**, 346 (1961).

[8] H. J. Borchers, Nuovo cimento **15**, 784 (1960).

[9] L. Gårding, (unpublished). This work is discussed by A. S. Wightman, *Dispersion Relations and Elementary Particles*, edited by C. de Witt and R. Omnes, (Hermann et Cie, Paris, France, 1960), pp 291–308.

[10] The analysis of local functions due to Jost and Lehmann, and Dyson; R. Jost and H. Lehmann, Nuovo cimento **5**, 1598 (1957); and F. J. Dyson, Phys. Rev. **110**, 1460 (1958), also leads to the results found here when the Jost-Lehmann-Dyson method is extended to treat more general regions of vanishing of $f(q)$ than were originally considered (private communication from H. J. Borchers and R. Stora).

[11] It is important to note that it is permissible to consider the distribution, $G(q, \sigma)$, restricted to a given value of σ, because Malgrange and Gårding have shown that a temperate distribution which satisfies the wave equation is C^∞ in its space-like variables when integrated with a test function in its time-like variable, and vice versa.[9] This theorem of Malgrange and Gårding also allows the discussion of an infinite zero in the commutator matrix elements of $\tilde{A}(p)$ which are considered in Sec. 3.

2. If G and its first normal derivative vanish on a space-like disk, then G vanishes in the double cone region subtended by that disk.[9]

3. A solution of the wave equation with prescribed initial data on a given space-like surface is analytic at any point which cannot be reached by characteristics from regions on the given space-like surface where the initial data is nonvanishing. (For an odd number of space-like dimensions the above statement holds in a stronger form with "vanishes" replacing "is analytic.") If the solution of the wave equation is known to vanish somewhere in the region of analyticity, then this third principle allows the region of vanishing to be extended to the whole region of analyticity.[12,13]

Statement 1 is remarkable since it states that the values of G in a solid five-dimensional region are controlled by the values of G and its derivatives on a single time-like line. Statement 2 is Huygens principle. It follows from Asgeirsson's lemma, however, that these two statements are equivalent.[9] Statement 3 follows from the form of the fundamental solutions of the wave equation.

3. PROOF THAT $A(x)$ IS A GENERALIZED FREE FIELD FOR CASES $\mathfrak{A}$, $\mathfrak{B}$, AND $\mathfrak{C}$

(a) Requirements of Quantum Field Theory

We list the usual requirements of local field theory[14]: I. relativistic transformation properties; II. unique, normalizable, invariant vacuum state Ψ_0, and no negative energy states or states of space-like momenta (spectrum); III. vanishing of the commutator $[A(x), A(y)]$ for $x - y$ space-like (locality); IV. completeness of the set of states obtained by applying polynomials in the smeared field operators to the vacuum state Ψ_0. In addition, we will strengthen property II to require that the states above the vacuum all have nonzero mass, i.e., for all (improper) energy-momentum eigenstates $\Psi_{p,\alpha}$, $p^2 \geq \mu^2 > 0$. We consider the case of a single neutral scalar field; however this simplifying assumption is not essential.

(b) Reduction to Proof of Vanishing of a Matrix Element of the Commutator

We will show below that for cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$,

$$(\Psi_{p,\alpha}, [A(x), A(y)]\Psi_0)$$

vanishes unless $\Psi_{p,\alpha}$ is the vacuum state [where $\{\Psi_{p,\alpha}\}$ is a complete set of (improper) energy-mementum eigenstates labeled by energy-momentum p and other quantum numbers α]. From this fact it follows that the commutator behaves like the c number

$$i\,\Delta'(x - y) \equiv (\Psi_0, [A(x), A(y)]\Psi_0)$$

acting on the vacuum state Ψ_0, where Δ' is an integral over masses of free field commutator functions. Then

$$\{[A(x), A(y)] - i\,\Delta'(x - y)\}\Psi_0 = 0,$$

and a fundamental argument due to Jost,[15] which we will not repeat in this note, leads to the conclusion that the commutator is a c number, i.e.,

$$[A(x), A(y)] = i\,\Delta'(x - y).$$

This c-number property of the commutator, together with the requirement (property II) of no negative energy states, implies that $A(x)$ is a generalized free field. The only assertion in this paragraph which remains to be proved is the first one,

$$\tilde{F}_{p,\alpha}(x, y) = (\Psi_{p,\alpha}, [A(x), A(y)]\Psi_0) = 0, \quad \text{unless} \quad \Psi_{p,\alpha} = \Psi_0. \tag{4}$$

We devote the next three subsections to proving Eq. (4) for cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$.

(c) Domains $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$

We will show that if, in addition to the general requirements of subsection 3(a), the Fourier transformed field $\tilde{A}(p)$ vanishes or has an infinite zero on any one of certain domains of momentum space, then $A(x)$ is a generalized free field. [The requirement that for p in some region $[\tilde{A}(p), A(x)]$ vanish for all x, which leads to the same results, is not really weaker than vanishing of $\tilde{A}(p)$.] In the cases where we require an infinite zero of $\tilde{A}(p)$ on a given domain, this requirement can be replaced by $\tilde{A}(p) = 0$ on any arbitrarily small open domain which includes the original one. The cases for which we show that $A(x)$ is a generalized free field are:

---

[12] We thank Dr. J. Peetre for bringing this third principle to our attention.

[13] We conjecture that a still stronger principle of vanishing holds: (2′) (Generalized Huygens principle). If from some point, P, all straight lines lying in or on one cone intersect one or more disconnected regions in which G and its first normal derivative vanish, then G vanishes at P. We can express this conjecture heuristically by saying that light must travel in straight lines.

[14] A. S. Wightman, Phys. Rev. **101**, 860 (1956); and "Problemes mathematiques de la theorie quantique des champs," University of Paris lecture notes, 1957 (unpublished).

[15] R. Jost, *Lectures on Field Theory and the Many-Body Problem*, edited by E. R. Caianiello (Academic Press Inc., New York, 1961), p. 136.

$\mathfrak{A}$: $\tilde{A}(p) = 0$, and $(\partial/\partial p^2)^n \tilde{A}(p) = 0$, n integral, for $p^2 = -a^2$.

$\mathfrak{B}$: $\tilde{A}(p) = 0$, $0 \leq p^2 < m^2$; and $\tilde{A}(p) = 0$, and $\mathcal{P}(\partial/\partial p^\mu)\tilde{A}(p) = 0$, for $p = 0$, where $\mathcal{P}$ is an arbitrary polynomial of differentiation.

$\mathfrak{C}$: $\tilde{A}(p) = 0$, $p^2 > M^2$.

(d) Requirements on the Commutator Function

For convenience in proving Eq. (4), we introduce the commutator function, $\tilde{f}_{p,\alpha}(x)$, from which the center-of-mass coordinate dependence has been removed

$$\tilde{f}_{p,\alpha}(x) \equiv (\Psi_{p,\alpha}, [A(x/2), A(-x/2)]\Psi_0),$$

and also the Fourier transforms, $F_p(k_1, k_2)$ and $f_p(q)$ of $\tilde{F}$ and $\tilde{f}$, where, for example,

$$\tilde{f}_p(x) = \int d^4q f_p(q) e^{-iq\cdot x}.$$

We have suppressed the quantum numbers α, since they play no part in our discussion. It is straightforward to find the relations between the large and small f's. They are

$$\tilde{F}_p(x, y) = e^{ip\cdot(1/2)(x+y)} \tilde{f}_p(x - y)$$

and

$$F_p(k_1, k_2) = \delta(p + k_1 + k_2) f_p(k_1 + \tfrac{1}{2}p).$$

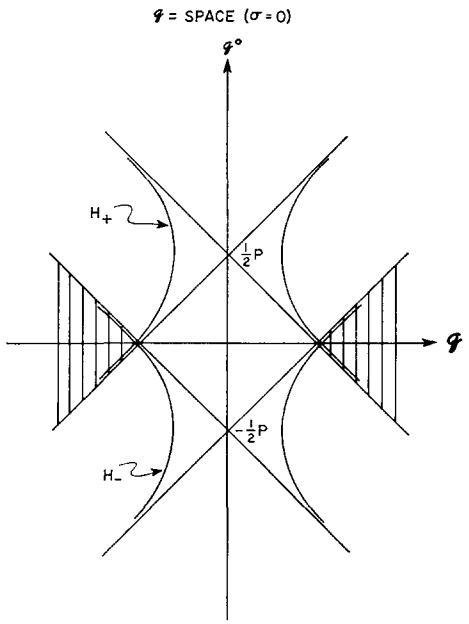


FIG. 1. Domains of vanishing of $f_p(q)$ for case $\mathfrak{A}$: $f_p(q) = 0$ in the shaded regions from property II, spectrum (the existence of a lowest mass state is not used for this region), and $f_p(q)$ has an infinite zero on the hyperboloids $H_\pm$: $(q \mp \frac{1}{2}p)^2 = -a^2$ from the assumption that $\tilde{A}(k)$ has an infinite zero on $k^2 = -a^2$.

The conditions on f_p are

$$\tilde{f}_p(x) = 0, \qquad x^2 < 0$$

from property III (locality),

$f_p(q) = 0$, outside the union of $q > -\frac{1}{2}p$ and $q < \frac{1}{2}p$, where $k_1 > k_2$ means $(k_1 - k_2)^2 \geq 0$, $k_1^0 - k_2^0 > 0$,

from property II (spectrum) and

$f_p(q) = 0$ on domains $\mathfrak{A}$, $\mathfrak{B}$, or $\mathfrak{C}$, one at a time from vanishing of $\tilde{A}(p)$ where, in detail, these conditions are

$\mathfrak{A}$: $f_p(k_1 + \frac{1}{2}p) = 0$, and $(\partial/\partial k_1^2)^n f_p(k_1 + \frac{1}{2}p) = 0$, n integral, for $k_1^2 = -a^2$; and the same with $+k_1$ replaced by $-k_2$, and $k_1 + \frac{1}{2}p$ replaced by $-k_2 - \frac{1}{2}p$.

$\mathfrak{B}$: $f_p(k_1 + \frac{1}{2}p) = 0$, $0 \leq k_1^2 < m^2$; $f_p(k_1 + \frac{1}{2}p) = 0$, and $\mathcal{P}(\partial/\partial k_1^\mu) f_p(k_1 + \frac{1}{2}p) = 0$, $k_1 = 0$; and the same with $k_1 \to -k_2$, and $k_1 + \frac{1}{2}p \to -k_2 - \frac{1}{2}p$.

$\mathfrak{C}$: $f_p(k_1 + \frac{1}{2}p) = 0$, $k_1^2 > M^2$; and the same with $k_1 \to -k_2$, and $k_1 + \frac{1}{2}p \to -k_2 - \frac{1}{2}p$. Since p is time-like here (spectrum) we choose $p = (\mathfrak{p}, \mathbf{0})$ with no loss of generality.

(See Figs. 1, 2, or 3, respectively.)

(e) Vanishing of f_p for Cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$

Now we show that $\tilde{f}_p(x)$ vanishes for cases $\mathfrak{A}$, $\mathfrak{B}$, and $\mathfrak{C}$. We refer again to Figs. 1, 2, and 3 in which the shaded regions and heavy lines are domains where $G_p(q, 0) = f_p(q)$ vanishes. We note that in cases $\mathfrak{A}$ and $\mathfrak{B}$ the only use we make of property II, spectrum, is to place p on the time axis without loss of generality. In case $\mathfrak{C}$ the only use we make of the strengthened spectral condition which prohibits light-like states is again to place p on the time axis. We consider the three cases in alphabetical order and show for each that principles 1, 2, and 3 allow the regions where G_p vanishes to be enlarged to the whole (q, σ) space.

This conclusion follows for case $\mathfrak{A}$ in two steps. First we use principle 1, which is valid for the family of time-like curves $\{T_\lambda\}$ lying in the hyperboloids $(q - \frac{1}{2}p)^2 = -a^2$ and $(q + \frac{1}{2}p)^2 = -a^2$, with end points $(\frac{1}{2}\mathfrak{p} + (\mathfrak{q}^2 - a^2)^{1/2}, \mathfrak{q}, 0)$ and $(-\frac{1}{2}\mathfrak{p} - (\mathfrak{q}^2 - a^2)^{1/2}, \mathfrak{q}, 0)$, for all $|\mathfrak{q}| = \lambda$, provided G_p and all its derivatives vanish on this family of curves. This last condition holds since (i) all derivatives along the surfaces of these hyperboloids vanish since G_p vanishes on these hyperbloids, (ii) all derivatives normal to the surface of these

hyperboloids, but in the plane $\sigma = 0$ vanish by the infinite zero hypothesis, (iii) all odd σ derivatives of G_p vanish because G_p is even in σ, Eq. (2), and finally (iv) all even σ derivatives of G_p, starting with $\partial^2/\partial\sigma^2$, vanish since G_p satisfies the wave equation Eq. (1). Using principle (1) for longer and longer families of lines $\{T_\lambda\}$ we find that $G_p(q, \sigma)$ vanishes in the union, $\mathcal{C}_\lambda$, of the family of double cones with vertices at $(\frac{1}{2}\mathfrak{p} + (\mathbf{q}^2 - a^2)^{1/2}, \mathbf{q}, 0)$ and $(-\frac{1}{2}\mathfrak{p} - (\mathbf{q}^2 - a^2)^{1/2}, \mathbf{q}, 0)$, for all $|\mathbf{q}| = \lambda$. For any finite point $P : (0, \mathbf{q}, \sigma)$ in the $q^0 = 0$ plane there is a $\lambda_0 \ni$ for $\lambda > \lambda_0$ P is included in C_λ,[16] and thus in the limit $\lambda \to \infty$, $G_p(q, \sigma)$ [and by similar arguments $(\partial/\partial q^0)G_p(q, \sigma)$] vanishes in this $q^0 = 0$ plane. Now from principle (2) we conclude that $G_p(q, \sigma)$ vanishes everywhere, and, therefore, so does $\tilde{f}_p(x)$, which completes the proof that $A(x)$ is a generalized free field for case $\mathfrak{A}$.

The same conclusion follows for case $\mathfrak{B}$ from similar arguments based on the region of vanishing for this case (Fig. 2).

For case $\mathfrak{C}$ we first use principle 1 which suffices to prove vanishing of $G_p(q, \sigma)$ in

$$\alpha : \{\cup_{(k)} V^+_{(k,0)} \mid (k + \tfrac{1}{2}p)^2 = M^2, k^0 + \tfrac{1}{2}\mathfrak{p} > 0\};$$

$$\beta : \{\cup_{(k)} V^-_{(k,0)} \mid (k - \tfrac{1}{2}p)^2 = M^2, k^0 - \tfrac{1}{2}\mathfrak{p} < 0\};$$

and

$$\gamma : \{\cup_{(k)} V^-_{(k,0)} \mid (k + \tfrac{1}{2}p)^2 = 0, k^0 + \tfrac{1}{2}\mathfrak{p} > 0, \times |\mathbf{k}| > \tfrac{1}{2}\mathfrak{p}\} \cap \{\cup_{(k)} V^+_{(k,0)} \mid (k - \tfrac{1}{2}p)^2 = 0, \times k^0 - \tfrac{1}{2}\mathfrak{p} < 0, |\mathbf{k}| > \tfrac{1}{2}\mathfrak{p}\},$$

where $V^{\pm}_{(q,\sigma)}$ is the five-dimensional forward (backward) cone with vertex at (q, σ). It is clear intuitively and is confirmed by an envelope calculation that these regions are

$$\alpha : V^+_{(-p/2,M)} \cap V^+_{(-p/2,-M)};$$

$$\beta : V^-_{(p/2,M)} \cap V^-_{(p/2,-M)}; \quad \text{and}$$

$$\gamma : (q, \sigma) \notin [V^+_{(-p/2,\tau)} \cup V^-_{(p/2,\tau)}], \ \tau \text{ arbitrary}.$$

Regions α and β are the intersections of five-dimensional cones; region γ is a cylinder, with base the original region of vanishing in the $\sigma = 0$ plane, and with developable sides orthogonal to this $\sigma = 0$ plane. Since α, β, and γ are disjoint, principle 2 leads to no further increase in the region of vanishing.

In order to prove that G_p vanishes identically, it suffices to connect α and β to γ. We do this by proving vanishing in the neighborhood of a point

[16] For $P = (0, \mathbf{q}, \sigma)$, $\lambda_0^2 = (p^0)^{-2}[\sigma^2 + a^2 - (\frac{1}{2}p^0)^2]^2 + a^2$.

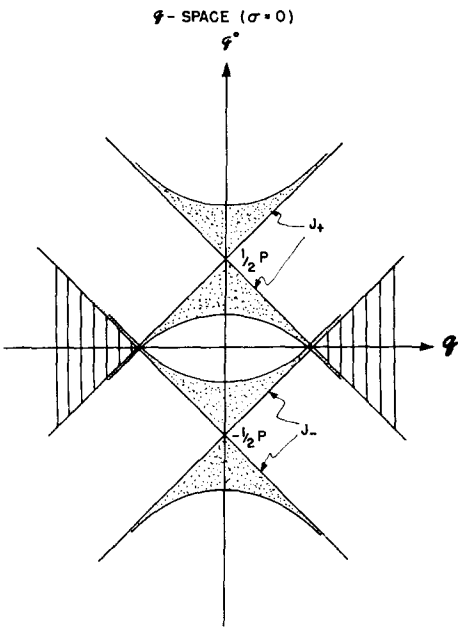


FIG. 2. Domains of vanishing of $f_p(q)$ for case $\mathfrak{B}$: $f_p(q) = 0$ in the dotted region bounded by $J_\pm$: $(q \mp \frac{1}{2}p)^2 = m^2$, $(\check{A}(k) = 0, 0 \leq k^2 < m^2)$, and $f_p(q)$ has an infinite zero at $q = \pm\frac{1}{2}p(\check{A}(k)$ has an infinite zero at $k = 0)$. The figure is drawn for the case $m^2 < (\frac{1}{2}p)^2$.

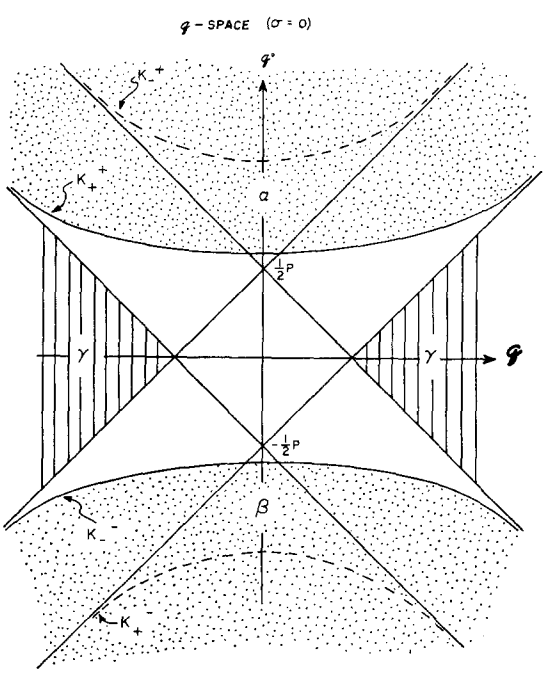


FIG. 3. Domains of vanishing of $f_p(q)$ for case $\mathfrak{C}$: $f_p(q) = 0$ in the vertically shaded regions γ (spectrum) and in the dotted regions α and β bounded by K_+^+: $(q \pm \frac{1}{2}p)^2 = M^2$, $q^0 + \frac{1}{2}p > 0$, and K_-^-: $(q - \frac{1}{2}p)^2 = M^2$, $q^0 - \frac{1}{2}p < 0$, respectively, $(\check{A}(k) = 0, k^2 > M^2)$. The figure is drawn for the (least favorable) case $M^2 > p^2$.

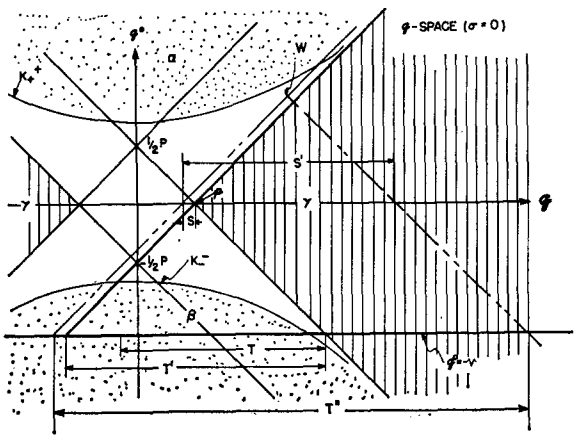


FIG. 4. Intersection in the $\sigma = 0$ plane of regions which occur in the discussion of case C.

$(w, 0)$ which lies in the $\sigma = 0$ plane in the region between the four-dimensional cone $V^+_{(p/2)}$ and the upper sheet of the hyperboloid, $(q + \frac{1}{2}p)^2 = M^2$, which approaches it asymptotically. (See Fig. 4.) The argument which we will give for vanishing of G_p near $(w, 0)$ is a simple geometrical one based on principles 2 and 3. However, to give formulas for all the regions involved is cumbersome, and may cloud the simplicity of the argument. For this reason, we first state the argument qualitatively, with reference to Figs. 4 to 7, and postpone the details of the argument to the next paragraph. Our objective is to show that G_p is analytic in a neighborhood of $(w, 0)$ which includes points in α and γ at which G_p vanishes. Then it follows that G_p vanishes in the neighborhood of $(w, 0)$, and this new region of vanishing connects the old regions α and γ. We show analyticity in this neighborhood of $(w, 0)$ by using principle 3 applied to the data in the intersection of the backward cone of dependence of $(w, 0)$ and some plane $q^0 = -v$, $v > 0$, region T''. (See Figs. 4, 5, 7.) Principle 3 will imply analyticity at $(w, 0)$ provided we can prove, from our information about regions of vanishing of G_p that $G_p(w, 0)$ cannot obtain a contribution from characteristics in the plane $q^0 = -v$. We use principle 2 in two steps to show this. First we consider the domain of dependence of $(w, 0)$ in the plane $q^0 = 0$, the region S'. (See Figs. 4, 6, 7.) We note that vanishing in γ restricts this region to S. The domain of dependence of $(w, 0)$ in the plane $q^0 = -v$ is then restricted to the domain of dependence of S in this plane, which we call T'. Vanishing in β restricts this domain of dependence of $(w, 0)$ still further to T. The crucial point is that for proper choice of w and v, T contains no characteristics which can reach a neighborhood of $(w, 0)$, which proves the desired analyticity at w.

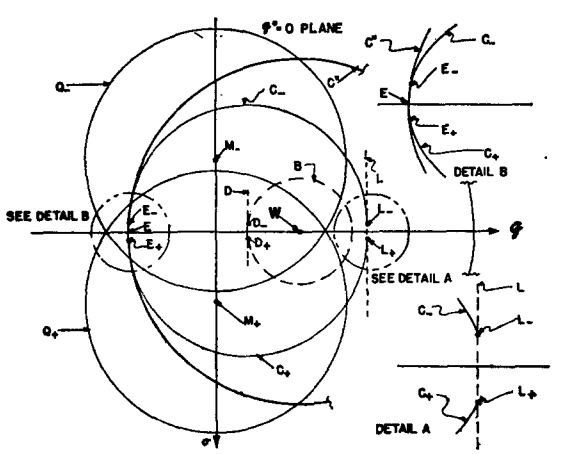


FIG. 5. Regions of interest for case C in the $q^0 = -v$ plane. The figure shows the intersection of the regions of interest with the two dimensional plane containing $\mathbf{q} = \mathbf{w}$ and σ. Formulas for the various regions are given in the text. T'' is the region inside the circle C''. T' is bounded on the left by an arc E of C'' with extrema $E_\pm$, above and below by arcs of the circles $C\mp$, whose centers $D\mp$ are given by $\mathbf{q} = \frac{1}{2}\mathfrak{p}\tilde{w}$, $\sigma = \mp\sigma_D = \mp[w^{0\frac{1}{2}} - (|\mathbf{w}|^2 - \frac{1}{2}\mathfrak{p})^2]^{1/2}$, and whose radii are both v, and at the right by a segment of the straight line L, whose end points $L_\pm$ are given be $\mathbf{q} = (\frac{1}{2}\mathfrak{p} + v)\tilde{w}$, $\sigma = \pm\sigma_D$. Q is the region inside the intersection of the circles $Q_\pm$. T is that part of T' which does not lie in Q. Note that no points of the intersection of the characteristic cone from $(w, 0)$ with the plane $q^0 = -v$, i.e., points on the circle C'', lie in T.

Now we go through this argument again giving formulas for the regions involved. We choose $(w, 0)$ to be any point such that

$$w^0 > M - \tfrac{1}{2}\mathfrak{p}, \qquad -\tfrac{1}{2}\mathfrak{p} + |\mathbf{w}| \leq w^0 \leq -\tfrac{1}{2}\mathfrak{p} + (\mathbf{w}^2 + M^2)^{1/2}.$$

The region T'' in the $q^0 = -v$ plane is bounded by the surface C'',

$$C'' : (\mathbf{q} - \mathbf{w})^2 + \sigma^2 = (w^0 + v)^2. \qquad \text{(See Figs. 5, 7)}$$

The region S' in the $q^0 = 0$ plane is bounded by the surface B,

$$B : (\mathbf{q} - \mathbf{w})^2 + \sigma^2 = w^{02}. \qquad \text{(See Figs. 6, 7)}$$

The region S in the $q^0 = 0$ plane is bounded by the surfaces B and D,

$$D : |\mathbf{q}| = \tfrac{1}{2}\mathfrak{p}. \qquad \text{(Figs. 6, 7)}$$

The region T' in the $q^0 = -v$ plane is bounded by the surfaces C'', $C_\pm$, and L, where $C_\pm$ and L are given by

$C_\pm$: envelope of $(\mathbf{q} - \frac{1}{2}\mathfrak{p}\hat{e})^2 + (\sigma \mp \sigma_{\hat{e}})^2 = v^2$, for all $\hat{e}$, where $\hat{e}$ is a unit 3-vector and (Figs. 5, 7)

$$\sigma_{\hat{e}} = [w^{02} - (\mathbf{w} - \tfrac{1}{2}\mathfrak{p}\hat{e})^2]^{1/2},$$

$$L : |\mathbf{q}| = \tfrac{1}{2}\mathfrak{p} + v. \qquad \text{(Figs. 5, 7)}$$

Finally, the region T in the $q^0 = -v$ plane is $T = T' - T' \cap Q$, where Q is bounded by the surfaces $Q_\pm$

$$Q_\pm : \mathbf{q}^2 + (\sigma \mp M)^2 = (v + \tfrac{1}{2}\mathfrak{p})^2.$$

The region T contains no characteristics which can reach $(w, 0)$ provided that we can choose w and v (subject to our initial conditions on w) so that all points $\{E\}$ in the intersection of the cone $V^-_{(w,0)}$ with the plane $q^0 = -v$ which lie in T' are in Q, and thus not in T. This requirement will be satisfied provided Q contains the extreme points of $\{E\}$, i.e., points with the largest σ coordinate, which we call $E_\pm$. The points $E_\pm$ have coordinates

$$E_\pm \begin{cases} \mathbf{q} = \mathbf{q}_0 = [\tfrac{1}{2}\mathfrak{p} - (v/w^0)(|\mathbf{w}| - \tfrac{1}{2}\mathfrak{p})]\hat{w} \\ \sigma = \pm\sigma_0 = \pm(1 + v/w^0)[w^{02} - (|\mathbf{w}|^2 - \tfrac{1}{2}\mathfrak{p})^2]^{1/2} \end{cases}$$

where $\hat{w}$ is a unit vector in the direction of $\mathbf{w}$. The condition that $E_\pm \in Q$ is

$$\mathbf{q}_0^2 + (\sigma_0 + M)^2 \leq (v + \tfrac{1}{2}\mathfrak{p})^2,$$

and is satisfied in the large $|\mathbf{w}|$ limit for a range of w^0 connecting α and γ provided $v > (2M^2)/\mathfrak{p}$, which can always be satisfied by proper choice of v. Thus, α and γ are connected by a region of vanishing, and use of principle 1 shows that $G_p(q, \sigma)$ vanishes identically. This concludes the proof that $A(x)$ is a generalized free field for case ℭ.

(f) Heuristic Remarks

If we consider that $\tilde{A}(p)$, for p off the mass shell, can be measured in some kind of "impulsive probe" experiment, then we might represent a matrix element in which $\tilde{A}(p)$ has space-like momentum, i.e., produces momentum transfer in interacting with some "B" particle, by a diagram like Fig. 8(a). A corresponding crossed matrix element where $\tilde{A}(p)$ carries time-like momentum would be represented by the diagram in Fig. 8(b). Case 𝔄 states that if all diagrams which correspond to $\tilde{A}(p)$ carrying momentum transfer have an infinite zero at some $-p^2 = \Delta^2 > 0$, then the crossed diagrams which correspond to $\tilde{A}(p)$ carrying mass also vanish; and in fact $\tilde{A}(p)$ leads to no physical effects at all. Similar remarks hold for case 𝔅. For case ℭ, by virtue of Borchers' result[2] that $\rho(p^2) = 0$, $p^2 > M^2$ suffices, the absence of high mass intermediate states in the propagator implies the absence of all physical effects. Note that the matrix element ${}_{\text{in}}\langle 2 \mid A(0) \mid 0\rangle$ can contribute to $\rho(p^2)$ for arbitrarily high masses, so that no conclusion can be drawn (on the basis of the present arguments) from the requirement ${}_{\text{in}}\langle n \mid A(0) \mid 0\rangle = 0$, $n > 2$, which is weaker than that of case ℭ.

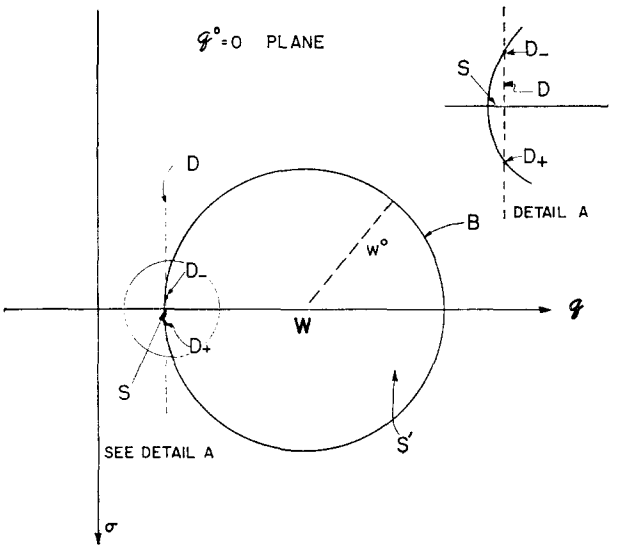


FIG. 6. Regions of interest for case ℭ in the $q^0 = 0$ plane. The figure shows the intersection of the regions of interest with the two-dimensional plane containing $\mathbf{q} = \mathbf{w}$ and σ. Formulas for the various regions are given in the text. S' is bounded by the circle B. S is bounded by an arc of B and a segment of the straight line D, whose end points $D_\pm$ are given by $\mathbf{q} = \tfrac{1}{2}\mathfrak{p}\hat{w}$, $\sigma = \pm\sigma_D$ (see Fig. 5 for σ_D).

4. COROLLARY FOR CASES 𝔄, 𝔅, AND ℭ

Next we consider two fields $A(x)$ and $B(x)$ which (a) describe theories which have properties I through IV above, (b) are in the same Borchers class 𝔅,

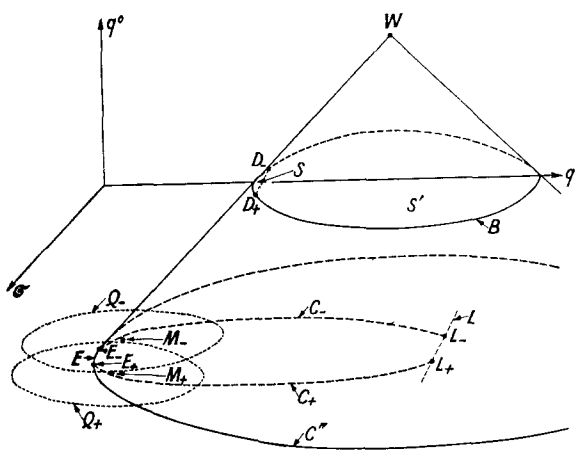


FIG. 7. Regions of interest for case ℭ in the three-dimensional subspace spanned by the σ axis, the q^0 axis, and the vector $\mathbf{q} = \mathbf{W}$. The circle C'' and the objects associated with it are in the $q^0 = -v$ plane; see the caption of Fig. 5 for a detailed description of these objects. The circle B and the objects associated with it are in the $q^0 = 0$ plane; see the caption of Fig. 6 for a detailed description of these objects.

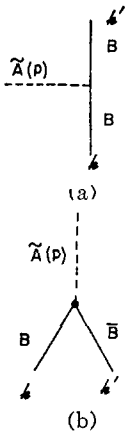


FIG. 8. (a) Heisenberg field $\tilde{A}(p)$ as a probe carrying space-like momentum, $-p^2 = \Delta^2 = (k - k')^2 > 0$. (b) Crossed diagram: Heisenberg field carrying time-like momentum $p^2 = \mu^2 = (k + k')^2 > 0$.

and (c) have Fourier transforms which agree[17] on the domains $\mathfrak{A}$, $\mathfrak{B}$, or $\mathfrak{C}$. For case $\mathfrak{C}$, vanishing of the Kallén-Lehmann weight of $A - B$ suffices; however, equality of the weights of A and B in $\mathfrak{C}$ *does not* suffice. Since $A(x) - B(x)$ satisfies the hypotheses of the discussion above, we conclude that A and B differ at most by a generalized free field in their Borchers class $\mathfrak{B}$.

5. COUNTER EXAMPLES FOR CASES $\mathfrak{D}$, AND $\mathfrak{E}$

Finally, we discuss cases $\mathfrak{D}$, $0 \leq M_1^2 < p^2 < M_2^2$, $p \neq 0$, the region between two double-sheeted hyperboloids excluding the origin; and $\mathfrak{E}$, $p = 0$, the origin, and exhibit counter-examples to show that $\tilde{A}(p) = 0$ in $\mathfrak{D}$ or $\mathfrak{E}$ does not imply that $A(x)$ is a generalized free field. The counter example for case $\mathfrak{D}$ is the Wightman polynomial[18] $A(x) = \phi_0(x) + g : \phi_0(x)^2 :$ where $\phi_0(x)$ is a free field of mass m. This example satisfies properties I through IV, and is not a generalized free field. However, the momentum space formula

$$\tilde{A}(k) = \tilde{\phi}_0(k)\ \delta(k^2 - m^2) + g \int d^4p : \tilde{\phi}_0(\tfrac{1}{2}k + p)\ \delta[(\tfrac{1}{2}k + p)^2 - m^2] \times \tilde{\phi}_0(\tfrac{1}{2}k - p)\ \delta[(\tfrac{1}{2}k - p)^2 - m^2]$$

shows that $\tilde{A}(k) = 0$, $0 \leq k^2 < m^2$, and $m^2 < k^2 < (2m)^2$, which finishes the discussion of this counter example. We conclude this note by remarking that the Wightman polynomial $A(x) = \phi_0(x) + g : \phi_0(x)^3$ has support everywhere in momentum space except at the origin and thus can serve as a counter-example for case $\mathfrak{E}$.

ACKNOWLEDGMENTS

We take pleasure in thanking Dr. R. Stora for helpful suggestions about an earlier draft of this work, and in thanking Dr. Stora, Dr. H. J. Borchers, Professor J. Diaz, Dr. J. Peetre, and Professor J. Sucher for valuable discussions of this work.

[17] By agree, we mean $\tilde{A}(p) = \tilde{B}(p)$ on open sets; $\tilde{A}(p) = \tilde{B}(p)$, and

$$(\partial/\partial p^2)^n \tilde{A}(p) = (\partial/\partial p^2)^n \tilde{B}(p)$$

n integral, on single hyperboloids; and $\tilde{A}(p) = \tilde{B}(p)$, and

$$\mathcal{P}(\partial/\partial p^\mu)\tilde{A}(p) = \mathcal{P}(\partial/\partial p^\mu)\tilde{B}(p)$$

at the origin.

[18] pp. 57–64 of lecture notes of reference 14.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Nonrelativistic S-Matrix Poles for Complex Angular Momenta

ROGER G. NEWTON*
Indiana University, Bloomington, Indiana
(Received April 18, 1962)

Regge's introduction of complex angular momenta is studied in more detail. The shape and number of trajectories of S-matrix poles as functions of the energy is investigated, with particular attention to the way they leave the real axis, and to their ends at $E \to \pm\infty$. The conditions are found under which the S matrix is meromorphic even in Re $l < -1/2$. Some properties of the S matrix in the left half-plane are discussed and so are its symmetry between left and right half-planes, its branch point at $E = 0$, and the residues at its poles.

1. INTRODUCTION

IT is one of the important consequences of quantum mechanics that the angular momentum of all particles must be an integral or half-integral multiple of $\hbar$. Nevertheless, lately it has become interesting in elementary particle physics to relax the quantum nature of the angular momentum mathematically and to consider it not only as a continuous, but even as a complex variable. In so doing, one makes use of the fact that nonrelativistic scattering theory may be formulated in terms of differential equations which, if analyzed by means of partial waves, contain the angular momentum merely as a parameter. The elimination of the angular aspects removes the necessity for the angular momentum to be quantized, and it is free to assume arbitrary values.

That such a relaxation of the discreteness of the angular momentum is interesting can be seen in various ways. For example, it is well known that if the nonrelativistic spherically symmetric potential between two particles goes down exponentially at infinity, then the individual S-matrix elements may have singularities in the complex plane beyond a certain distance from the real axis. However, if these elements are summed up to the forward scattering amplitude, then the result no longer contains any singularities in the upper half of the complex k plane, a fact which is commonly expressed as a dispersion relation. There must, therefore, be strong correlations between the singularities of various S-matrix elements in the complex k plane that assure that in the sum making up the forward amplitude (or even some nonforward amplitudes), the singularities all manage to cancel. These correlations should in principle be comprehensible by considering the various S-matrix elements as special values of a continuous function of the angular momentum.

* Supported in part by the National Science Foundation.

Another reason, and one which induces the consideration not only of continuous real but even of complex angular momenta, is the one that led Regge[1] when he first proposed it.

The scattering amplitude of two spinless particles interacting via a spherically symmetric potential,

$$f(\theta) = (2ik)^{-1} \sum_l (2l + 1)(e^{2i\delta_l} - 1)P_l(\cos\theta), \quad (1.1)$$

may be written formally as a Watson[1a] contour integral

$$f(\theta) = (2\pi k)^{-1} \times \int_C d\lambda\lambda P_{\lambda-1/2}(-\cos\theta)(S_\lambda - 1)/\cos\pi\lambda, \quad (1.2)$$

provided that $\exp(2i\delta_l)$ can be considered the value of an analytic function S_λ at $\lambda = l + \frac{1}{2}$. The contour C is indicated in Fig. 1. Further, assuming that S_λ is regular in the right-hand half-plane

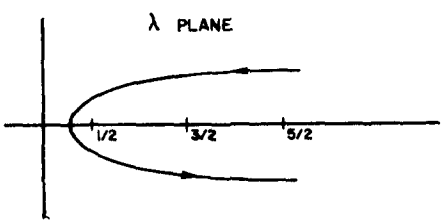


FIG. 1. The contour C.

[1] T. Regge, Nuovo cimento **14**, 951 (1959).

[1a] *Added in proof.* In view of some confusion in the references to the historical origin of the replacement of this series by a contour integral it may be well to straighten out the record. The method is based on those introduced in the context of bending of electromagnetic waves by a sphere by H. Poincaré [Rendiconti Circolo Mat. Palermo **29**, 169 (1910)] and J. W. Nicholson [Phil. Mag. **19**, 516 (1910); **20**, 157 (1910); Messenger Math. **37**, 84 (1907)]. However, the exact form of the general method as it is used at present first appeared in the work of G. N. Watson [Proc. Roy. Soc. (London) **95**, 83 (1918)]. A. Sommerfeld [*Partial Differential Equations in Physics* (Academic Press Inc., New York, 1949), pp. 282 *ff.*] deserves credit for resurrecting it. I am indebted to Dr. E. Guth for bringing the Poincaré reference to my attention.

Re $\lambda \geq 0$, except for isolated poles at $\lambda = \alpha_n + \frac{1}{2}$, and that $(S_\lambda - 1)$ vanishes sufficiently rapidly at infinity, the contour can be shifted to the imaginary axis and we get

$$f(\theta) = (2\pi k)^{-1} \times \int_{-\infty}^{\infty} d\lambda' \lambda' [S(i\lambda') - 1] P_{i\lambda'-1/2}(-\cos\theta)/\cosh \pi\lambda' - ik^{-1} \sum_n (\alpha_n + \tfrac{1}{2}) S_n P_{\alpha_n}(-\cos\theta)/\sin \pi\alpha_n. \quad (1.3)$$

The necessary convergence conditions for large $|\lambda|$ have been proved by Regge *et al.*[1-3]

The virtues of the representation (1.3) are the following. For the purpose of double dispersion relations one must know the behavior of $f(\theta)$ at fixed energy and large momentum transfer Δ. Since the latter is related to the scattering angle by

$$\cos\theta = 1 - \Delta^2/2k^2,$$

letting Δ go to infinity leads to the unphysical "angle" $|\cos\theta| \to \infty$. Now Bottino *et al.*[3] have shown that the integral in (1.3) vanishes in the limit as $|\cos\theta| \to \infty$. Since for large $|\cos\theta|$, $P_{\alpha_n}(\cos\theta)$ goes as $|\cos\theta|^{\mathrm{Re}\,\alpha_n}$, the behavior of $f(\theta)$ is thus determined by the pole term with the largest Re α_n, which in turn fixes the number of subtractions necessary in the Mandelstam representation. It is important, therefore, to establish that the poles don't diffuse on the right all the way to infinity. Furthermore, since their positions are functions of the energy, it becomes of interest to study the trajectories described by the poles of S in the complex λ plane as a function of the energy.

The pole terms in (1.3) can be thought of in specific physical terms. We may recover the physical S-matrix elements from $f(\theta)$ by

$$(S_l - 1)/ik = \int_{-1}^{1} d\cos\theta f(\theta) P_l(\cos\theta). \quad (1.4)$$

If we insert (1.3) in this then the contribution from the nth pole term in (1.3) is[4]

$$S_l^{(n)}(E) = \frac{(2/\pi)(\alpha_n + \frac{1}{2})S_n}{(l - \alpha_n)(l + \alpha_n + 1)}, \quad (1.5)$$

because[4]

$$\int_1^1 dz P_l(z) P_\alpha(-z) = \frac{2}{\pi} \frac{\sin\pi\alpha}{(l-\alpha)(l+\alpha+1)}.$$

This will be the most strongly varying contribution to S near an energy E_0, where Re $\alpha_n = l$ and Im α_n is small. There is a resonance. We may describe it in terms of a Breit-Wigner formula by expanding $\alpha_n(E)$ about $E = E_0$. Then

$$\frac{1}{l - \alpha_n(E)} \approx \frac{1}{\alpha_n'(E_0)} \frac{1}{E - E_0 + \Delta E + (i/2)\Gamma}, \quad (1.6)$$

where

$$\Delta E = \frac{\mathrm{Im}\,\alpha_n \,\mathrm{Im}\,\alpha_n'}{(\mathrm{Re}\,\alpha_n')^2 + (\mathrm{Im}\,\alpha_n')^2}, \quad (1.7)$$

$$\Gamma = \frac{2\,\mathrm{Im}\,\alpha_n \,\mathrm{Re}\,\alpha_n'}{(\mathrm{Re}\,\alpha_n')^2 + (\mathrm{Im}\,\alpha_n')^2}, \quad (1.8)$$

everything being evaluated at $E = E_0$, and the prime indicating differentiation with respect at E.[5]

The picture that emerges from a consideration of the pole trajectories as functions of E in the complex λ (or l) plane is this: For negative E, the pole moves along the real axis to the right. When it goes through an integral l value it causes a physical bound state. Depending on the strength of the interparticle force, it may go through one or several integral values of l, or it may never even reach $l = 0$ before, at $E = 0$, it turns away from the real axis into the upper half of the complex l plane. If it leaves at a point $l \geq \frac{1}{2}$, then it does so in the forward direction with zero slope and can therefore be expected to cause a resonance as its projection onto the real axis passes an integral l value. Under certain conditions on the potential it will eventually turn around and pass back into the region Re $l < -\frac{1}{2}$. Thus, one has the physically very appealing view in which several bound states and possibly resonances are different manifestations of the same fundamental cause. They originate from the "same" pole of the S matrix.

Section 2 is concerned with preparatory definitions of wavefunctions and the S matrix. Section 3 contains a discussion of the physical consequences of inequality (3.4) for successive phase shifts at the same energy, first derived by Regge.[1] In Sec. 4 some details of the shape of a pole trajectory in the region Re $l \geq -\frac{1}{2}$ are discussed. One of the

[2] T. Regge, Nuovo cimento **18**, 947 (1960).
[3] A. Bottino, A. M. Longoni, and T. Regge, Nuovo cimento **23**, 954 (1962).
[4] G. F. Chew, S. C. Frautschi, and S. Mandelstam, Phys. Rev. **126**, 1202 (1962).

[5] The width of the resonance thus depends not only on the nearness of the pole to the real axis but also on its energy dependence. Equation (1.8) differs from (1.6) of reference 4 which failed to take the energy dependence of Im α into account. It should be remembered in addition, as in all resonance theories, that to speak of a "resonance" is observationally meaningful only if it is sharp relative to other energy dependencies in the amplitude. Otherwise the resonance denominator need not lead to a maximal partial cross section. Furthermore, the energy shift ΔE must also be small in order for the result to be meaningful. If ΔE is not small then no resonance occurs near E_0, and near $E_0 - \Delta E$ the neglected terms in the expansion of $\alpha(E)$ become important.

results is the angle under which a pole leaves the real axis at $E = 0$, given by (4.8) and the subsequent discussion. Section 5 deals with the number of pole trajectories. The main result is embodied in the inequality (5.9) for the number of pole trajectories that enter the complex l plane at $l \geq -\frac{1}{2}$.

Up to this point all considerations were limited to the region Re $l \geq -\frac{1}{2}$; i.e., Re $\lambda \geq 0$. From the point of view of direct solution of the integral equation for the regular wave function, the left-hand half of the λ plane is quite inaccessible. The "regular" solution there becomes "irregular" at $r = 0$ and the integrals no longer converge absolutely. In order to say anything about the S matrix in this region, an analytic continuation of the wave function is necessary. The possibility of doing that depends only on the behavior of the potential at $r = 0$, and it is shown in Sec. 6 that if rV and its first m derivatives are finite at $r = 0$, then the S matrix is meromorphic for Re $l \geq -m/2 - 1$. In Sec. 7 the analytic continuation to the left-hand λ plane is exploited for a discussion of certain symmetries that exist between poles on the left and on the right. In Sec. 8 it is shown that for Re $\lambda < 0$, also, the first Born approximation is good at $E \to \pm\infty$. As one of the consequences, it is proved that as $E \to \pm\infty$, a trajectory can end only either at infinity or else on a negative integral value of l (if rV and all its derivatives exist at $r = 0$). There are two Appendices; Appendix A is concerned with the branch point of S at $E = 0$; Appendix B, with the residue of S at a pole.

2. PRELIMINARIES

We want to consider the radial Schrödinger equation[6]

$$-\psi'' + (\lambda^2 - \tfrac{1}{4})r^{-2}\psi + V\psi = k^2\psi \qquad (2.1)$$

where $\lambda = l + \frac{1}{2}$. An irregular solution $f(\lambda, k, r)$ is defined by the boundary condition

$$\lim_{r\to\infty} f(\lambda, k, r)e^{ikr} = 1. \qquad (2.2)$$

It follows that

$$f(-\lambda, k, r) = f(\lambda, k, r) \qquad (2.3)$$

and in any region of analyticity connected with the real λ and k axes,

$$f^*(\lambda^*, -k^*, r) = f(\lambda, k, r). \qquad (2.4)$$

In the absence of a potential, f reduces to

$$f_0(\lambda, k, r) = e^{i(\pi/2)(\lambda+1/2)}(\pi kr/2)^{1/2}H_\lambda^{(2)}(kr), \qquad (2.5)$$

$H_\lambda^{(2)}$ being the Hankel function of the second kind. With the potential V present, f satisfies the integral equation

$$f(\lambda, k, r) = f_0(\lambda, k, r) - \int_r^\infty dr' g_\lambda(k; r, r')V(r')f(\lambda, k, r') \qquad (2.6)$$

with

$$g_\lambda(k; r, r') = \tfrac{1}{2}\pi(rr')^{1/2}[J_\lambda(kr')Y_\lambda(kr) - J_\lambda(kr)Y_\lambda(kr')], \qquad (2.7)$$

where J_λ and Y_λ are the Bessel functions of the first and second kind. The standard proof[7] of the convergence of the series of successive approximations to (2.6) goes through irrespective of whether λ is a positive half-integer or not, provided that the first and second absolute moments of V are finite. Everything that is known of f as a function of k depending on various assumptions about the potential can be easily extended to arbitrary complex values of λ. For fixed $r \neq 0$ and $k \neq 0$, f is, by similar arguments, seen to be an entire function of λ.

A regular function φ is defined for Re $\lambda \geq 0$ by the boundary condition

$$\lim_{r\to 0} r^{-1/2-\lambda}\varphi(\lambda, k, r) = 2^{\lambda-1/2}\Gamma(\tfrac{1}{2} + \lambda)/\Gamma(1 + 2\lambda). \qquad (2.8)$$

In the absence of a potential we have

$$\varphi_0(\lambda, k, r) = (\tfrac{1}{2}\pi r)^{1/2}k^{-\lambda}J_\lambda(kr), \qquad (2.9)$$

while in the presence of V, φ obeys the integral equation

$$\varphi(\lambda, k, r) = \varphi_0(\lambda, k, r) + \int_0^r dr' g_\lambda(k; r, r')V(r')\varphi(\lambda, k, r'). \qquad (2.10)$$

Clearly φ satisfies

$$\varphi(\lambda, -k, r) = \varphi(\lambda, k, r) \qquad (2.11)$$

and

$$\varphi^*(\lambda^*, k^*, r) = \varphi(\lambda, k, r) \qquad (2.12)$$

in any region in the complex λ and k planes connected with the real axes. Again, the proof of the convergence of the series of successive approximations goes through as in the case of integral l, and the analyticity properties of φ as a function of k are the same. Furthermore, for each fixed k and r, φ is an analytic function of λ regular in Re $\lambda \geq 0$.

[6] The following is entirely analogous to what is done for integral values of l and closely follows the procedure of reference 7.

[7] R. G. Newton, J. Math. Phys. 1, 319 (1960).

The integral equation as it stands breaks down for Re $\lambda < 0$. Its analytic continuations to that region will be discussed in Sec. 6.

We next express φ in terms of the two linearly independent solutions $f(\lambda, k, r)$ and $f(\lambda, -k, r)$:

$$\varphi(\lambda, k, r) = [F(\lambda, k)f(\lambda, -k, r) - F(\lambda, -k)f(\lambda, k, r)]/2ik, \quad (2.13)$$

where F is the Wronskian

$$F(\lambda, k) = f\varphi' - f'\varphi. \quad (2.14)$$

For large r, therefore,

$$2ik\varphi \sim F(\lambda, k)e^{ikr} - F(\lambda, -k)e^{-ikr}$$

and hence we define the S matrix as

$$S(\lambda, k) \equiv e^{i\pi(\lambda-1/2)}F(\lambda, k)/F(\lambda, -k). \quad (2.15)$$

The function F is a generalization of the Jost function, the simplest one for the present purpose. In Sec. 8 and the Appendix we shall define another possible generalization that is more suitable for other occasions. The definition (2.15), which for integral l goes over into the customary one, is shown in the Appendix to be natural, the exponential factor being necessary in order that S be unitary and tend to unity at $E \to \infty$ even for nonintegral l values.

As a function of k, F and S have all the essential properties known for integral l values, so long as Re $\lambda \geq 0$. The region Re $\lambda < 0$ is not accessible until the existence of φ has been proved there. In addition, F and S are, for fixed $k \neq 0$ in the region of regularity, analytic functions of λ regular for Re $\lambda \geq 0$. We also have in any region of analyticity connected with the real axes of λ and k,

$$F^*(\lambda^*, -k^*) = F(\lambda, k) \quad (2.16)$$

as well as

$$S^*(\lambda^*, -k^*) = S(\lambda, k). \quad (2.17)$$

3. THE PHASE-SHIFT DERIVATIVE

We want to discuss briefly a simple result first obtained by Regge.[1] If we define

$$\Delta \equiv \arg F, \qquad 2\delta \equiv \arg S, \quad (3.1)$$

then by (2.15)

$$\delta = \Delta + (i\pi/2)(\lambda - \tfrac{1}{2}). \quad (3.2)$$

Differentiation of (2.1) with respect to λ, multiplication by φ and subsequent subtraction of (2.1) multiplied by $\partial\varphi/\partial\lambda$ yields

$$\frac{d}{dr}\left(\varphi\frac{\partial\varphi'}{\partial\lambda} - \varphi'\frac{\partial\varphi}{\partial\lambda}\right) = 2\lambda r^{-2}\varphi^2.$$

Integration from zero to infinity gives for real k and $\lambda > 0$

$$-k^{-1}|F|^2\,\partial\Delta/\partial\lambda = 2\lambda\int_0^\infty dr r^{-2}\varphi^2 > 0.$$

We conclude that for $k > 0$

$$\partial\Delta/\partial\lambda > 0$$

and hence by (3.2),[8]

$$\partial\delta/\partial\lambda < \pi/2, \quad (3.3)$$

which implies that

$$\delta_{l+1} - \delta_l < \pi/2. \quad (3.4)$$

We first observe that, δ being defined only *modulo* π, the inequality (3.4) imposes no restriction whatever on the values of successive phase shifts at a fixed energy. The restriction becomes operative only when put together with a demand of continuity of the phase shifts as functions of k and with their behavior at $k = 0$ and $k = \infty$.

The first question to which (3.4) gives a partial answer is this. For individual phase shifts (assumed continuous as functions of k for $k > 0$) it is customary to define $\delta_l(\infty) = 0$ and then to let $\delta_l(0)$ be determined by the Levinson theorem[9]:

$$\delta_l(0) = n_l\pi,$$

where n_l is the number of bound states of angular momentum l. Alternatively, one could define $\delta_l(0) = 0$ in order to make it continuous across $k = 0$. Now that we demand a continuous connection between phase shifts of different l values a question of consistency arises. If we define $\delta_0(\infty) = 0$, then we are no longer free to dispose of the other phase shifts.

Suppose we were to define $\delta_0(0) = 0$ and let the Levinson theorem determine $\delta_0(\infty)$; and suppose there is an s-wave bound state and none for the p wave. Then we would have $\delta_0(\infty) = -\pi$ and hence, by (3.4), $\delta_1(0) = \delta_1(\infty) < -\pi/2$. Since this entails $\delta_1(0) \leq -\pi$, (3.4) also shows that $\delta_2(0) \leq -\pi$, etc. As a result, the s-wave bound state would force all other phase shifts to be at most

[8] From a semi-classical point of view the inequality (3.3) is well known. Since the classical deflection function Θ is connected with the WKB phase shift by $\Theta = 2d\delta/dl$, (3.3) merely states the fact that $\Theta < \pi$. ("Orbiting" implies a large *negative* Θ.) I am indebted to Dr. Joseph W. Weinberg for calling the connection with the classical result to my attention.

[9] N. Levinson, Kgl. Danske Videnskab. Selskab, Mat.-fys. Medd. **25**, No. 9 (1949).

$-\pi$ at $E = 0$, which makes no sense. On the other hand, if we define $\delta_0(\infty) = 0$, then an s-wave bound state means that $\delta_0(0) = \pi$, and (3.4) is compatible with all $\delta_l(\infty) = 0$. We conclude from this that $\delta_0(\infty) = 0$ is the most natural choice and that it may be expected that $\delta_l(\infty) = 0$ follows for all other l, although a proof of this is still missing.

Consider now a case in which the p wave has a sharp resonance at some energy E_0 in the vicinity of which the s and the d waves are smoothly varying functions of E, and below which the p phase shift is positive but less than $\pi/2$.[10] The p-wave phase shift then looks as indicated in Fig. 2 near $E = E_0$. Applying (3.4) to $l = 0$ just above E_0, we find that $\delta_0(E_0) > A > \pi/2$. Consequently, either there is an s-wave bound state, or else the s wave must have had a resonance at some lower energy. This physically quite plausible result is a simple extension of the well-known fact that the p wave cannot have a bound state unless the s wave has one of lower energy.[11]

We may also use (3.4) for $l = 1$ under the stated resonance assumptions, and we then take E just below E_0. That shows that at $E \approx E_0$, $\delta_2 < A$. These two inequalities, for $l = 0$ and $l = 2$, may serve to determine whether an experimentally given resonance is described by a local, single-channel potential or not.

4. THE POLE TRAJECTORIES

We are now going to investigate the poles of S for k either real or positive imaginary. Assuming that the potential is of such a nature that F can be analytically continued into the relevant region in the upper half of the complex k plane, the poles of S are due to zeros of F. An acceptable class of potentials would be those that vanish beyond some finite distance; another would be a (continuous) superposition of Yukawa potentials, for which F is regular in the entire k plane cut along the positive imaginary axis from some μ to infinity.

What we are after is the motion of the zeros of $F(\lambda, k)$ in the complex λ plane as a function of k. The energy is kept real and hence for $E > 0$, k is real, and for $E < 0$, k is purely imaginary. The "physical sheet" of the Riemann surface of S as a function of E corresponds to the upper half of the k plane. In view of (3.14), this means that for $E < 0$

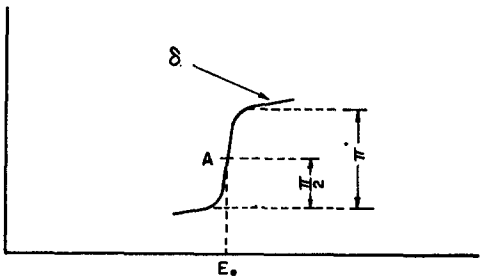


FIG. 2. The p-wave phase shift.

we are looking for the zero of F on the *negative* imaginary axis, where for integral l, they denote bound states; for $E > 0$ we take the upper rim of the cut, i.e., $k > 0$, and hence in F, $k < 0$.

Taking the complex conjugate of (2.1) for φ, multiplying by φ, and subtracting (2.1) for φ multiplied by φ^* we obtain for real k^2

$$(d/dr)(\varphi^*\varphi' - \varphi'^*\varphi) = 2i \text{ Im } \lambda^2 r^{-2} |\varphi|^2$$

and by integration

$$2i \text{ Im } \lambda^2 \int_0^\infty dr r^{-2} |\varphi|^2 = \lim_{r\to\infty} (\varphi^*\varphi' - \varphi'^*\varphi). \quad (4.1)$$

First consider the case of purely negative imaginary k and assume that for $k = k_0$, $\lambda = \lambda_0$, $F(\lambda_0, k_0) = 0$. Then φ is a multiple of f and thus vanishes exponentially at infinity. Consequently (4.1) reads

$$\text{Im } \lambda^2 \int_0^\infty dr r^{-2} |\varphi|^2 = 0. \quad (4.2)$$

This means that for $E < 0$, a zero of F can occur only on the imaginary, or on the real λ axis. It will be shown in Sec. 7 that it cannot occur on the imaginary λ axis (except at $\lambda = 0$). Hence for $E < 0$ an S-matrix pole in the region Re $\lambda \geq 0$ must lie on the real λ axis. When it goes through a half-integer (integral l), then it means there is a bound state.

A little more may be said about the motion of a zero of F along the real positive λ axis by differentiating (2.1) with respect to k, assuming k^2 and λ real and all along $F(\lambda_0, k_0) = 0$ so that λ_0 becomes a function of k_0. In the standard way we then get

$$(d/dr)(\varphi'\varphi_k - \varphi_k'\varphi) = [1 - (\lambda_0^2)_k r^{-2}]\varphi^2,$$

where the subscript k indicates differentiation with respect to k_0. Since φ vanishes both at $r = 0$ (as $r^{1/2+\lambda}$) and at $r \to \infty$, we get by integrating

$$\int_0^\infty dr \varphi^2 [1 - (\lambda_0^2)_k r^{-2}] = 0$$

[10] In other words $\sin^2 \delta_1 = 1$ *before* $\sin^2 \delta_1 = 0$.

[11] It should be noticed that the foregoing result refers directly to observable resonances in the sense that $\sin^2 \delta = 1$, and not merely to S-matrix poles, which may or may not entail observable resonances.

or[12]

$$d\lambda_0^2/dk_0^2 = 1/\langle r^{-2}\rangle. \tag{4.3}$$

It will be of special use to us later to know how (4.3) behaves as $E \to 0_-$. For that purpose it is simplest to assume that V has a finite range.[13] The relevant contribution to be investigated is that of the "outside" wave function to the normalization integral in (4.3). The appropriate normalization that assures that the (unnormalized) matrix element of r^{-2} remains finite and different from zero as $k \to 0$, is $k^{1/2+\lambda} f_0(\lambda, k, r)$ for the outside wave function. Then for Im $k < 0$, as $k \to 0$,

$$\int_R^\infty dr[k^{1/2+\lambda} f_0(\lambda, k, r)]^2 = \begin{cases} O(1) & \text{for } l > \frac{1}{2} \\ O(\ln|k|) & \text{for } l = \frac{1}{2} \\ O(k^{2l-1}) & \text{for } l < \frac{1}{2} \end{cases}$$

This is, therefore, the behavior of $d\lambda_0^2/dE_0$ as $E \to 0_-$:

$$\frac{d\lambda_0^2}{dE_0} = \begin{cases} O(1) & \text{for } l_0 > \frac{1}{2}\ (\lambda_0 > 1) \\ O(\ln|E_0|) & \text{for } l_0 = \frac{1}{2}\ (\lambda_0 = 1) \\ O(E_0^{l-1/2}) & \text{for } -\frac{1}{2} < l_0 < \frac{1}{2}\ (0 < \lambda_0 < 1) \end{cases} \tag{4.4}$$

Let us now consider k real $\neq 0$. Then $F(\lambda, k)$ cannot vanish on the real positive λ axis. That is because (2.14) shows that if $F(\lambda, k) = 0$ for real positive λ and k, then $F(\lambda, -k) = 0$ too; but (2.13) implies that then $\varphi \equiv 0$, which contradicts the boundary conditions (2.8). As a result the zero must get off the real λ axis for $E > 0$. Where it goes can be seen from (4.1) for real k. In that case use of (2.13) and of the boundary condition (2.2) lead to

$$8k \operatorname{Re}\lambda \operatorname{Im}\lambda \int_0^\infty dr r^{-2} |\varphi|^2 = |F(\lambda, k)|^2 - |F(\lambda, -k)|^2. \tag{4.5}$$

The implication is that for $k < 0$ and Im $\lambda < 0$

$$|F(\lambda, k)|^2 > |F(\lambda, -k)|^2 = |F(\lambda^*, k)|^2$$

and so $F(\lambda, k)$ cannot vanish. For k negative, F can vanish only in the upper half of the λ plane, and for k positive, in the lower.[14] In terms of the S matrix this means that on the upper rim of the cut along the positive E axis the poles of S are in the upper half of the λ plane, on the lower rim they are in the lower half.

It should now be remembered that where $F(\lambda, k)$ is an analytic function of λ and k and where its derivative with respect to λ is different from zero, there $F(\lambda, k) = 0$ defines λ as an analytic function of k. Consequently, $\lambda(k)$ is a mapping of the k to the λ plane which preserves angles in the small. So as we let k approach zero along the negative imaginary axis and then move on along the real axis, if F were an analytic function of k at $k = 0$ and if $\partial F/\partial\lambda \neq 0$, the trajectory of the zero of F would have to leave the real axis at right angles. As it is, this is generally not so.

Indeed, let us assume $F(\lambda_0, k_0) = 0$. Then (4.5) together with (2.13) shows that

$$2 \operatorname{Re}\lambda_0 \operatorname{Im}\lambda_0 \int_0^\infty dr r^{-2} |f|^2 = -k_0. \tag{4.6}$$

The integral exists since now f is proportional to φ. In order to learn how Im λ_0 approaches zero as $k_0 \to 0$ we assume again that $V = 0$ for $r > R$.[15] Just as in the discussion of (4.3), the appropriate wave function which remains finite in the outside region as $k \to 0$ differs from f by a factor of k^l. So

$$\int_0^\infty dr r^{-2} |f|^2 = k^{1-2\lambda} \int_0^\infty dr r^{-2} |k^{\lambda-1/2} f|^2 = O(k^{1-2\lambda}) \quad \text{as} \quad k \to 0,$$

since the integral remains finite. Division of (4.6) by k_0^2 therefore tells us that

$$d \operatorname{Im}\lambda_0/dk_0 = O(k_0^{2l_0}) \quad \text{as} \quad k_0 \to 0. \tag{4.7}$$

In order to obtain the angle γ_0 in the λ plane at which the zero trajectory leaves the real axis at $E = 0$, we use (4.7) together with (4.4) in

$$\cot\gamma_0 = \lim_{E_0 \to 0+} \frac{\partial \operatorname{Re}\lambda_0/\partial k_0}{\partial \operatorname{Im}\lambda_0/\partial k_0}.$$

Now (4.4) was derived for k_0 on the negative imaginary axis. However, it will hold for k_0 approaching zero along any path in the lower half of the k plane, and specifically just below the real axis. Since Im $\lambda_0 \to 0$ we conclude that (4.4) holds for

[12] Since the right-hand side is positive, this proves the physically obvious fact that a zero on the real positive axis always moves to the right as the energy increases. The expectation value of r^{-2} depends, of course, both on k and on λ, and (4.3) must not be mistaken for a constant. This equation was first obtained by Regge.[2] However, in contrast to his use of (5.1) in reference 2, it should be remembered that (4.3) cannot be used when $E > 0$, since then φ is not normalizable.

[13] The result will hold under much more general conditions, but for a potential that vanishes asymptotically as a power of r, it breaks down if at $E_0 = 0$, λ_0 is too large.

[14] These restrictions do not hold for Re $\lambda < 0$, though. *Added in proof*: Numerical work was meanwhile shown that for a Yukawa potential, zeros occur both in the third and in the fourth quadrant; see A. Ahmadzadeh, P. G. Burke, and C. Tat (to be published).

[15] Again the result holds under much more general conditions.

the derivative of Re λ_0 as $k_0 \to 0$ along the real axis. Thus we get

$$\cot \gamma_0 = \begin{cases} O(k_0^{1-2l_0}) = \infty & \text{for } l_0 > \frac{1}{2} \ (\lambda_0 > 1) \\ O(\ln |k_0|) = \infty & \text{for } l_0 = \frac{1}{2} \ (\lambda_0 = 1) \\ O(1) & \text{for } l_0 < \frac{1}{2} \ (\lambda_0 < 1). \end{cases} \quad (4.8)$$

When $l_0 < \frac{1}{2}$ we may learn more by looking at (4.3) and (4.4) in more detail. Near $k_0 = 0$ we have

$$d\lambda_0/dk_0^2 \approx ae^{i\pi(l-1/2)}k_0^{2l-1},$$

where a is positive so that the left-hand side is positive when k_0 is on the negative imaginary axis. Hence, when k_0 is on the real axis, then

$$d \operatorname{Re} \lambda_0/dk_0 \approx ak_0^{2l} \cos \pi(l - \tfrac{1}{2}).$$

This is positive for $l > 0$ and negative for $l < 0$; for $l = 0$ it vanishes. Thus we have the following result.

If the trajectory meets the real axis at a point $-\frac{1}{2} < l < 0$ then it does so at a finite angle pointing *backwards*; if it meets it at $l = 0$, it does so at right angles; if it meets it at $0 < l < \frac{1}{2}$, it does so at a finite angle pointing *forward*; if it meets it at $l > \frac{1}{2}$, it does so at a zero angle, i.e., it osculates it and it hugs the real axis more and more closely the higher the angular momentum at which it leaves.[16] That for $l < 0$ the trajectory leaves the real axis in the backward direction explains, from the present point of view, the well-known fact that an "almost bound" s state causes no low energy resonance, while "almost bound" states of higher angular momentum do.

The observable effect of the trajectory is of course the resonance, and specifically its width, produced when the pole projection on the real axis passes an integer. If this happens at small energy then (4.4), (4.7), and (1.8) imply the well-known fact that $\Gamma = O(E^{l+1/2})$.[17]

What happens to a trajectory after it leaves the real axis? We may use (4.5) to get a rough limitation on the position of a zero in the complex λ plane. If the potential vanishes for $r > R$ then for $r > R$ we have $f = f_0$; hence,

$$\int_0^\infty dr r^{-2} |f|^2 > \int_R^\infty dr r^{-2} |f_0|^2.$$

The boundary condition (2.2) implies that as $r \to \infty$, $|f| \to 1$ (k is real now). A closer examination of f shows that for real λ it approaches 1 from *above*; for complex λ it dips below 1 for large r, but not below $\frac{1}{2}$ after the last trip below 1. As a result one may expect, at least if Im λ is not too large, that $|f|^2$ is on the average of the order of magnitude 1 for $r > R$. Therefore,

$$\int_R^\infty dr r^{-2} |f|^2 \approx \int_R^\infty dr r^{-2} = R^{-1},$$

and thus from (4.5)

$$2 \operatorname{Re} \lambda_0 \operatorname{Im} \lambda_0 \lesssim R |k_0|. \quad (4.9)$$

This inequality is to be considered at best a rough estimate. But so long as Im λ_0 is not too large it ought to be reasonably reliable. It should be noticed that it is independent of the potential strength.

A more exact limit on the trajectory was derived by Regge under the assumption that the potential can be analytically continued into the complex r plane, to the whole imaginary r axis. In that case the zero trajectory must, for sufficiently large E, cross the straight line Im λ = Re λ and remain above.[2] Furthermore, if everywhere on the imaginary r axis

$$|\operatorname{Im} V(iy)| < M/y^2$$

then[1]

$$\operatorname{Re} \lambda_0 \operatorname{Im} \lambda_0 < M/2$$

so that the zero trajectory must always remain between the real axis and the hyperbola Re λ Im λ = $M/2$. Thus, it is forced to turn around eventually and move toward the imaginary λ axis. The same was shown under somewhat weaker conditions applicable to the Yukawa potential.[2] The important implication of this is, of course, that there exists for each trajectory a value $\bar{\lambda}$ beyond which Re λ_0 cannot go. This is what limits the asymptotic behavior of the scattering amplitude as a function of the momentum transfer or angle, as seen in (1.3).

What happens at the turning point of a trajectory? Equation (1.8) shows that the effective width of the corresponding resonance (if the turning point happens to occur sufficiently close to the real axis, and at a half integral values of Re λ_0 so that it really corresponds to a resonance) then goes through zero and changes sign. Any apparent resonance caused by the trajectory on the return journey is thus not a resonance at all because it

[16] After this work was finished, the author learned that Barut and Zwanziger have come to similar conclusions; A. O. Barut and D. E. Zwanziger (to be published).

[17] The present derivation of this result implies that the l value in it is not that of the resonance but that at which the trajectory meets the real axis at $E = 0$. If that happens at $0 < l < \frac{1}{2}$, then $\Gamma = O(E)$; but it is then quite unlikely that the trajectory causes a p-wave resonance at sufficiently low energy for the formula to be applicable. In general it is not likely to be applicable unless the l values of the resonance and of the point $E = 0$ are nearly equal.

corresponds to a *downward* passage of the phase shift through $\pi/2$ and thus is associated not with a delayed outgoing signal, but with an advanced one. It is at best associated with the downward passage through $\pi/2$ that a phase shift *must* experience if it ever goes through $\pi/2$ upwards (if it is caused by a suitably well-behaved local one-channel potential).[4]

One may like to associate a zero of $F(\lambda, k)$ for real k and complex λ with one for real λ and complex k. This could be done, for example, by keeping the real parts of k and λ fixed but allowing the imaginary part of k to vary while taking the imaginary part of λ to zero. We then have a one-to-one correspondence of zeros with positive λ in the upper half of the k plane and zeros with negative k in the upper half of the λ plane. The fact that a zero cannot get beyond a certain value of Re λ presumably means that for that value of λ the corresponding zero in the complex k plane has moved to infinity in the imaginary direction.[18]

After the trajectory has turned back it may or may not cross the imaginary λ axis and pass into the left-hand half plane. Equation (4.6) must not be mistaken for a stricture against Re $\lambda_0 = 0$. When the real part of λ vanishes then $|f|$ goes as $r^{1/2}$ at the origin and the integral diverges. On the other hand, there appears to be no general reason why the trajectory *must* always cross back beyond Re $l = -\frac{1}{2}$.[18a] In the special case of a Yukawa potential it is explicitly known to do so.[19-21] In the Coulomb case it crosses over at infinity (at $E = 0$).[22] Nor is there any known general reason why a trajectory cannot oscillate or spiral, thereby causing several resonances of the same angular momentum.

5. THE NUMBER OF TRAJECTORIES

It is of obvious interest to raise the question: How many zero trajectories are there? Since we are presently concerned with Re $\lambda \geq 0$ only, we first ask the more restricted question: How many are there in the right-hand half plane?

Considering the fact that as we increase the strength of an attractive potential,[23] an unlimited number of bound states are newly introduced, and there must be at least as many trajectories as there are s-wave bound states, we conclude that there must (at least for a somewhere attractive potential) be an unlimited supply of zeros somewhere in the λ plane. We may fix our attention on the zeros at $E = 0$; according to the results of Sec. 4 they must for Re $\lambda \geq 0$ be on the real or on the imaginary λ axis; for $E > 0$ they then enter the upper right-hand quarter plane. Since increase of the potential strength moves more and more such zeros past any fixed point $\lambda > 0$, there are only three alternatives for the location of the supply of zeros at $E = 0$: (1) They all lie on the imaginary axis and as the potential strength is increased, they move down and onto the real axis; (2) there are infinitely many of them on the positive real axis, with an accumulation point at $\lambda = 0$; (3) there are infinitely many of them in the left-hand half of the λ plane (or they are "generated" there) and as the potential strength increases they move through $\lambda = 0$ onto the positive real axis.

If we were to think about these matters in "physical" terms we would be led to believe in alternative (1). The reason is that if the centrifugal r^{-2} term is cut off at small distances we may think of the centrifugal barrier as another potential. For $\lambda^2 < \frac{1}{4}$ that "barrier" becomes attractive. The more imaginary we take λ, the deeper that attractive potential becomes and the more bound states it produces. Hence, with the cutoff present there must be infinitely many zeros of F (which then is an even function of λ) on the imaginary λ axis for any negative energy. Such reasoning, however, is entirely misleading. It will be proved in Sec. 7 that there can in fact be *no* $E \leq 0$ zeros on the imaginary λ axis. In other words, the analytic structure of $F(\lambda, k)$ is changed so violently by the cutoff that even such gross features as zero distributions are changed radically.

Alternative (2) can be eliminated if we can estimate the number of zeros of F at $E = 0$ on the positive real λ axis, and we find that it is finite. The appropriate means for that is Bargmann's inequality[24] for the number of bound states of angular momentum l:

[18] This suggestion appears to be contradicted by the fact that as Re $\alpha/dE \to 0$, $\Gamma \to 0$. The resolution of this quandary is presumably that the turning point of a trajectory cannot have a small imaginary part. If it does not, then the shift term ΔE is not small and the resonance formula has no significance. See footnote 5.

[18a] *Added in proof.* In the square well case it is now known *not* to do so; see A. D. Barut and F. Colagero (to be published).

[19] R. Blankenbecler and M. L. Goldberger (to be published).

[20] M. Froissart, private communication from M. L. Goldberger.

[21] A. Ahmadzadeh, P. G. Burke, and C. Tate (to be published).

[22] V. Singh (preprint).

[23] For the purpose of this argument, it is sufficient that the potential be negative *somewhere*. We may then multiply the attractive piece by a positive parameter and increase it.

[24] V. Bargmann, Proc. Natl. Acad. Sci. U. S. **38**, 961 (1952).

$$n_l < \int_0^\infty dr\, r\, |V(r)|/(2l+1). \tag{5.1}$$

Since the number of trajectories that enter the complex plane to the right of l is equal to the number that has passed through l at some negative energy and hence is equal to the number of bound states of angular momentum l, (5.1) is the kind of estimate we want. However, as $l \to -\frac{1}{2}$, i.e., $\lambda \to 0$, (5.1) diverges and hence tells us nothing. We must sharpen it expecially to get near $\lambda = 0$.[25]

Following Schwinger's derivation[26] of (5.1), we increase the number of bound states at $\lambda = l + \frac{1}{2}$ by replacing V by $-\mathcal{U}$, where

$$\mathcal{U} \equiv \begin{cases} -V & \text{where} \quad V < 0, \\ 0 & \text{where} \quad V \geq 0. \end{cases}$$

The number of bound states of $-\mathcal{U}$ is equal to the number of $E = 0$ bound states that are introduced when $-\mathcal{U}$ is replaced by $-\sigma\mathcal{U}$ and σ is increased from zero to one. Thus, we are looking for the number of values that allow us to solve

$$\phi_\lambda(r) = \sigma_\lambda \int_0^\infty dr' G_\lambda(r, r')\mathcal{U}(r')\phi_\lambda(r') \tag{5.2}$$

with $0 \leq \sigma_\lambda \leq 1$, where G_λ is the symmetric Green's function of (2.1) at $E = 0$:

$$2\lambda G_\lambda(r, r') = r_<^{1/2+\lambda} r_>^{1/2-\lambda}. \tag{5.3}$$

Since $\mathcal{U} \geq 0$ we may define

$$\psi_\lambda \equiv \mathcal{U}^{1/2}\phi_\lambda$$

$$K_\lambda(r, r') \equiv [\mathcal{U}(r)\mathcal{U}(r')]^{1/2} g_\lambda(r, r')$$

so that

$$\sigma_\lambda^{-1}\psi_\lambda(r) = \int_0^\infty dr' K_\lambda(r, r')\psi_\lambda(r'). \tag{5.4}$$

Hence, the strengths $\sigma_\lambda^{(i)}$ at which new bound states appear are the inverses of the eigenvalues of the real, symmetric, positive semi-definite kernel K_λ. Therefore,

$$\sum_{i=1}^\infty \sigma_\lambda^{(i)-1} = \operatorname{Tr} K_\lambda = \int_0^\infty dr K_\lambda(r, r)$$

$$= \int_0^\infty dr\, r\mathcal{U}(r)/2\lambda. \tag{5.5}$$

But n_λ is given by the requirement that $\sigma_\lambda^{(i)} \leq 1$ for $i \leq n_\lambda$ and $\sigma_\lambda^{(i)} > 1$ for $i > n_\lambda$; consequently

$$\sum_{i=1}^\infty \sigma_\lambda^{(i)-1} \geq \sum_1^{n_\lambda} \sigma_\lambda^{(i)-1} \geq n_\lambda,$$

and a slightly strengthened version of (5.1) follows from (5.5):

$$n_\lambda \leq \int_0^\infty dr\, r\mathcal{U}(r)/2\lambda. \tag{5.1'}$$

In order to understand what happens at $\lambda \approx 0$ we stop before the last step and write instead

$$\sum_{i=1}^\infty \sigma_\lambda^{(i)-1} \geq \sum_1^{n_\lambda} \sigma_\lambda^{(i)-1} \geq \sigma_\lambda^{(1)-1} + n_\lambda - 1, \tag{5.6}$$

so that (5.5) yields

$$n_\lambda \leq (2\lambda)^{-1} \int_0^\infty dr\, r\mathcal{U}(r) + 1 - \sigma_\lambda^{(1)-1}. \tag{5.7}$$

We now want to find an estimate of $\sigma_\lambda^{(1)}$, the *least* strength that introduces a bound state for $\lambda \ll 1$.

We define

$$\mathfrak{K}_\lambda \equiv 2\lambda K_\lambda, \qquad \gamma_\lambda^{(i)} \equiv 2\lambda\sigma_\lambda^{(i)-1}$$

so that (5.4) reads

$$\gamma_\lambda^{(i)}\psi_\lambda(r) = \int_0^\infty dr' \mathfrak{K}_\lambda(r, r')\psi_\lambda(r'). \tag{5.8}$$

As $\lambda \to 0$ we have

$$2\lambda G_\lambda(r, r') \to (rr')^{1/2}$$

so that in the limit as $\lambda \to 0$ (5.8) approaches

$$\gamma_0^{(i)}\psi_0(r) = [r\mathcal{U}r)]^{1/2} \int_0^\infty dr'[r'\mathcal{U}(r')]^{1/2}\psi_0(r').$$

The kernel now being separable this equation has a single nonzero eigenvalue obtained by multiplying by $[\mathcal{U}(r)\, r]^{1/2}$ and integrating:

$$\gamma_0^{(1)} = \int_0^\infty dr\, r\mathcal{U}(r).$$

The normalized eigenfunction is

$$\psi_0^{(1)}(r) = [r\mathcal{U}(r)]^{1/2} \Big/ \left[\int_0^\infty dr' r'\mathcal{U}(r')\right]^{1/2}.$$

In order to find $\gamma_\lambda^{(1)}$ for $\lambda \ll 1$ we now calculate it from $\gamma_0^{(1)}$ by perturbation theory, using

$$2\lambda G_\lambda(r, r') = (rr')^{1/2} + (rr')^{1/2}[(r_</r_>)^\lambda - 1]$$

$$\approx (rr')^{1/2} - \lambda(rr')^{1/2} \ln (r_>/r_<)$$

to first order in λ. The perturbation of the eigenvalue is therefore given by

[25] Bargmann has shown that for fixed l, (5.1) cannot be improved without special restrictions on the potential. For given l there always exists a potential that causes (5.1) to be as near to equality as one pleases. Here, however, we are interested in improving (5.1) if we fix the potential and let l approach $-\frac{1}{2}$. There is no contradiction.

[26] J. Schwinger, Proc. Natl. Acad. Sci. U. S. 47, 122 (1961).

$$\int_0^\infty dr r\mathcal{U}(r)\,\Delta\gamma_\lambda^{(1)}$$

$$= -\lambda \int_0^\infty dr \int_0^\infty dr' r r' \mathcal{U}(r)\mathcal{U}(r') \ln (r_>/r_<)$$

$$= -2\lambda \int_0^\infty dr \int_0^r dr' r r' \mathcal{U}(r)\mathcal{U}(r') \ln (r/r')$$

and thus the lowest eigenvalue is to first order in λ

$$\gamma_\lambda^{(1)} = \int_0^\infty dr r\mathcal{U}(r) - 2\lambda \frac{\int_0^\infty dr \int_0^r dr' r r' \mathcal{U}(r)\mathcal{U}(r') \ln (r/r')}{\int_0^\infty dr r\mathcal{U}(r)}.$$

Insertion of this result in (5.7) yields the desired estimate for $\lambda = 0$, in which limit the perturbation theory becomes exact

$$n_0 \leq 1 + \frac{\int_0^\infty dr \int_0^r dr' r r' \mathcal{U}(r)\mathcal{U}(r') \ln (r/r')}{\int_0^\infty dr r\mathcal{U}(r)}. \qquad (5.9)$$

The zero-order term in $\sigma_\lambda^{(1)-1}$ has canceled the first term in (5.7). In other words, the reason why the estimate (5.1′) fails to tell us anything as $\lambda \to 0$ is not that there are in fact infinitely many eigenvalues, but that, no matter how weakly attractive the potential is, there is *always* at least one eigenvalue near $\lambda = 0$.[27,27a] Indeed there is one at $\lambda = 0$ for $V = 0$ since in that case (for $E = 0$) the solution $r^{1/2}$ which is more "regular" at $r = 0$ (the other one is $r^{1/2} \ln r$) is also more "regular" at infinity.[28]

As a result of (5.9), alternative (2) for the distribution of the zeros of F at $E = 0$ is eliminated for a wide class of potentials. Since alternative (1) will be eliminated later, we are left with (3). The zeros come from the left-hand half plane. We now want to investigate that region of the complex λ plane.

6. THE LEFT-HAND HALF PLANE

Let us first consider for orientation purposes what region of the potential may determine the analyticity of F for Re $\lambda < 0$. The Wronskian (2.14) may be evaluated for any r value you please; $f(\lambda, k, r)$ is, for fixed $k \neq 0$ and $r \neq 0$, an entire function of λ. Since $\varphi(\lambda, k, r)$ depends on the potential only in the region less than r, we may expect that evaluation of (2.14) near $r = 0$ leads to a criterion of analyticity in Re $\lambda < 0$ that depends critically on the behavior of V near $r = 0$.[29] That expectation will be borne out.

The crucial tool for the analytic continuation of φ into the region Re $\lambda < 0$ will be the following *lemma*:

Let $f(x)$ and its first m derivatives exist for $0 \leq x \leq x_0$. Then,

$$g(\alpha, y) = y^{-\alpha-1} \int_0^y dx\, x^\alpha f(x)$$

and its first m derivatives with respect to y are analytic functions of α for $0 \leq y \leq x_0$, regular in the region Re $\alpha > -m - 1$, except for simple poles at $\alpha = -1, -2, \cdots$. Furthermore, the residue of the pole at[30] $\alpha = -N$ is proportional to y^{N-1}.

Proof: The existence of f and of its first m derivaties implies that for $0 \leq x \leq x_0$ it can be written

$$f(x) = \sum_{N=0}^{m-1} C_N x^N + x^m \bar{f}(x),$$

where

$$x^m \bar{f}(x) = \int_0^x dx_1 \int_0^{x_1} dx_2 \cdots \int_0^{x_{m-1}} dx_m f^{(m)}(x_m),$$

and hence $\bar{f}(x)$ is bounded in $0 \leq x \leq x_0$. Insertion in g yields

$$g(\alpha, y) = \sum_{N=0}^{m-1} \frac{C_N y^N}{\alpha + N + 1} + y^{-\alpha-1} \int_0^y dx x^{m+\alpha} \bar{f}(x).$$

The first terms and all its y derivatives are analytic functions of α everywhere, except for simple poles at the negative integers; furthermore, the residues there are proportional to y^N. The second term and its first m y-derivatives are analytic functions of α regular for Re $\alpha > -m - 1$ and for all y in $0 \leq y \leq x_0$. Q.E.D.

We now want to prove that $\varphi(\lambda, k, r)$ may be analytically continued to Re $\lambda < 0$. We assume that $rV \equiv U$ is m times differentiable in $0 \leq r \leq r_0$ for some $r_0 > 0$. In the integral equation (2.10)

[27] It must be remembered that for $E = 0$ and small λ the eigenvalue criterion is quite weak. The demand is merely that the solution which is dominant ("irregular") at $r \to 0$ also be dominant at $r \to \infty$. That implies that the solution which is more "regular" at the origin is also more "regular" at infinity. "More regular" here simply means "asymptotically smaller."

[27a] *Added in proof.* S. C. Frautschi, M. Gell-Mann, and F. Zachariasen, Phys. Rev. **126**, 2204 (1962) have come to the same conclusion.

[28] It can be shown directly for an attractive square well potential that, as the depth tends to naught, the angular momentum of the $E = 0$ eigenvalue tends to $l = -\frac{1}{2}$.

[29] In addition, of course, to the general assumptions that are necessary for the considerations of Re $\lambda \geq 0$.

[30] In the following N always stands for a positive integer, unless otherwise indicated.

for φ, let us write (not explicitly indicating the k dependence)

$$\psi_\lambda(r) \equiv r^{-1/2-\lambda}\varphi(\lambda, k, r),$$

$$h_\lambda(r, r') \equiv r^{-1/2-\lambda}r'^{-1/2+\lambda}g_\lambda(k; r, r')$$
$$= (r'/r)^{2\lambda}\mathfrak{U}_\lambda(kr')\mathfrak{V}_\lambda(kr) - \mathfrak{U}_\lambda(kr)\mathfrak{V}_\lambda(kr'),$$

$$\mathfrak{U}_\lambda(kr) \equiv (\tfrac{1}{2}\pi)^{1/2}(kr)^{-\lambda}J_\lambda(kr),$$

$$\mathfrak{V}_\lambda(kr) \equiv (\tfrac{1}{2}\pi)^{1/2}(kr)^{\lambda}Y_\lambda(kr),$$

so that all the r dependence near $r = 0$ is explicitly visible. The functions $\mathfrak{U}_\lambda$ and $\mathfrak{V}_\lambda$ are now finite at $r = 0$ for all λ and can be expanded in a power series in r. They are entire function of λ for all r.

ψ_λ satisfies the integral equation

$$\psi_\lambda(r) = \mathfrak{U}_\lambda(kr) + \int_0^r dr' U(r')h_\lambda(r, r')\psi_\lambda(r') \qquad (6.1)$$

which we solve by successive approximations:

$$\psi_\lambda = \sum_n \psi_\lambda^{(n)}$$

$$\psi_\lambda^{(0)}(r) = \mathfrak{U}_\lambda(kr)$$

$$\psi_\lambda^{(n+1)}(r) = \int_0^r dr' U(r')h_\lambda(r, r')\psi_\lambda^{(n)}(r').$$

So for $0 \le r \le r_0$ according to our lemma,

$$\bar{\psi}_\lambda^{(1)}(r) \equiv r^{-1}\psi_\lambda^{(1)}(r)$$
$$= r^{-1}\int_0^r dr' U(r')[(r'/r)^{2\lambda}\mathfrak{U}_\lambda(kr')\mathfrak{V}_\lambda(kr)$$
$$- \mathfrak{U}_\lambda(kr)\mathfrak{V}_\lambda(kr')]\mathfrak{U}_\lambda(kr')$$

and its first m r-derivatives, are analytic functions of λ regular for $\mathrm{Re}\,\lambda > -\frac{1}{2}m - \frac{1}{2}$, except possibly for poles at $\lambda = -\frac{1}{2}, -1, -\frac{3}{2}, \cdots$. However, as λ approaches a negative integer $-N$, the first $2N$ terms in the expansion of $\mathfrak{U}_\lambda$ in powers of (kr') have simple zeros as functions of λ, and at $\lambda = -N$, $\mathfrak{U}_\lambda(kr')$ is of order r^{2N} as $r \to 0$. Consequently, $\bar{\psi}_\lambda^{(1)}$ has *no* poles at $\lambda = -1, -2, \cdots$. Moreover, owing to the presence of *two* factors of $\mathfrak{U}_\lambda$ in the troublesome term, $\bar{\psi}_\lambda^{(1)}(r)$ is of order[31] r^{2N}. The residue of the pole of $\bar{\psi}_\lambda^{(1)}$ at $\lambda = -N - \frac{1}{2}$ is proportional to r^{2N}.

We proceed by induction. Assume that $\bar{\psi}_\lambda^{(n)} \equiv r^{-n}\psi_\lambda^{(n)}$ and its first m r-derivatives are analytic functions of λ for $\mathrm{Re}\,\lambda > -\frac{1}{2}m - \frac{1}{2}$, (for $0 \le r \le r_0$) except for simple poles at $\lambda = -N - \frac{1}{2}$, $N = 0, 1, 2, \cdots$, where the residue of $\bar{\psi}_\lambda^{(n)}$ is proportional to r^{2N}; and that at $\lambda = -N$ $\bar{\psi}_\lambda^{(n)}$ is of order r^{2N}. Then for $n \ge 1$

$$\bar{\psi}_\lambda^{(n+1)}(r) \equiv r^{-n-1}\psi_\lambda^{(n+1)}(r)$$
$$= r^{-n-1}\int_0^r dr' U(r')r'^n[(r'/r)^{2\lambda}\mathfrak{U}_\lambda(kr')\mathfrak{V}_\lambda(kr)$$
$$- \mathfrak{U}_\lambda(kr)\mathfrak{V}_\lambda(kr')]\bar{\psi}_\lambda^{(n)}(r')$$

and its first m r-derivatives are analytic functions of λ regular for $\mathrm{Re}\,\lambda > -\frac{1}{2}m - \frac{1}{2}$ except for poles at $\lambda = -\frac{1}{2}, -\frac{3}{2}, \cdots$. The poles for $-\lambda \le \frac{1}{2}n$ are simple because they come from $\bar{\psi}_\lambda^{(n)}$ only; those for $-\lambda > \frac{1}{2}n$ appear to be double, coming as they do both from $\bar{\psi}_\lambda^{(n)}$ and from the integral. But since the residue of $\bar{\psi}_\lambda^{(n)}$ at $\lambda = -N - \frac{1}{2}$, $N = 0, 1, 2, \cdots$, is proportional to r'^{2N}, the integration does not introduce a further divergence as $\lambda \to -N - \frac{1}{2}$. The poles are thus all simple. Moreover, the residue of $\bar{\psi}_\lambda^{(n+1)}$ at the pole at $\lambda = -N - \frac{1}{2}$ is proportional to r^{2N}. There are no poles at $\lambda = -N$ because the first $2N$ terms of $\mathfrak{U}_\lambda$ have simple zeros there as functions of λ. Since furthermore $\bar{\psi}_\lambda^{(n)}$ is of order r^{2N} at $\lambda = -N$, $\bar{\psi}_\lambda^{(n+1)}$ is of order r^{2N} there.

This proves that each term in the series of successive approximations to φ is an analytic function of λ for $\mathrm{Re}\,\lambda > -\frac{1}{2}m - \frac{1}{2}$ except for simple poles at $\lambda = -\frac{1}{2}, -\frac{3}{2}, \cdots$. We must now consider the convergence of the series.

We keep $\psi_\lambda^{(0)} + \cdots + \psi_\lambda^{(m)}$ explicitly and start summing the series for $n > m$. Then for $n > m$

$$\Psi_\lambda^{(n+1)}(r) \equiv r^{-m-1}\psi_\lambda^{(n+1)}(r)$$
$$= \int_0^r dr_n U(r_n)(r_n/r)^{m+1}[(r_n/r)^{2\lambda}\mathfrak{U}_\lambda(kr_n)\mathfrak{V}_\lambda(kr)$$
$$- \mathfrak{U}_\lambda(kr)\mathfrak{V}_\lambda(kr_n)]\Psi_\lambda^{(n)}(r_n)$$

and therefore for $\mathrm{Re}\,2\lambda > -m - 1$,

$$|\Psi_\lambda^{(n+1)}(r)| \le C\int_0^r dr_n\,|U(r_n)|\,|\Psi_\lambda^{(n)}(r_n)| \le \cdots,$$

$$|\Psi_\lambda^{(m+1+p)}(r)| \le C^p\int_0^r dr_1\int_0^{r_1} dr_2 \cdots$$
$$\int_0^{r_{p-1}} dr_p\,|U(r_1)| \cdots |U(r_p)|\,|\Psi_\lambda^{(m+1)}(r_p)|$$
$$\le C\left[C\int_0^r dr'\,|U(r')|\right]^p \Gamma(\tfrac{1}{2} + \lambda)/p!$$

since $\Psi^{(m+1)}(r)/\Gamma(\frac{1}{2} + \lambda)$ is bounded in $0 \le r \le r_0$. It follows that the series converges absolutely. Thus, $\varphi(\lambda, k, r)$ is for each k and $r \le r_0$ an analytic function of λ regular for $\mathrm{Re}\,\lambda > -\frac{1}{2}m - \frac{1}{2}$, except for simple poles at negative half integral values of λ (negative integral l).

[31] Strictly speaking, $\mathfrak{V}_{-N}$ contains a logarithmic term so that $\bar{\psi}_\lambda^{(1)}$ appears to be of order $r^{2N}\ln r$. The ln term generally cancels, though. Whether it is there or not is of no consequence in the following.

In order to eliminate the poles of φ we define a new function

$$\bar{\varphi}(\lambda, k, r) \equiv \varphi(\lambda, k, r)/\Gamma(\tfrac{1}{2} + \lambda) \qquad (6.2)$$

which has no singularities for Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$. However, at $\lambda = -\frac{1}{2}, -1, -\frac{3}{2}, \cdots$ $\bar{\varphi}(\lambda, k, r)$ is linearly related to $\bar{\varphi}(-\lambda, k, r)$. We see this from the discussion of the expansion of φ as well as from the value of the Wronskian

$$\bar{\varphi}(\lambda, k, r)\bar{\varphi}'(-\lambda, k, r) - \bar{\varphi}'(\lambda, k, r)\bar{\varphi}(-\lambda, k, r) = -\sin(2\pi\lambda)/2\pi \qquad (6.3)$$

which is obtained from the boundary condition (2.8) and (6.2). So we have

$$\bar{\varphi}(-\tfrac{1}{2}N, k, r) = C_N(k^2)\bar{\varphi}(\tfrac{1}{2}N, k, r). \qquad (6.4)$$

For special values of k, however, $C_N(k^2)$ may vanish. In other words, in contrast to $\bar{\varphi}(\lambda, k, r)$ for $\lambda \neq -\frac{1}{2}N$, $\bar{\varphi}(-\frac{1}{2}N, k, r)$ is *not* guaranteed by any boundary condition not to vanish identically in r.

We may cite two examples in which the above behavior is well known. One is φ_0, i.e., the Bessel function. In that case, $\varphi(-N, k, r)$ is a multiple of $\varphi(N, k, r)$, but there happen to be no poles at $\lambda = -N + \frac{1}{2}$; hence $\bar{\varphi}(-N + \frac{1}{2}, k, r) = 0$. The other example is that of the Coulomb potential. Then

$$\bar{\varphi}(\lambda, k, r) = 2^{\lambda-1/2}r^{1/2+\lambda}e^{ikr} F(\tfrac{1}{2} + \lambda + i\eta, 1 + 2\lambda, -2ikr)/\Gamma(1 + 2\lambda),$$

where $\eta = e^2/2k$ and F is the confluent hypergeometric function. F has simple poles at $2\lambda = -N$, and for $2\lambda \to -N$, F/Γ is a multiple of its value at $2\lambda = +N$. However, if $\frac{1}{2} + \lambda + i\eta$ assumes a negative integral value too, then F/Γ vanishes, i.e., $C_N = 0$.

The analytic continuation of φ to the left-hand λ plane allows us now to perform the same continuation for the Jost function $F(\lambda, k)$ via (2.14). The function $f(\lambda, k, r)$, being even in λ, offers no difficulty; it is everywhere regular as a function of λ. As a result, if $U = rV$ is m times differentiable at $r = 0$, then $F(\lambda, k)$ for fixed $k \neq 0$, Im $k \leq 0$, is an analytic function of λ, regular for Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$ except for simple poles at $\lambda = -\frac{1}{2}, -\frac{3}{2}, \cdots$.[32] The function

$$\bar{F}(\lambda, k) \equiv F(\lambda, k)/\Gamma(\tfrac{1}{2} + \lambda) = f\bar{\varphi}' - f'\bar{\varphi} \qquad (6.5)$$

then contains none of these poles. It serves as well as F for the S matrix. By (2.15)

$$S(\lambda, k) = e^{i\pi(\lambda-1/2)}\bar{F}(\lambda, k)/\bar{F}(\lambda, -k). \qquad (6.6)$$

Some remarks are in order concerning the conditions under which the analytic continuation of φ, F, and S to the left-hand λ plane has been proved. If the potential is a superposition of Yukawa potentials,

$$V = \int_\mu^\infty d\mu'\sigma(\mu')e^{-\mu' r}/r$$

then the existence of rV and of its first m derivatives at $r = 0$ follows from the existence of the first $m + 1$ absolute moments of σ:

$$\int_\mu^\infty d\mu'\mu'^p\,|\sigma(\mu')| < \infty, \qquad 0 \leq p \leq m.$$

The important criterion is therefore the behavior of σ for large μ'.

At the same time it should be realized that while the criterion of differentiability is sufficient, it is not necessary. Indeed, the same type of proof would clearly be applicable if rV near the origin, say, went like $r^{m+\beta}$, where $0 < \beta < 1$. Then the $(m + 1)$st derivative would fail to exist at $r = 0$, but the effect would merely be to introduce additional fixed poles on the negative λ axis, not at the integers or half integers. Such reasoning can then be extended to a very much larger class of potentials. The general conclusion, at any rate, that emerges from this is that the continuability of the S matrix to the left-hand λ plane depends primarily on the detailed behavior of the interaction at small distances.

7. CONSEQUENCES OF THE ANALYTIC CONTINUATION

We may insert (2.13) with (6.2) and (6.5) in (6.3). The result is

$$\bar{F}(\lambda, k)\bar{F}(-\lambda, -k) - \bar{F}(\lambda, -k)\bar{F}(-\lambda, k) = -i(k/\pi)\sin 2\pi\lambda. \qquad (7.1)$$

For k real and $\lambda = i\lambda'$ imaginary this says by (2.16)

$$|\bar{F}(i\lambda', k)|^2 - |\bar{F}(-i\lambda', k)|^2 = (k/\pi)\sinh 2\pi\lambda',$$

which merely corroborates the previous finding that $\bar{F}$ cannot vanish for positive (negative) k on the positive (negative) imaginary axis.

For $\lambda = i\lambda'$ and $k = ik'$ both imaginary, (7.1) implies by (2.16) that

$$\frac{1}{2}\left[\frac{\bar{F}(\lambda, k)}{\bar{F}^*(\lambda, k)} - \frac{\bar{F}(\lambda, -k)}{\bar{F}^*(\lambda, -k)}\right] = \frac{ik'\sinh 2\pi\lambda'}{2\pi\bar{F}^*(\lambda, k)\bar{F}^*(\lambda, -k)}.$$

Since the magnitude of the left-hand side cannot exceed unity, we must have

[32] Some or all of these poles may be absent in special cases. For example, if V remains bounded as $r \to 0$ then there is no pole at $\lambda = -\frac{1}{2}$. For a square well potential of radius R, φ has no pole for $r < R$; it is essentially a Bessel function.

$$|\bar{F}(\lambda, k)\bar{F}(\lambda, -k)| \geq |(k'/2\pi)\sinh 2\pi\lambda'|. \quad (7.2)$$

This implies that for negative energy $\bar{F}$ cannot vanish on the imaginary λ axis (except at $\lambda = 0$) thus eliminating alternative (1) of Sec. 5 and the possibility, still left open in Sec. 4, that as E decreases, a zero of $\bar{F}$ moves down or up the imaginary axis. There are then two alternatives for the motion of a zero at $E < 0$ in $\mathrm{Re}\,\lambda > 0$ as the energy decreases. Either it approaches $\lambda = 0$ as $E \to -\infty$ or else it passes into the left-hand half plane. The first alternative will be eliminated in Sec. 8.

Once a zero of $\bar{F}$ moves into the region $\mathrm{Re}\,\lambda < 0$ for negative E, it is no longer constrained by the previous reason to remain on the real axis; Eq. (4.2) cannot be analytically continued, $|\varphi|^2$ not being an analytic function. However, (2.16) still being valid, if $\bar{F}(\lambda, -ik') = 0$ for real k then $\bar{F}(\lambda^*, -ik') = 0$ too. That implies that if a zero were to pass into the complex plane then it would have to split in two, one passing into the upper, the other, into the lower half plane. Such a situation is unlikely to occur, but there appears to be no reason why it couldn't. We shall assume that it is at worst an exceptional possibility; that for $E < 0$ a zero in the left half of the λ plane remains generally on the real axis.[32a]

Suppose that a zero of $\bar{F}(\lambda, k)$ for $E < 0$ lands on a positive integral or half integral value $\lambda = \frac{1}{2}N$. Then (7.1) shows that

$$\bar{F}(\tfrac{1}{2}N, -k)\bar{F}(-\tfrac{1}{2}N, k) = 0.$$

But $\bar{F}(\lambda, k)$ and $\bar{F}(\lambda, -k)$ cannot both vanish for the same k and λ, because otherwise (2.13) would imply that $\varphi \equiv 0$, contradicting (2.8). Consequently we must have $\bar{F}(-\frac{1}{2}N, k) = 0$. If a zero passes through an integral or half integral value of λ then there must be a zero on the left passing at the same energy through the symmetrical point. This conclusion, however, cannot be inverted. If $\bar{F}$ vanishes at $\lambda = -\frac{1}{2}N$ then (7.1) says that

$$\bar{F}(\tfrac{1}{2}N, k)\bar{F}(-\tfrac{1}{2}N, -k) = 0$$

for the same k. But now we don't know that $\bar{F}(-\frac{1}{2}N, -k) \neq 0$, because $\lambda = -\frac{1}{2}N$ are just the values at which $\bar{\varphi}$ may vanish identically in r. If $\bar{F}(-\frac{1}{2}N, -k) = 0$ then there need be no zero at $\lambda = \frac{1}{2}N$. Such a situation means that S has a pole and a zero which meet at $\lambda = -\frac{1}{2}N$ and they annihilate one another. Thus there is neither a pole nor a zero of S at $\lambda = -\frac{1}{2}N$, although there are both nearby at neighboring energies. S then need not have a pole at $\lambda = \frac{1}{2}N$.

An example of such a situation is the Coulomb potential. In the attractive case we have for $k = i|k|$

$$S = \Gamma(\tfrac{1}{2} + \lambda - |\eta|)/\Gamma(\tfrac{1}{2} + \lambda + |\eta|).$$

The poles occur at $\frac{1}{2} + \lambda - |\eta| = -N$; if that happens when $\lambda = \frac{1}{2}m$, $m = 1, 2, \cdots$, then $\frac{1}{2} - \lambda - |\eta| = -N - m$ is a negative integer too and S has a pole also at $\lambda = -\frac{1}{2}m$. *Vice versa*, if the pole occurs at $\lambda = -\frac{1}{2}m$ then $\frac{1}{2} - \lambda - |\eta| = -N + m$ is a nonpositive integer only if $N \geq m$. Hence, for $m > N$ there is no pole on the right corresponding to the one on the left. Instead $\frac{1}{2} + \lambda + |\eta| = N + 1 - m$ is a nonpositive integer and a zero of S coincides with the pole. Similarly for the repulsive Coulomb case:

$$S = \Gamma(\tfrac{1}{2} + \lambda + |\eta|)/\Gamma(\tfrac{1}{2} + \lambda - |\eta|)$$

for $k = i|k|$. There are, of course, no poles for $\lambda > 0$. For $|\eta| = -N - \lambda - \frac{1}{2}$ there are poles; if such a pole occurs at $\lambda = -\frac{1}{2}m$, then $m > 2N + 1$, and $\frac{1}{2} + \lambda - |\eta| = N + 1 - m$ is a negative integer too and a zero and a pole of S coincide.[33]

It is worthwhile to rewrite (7.1) directly as an equation for the S matrix. It then reads

$$e^{-i\pi\lambda}S(\lambda, k) - e^{i\pi\lambda}S(-\lambda, k) = -\frac{k}{\pi}\frac{\sin 2\pi\lambda}{\bar{F}(\lambda, -k)\bar{F}(-\lambda, -k)}. \quad (7.3)$$

From this one may conclude that usually

$$S(-\tfrac{1}{2}N, k) = (-)^N S(\tfrac{1}{2}N, k) \quad (7.4)$$

except when $\bar{F}(-\frac{1}{2}N, -k) = 0$ or $\bar{F}(+\frac{1}{2}N, -k) = 0$. The exceptional cases are just the ones in which a pole and a zero of S coincide. S then has poles neither at $\lambda = \frac{1}{2}N$ nor at $\lambda = -\frac{1}{2}N$, and in addition (7.4) breaks down. On the other hand, if S does have a pole at $\lambda = \frac{1}{2}N$, then (7.4) holds for the residues.

8. HIGH-ENERGY BEHAVIOR

For integral values of l it is a well-established fact that φ and f approach their "unperturbed" values as $E \to \infty$ (for f, with k in the lower half plane or real). The proof is obviously extended easily to all λ with $\mathrm{Re}\,\lambda > 0$. In the left half plane,

[32a] *Added in proof.* Meanwhile, numerical work by A. O. Barut and F. Calogero (to be published) indicates that for repulsive square-well potentials, complex negative energy zeros *do* occur. This will be discussed in more detail in a forthcoming publication with B. R. Desai.

[33] It would be a mistake to believe that such a coincidence of zeros and poles at $\lambda = -\frac{1}{2}N$ may be a result of the anomalous tail of the Coulomb field. It must happen, for example, for the Yukawa potential too. Otherwise there would be a prohibition against a pole moving through $\lambda = -\frac{1}{2}N$ for large enough N.

the extension for f is also immediate since f is an even function of λ. For φ, however, the proof needs a little elaboration because of the need to continue the integral equation (2.10) analytically.

If we define

$$\zeta(\lambda, x) \equiv |k|^{1/2} k^{\lambda} \varphi(\lambda, k, x/|k|)$$

then ζ satisfies the integral equation

$$\zeta(\lambda, x) = \zeta_0(\lambda, x) + \int_0^x dx' g_\lambda(x, x') |k|^{-2} V(x'/|k|)\zeta(\lambda, x'), \qquad (8.1)$$

where

$$\zeta_0(\lambda, x) = (\tfrac{1}{2}\pi x)^{1/2} J_\lambda(xe^{i\varphi}),$$

$$g_\lambda(x, x') = \tfrac{1}{2}\pi(xx')^{1/2} [J_\lambda(x'e^{i\varphi}) Y_\lambda(xe^{i\varphi}) - J_\lambda(xe^{i\varphi}) Y_\lambda(x'e^{i\varphi})],$$

$$k = |k| e^{i\varphi}.$$

We have seen in Sec. 6 that (2.10) and therefore (8.1) can be solved by successive approximations and each term can be analytically continued to Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$ under appropriate assumptions on the potential, which include existence of $rV(r) = |k|^{-1}xV(x/|k|)$ at $r \to 0$. It is clear then from (8.1) that as $|k| \to \infty$

$$\zeta(\lambda, x) \to \zeta_0(\lambda, x)$$

for Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$.

It follows from the fact that φ and f approach f_0 and φ_0, respectively, as $|k| \to \infty$, that S approaches unity. Since, for large k, F does not go to unity, it is more convenient to discuss the function

$$\mathfrak{f}(\lambda, k) \equiv e^{(i\pi/2)(\lambda - 1/2)} k^{\lambda - 1/2} F(\lambda, k) \qquad (8.2)$$

which is treated in more detail in Appendix A. In terms of it we have directly

$$S(\lambda, k) = \mathfrak{f}(\lambda, k)/\mathfrak{f}(\lambda, -k) \qquad (8.3)$$

and it has the integral representation[34]

$$\mathfrak{f}(\lambda, k) = 1 - i(\tfrac{1}{2}\pi)^{1/2} k^{\lambda} \int_0^\infty dr r^{1/2} V(r) H_\lambda^{(2)}(kr)\varphi(\lambda, k, r)$$

$$= 1 - i(\tfrac{1}{2}\pi)^{1/2} \int_0^\infty dx x^{1/2} |k|^{-2} V(x/|k|) H_\lambda^{(2)}(x)\zeta(\lambda, x). \qquad (8.4)$$

Divide the integral into two pieces

$$\int_0^\infty dx = \int_0^X dx + \int_X^\infty dx.$$

The second piece requires no analytic continuation to Re $\lambda < 0$; it can be handled directly and for fixed λ its contribution vanishes as $|k| \to \infty$. In the first integral we insert the series of successive approximations to ζ and analytically continue each term to Re $\lambda < 0$. We then keep $\lambda \neq -N + \frac{1}{2}$ fixed and let $|k| \to \infty$. Clearly all the integrals tend to naught provided that $|k|^{-2}V(x/|k|) \to 0$, i.e., $r^2V(r) \to 0$ as $r \to 0$.[35] As a result we find that for each fixed $\lambda \neq -N + \frac{1}{2}$ in Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$

$$\mathfrak{f}(\lambda, k) \to 1 \quad \text{as} \quad |k| \to \infty \qquad (8.5)$$

with Im $k \leq 0$ (because of the $H_\lambda^{(2)}$).

If (8.5) holds for each fixed $\lambda \neq -N + \frac{1}{2}$ in the region Re $\lambda > -\frac{1}{2}m - \frac{1}{2}$, then it must hold uniformly in any closed region in the λ plane whose outer boundary is to the right of Re $\lambda = -\frac{1}{2}m - \frac{1}{2}$, but which is otherwise as large as you please, and which excludes small circles around the points $\lambda = -N + \frac{1}{2}$. That implies that there exists a number K so that for $|k| > K$, $\mathfrak{f}(\lambda, k)$ can have no more zeros in that region. The same then holds for $\bar{F}(\lambda, k)$. As $E \to \pm\infty$ all the zeros of $\bar{F}$ must move out of the region stated. That implies that a zero must move either to Re $\lambda \leq -\frac{1}{2}m - \frac{1}{2}$, or to infinity, or else to $\lambda = -N + \frac{1}{2}$. If all the derivaties of Vr exist at $r = 0$, then a zero trajectory of $\bar{F}$ can end (at $|E| = \infty$) only on the negative integral values of l, or else at $|l| = \infty$.[36,37] (That disposes of the possibility of ending at $\lambda = 0$, mentioned in Sec. 7.)

An example in which the trajectories all end on negative integral values of l is the Coulomb field. Another example in which this is known to happen is the Yukawa potential.[19,20] On the other hand, for the square well potential all trajectories must lead to $|l| = \infty$. That follows from the fact that φ has no poles in that case.[32,38]

---

[34] This is Eq. (4.4) of reference 7.

[35] The integral in (8.4) does not diverge as $\lambda \to -N$ because we have seen that there the first $2N$ terms in φ(or ζ) have simple zeros and ζ goes as $x^{N+1/2}$.

[36] Nothing is known that would prohibit one end of a trajectory from being at $l = -N$ and the other at $l = -N'$, $N' \neq N$, or at $|l| = \infty$.

[37] It was noted at the end of Sec. 6 that the proof of the meromorphic character of S can be extended to cases in which rV goes as r^α, say, near the origin. It is clear in view of the foregoing considerations that in such a case the ends of trajectories may lie at nonintegral values of $-l$. The possible end points of trajectories are the poles of φ, i.e., of F. These are easily obtainable from α.

[38] This was also shown directly by B. R. Desai (private communication). It is clear from our general reason that this is not due to the sharp cutoff at R. For any potential that is constant over a finite region starting at $r = 0$, the trajectories must end at infinity in the λ plane. In that case φ has no poles.

ACKNOWLEDGMENT

The author is happy to acknowledge some stimulating discussions with Dr. B. R. Desai.

APPENDIX A. THE BRANCH POINT AT $k = 0$

The Bessel function $J_\lambda(z)$ has a branch point of a simple nature at $z = 0$; $z^{-\lambda}J_\lambda(z)$ is analytic there. The Hankel function, on the other hand, does not have such a simple property. The function $z^\lambda H_\lambda^{(2)}(z)$ is finite at $z = 0$, yet there is still a branch point. It is readily shown from

$$H_\lambda^{(2)}(z) = \frac{e^{i\pi\lambda}J_\lambda(z) - J_{-\lambda}(z)}{i \sin \pi\lambda}$$

that

$$H_\lambda^{(2)}(ze^{-2\pi i}) = -H_\lambda^{(2)}(z) + 2 \cos \pi\lambda H_\lambda^{(2)}(ze^{-i\pi}). \quad \text{(A1)}$$

As a result, we have by (2.5)[39]

$$f_0(\lambda, ke^{-2\pi i}, r) = f_0(\lambda, k, r) - 2i \cos \pi\lambda f_0(\lambda, -k, r) \quad \text{(A2)}$$

since we want to mean $f(\lambda, -k, r) = f(\lambda, ke^{-i\pi}, r)$ so that the connection from the positive to the negative k axis goes via the lower half plane. The Green's function in (2.6) being analytic at $k = 0$, we get the same result for f:

$$f(\lambda, ke^{-2\pi i}, r) = f(\lambda, k, r) - 2i \cos \pi\lambda f(\lambda, -k, r), \quad \text{(A3)}$$

and similarly, for F,

$$F(\lambda, ke^{-2\pi i}) = F(\lambda, k) - 2i \cos \pi\lambda F(\lambda, -k). \quad \text{(A4)}$$

Since at $k \approx 0$ and $\operatorname{Re} \lambda > 0$, f is of order $k^{1/2-\lambda}$, F is also of the same order there and it is convenient to define

$$\mathfrak{f}(\lambda, k) \equiv e^{(i\pi/2)(\lambda-1/2)}k^{\lambda-1/2}F(\lambda, k). \quad \text{(A5)}$$

The exponential factor assures that for real k and λ

$$\mathfrak{f}(\lambda, -k) \equiv \mathfrak{f}(\lambda, ke^{-i\pi}) = \mathfrak{f}^*(\lambda, k). \quad \text{(A6)}$$

Insertion in the definition (2.15) of the S matrix then gives

$$S(\lambda, k) = \mathfrak{f}(\lambda, k)/\mathfrak{f}(\lambda, -k). \quad \text{(A7)}$$

The function $\mathfrak{f}$ is the direct extension of the Jost function, as a comparison with (4.3) and (3.4) of reference 7 shows. It tends to unity as $|k| \to \infty$ in $\operatorname{Im} k \leq 0$ and it has the integral representation (8.4). The exponential factor in (2.15) is therefore necessary in order that even for nonintegral l values $S(\lambda, k)$ tend to unity as $k \to \pm\infty$ and that S be unitary.

The function $\mathfrak{f}$ still has a branch point at $k = 0$, even though it is finite there. From (A4) and (A5) we find

$$\mathfrak{f}(\lambda, ke^{-2\pi i}) = -e^{-2\pi i\lambda}\mathfrak{f}(\lambda, k) + (1 + e^{-2\pi i\lambda})\mathfrak{f}(\lambda, -k). \quad \text{(A8)}$$

Consequently the S matrix has a branch point at $k = 0$. From (A7) and (A8) we get[39]

$$S(\lambda, ke^{-i\pi}) \equiv S(\lambda, -k) = e^{2\pi i\lambda}/[1 + e^{2\pi i\lambda} - S(\lambda, k)], \quad \text{(A9)}$$

which for half integral λ (integral l) goes over into the usual

$$S(\lambda, -k) = 1/S(\lambda, k).$$

Repetition of (A9) yields[39]

$$S(\lambda, ke^{-2\pi i}) = \frac{S(\lambda, k) - 1 - e^{2\pi i\lambda}}{(1 + e^{-2\pi i\lambda})S(\lambda, k) - 1 - e^{2\pi i\lambda} - e^{-2\pi i\lambda}}. \quad \text{(A10)}$$

Therefore, when $S(\lambda, k)$ has a pole then

$$S(\lambda, ke^{-2\pi i}) = (1 + e^{-2\pi i\lambda})^{-1}. \quad \text{(A11)}$$

The direction of the branch cut from $k = 0$ is arbitrary. We may take it along the negative real k axis, so that both the upper and the lower half plane are directly accessible from the positive k axis. As a function of E, S of course always has a cut along the positive axis. So long as we stay on the "physical sheet," the additional cut along the lower edge of the positive E axis never comes into play.

The equation $F(\lambda, k) = 0$ defines λ as an analytic function of k. Since F has a branch cut along the negative k axis and a zero of F on one sheet is by (A4) not in general also a zero of F on another sheet, $\lambda(k)$ also has a branch cut along the negative k axis. The same remark as for $S(\lambda, k)$ applies. Regarding λ as a function of E, if we stay on the "physical sheet," the additional cut never matters.

Added in proof. It has become customary to follow Bottino *et al.*[3] and to place the "kinematic" branch cut of $F(\lambda, k)$ along the positive imaginary k axis, where, starting at some finite value, the "Yukawa cut" lies. The corresponding cut of S in the E plane then runs along the negative real axis (on both sheets) and is referred to as the "left-hand cut." Since the pole position $\alpha(k) = \lambda(k) - \frac{1}{2}$ is defined by $F(\lambda, -k) = 0$ and $F(\lambda, -k)$ has its left-hand cut on the *second* sheet as a function of E, $\alpha(E)$ has no left-hand cut on the first sheet, but only on the second.

[39] This was shown also in reference 3.

APPENDIX B. THE RESIDUES OF S

The residue of the S matrix at a point λ_0, k_0 where $F(\lambda_0, -k_0) = 0$ (as a function of λ) is by (2.15)

$$S_n = e^{i\pi(\lambda-1/2)}F(\lambda_0, k_0)/[\partial F(\lambda_0, -k_0)/\partial\lambda_0]. \quad \text{(B1)}$$

By the same technique used in Sec. 3, we get when $F(\lambda_0, -k_0) = 0$

$$\partial F(\lambda_0, -k_0)/\partial\lambda_0 = 4i\lambda_0 k_0 \int_0^\infty dr r^{-2}\varphi^2(\lambda_0, k_0, r)/F(\lambda_0, k_0).$$

But by (2.13) we then have

$$\varphi(\lambda_0, k_0, r) = (2ik_0)^{-1}F(\lambda_0, k_0)f(\lambda_0, -k_0, r). \quad \text{(B2)}$$

As a result

$$S_n(k_0) = \frac{e^{i\pi(\lambda_0-1/2)}ik_0/\lambda_0}{\int_0^\infty dr r^{-2}f^2(\lambda_0, -k_0, r)}. \quad \text{(B3)}$$

When $k_0 \to 0$ we find that

$$S_n(k_0) = O(k_0^{2l_0+1}). \quad \text{(B4)}$$

Added in proof. It follows from (B2) and from the fact that neither $\varphi(\lambda, k, r)$, nor $f(\lambda, -k, r)$, nor $\alpha(k)$ has a cut along the *positive* imaginary k-axis, that $F[\lambda_0(k), k]$ has no such cut either. Consequently the residue $S_n(k)$ by (B1) has no "left-hand cut." This fact was explicitly pointed out by J. R. Taylor (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Spectral Theory of Dirac's Radial Relativistic Wave Equation

B. W. Roos and W. C. Sangren*
John Jay Hopkins Laboratory for Pure and Applied Science,
General Atomic Division of General Dynamics Corporation,
San Diego, California
(Received March 9, 1962)

The analytical methods developed by Weyl and Titchmarsh for the study of the Sturm-Liouville equation are extended to the investigation of the spectral properties of the Dirac radial wave equation. It is shown how Weyl's limit-point, limit-circle theorem may be generalized to include the singular cases of a system of two first-order differential equations. A transformation is introduced and order properties are established for the solution of the corresponding integral equations. The nature of the spectrum is discussed for specific singular potentials.

I. INTRODUCTION

THE analytic methods developed by Weyl,[1] Titchmarsh,[2] and many others for the solution of the Sturm-Liouville equation

$$x''(r) + [\lambda - V(r)]x(r) = 0 \qquad (' \equiv d/dr) \quad \text{(1.1)}$$

have proven to be powerful tools for the study of the spectral properties of the singular cases of the second-order differential equations of mathematical physics. For example, since Eq. (1.1) corresponds to the radial wave equation of a nonrelativistic particle in a central field, the methods referred to may be advantageously applied to determine the spectral properties of the Schrödinger operator for singular potentials. In the case of the Schrödinger equation such potentials behave near the origin as r^{-2} or, in general as r^{-n} with $n \geq 2$. The spectral properties and the solutions of wave equations with singular potentials of this type and of a more general character have been studied by Case.[3]

It is the purpose of this paper to generalize the methods of Weyl and Titchmarsh in order to investigate the spectral behavior of the system of two first-order differential equations,

$$\begin{aligned} x_1'(r) - [\lambda a(r) + b(r)]x_2(r) &= 0 \\ x_2'(r) + [\lambda c(r) + d(r)]x_1(r) &= 0, \end{aligned} \quad \text{(1.2)}$$

with singular coefficients.

Because the system (1.2) corresponds to Dirac's radial relativistic wave equation for a particle in a central field, these generalizations enable us to

* Present address: Computer Applications, Inc., San Diego, California.

[1] H. Weyl, Math. Annal. **68**, 220 (1910).

[2] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Differential Equations* (Clarendon Press, Oxford, England, 1946).

[3] K. Case, Phys. Rev. **80**, 797 (1950).

APPENDIX B. THE RESIDUES OF S

The residue of the S matrix at a point λ_0, k_0 where $F(\lambda_0, -k_0) = 0$ (as a function of λ) is by (2.15)

$$S_n = e^{i\pi(\lambda-1/2)}F(\lambda_0, k_0)/[\partial F(\lambda_0, -k_0)/\partial\lambda_0]. \quad \text{(B1)}$$

By the same technique used in Sec. 3, we get when $F(\lambda_0, -k_0) = 0$

$$\partial F(\lambda_0, -k_0)/\partial\lambda_0 = 4i\lambda_0 k_0 \int_0^\infty dr r^{-2}\varphi^2(\lambda_0, k_0, r)/F(\lambda_0, k_0).$$

But by (2.13) we then have

$$\varphi(\lambda_0, k_0, r) = (2ik_0)^{-1}F(\lambda_0, k_0)f(\lambda_0, -k_0, r). \quad \text{(B2)}$$

As a result

$$S_n(k_0) = \frac{e^{i\pi(\lambda_0-1/2)}ik_0/\lambda_0}{\int_0^\infty dr r^{-2}f^2(\lambda_0, -k_0, r)}. \quad \text{(B3)}$$

When $k_0 \to 0$ we find that

$$S_n(k_0) = O(k_0^{2l_0+1}). \quad \text{(B4)}$$

Added in proof. It follows from (B2) and from the fact that neither $\varphi(\lambda, k, r)$, nor $f(\lambda, -k, r)$, nor $\alpha(k)$ has a cut along the *positive* imaginary k-axis, that $F[\lambda_0(k), k]$ has no such cut either. Consequently the residue $S_n(k)$ by (B1) has no "left-hand cut." This fact was explicitly pointed out by J. R. Taylor (to be published).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Spectral Theory of Dirac's Radial Relativistic Wave Equation

B. W. Roos and W. C. Sangren*
John Jay Hopkins Laboratory for Pure and Applied Science,
General Atomic Division of General Dynamics Corporation,
San Diego, California
(Received March 9, 1962)

The analytical methods developed by Weyl and Titchmarsh for the study of the Sturm-Liouville equation are extended to the investigation of the spectral properties of the Dirac radial wave equation. It is shown how Weyl's limit-point, limit-circle theorem may be generalized to include the singular cases of a system of two first-order differential equations. A transformation is introduced and order properties are established for the solution of the corresponding integral equations. The nature of the spectrum is discussed for specific singular potentials.

I. INTRODUCTION

THE analytic methods developed by Weyl,[1] Titchmarsh,[2] and many others for the solution of the Sturm-Liouville equation

$$x''(r) + [\lambda - V(r)]x(r) = 0 \qquad (' \equiv d/dr) \quad \text{(1.1)}$$

have proven to be powerful tools for the study of the spectral properties of the singular cases of the second-order differential equations of mathematical physics. For example, since Eq. (1.1) corresponds to the radial wave equation of a nonrelativistic particle in a central field, the methods referred to may be advantageously applied to determine the spectral properties of the Schrödinger operator for singular potentials. In the case of the Schrödinger equation such potentials behave near the origin as r^{-2} or, in general as r^{-n} with $n \geq 2$. The spectral properties and the solutions of wave equations with singular potentials of this type and of a more general character have been studied by Case.[3]

It is the purpose of this paper to generalize the methods of Weyl and Titchmarsh in order to investigate the spectral behavior of the system of two first-order differential equations,

$$\begin{aligned} x_1'(r) - [\lambda a(r) + b(r)]x_2(r) &= 0 \\ x_2'(r) + [\lambda c(r) + d(r)]x_1(r) &= 0, \end{aligned} \quad \text{(1.2)}$$

with singular coefficients.

Because the system (1.2) corresponds to Dirac's radial relativistic wave equation for a particle in a central field, these generalizations enable us to

* Present address: Computer Applications, Inc., San Diego, California.

[1] H. Weyl, Math. Annal. **68**, 220 (1910).

[2] E. C. Titchmarsh, *Eigenfunction Expansions Associated with Second-Order Differential Equations* (Clarendon Press, Oxford, England, 1946).

[3] K. Case, Phys. Rev. **80**, 797 (1950).

investigate the asymptotic behavior of the solutions of this equation and the spectral properties of its operator. Consequently, system (1.2) is discussed here for the case that the coefficients a, b, c, and d take the form required in the Dirac equation.

For additional information about system (1.2), the reader is referred to some other papers and reports by the authors.[4-6]

II. THE DIRAC EQUATION FOR A PARTICLE IN A CENTRAL FIELD

The relativistic wave equation as proposed by Dirac may be written in the Hamiltonian form

$$i\hbar(\partial/\partial t)\psi(\mathbf{r}, t) = H\psi(\mathbf{r}, t). \tag{2.1}$$

For the wave equation of a Dirac particle in the central field of a fictitious nucleus, the Hamiltonian takes the form

$$H = -c\boldsymbol{\alpha}\cdot\mathbf{p} - \beta mc^2 + V(r). \tag{2.2}$$

The solution of the system of four simultaneous linear differential equations of the first order (2.1) is a column vector with four components. The coefficients β, $\boldsymbol{\alpha}$, are anticommutative 4×4 Hermitian matrices which find their ultimate form in the well-known Pauli spin matrices.

Since the potential function $V(r)$ in (2.2) is spherically symmetric, the Dirac equation for a particle in a central field can be separated in spherical coordinates. We will not go into the details of the separation process, but will instead refer to the many excellent references.[7,8] It may be shown that the resulting radial part of the solution of the wave equation has two components, $f(r) = [r^{-1}f_1(r), r^{-1}f_2(r)]$, and that the investigation of the radial part is sufficient for the determination of the spectral properties of the equations. Separating out the angular dependence, the equations for the radial wave functions may be written in the form given by Bethe[9]:

$$\begin{aligned}
&[E + mc^2 - V(r)]f_1(r) \\
&\quad - \hbar c[df_2(r)/dr] - (\hbar ck/r)f_2(r) = 0 \\
&[E - mc^2 - V(r)]f_2(r) \\
&\quad + \hbar c[df_1(r)/dr] - (\hbar ck/r)f_1(r) = 0.
\end{aligned} \tag{2.3}$$

Here the parameter k can take the values ± 1, ± 2, $\cdots$. At this point it is convenient to adopt the system of rational relativistic units in which $\hbar = m = c = 1$. In order to have the system of Eqs. (2.3) conform with the system (1.2), we make the substitution $x_1(r) = f_2(r)r^k$ and $x_2(r) = f_2(r)r^{-k}$. Hence,

$$\begin{aligned}
x_1'(r) - r^{2k}[E + 1 - V(r)]x_2(r) &= 0 \\
x_2'(r) + r^{-2k}[E - 1 - V(r)]x_1(r) &= 0
\end{aligned} \tag{2.4}$$

or as given in Eqs. (1.2)

$$\begin{aligned}
x_1'(r) - [\lambda a(r) + b(r)]x_2(r) &= 0 \\
x_2'(r) + [\lambda c(r) + d(r)]x_1(r) &= 0,
\end{aligned} \tag{1.2}$$

where

$$\begin{aligned}
\lambda = E,\ a(r) = r^{2k},\quad b(r) = r^{2k}[1 - V(r)], \\
c(r) = r^{-2k},\ d(r) = -r^{-2k}[1 + V(r)].
\end{aligned} \tag{2.5}$$

In the following sections we investigate the spectral properties of system (1.2) for the coefficients (2.5). After making some preliminary information available in Sec. 3, we discuss the analog of Weyl's limit-point, limit-circle theorem in Sec. 4. Section 5 introduces a transformation of the basic equations and in Secs. 5 through 9 the asymptotic properties of the relevant parameters and the solutions of the transformed equations are investigated. In Sec. 10 the spectral properties of the Dirac equations are discussed.

III. PRELIMINARIES

On the finite interval (r_0, r^0) let $v(r, \lambda) = [v_1(r, \lambda), v_2(r, \lambda)]$, $w(r, \lambda) = [w_1(r, \lambda), w_2(r, \lambda)]$ be two vector solutions of (1.2) that satisfy the conditions

$$\begin{aligned}
v_1(r_0) &= -\sin\beta, \qquad v_2(r_0) = +\cos\beta, \\
w_1(r_0) &= -\cos\beta, \qquad w_2(r_0) = -\sin\beta.
\end{aligned}$$

The Wronskian of v and w is defined by

$$W_r(v, w) = v_1(r)w_2(r) - v_2(r)w_1(r).$$

Since $W_r(v, w)$ is independent of r and $W_0(v, w) = 1$, $W_r(v, w) = 1$ and v and w are linearly independent solutions. The general solution of (1.2) may be written as

$$w(t, \lambda) + l(\lambda)v(t, \lambda).$$

---

[4] B. W. Roos and W. C. Sangren, Proc. Am. Math. Soc. 12, 468 (1961).
[5] B. W. Roos and W. C. Sangren, Pacific J. Math. (to be published).
[6] B. W. Roos and W. C. Sangren, "Asymptotic Solutions and an Equiconvergence Theorem for a Pair of First-Order Differential Equations," presented at the 586th Meeting of the American Mathematical Society (unpublished).
[7] L. I. Schiff, *Quantum Mechanics* (McGraw-Hill Book Company, Inc., 1950).
[8] P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, New York, 1935).
[9] H. A. Bethe, *Handbuch der Physik*, edited by Geiger Steele (Verlag Julius Springer, Berlin, 1933), p. 1, 313.

It is known[10,11] that if the general solution satisfies Sturmian boundary conditions at $r = r_1$, the eigenvalues will be real, nondegenerate, discrete, and extend from $\lambda = -\infty$ to $\lambda = +\infty$. The corresponding eigenfunctions are real functions of r. For the singular case the spectrum can be investigated by taking the limit of the general solution as $r^0 \to \infty$. As in the case of singular second-order differential equations, it can be shown (Sec. 4) by a limit-point, limit-circle argument that for $\operatorname{Im} \lambda \neq 0$ the system (1.2) will have a vector solution:

$$z(r, \lambda) = w(r, \lambda) + m(\lambda)v(r, \lambda)$$

belonging to the class of square integrable functions $L^2(r_0, \infty)$. The function $m(\lambda)$ depends upon the limit of circles in the complex λ plane and, for $r^0 \to \infty$, is either a limit-point or a point on a limit-circle. In the limit-circle case all solutions are in the class $L^2(r_0, \infty)$. Furthermore, $m(\lambda)$ is analytic for $\operatorname{Im} \lambda \neq 0$ and $m(\bar{\lambda}) = \overline{m(\lambda)}$.

The spectral distribution function[10,11] is determined by the imaginary part of $m(\lambda)$ and the spectrum associated with a problem for which the spectral distribution function is uniquely determined is the set of nonconstancy points of this distribution function. It may be shown that this spectrum is a closed set. The set of all discontinuity points of the spectral distribution function comprises the point spectrum. The points in the point spectrum are better known as the eigenvalues and the solutions of the differential equations corresponding to such points as the eigenfunctions. The continuous spectrum is the set of continuity points of the spectral distribution function which are in the spectrum. In the case where the spectrum corresponds to the energy states of a physical system, the continuous states of the system are to be found in the corresponding continuous spectrum of the differential equations describing the system. In the $m(\lambda)$ language the point spectrum is the set of the poles of $m(\lambda)$ if $m(\lambda)$ is a meromorphic function.

IV. LIMIT-POINT AND LIMIT-CIRCLE THEOREM

Before turning specifically to the singular cases, a theorem which is similar to that originally given by Weyl[1] is developed. This theorem deals with the solutions of Eqs. (1.2) in the interval (r_0, r^0). The coefficients $a(r)$, $b(r)$, $c(r)$, $d(r)$ are assumed to satisfy the conditions

$a(r)$, $b(r)$, $c(r)$, and $d(r)$ are real-valued functions and $a(r) > 0$, $c(r) > 0$, (4.1.a)

$a(r)$, $b(r)$, $c(r)$, and $d(r)$ are sectionally continuous functions. (4.1.b)

Let r' be an interior point of the interval (r_0, r^0). Consider two solutions $v(r) = [v_1(r), v_2(r)]$ and $w(r) = [w_1(r), w_2(r)]$ of Eqs. (1.2), such that

$$v_1(r') = -\sin \beta, \qquad v_2(r') = +\cos \beta,$$

$$w_1(r') = -\cos \beta, \qquad w_2(r') = -\sin \beta,$$

where β is real. Because the Wronskian

$$W_r(v, w) = W_{r'}(v, w) = 1,$$

the two solutions are linearly independent. The general solution of the system (1.2) has the form

$$z(r) = w(r) + m(\lambda)v(r),$$

where $m(\lambda)$ is a complex number. By means of elementary operations and the assumption that $z(r)$ is a solution of (1.2) and $\overline{z(r)}$ is a solution of the system

$$\bar{z}_1'(r) - [\bar{\lambda}a(r) + b(r)]\bar{z}_2(r) = 0$$

$$\bar{z}_2'(r) + [\bar{\lambda}c(r) + d(r)]\bar{z}_1(r) = 0,$$

one obtains

$$\bar{z}_2(r)z_1'(r) - z_2(r)\bar{z}_1'(r) = (\lambda - \bar{\lambda})a(r)z_2(r)\bar{z}_2(r) = 2i\tau a(r)\,|z_2(r)|^2,$$

$$-\bar{z}_1(r)z_2'(r) + z_1(r)\bar{z}_2'(r) = (\lambda - \bar{\lambda})c(r)z_1(r)\bar{z}_1(r) = 2i\tau c(r)\,|z_1(r)|^2,$$

where it is assumed that λ is a fixed number, such that $\tau = \operatorname{Im} \lambda \neq 0$. The result of adding first and then integrating from r' to r yields

$$2i\tau \int_{r'}^{r} [c(s)\,|z_1(s)|^2 + a(s)\,|z_2(s)|^2]\,ds = [z_1(r)\bar{z}_2(r) - z_2(r)\bar{z}_1(r)]_{r'}^{r} = W_r(z, \bar{z}) - W_{r'}(z, \bar{z}).$$

For convenience, we now introduce the quantity

$$2i\tau V_r(f, g) = W_r(f, g)$$

and we obtain

$$\int_{r'}^{r} [c(s)\,|z_1(s)|^2 + a(s)\,|z_2(s)|^2]\,ds = V_r(z, \bar{z}) - V_{r'}(z, \bar{z}). \tag{4.2}$$

The integral on the left-hand side must now be

[10] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1955).

[11] E. L. Ince, *Ordinary Differential Equations* (Longmans Green, London, 1927).

investigated for r tending to r_0 or r^0. However, since these two cases are similar it will only be necessary to consider the case that r tends to r^0. First, consider $V_{r'}(z, \bar{z})$. We have

$$V_{r'}(v, \bar{v}) = 0,\ V_{r'}(w, \bar{w}) = 0,\ \text{and}\ V_{r'}(w, \bar{v}) = 1/2i\tau$$

and, consequently,

$$V_{r'}(z, \bar{z}) = \frac{m(\lambda) - \overline{m(\lambda)}}{2i\tau} = \frac{\text{Im}\ \{m(\lambda)\}}{\tau}$$

will be a real number. If it would be possible to choose $m(\lambda)$ in such a way that $V_r(z, \bar{z}) \leq 0$ for all r in (r_0, r^0), then

$$\int_{r'}^{r} [c(s)\ |z_1(s)|^2 + a(s)\ |z_2(s)|^2]\ ds \leq -V_{r'}(z, \bar{z})$$

and, therefore,

$$\int_{r'}^{r} [c(s)\ |z_1(s)|^2 + a(s)\ |z_1(s)|^2]\ ds < \infty.$$

Now, since $z(r) = w(r) + m(\lambda)v(r)$,

$$V_r(z, \bar{z}) = m(\lambda)\overline{m(\lambda)}\, V_r(v, \bar{v}) + m(\lambda)W_r(v, \bar{w}) + \overline{m(\lambda)}W_r(w, \bar{v}) + W_r(w, \bar{w}).$$

Consequently,

$$V_r(z, \bar{z}) = V_r(v, \bar{v}) \times \left\{\left[m(\lambda) + \frac{V_r(v, \bar{w})}{V_r(v, \bar{v})}\right]\left[\overline{m(\lambda)} + \frac{V_r(w, \bar{v})}{V_r(v, \bar{v})}\right] - \frac{V_r(v, \bar{w})\, V_r(w, \bar{v}) - V_r(v, \bar{v})\, V_r(w, \bar{w})}{[V_r(v, \bar{v})]^2}\right\} \tag{4.3}$$

and from this relation it follows that for fixed r the set of points in the $m(\lambda)$ plane which satisfies the equation

$$V_r(z, \bar{z}) = 0$$

will be a circle with center

$$C(r) = -V_r(v, \bar{w})/V_r(v, \bar{v})$$

provided

$$V \equiv V_r(v, \bar{w})\, V_r(w, \bar{v}) - V_r(v, \bar{v})\, V_r(w, \bar{w})$$

is a positive number. Now,

$$V = V_r(v, \bar{w})\, V_r(w, \bar{v}) - V_r(v, \bar{v})\, V_r(w, \bar{w}) = \tau^{-2}\, |W_r(w, v)|^2 = \tau^{-2};$$

therefore V is a positive quantity. The radius of the circle is therefore

$$R(r) = 1/|\tau|\ V_r(v, \bar{v}). \tag{4.4}$$

When $m(\lambda) = C(r)$, $V_r(z, \bar{z})$ is negative and it follows that the set of points $m(\lambda)$ interior to this circle corresponds to $V_r(z, \bar{z}) < 0$. If we now fix a point $m(\lambda)$ in (4.2) and allow r to increase ($r' \leq r \leq r^0$), it follows that $V_r(z, \bar{z})$ is an increasing function of r.

Consider next the family of circles in the $m(\lambda)$ plane, which are represented by $V_r(z, \bar{z}) \leq 0$ when $r' \leq r < r^0$. Let $r' \leq r_{\cdot\cdot} < r'' < r^0$, then let $V_{r_{\cdot\cdot}}(z, \bar{z}) \leq 0$ for some point $m(\lambda)$. It then follows that $V_{r''}(z, \bar{z}) \leq 0$ for this same $m(\lambda)$, since $V_r(z, \bar{z})$ is an increasing function of r. Geometrically, this means that the circle corresponding to r'' is contained in the circle corresponding to $r_{\cdot\cdot}$. Consequently, there exist two possibilities as r tends to r^0. Either the circles $V_r(z, \bar{z}) = 0$ converge to a point, called the limit-point, or they converge to a circle, called the limit-circle. The limit-point case occurs when

$$\lim_{r \to r^0} V_r(v, \bar{v}) = \int_{r'}^{r^0} [c(s)\ |v_1(s)|^2 + a(s)\ |v_2(s)|^2]\ ds = \infty.$$

This follows immediately from the defining relation for the radius $R(r)$ of the circle (4.4). Let this limit-point be denoted by $m(r^0, \lambda)$ and let $z(r) = w(r) + m(r^0, \lambda)v(r)$. Since this point $m(r^0, \lambda)$ is contained in all the circles $V_r(z, \bar{z}) \leq 0$, we obtain the important result that

$$\int_{r'}^{r^0} [c(s)\ |z_1(s)|^2 + a(s)\ |z_2(s)|^2]\ ds < \infty.$$

Furthermore, from Eq. (4.3) it follows that

$$V_r(z, \bar{z}) \geq -V_r(v, \bar{v})R^2(r) = -\frac{1}{\tau^2 V_r(v, \bar{v})}.$$

Now, since $\lim_{r \to r^0} V_r(v, \bar{v}) = \infty$, $V_r(z, \bar{z}) \geq 0$. However, we also have $V_r(z, \bar{z}) \leq 0$ and therefore

$$\lim_{r \to r^0} V_r(z, \bar{z}) = 0$$

or

$$\lim_{r \to r^0} [z_1(r)\bar{z}_2(r) - z_2(r)\bar{z}_1(r)] = 0.$$

The limit-circle case occurs when

$$\lim_{r \to r^0} V_r(v, \bar{v}) = \int_{r'}^{r} [c(s)\ |v_1(s)|^2 + a(s)\ |v_2(s)|^2]\ ds < \infty.$$

The radius of this circle, using relation (4.4) is given by

$$\lim_{r \to r^0} R(r) = \lim_{r \to r^0} \frac{1}{|\tau|\ V_r(v, \bar{v})}.$$

Now, consider any $m(\lambda)$ contained in the limit-circle.

For such an $m(\lambda)$, $V_r(z, \bar{z}) \leq 0$ for all r in $[r', r^0]$ and, consequently,

$$\int_{r'}^{r^0} [c(s)\ |z_1(s)|^2 + a(s)\ |z_2(s)|^2]\ ds < \infty.$$

It also follows then, in this case, that every solution $x(r) = [x_1(r), x_2(r)]$ of the system (1.2) is such that

$$\int_{r'}^{r^0} [c(s)\ |x_1(s)|^2 + a(s)\ |x_2(s)|^2]\ ds < \infty.$$

Equation (4.3) has the form

$$V_r(z, \bar{z}) = V_r(v, \bar{v})[|m(\lambda) - C(r)|^2 - R^2(r)],$$

where $C(r)$ and $R(r)$ are the center and radius of the circle, respectively, corresponding to $V_r(z, \bar{z}) = 0$. If r tends to r^0, then

$$\lim_{r \to r^0} V_r(z, \bar{z}) = \lim_{r \to r^0} V_r(v, \bar{v})[|m(\lambda) - C(r^0)|^2 - R^2(r^0)],$$

where $C(r^0)$ and $R(r^0)$ are now the center and radius of the limit-circle, respectively. Now for a point on the limit-circle,

$$m(\lambda) = C(r^0) + R(r^0)e^{i\alpha},$$

where $0 \leq \alpha \leq 2\pi$, and hence,

$$\lim_{r \to r^0} V_r(z, \bar{z}) = \lim_{r \to r^0} [z_1(r)\bar{z}_2(r) - z_2(r)\bar{z}_1(r)] = 0.$$

The results may be summarized in the following theorem:

Theorem 1. (a) *For every value* λ *other than real values, there exists a solution* $z(r)$ *of the system (1.2) such that*

$$\int_{r'}^{r^0} [c(s)\ |z_1(s)|^2 + a(s)\ |z_2(s)|^2]\ ds < \infty$$

for some r' *in* (r_0, r^0).

(b) *If there exists at least one solution* $x(r) = [x_1(r), x_2(r)]$ *such that*

$$\int_{r'}^{r^0} [c(s)\ |x_1(s)|^2 + a(s)\ |x_2(s)|^2]\ ds = \infty,$$

then for any solution y, *such that*

$$\int_{r'}^{r^0} [c(s)\ |y_1(s)|^2 + a(s)\ |y_2(s)|^2]\ ds < \infty,$$

$$\lim_{r \to r^0} [y_1(r)\bar{y}_2(r) - y_2(r)\bar{y}_1(r)] = 0.$$

(c) *If for all solutions* $x(r) = [x_1(r), x_2(r)]$ *such that*

$$\int_{r'}^{r} [c(s)\ |x_1(r)|^2 + a(s)\ |x_2(s)|^2]\ ds < \infty,$$

then there exist two linearly independent solutions $v(r)$ *and* $w(r)$ *and a circle* $m(r^0, \lambda) = C(r^0) + R(r^0)e^{i\alpha}$ *in the complex plane, such that*

$$\lim_{r \to r^0} [z_1(r)\bar{z}_2(r) - z_2(r)\bar{z}_1(r)] = 0$$

for $z(r) = w(r) + m(\lambda)v(r)$ *when* $m(\lambda) = C(r^0) + R(r^0)e^{i\alpha}$.

It may be shown that the system (1.2) is uniquely in the limit-point or limit-circle situation.[12]

V. THE GENERAL TRANSFORMATION OF THE BASIC EQUATIONS

The asymptotic behavior of the solutions of system (1.2) for large λ and for large values of a coefficient when the independent variable approaches a singular point can be conveniently obtained by first using the following transformation. Let the independent variable r be replaced by

$$\alpha(r) = \int_{r_0}^{r} \{[\lambda a(s) + b(s)][\lambda c(s) + d(s)]\}^{1/2}\ ds \qquad (5.1)$$

and the dependent functions $x_1(r)$ and $x_2(r)$ by

$$\begin{aligned} u_1(r) &= F(r)x_2(r) \\ \text{and} \qquad & \qquad (5.2) \\ u_2(r) &= -F^{-1}(r)x_1(r) + G(r)x_2(r), \end{aligned}$$

where

$$F(r, \lambda) = [\lambda a(r) + b(r)]^{\frac{1}{4}}[\lambda c(r) + d(r)]^{-1/4}$$

$$G(r, \lambda) = F'(\alpha')^{-1}.$$

This transformation gives formally

$$\begin{aligned} du_1/d\alpha &= u_2, \\ du_2/d\alpha &= -u_1 + R(\lambda, a, b, c, d)u_1, \end{aligned} \qquad (5.3)$$

where

$$R(\lambda, a, b, c, d) = G'(\alpha')^{-1}F^{-1}.$$

Although the transformed equations are similar in form to Eqs. (1.2), the coefficients on the right-hand side remain bounded for large λ and for certain important cases when a, b, c, or d become large.

Now consider the following integral equations

$$\begin{aligned} U_1(r, \lambda) &= U_2(r_0) \sin \alpha(r) + U_1(r_0) \cos \alpha(r) \\ &\quad + \int_{r_0}^{r} U_1(s)S(s) \sin [\alpha(r) - \alpha(s)]\ ds \\ U_2(r, \lambda) &= -U_1(r_0) \sin \alpha(r) + U_2(r_0) \cos \alpha(r) \\ &\quad + \int_{r_0}^{r} U_1(s)S(s) \cos [\alpha(r) - \alpha(s)]\ ds, \end{aligned} \qquad (5.4)$$

[12] B. W. Roos and W. C. Sangren, "Spectral Theory of a Pair of First Order Differential Operators," presented at the 574th Meeting of the American Mathematical Society (unpublished).

where $S(r) = G'(r)F^{-1}(r)$. It is easily verified by differentiation of U_1 and U_2 with respect to $\alpha(r)$ that a solution of Eqs. (5.3) satisfies the Eqs. (5.4).

VI. THE $S(r)$ FUNCTION FOR THE DIRAC EQUATION

It is important for later use to determine the order properties of the $S(r)$ function. In particular, we will have to determine whether or not the function $S(r)$ is integrable. $S(r)$ can be written as

$$\begin{aligned} S(r) &= (1/4)[(\lambda a'' + b'')(\lambda a + b)^{-3/2}(\lambda c + d)^{-1/2} \\ &- (\lambda c'' + d'')(\lambda a + b)^{-1/2}(\lambda c + d)^{-3/2} \\ &- (5/4)(\lambda a' + b')^2(\lambda a + b)^{-5/2}(\lambda c + d)^{-1/2} \\ &+ (7/4)(\lambda c' + d')^2(\lambda a + b)^{-1/2}(\lambda c + d)^{-3/2} \\ &- (1/2)(\lambda a' + b')(\lambda c' + d')(\lambda a + b)^{-3/2}(\lambda c + d)^{-3/2}], \end{aligned} \tag{6.1}$$

where

$$\lambda = E, \quad a(r) = r^{2k}, \quad b(r) = [1 - V(r)]r^{2k}$$
$$c(r) = r^{-2k}, \quad d(r) = -[1 + V(r)]r^{2k}.$$

Case 1. Let $r \to \infty$. Various behaviors of the potential function $V(r)$ may be assumed when $r \to \infty$. Consider, for example, the potential

$$V(r) = O(r^{-\delta}) \quad \text{for} \quad r \to \infty.$$

It is not difficult to verify by (6.1) that in this case

$$S(r) = O(r^{-2}) \quad \text{for} \quad \delta \geq 0 \quad \text{and} \quad r \to \infty$$
$$S(r) = O(r^{-2+\delta}) \quad \text{for} \quad \delta \leq 0 \quad \text{and} \quad r \to \infty$$

and that $S(r)$ is integrable $L(r_0, \infty)$. In the same fashion, when $V(r) \to 0$ for $r \to \infty$, $S(r) = O(r^{-2})$ and is again integrable $L(r_0, \infty)$. In numerous cases it can also be shown that $S(r)$ is integrable $L(r_0, \infty)$ when $|V(r)| \to \infty$ for $r \to \infty$.

Case 2. Let $r \to 0$, and let it be assumed that $V(r)$ behaves like $r^{-\gamma}$ for $r \to 0$. It may now be shown that

$$S(r) = O(r^{-2+\gamma}) \quad \text{for} \quad \gamma \geq 0 \quad \text{and} \quad r \to 0$$
$$S(r) = O(r^{-2}) \quad \text{for} \quad \gamma \leq 0 \quad \text{and} \quad r \to 0.$$

Similarly, if $V(r) \to 0$ for $r \to 0$, $S(r) = O(r^{-2})$. $S(r)$ is therefore only integrable $L(0, r_0)$ when $\gamma > 1$. Again, it may be expected that $S(r)$ is integrable $L(0, r_0)$ for many situations in which $|V(r)| \to \infty$.

VII. THE CHARACTER OF $\alpha(r)$ FOR THE RADIAL DIRAC EQUATIONS

The character of $\alpha(r)$ for large values of the parameter λ and at a singularity of the potential function must be investigated before determining the corresponding order properties of the solutions of the integral Eqs. (5.4). In the integral Eqs. (5.4), $\alpha(r)$ is present explicitly in the terms exp $[i\alpha(r)]$ and exp $[-i\alpha(r)]$, which consitute the functions sin $\alpha(r)$ and cos $\alpha(r)$, and is present implicity in the function $S(r)$. By definition, and because $a(r)$ and $c(r)$ are real and positive,

$$\alpha(r, \lambda) = \int_{r_0}^{r} [a(s)c(s)]^{1/2}\left\{\left[\lambda + \frac{b(s)}{a(s)}\right]\left[\lambda + \frac{d(s)}{c(s)}\right]\right\}^{1/2} ds.$$

First, consider the character of $\alpha(r)$ for large λ. For any fixed r interior to the interval under consideration, it may be shown by using the binomial expansion that

$$\alpha(r, \lambda) = \lambda h(r) + g(r) + O(\lambda^{-1}),$$

where

$$h(r) = \int_{r_0}^{r} [a(s)c(s)]^{1/2}\, ds,$$

$$g(r) = \frac{1}{2}\int_{r_0}^{r} \frac{b(s)c(s) + a(s)d(s)}{[a(s)c(s)]^{1/2}}\, ds.$$

Next, consider the behavior of $\alpha(r)$ near either of the end points of the interval, for example, near $r = \infty$. It is assumed that λ is fixed and is given by $\lambda = \sigma + i\tau$, where $\tau > 0$. The behavior of $\alpha(r)$ near $r = \infty$ is necessarily quite varied because it depends strongly on the behavior of the four functions $a(r)$, $b(r)$, $c(r)$, and $d(r)$ in the neighborhood of $r = \infty$. We will not attempt to present a comprehensive discussion of the varied behavior here.[4] However, the following general comments can be made.

When $S(r)$ is integrable and Im $\alpha(r)$ tends to plus or minus infinity, the limit-point case can be expected. If $S(r)$ is integrable, but Im $\alpha(r)$ is bounded, then the limit-circle case can be expected. In the limit-circle case the spectrum is discrete whereas in the limit-point case the spectrum may be either continuous or discrete or both.

Let us now return to the Dirac equations. Here

$$\alpha(r) = \int_{r_0}^{r} \{[\lambda + 1 - V(s)][\lambda - 1 - V(s)]\}^{1/2}\, ds$$

and it may be noted that this $\alpha(r)$ is a particular case of the one discussed in a previous paper.[4] Here, $1 - V(r)$ and $-[1 + V(r)]$ correspond, respectively, to the $q_1(r)$ and $q_2(r)$ of that paper. The results of that study can therefore be used immediately at $r = \infty$. Specifically, consider the situation where $V(r) \to 0$ for $r \to \infty$, as in the case

when $V(r) = O(r^{-\delta})$ and $\delta > 0$ for $r \to \infty$. This is an example of case 9 of the previous paper and it was shown, for real λ and when $\lambda^2 > 1$, that $\alpha(r)$ is real and $\alpha(r) \to +\infty$, whereas if $\lambda^2 < 1$, $\alpha(r)$ will be imaginary and Im $\alpha(r) \to +\infty$. For $\delta < 0$, or more generally, if $|V(r)| \to \infty$, either case 3 or case 4, of the paper mentioned, hold. In either case, for real λ, $\alpha(r)$ is real and $|\alpha(r)| \to \infty$.

When $r \to 0$ and $V(r)$ behaves like $r^{-\gamma}$, where $\gamma > 0$, it is not difficult to show for real λ, that $\alpha(r) = 0(1)$ and is real when $0 < \gamma < 1$, and that $|\alpha(r)| \to \infty$ when $\gamma > 1$. This last behavior can be expected whenever $|rV(r)| \to \infty$. Since $S(r)$ does not satisfy the crucial condition for our study that it is integrable $L(0, r_0)$ when $V(r) \to 0$, we will not consider this situation here.

VIII. THE ORDER PROPERTIES OF $U(r,\lambda)$ IN THE NEIGHBORHOOD OF SINGULAR POINTS

In the case of the Dirac radial wave equations, the order properties of the solutions of the integral Eqs. (5.4) are directly obtained from the results of cases 3, 4, and 9 of reference 4. When Im $\alpha(r) \to \infty$, it was shown that

$$U_1(r, \lambda) = e^{-i\alpha(r)}[M^+(\lambda) + o(1)],$$
$$U_2(r, \lambda) = e^{-i\alpha(r)}[N^+(\lambda) + o(1)],$$

where

$$M^+(\lambda) = -(1/2i)U_2(r_0) + \tfrac{1}{2}U_1(r_0) - \frac{1}{2i}\int_{r_0}^{\infty} e^{i\alpha(s)}S(s)U_1(s)\,ds$$

$$N^+(\lambda) = (1/2i)U_1(r_0) + \tfrac{1}{2}U_2(r_0) + \frac{1}{2}\int_{0}^{\infty} e^{i\alpha(s)}S(s)U_1(s)\,ds.$$

Similarly, when Im $\alpha(r) \to -\infty$,

$$U_1(r, \lambda) = e^{i\alpha(r)}[M^-(\lambda) + o(1)],$$
$$U_2(r, \lambda) = e^{i\alpha(r)}[N^-(\lambda) + o(1)],$$

where

$$M^-(\lambda) = (1/2i)U_2(r_0) + \tfrac{1}{2}U_1(r_0) + \frac{1}{2i}\int_{r_0}^{\infty} e^{-i\alpha(s)}S(s)U_1(s)\,ds$$

$$N^-(\lambda) = -(1/2i)U_1(r_0) + \tfrac{1}{2}U_2(r_0) + \frac{1}{2}\int_{r_0}^{\infty} e^{-i\alpha(s)}S(s)U_1(s)\,ds.$$

From the basic transformations it follows that

$$U_1(r_0) = F(r_0)x_2(r_0),$$
$$U_2(r_0) = -F^{-1}(r_0)x_1(r_0) + G(r_0)x_2(r_0).$$

The associated initial conditions for $U(r, \lambda)$, corresponding to the two linear independent solutions $w(r, \lambda)$ and $v(r, \lambda)$ of the system (1.2), are

$$U_{1v}(r_0) = F(r_0)\cos\beta, \qquad U_{2v}(r_0) = F^{-1}(r_0) \times \sin\beta + G(r_0)\cos\beta$$

$$U_{1w}(r_0) = -F(r_0)\sin\beta, \quad U_{2w}(r_0) = F^{-1}(r_0) \times \cos\beta - G(r_0)\sin\beta.$$

From the definition of $F(r_0)$, $G(r_0)$, it follows that $U_{1v}(r_0)$, $U_{2v}(r_0)$, $U_{1w}(r_0)$, $U_{2w}(r_0)$ are bounded provided λ is such that

$$|\lambda a(r_0) + b(r_0)|\ |\lambda c(r_0) + d(r_0)| \geq \epsilon > 0.$$

Consider those cases where λ is real and $\alpha(r)$ is real; $U_1(r)$ and $U_2(r)$ are therefore bounded. Furthermore, from Eqs. (5.4) it follows that

$$U_1(r, \lambda) = \mu(\lambda)\sin\alpha(r) + \nu(\lambda)\cos\alpha(r) + o(1)$$
$$U_2(r, \lambda) = -\nu(\lambda)\sin\alpha(r) + \mu(\lambda)\cos\alpha(r) + o(1),$$

where

$$\mu(\lambda) = U_2(r_0) + \int_{r_0}^{\infty} U_1(s)S(s)\cos\alpha(s)\,ds$$

$$\nu(\lambda) = U_1(r_0) - \int_{r_0}^{\infty} U_1(s)S(s)\sin\alpha(s)\,ds.$$

The integrals in the last two relations converge uniformly in λ, and hence, $\mu(\lambda)$ and $\nu(\lambda)$ are continuous and bounded functions of λ. It was assumed here that $S(r)$ is integrable $L(r_0, \infty)$. It may now be verified that when λ is real and $r \to \infty$,

$$\mu_w(\lambda)\nu_v(\lambda) - \mu_v(\lambda)\nu_w(\lambda) = 1,$$

and it follows that neither μ_w, ν_w nor μ_v, ν_v can both vanish simultaneously for the same λ.

For $r \to 0$, $U_1(r, \lambda)$ and $U_2(r, \lambda)$ are bounded functions in the Dirac case. It was noted in the previous section that Im $\alpha(r)$ is 0(1) for complex λ when $r \to 0$. Consequently, from Eqs. (5.4) it follows for small r_0 that

$$|U(r, \lambda)| \leq K + \int_{0}^{r_0} |U_1(s, \lambda)|\,|S(s)|\,ds$$

and by a lemma in a previous paper,[4] $U(r, \lambda)$ is bounded provided that $S(r)$ is integrable $L(0, r_0)$.

IX. THE CHARACTER OF $m(\lambda)$

It was noted in Sec. 5, that system (1.2) has a solution $z(r) = w(r) + m(\lambda)V(r)$ which for nonreal values of λ is Lebesgue integrable square at each

end of the interval. Therefore, there exists a solution

$$z_0(r, \lambda) = w(r) + m_0(\lambda) V(r)$$

belonging to $L^2(0, r_0)$ and a solution

$$z_\infty(r, \lambda) = w(r) + m_\infty(\lambda) V(r)$$

belonging to $L^2(r_0, \infty)$. First consider the solution $z_\infty(r, \lambda)$. The functions $w(r)$ and $V(r)$ do not belong to $L^2(r_0, \infty)$. This nonintegrability can be verified by observing that

$$F^{-1}(r) = [\lambda a(r) + b(r)]^{\frac{1}{4}}[\lambda c(r) + d(r)]^{-1/4}$$

does not vanish for any general behavior of $V(r)$ in a manner that will compensate for the growth in the term $\exp[-i\alpha(r)]$ when Im $\alpha(r) \to +\infty$ or the term $\exp[i\alpha(r)]$ when Im $\alpha(r) \to -\infty$. Consequently,[4]

$$w_2^+(r, \lambda) = [F^{-1}(r)M_w^+(\lambda) + o(1)]e^{-i\alpha(r)},$$
$$w_2^-(r, \lambda) = [F^{-1}(r)M_w^-(\lambda) + o(1)]e^{i\alpha(r)},$$
$$v_2^+(r, \lambda) = [F^{-1}(r)M_v^+(\lambda) + o(1)]e^{-i\alpha'(r)},$$
$$v_2^-(r, \lambda) = [F^{-1}(r)M_v^-(\lambda) + o(1)]e^{i\alpha(r)}.$$

The right-hand boundary point at ∞ is thus seen to give rise to the limit-point situation. It may be verified that if $z_\infty(r, \lambda)$ is to be in $L^2(r_0, \infty)$ then

$$m_\infty^*(\lambda) = -M_w^*(\lambda)/M_v^*(\lambda) = -N_w^*(\lambda)/N_v^*(\lambda)$$

and it is well to remember that the quantities $M(\lambda)$ and $N(\lambda)$, which constitute $m_\infty^*(\lambda)$, contain the quantities $S(r)$, $U_1(r_0)$, $U_2(r_0)$. In turn, $U_1(r_0)$ and $U_2(r_0)$ are composed of the quantities $F(r_0)$, $F^-(r_0)$, and $G(r_0)$. From the definitions of $F(r)$ and $G(r)$ [Eqs. (5.3)], it is not difficult to verify that when λ is real and $[\lambda a(r_0) + b(r_0)][\lambda c(r_0) + d(r_0)] > 0$, $F(r_0)$, $F^-(r_0)$, and $G(r_0)$ are real. When λ is real and $[\lambda a(r_0) + b(r_0)][\lambda c(r_0) + d(r_0)] < 0$, it is apparent that $F(r_0)$, $F^-(r_0)$, and $G(r_0)$ are complex with argument $\pi/4$ or $-\pi/4$. Similarly, $S(r_0)$ is real provided $[\lambda a(r_0) + b(r_0)][\lambda c(r_0) + d(r_0)] > 0$ and is imaginary if $[\lambda a(r_0) + b(r_0)][\lambda c(r_0) + d(r_0)] < 0$.

In the Dirac case, $[\lambda a(r) + b(r)][\lambda c(r) + d(r)] = [\lambda + 1 - V(r)][\lambda - 1 - V(r)]$. If $V(r) \to 0$ for $r \to \infty$, then $[\lambda + 1 - V(r)][\lambda - 1 - V(r)]$ behaves like $\lambda^2 - 1$. Consequently, for $\lambda^2 - 1 > 0$ it is possible to choose an r_0 such that $S(r)$, $\alpha(r)$, $F(r)$, and $G(r)$ are simultaneously real for $r > r_0$ and for real λ. It follows from the definitions that

$$m_\infty^+(\lambda) = -\frac{M_w^+(\lambda)}{M_v^+(\lambda)} = -\frac{[\mu_w(\lambda)u_v(\lambda) + \nu_w(\lambda)\nu_v(\lambda)] + i}{\mu_v^2(\lambda) + \nu_v^2(\lambda)} = \overline{m_\infty^-(\lambda)}.$$

Consequently, Im $m^+(\lambda)$ and Im $m^-(\lambda)$ are nonvanishing. If $V(r) \to 0$ for $r \to \infty$ and $\lambda^2 < 1$, $S(r)$ and $\alpha(r)$ are imaginary and $F(r)$ and $G(r)$ are complex with argument $\pi/4$ or $-\pi/4$. The M's and N's which constitute $m_\infty(\lambda)$ can, when multiplied by $e^{i\pi/4}$ or $e^{-i\pi/4}$, be shown to be real, and since the M's and N's are holomorphic in the whole λ plane, the $m^+(\lambda)$ and the $m^-(\lambda)$ are meromorphic functions of λ.

If $V(r)$ behaves like $r^{-\gamma}$ when $r \to 0$ and $\gamma > 0$ then $U(r, \lambda)$ was seen to be bounded for complex λ and $S(r)$ integrable $L(0, r_0)$. However, $S(r)$ is $L(0, r_0)$ only when $\gamma > 1$. By the transformation (5.2) we have $x_2(r) = F^{-1}(r)U_1(r, \lambda)$ and $x_1(r) = -F(r)U_2(r, \lambda) + G(r)U_1(r, \lambda)$. Therefore, as $r \to 0$, $F(r) = O(r^k)$ and $G(r) = O(r^{\gamma+k-1})$. Consequently, $x_1(r) = O(r^k) + O(r^{\gamma+k-1})$ and $x_2(r) = O(r^{-k})$. Although for an integer k either $x_1(r, \lambda)$ or $x_2(r, \lambda)$ is not integrable, the limit-point, limit-circle theory indicates that the relevant issue is whether $\int [a(s)\,|x_2(s)|^2 + c(s)\,|x_1(s)|^2]\,ds$ exists. Now

$$a(x)\,|x_2(r)|^2 = O(1) \quad \text{and} \quad c(r)\,|x_1(r)|^2 = O(1) + O(r^{2(\gamma-1)})$$

and, consequently, for $\gamma > 1$, all solutions of the system (1.2) satisfy the integrability requirements and the limit-circle case is present.

It is not proved here, but it can be shown in a straightforward manner, that if any point in the limit-circle is selected, the associated $m_0(\lambda)$ is a meromorphic function of λ, is real for real λ, and has poles and zeros which are real and simple.[2]

X. SPECTRA FOR THE DIRAC CASE

In the Dirac case, where singularities exist at both ends of the interval, the spectra is determined by the three functions[2]

$$\frac{1}{m_0(\lambda) - m_\infty(\lambda)}, \quad \frac{m_0(\lambda)}{m_0(\lambda) - m_\infty(\lambda)}, \quad \text{and} \quad \frac{m_0(\lambda)m_\infty(\lambda)}{m_0(\lambda) - m_\infty(\lambda)}.$$

In a given λ interval of the real axis, three situations need to be considered. First, $m_0(\lambda)$ and $m_\infty(\lambda)$ may be meromorphic functions. Second, the imaginary parts of $m_0(\lambda)$ and $m_\infty(\lambda)$ may be nonvanishing. Third, only one of the $m_0(\lambda)$ and $m_\infty(\lambda)$ is meromorphic whereas the other one has a nonvanishing imaginary part.

In the first situation where $m_0(\lambda)$ and $m_\infty(\lambda)$ are meromorphic it is apparent that the three functions

are also meromorphic. It then would follow that the spectrum is discrete for the associated λ interval.

For the second situation it is easily shown that the imaginary part of $m_0(\lambda) - m_\infty(\lambda)$ does not vanish. Consequently, since $m_0(\lambda)$ and $m_\infty(\lambda)$ are bounded, it follows that the spectrum is continuous for the associated λ interval.

In the third situation, where either $m_0(\lambda)$ or $m_\infty(\lambda)$ is meromorphic, Im $[m_0(\lambda) - m_\infty(\lambda)]$ tends to a finite limit in the whole interval except at certain discrete points. Therefore, the imaginary parts of the three functions tend to finite limits which can vanish at most at discrete points. The associated spectrum is therefore continuous.

The situation where $V(r) \to 0$ as $x \to \infty$ and $V(r)$ behaves like $r^{-\gamma}$ for $\gamma > 1$ results in the interesting spectral properties from a physical standpoint. In this situation $m_0(\lambda)$ and $m_\infty(\lambda)$ are meromorphic for $\lambda^2 < 1$ and $m_0(\lambda)$ is meromorphic, but Im $m_\infty(\lambda)$ is nonvanishing for $\lambda^2 > 1$. Consequently, the spectrum is discrete for $\lambda^2 < 1$ and continuous for $\lambda^2 > 1$. This result could likewise be expected for numerous other cases when $|V(r)| \to \infty$ as $r \to 0$ and $V(r) \to 0$ as $r \to \infty$. If, however, $|V(r)| \to \infty$ as $r \to \infty$, only a continuous spectrum can be expected.

We have seen that with the formalism described above, the well-known result that the discrete levels are restricted to a region between $\pm mc^2$ is reproducible. However, we required the potential $V(r)$ to behave better than $r^{-\gamma}$ near the origin so that a Coulomb field, for example, was excluded. However, the Coulomb field is just the case that is amenable to a rigorous analysis by series expansion methods and the spectrum is easily determined.

ACKNOWLEDGMENT

It is a pleasure to thank Professor K. Case for his encouragement and a highly stimulating discussion on the topics covered in this paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Behavior of Asymptotically Flat Empty Spaces*

EZRA T. NEWMAN AND THEODORE W. J. UNTI
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received April 27, 1962)

The asymptotic behavior of the Weyl tensor and metric tensor is investigated for probably all asymptotically flat solutions of the empty space Einstein field equations. The systematic investigation utilizes a set of first order differential equations which are equivalent to the empty space Einstein equations. These are solved asymptotically, subject to a condition imposed on a tetrad component of the Riemann tensor Ψ_0 which ensures the approach to flatness at spatial infinity of the space-time. If Ψ_0 is assumed to be an analytic function of a suitably defined radial coordinate, uniqueness of the solutions can be proved. However, this paper makes considerable progress toward establishing a rigorous proof of uniqueness in the nonanalytic case. A brief discussion of the remaining coordinate freedom, with certain topological aspects, is also included.

I. INTRODUCTION

THE problem of the behavior of the metric tensor and Riemann tensor at spatial infinity in an asymptotically flat space-time has recently received considerable attention for a variety of reasons. For example, Trautman and others[1] have used the asymptotic behavior to study conservation of energy and momentum as well as gravitational radiation. Bergmann[2] has been interested in the group properties of the transformation at spatial infinity in order to shed light on quantization problems.

Until recently, most of these approaches were based on reasonable guesses for the behavior of $g_{\mu\nu}$ at infinity. Bondi[3] and Bondi and Sachs[4] have improved this situation with some beautiful theorems on the asymptotic behavior of the metric. Their work, however, contained several restrictive assumptions; the analytic behavior of all expressions as functions of r^{-1} (r is a suitably defined radial coordinate), and certain simplifying topological restrictions.

In the present paper, these assumptions are dropped, and the techniques developed by Newman and Penrose[5] (this reference will be denoted by NP) are applied to the general problem of the asymptotic behavior of the Weyl tensor and metric tensor for probably all asymptotically flat solutions to the empty space Einstein field equations.

In a hyperbolic Riemannian manifold a family of null hypersurfaces may always be introduced. This being done, in Sec. II a coordinate system and tetrad are associated with the given surfaces, bringing the metric into a certain canonical form. Then a set of equations equivalent to the empty space Einstein field equations is shown, and the variables occurring in these equations are defined. From these equations the asymptotic behavior of the field can be investigated, systematically and in detail, subject to a condition of approach to flatness at infinity on the space-time. This condition of asymptotic flatness is not imposed on the metric, as is usually done, but is instead imposed on the empty-space Riemann tensor. More explicitly, using the orthonormal tetrad defined in Sec. II, there are only five independent (complex) tetrad components of the Weyl tensor, which we denote by Ψ_A. One of these components Ψ_0, is specified asymptotically as $O(r^{-5})$,[6] which guarantees the asymptotic flatness of the space.

In Sec. III the field equations are solved asymptotically, with the above condition on Ψ_0. In order to do this, the field equations are divided into three groups. The first group is integrated to find the radial dependence of all the variables, up to an appropriate order of magnitude. Each integration produces an arbitrary function of the remaining three (nonradial) coordinates. The second group of equations sets up relations among these "constants" of integration, which allow most of the functions to be expressed in terms of two basic functions. The third group of equations, part of the Bianchi identities, determines the propagation of components of the Weyl tensor off the hypersurface.

* This work has been supported by the Office of Aerospace Research, U.S. Air Force.

[1] A. Trautman, "Conservation Laws in General Relativity" (to be published); This is also an excellent source of references to recent works on conservation laws.

[2] P. Bergmann, Phys. Rev. **124**, 274 (1961).

[3] H. Bondi, Proc. Roy. Soc. (London) (to be published).

[4] H. Bondi and R. Sachs (private communication).

[5] E. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[6] $f(r, u, x^k) = O(g(r))$ means $|f(r, u, x^k)| < g(r)F(u, x^k)$ for some function F independent of r and for all r, sufficiently large.

In Sec. IV the general class of coordinate transformations which preserves the form of the metric is found. Some of the coordinate transformations change the system of hypersurfaces, others do not. Most of this coordinate freedom is eliminated in achieving a simplification of one of the basic variables.

In Sec. V we summarize the results already obtained and discuss the independent data that can be specified to determine a solution of the field equations. These data consist of a complex function defined from the shear of a null congruence (Bondi[3] calls it the "news function"), a two-dimensional metric tensor for the surface r = const (Bondi chooses this surface to be a sphere), and certain Weyl tensor parts, including Ψ_0. In order to obtain explicit information from the third group of equations (the equations which propagate the Weyl tensor off the hypersurface), one must specify in more detail the dependence of Ψ_0 on r^{-1}, other than $\Psi_0 = O(r^{-5})$. For example, if one writes

$$\Psi_0 = \Psi_0^o r^{-5} + O(r^{-6}), \qquad \partial\Psi_0^o/\partial r = 0,$$

then the propagation of Ψ_0^o off the surface can be explicitly worked out.

If one assumes that Ψ_0 is an analytic function of r^{-1}, it can be shown that the solution of the field equations is uniquely determined by the data that have been set and the field equations themselves.[4] However, it now seems to be generally agreed that the assumption of analyticity is probably not essential in showing that the solution is unique. It would be very desirable to have a rigorous proof of uniqueness in the nonanalytic case. Our work carries this proof almost to completeness; the final step would be to show that the propagation of the nonanalytic Ψ_0 off the surface is unique. Finally, the remaining coordinate freedom is discussed.

In a recent paper, Sachs[7] discusses the geometric interpretation of several of the terms and first order differential equations listed in Sec. II of this paper.

The range and summation conventions used here are: lower case Greek indices 1, 2, 3, 4; lower case Latin indices 3, 4.

II. THE FIELD EQUATIONS

In this section a special coordinate system, with an associated tetrad, is constructed as in Robinson-Trautman[8] and NP. The pertinent results of NP are presented, including a set of equations which are equivalent to the Einstein empty space field equations.

We begin by introducing a family of null hypersurfaces, which is always possible in a normal hyperbolic Riemannian space. (In flat space-time a particular family might be the null cones emanating from a timelike world line.) They may be designated by a parameter u = const, so that

$$g^{\mu\nu}u_{,\mu}u_{,\nu} = 0. \tag{1}$$

The first tetrad vector will be chosen to be orthogonal to the hypersurfaces,

$$l_\alpha = u_{,\alpha}. \tag{2}$$

Since the hypersurfaces are null, the vectors l_α will also be tangent to a family of curves that lie within the surfaces. These curves are null geodesics; their tangent vectors l^α satisfy

$$l^\mu{}_{;\nu}l^\nu = 0.$$

Using this geometric approach, it is convenient to take u as coordinate x^1. Then Eq. (2) becomes

$$l_\alpha = \delta^1_\alpha.$$

An affine parameter, defined up to a linear transformation, can be associated with the null geodesics lying in the hypersurfaces. This affine parameter r will be the coordinate x^2. The two remaining coordinates x^k will label the geodesics on each hypersurface u = const.

The tangents to the geodesics are given by

$$l^\mu = dx^\mu/dr = g^{\mu\nu}u_{,\nu} = g^{\mu 1}. \tag{3}$$

With our choice of coordinates, r being the affine parameter x^2, Eq. (3) becomes

$$l^\mu = \delta^\mu_2 = g^{\mu 1}.$$

Hence, the metric assumes the form

$$g^{\mu\nu} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & g^{22} & g^{2k} & \\ 0 & g^{2k} & g^{mn} & \\ 0 & & & \end{bmatrix}. \tag{4}$$

In addition to the vector l^μ, we define another null vector n^μ normalized by $l_\mu n^\mu = 1$, and two unit spacelike vectors ζ^μ and ρ^μ orthogonal to l^μ, n^μ, and each other. Instead of the real spacelike vectors ρ^μ and ζ^μ it is convenient to use the following complex vectors

$$m^\mu = (\zeta^\mu - i\rho^\mu)/\sqrt{2}$$
$$\bar{m}^\mu = (\zeta^\mu + i\rho^\mu)/\sqrt{2}.$$

[7] R. Sachs, Proc. Roy. Soc. (London) **264**, 309 (1961).
[8] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **265**, 463 (1962).

The four tetrad vectors l^μ, n^μ, m^μ, and $\bar{m}^\mu$ are null and satisfy the following orthonormality relations

$$l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1,$$
$$l_\mu l^\mu = n_\mu n^\mu = m_\mu m^\mu = \bar{m}_\mu \bar{m}^\mu = l_\mu m^\mu = l_\mu \bar{m}^\mu = n_\mu m^\mu = n_\mu \bar{m}^\mu = 0. \quad (5)$$

In order to satisfy Eqs. (5), the vectors m^μ and n^μ must have the form

$$m^\mu = \omega \delta_2^\mu + \xi^k \delta_k^\mu$$
$$n^\mu = \delta_1^\mu + U\delta_2^\mu + X^k \delta_k^\mu,$$

where ω, ξ^k, U, and X^k are arbitrary functions of the coordinates. The completeness relation

$$g^{\mu\nu} = l^\mu n^\nu + n^\mu l^\nu - m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu$$

is a consequence of the orthonormality relations Eqs. (5). It permits us to express the metric in terms of the tetrad components;

$$g^{22} = 2(U - \omega\bar{\omega})$$
$$g^{2k} = X^k - (\xi^k \bar{\omega} + \bar{\xi}^k \omega) \quad (6)$$
$$g^{mn} = -(\xi^m \bar{\xi}^n + \bar{\xi}^m \xi^n).$$

There remains the following freedom in the choice of the tetrad; the spatial rotation

$$l^{\mu'} = l^\mu, \quad n^{\mu'} = n^\mu, \quad m^{\mu'} = m^\mu e^{iC}, \quad (C \text{ real}) \quad (7)$$

which depends on one parameter, and a 2-parameter group, the so-called null rotations

$$l^{\mu'} = l^\mu, \quad n^{\mu'} = n^\mu + \bar{B}m^\mu + B\bar{m}^\mu + B\bar{B}l^\mu \quad (8)$$
$$m^{\mu'} = m^\mu + Bl^\mu,$$

(B complex).

These tetrad transformations leave the direction of l_μ fixed. Later (Sec. IV) we will be interested in coordinate transformations which necessitate the choice of an entirely new family of hypersurfaces and tetrad system. But at this point the restricted 3-parameter group Eqs. (7) and (8) can be used to simplify some of the calculations. We will demand parallel propagation of the tetrad along the geodesics, restricting the functions C and B to be independent of r.

With the above preparations, we are in a position to derive the asymptotic behavior of the metric and Riemann tensors for empty space. In NP a set of equations was developed which are equivalent to the Einstein empty space field equations. Although many more in number, these equations are of first order, most of them being linear. They are essentially linear combinations of the equations for the Riemann tensor expressed in terms of either Ricci rotation coefficients, or in terms of the spinor affine connection, many of the equations having straightforward geometric interpretations.[7]

The field equations [using our coordinate conditions, Eq. (4)] may be divided into three groups. The first group is distinguished by a simple $D \equiv \partial/\partial r$ derivative. Integration of these equations yields the r dependence of the variables, with (in general) a "constant" of integration, a function of u, x^3, x^4, for each variable.

I. Radial equations:

$$D\xi^i = \rho\xi^i + \sigma\bar{\xi}^i, \quad (9a)$$
$$D\omega = \rho\omega + \sigma\bar{\omega} - (\bar{\alpha} + \beta), \quad (9b)$$
$$DX^i = (\bar{\alpha} + \beta)\bar{\xi}^i + (\alpha + \bar{\beta})\xi^i, \quad (9c)$$
$$DU = (\bar{\alpha} + \beta)\bar{\omega} + (\alpha + \bar{\beta})\omega - (\gamma + \bar{\gamma}), \quad (9d)$$
$$D\rho = \rho^2 + \sigma\bar{\sigma}, \quad (9e)$$
$$D\sigma = 2\rho\sigma + \Psi_0, \quad (9f)$$
$$D\tau = \tau\rho + \bar{\tau}\sigma + \Psi_1, \quad (9g)$$
$$D\alpha = \alpha\rho + \beta\bar{\sigma}, \quad (9h)$$
$$D\beta = \beta\rho + \alpha\sigma + \Psi_1, \quad (9i)$$
$$D\gamma = \tau\alpha + \bar{\tau}\beta + \Psi_2, \quad (9j)$$
$$D\lambda = \lambda\rho + \mu\bar{\sigma}, \quad (9k)$$
$$D\mu = \mu\rho + \lambda\sigma + \Psi_2, \quad (9l)$$
$$D\nu = \tau\lambda + \bar{\tau}\mu + \Psi_3, \quad (9m)$$
$$D\Psi_1 - \bar{\delta}\Psi_0 = 4\rho\Psi_1 - 4\alpha\Psi_0, \quad (9n)$$
$$D\Psi_2 - \bar{\delta}\Psi_1 = 3\rho\Psi_2 - 2\alpha\Psi_1 - \lambda\Psi_0, \quad (9o)$$
$$D\Psi_3 - \bar{\delta}\Psi_2 = 2\rho\Psi_3 - 2\lambda\Psi_1, \quad (9p)$$
$$D\Psi_4 - \bar{\delta}\Psi_3 = \rho\Psi_4 + 2\alpha\Psi_3 - 3\lambda\Psi_2, \quad (9q)$$

The second group have more complicated derivatives,

$$\Delta = U(\partial/\partial r) + (\partial/\partial u) + X^k(\partial/\partial x^k)$$
$$\delta = \omega(\partial/\partial r) + \xi^k(\partial/\partial x^k).$$

From this group we obtain relations between the integration "constants" found above.

II. Nonradial equations:

$$\delta X^i - \Delta\xi^i = (\mu + \bar{\gamma} - \gamma)\xi^i + \bar{\lambda}\bar{\xi}^i, \quad (10a)$$
$$\delta\bar{\xi}^i - \bar{\delta}\xi^i = (\bar{\beta} - \alpha)\xi^i + (\bar{\alpha} - \beta)\bar{\xi}^i, \quad (10b)$$
$$\delta\bar{\omega} - \bar{\delta}\omega = (\bar{\beta} - \alpha)\omega + (\bar{\alpha} - \beta)\bar{\omega} + (\mu - \bar{\mu}), \quad (10c)$$
$$\delta U - \Delta\omega = (\mu + \bar{\gamma} - \gamma)\omega + \bar{\lambda}\bar{\omega} - \bar{\nu}, \quad (10d)$$

$$\Delta\lambda - \bar{\delta}\nu = 2\alpha\nu + (\bar{\gamma} - 3\gamma - \mu - \bar{\mu})\lambda - \Psi_4, \quad (10e)$$

$$\delta\rho - \bar{\delta}\sigma = (\beta + \bar{\alpha})\rho + (\bar{\beta} - 3\alpha)\sigma - \Psi_1, \quad (10f)$$

$$\delta\alpha - \bar{\delta}\beta = \mu\rho - \lambda\sigma - 2\alpha\beta + \alpha\bar{\alpha} + \beta\bar{\beta} - \Psi_2, \quad (10g)$$

$$\delta\lambda - \bar{\delta}\mu = (\alpha + \bar{\beta})\mu + (\bar{\alpha} - 3\beta)\lambda - \Psi_3, \quad (10h)$$

$$\delta\nu - \Delta\mu = \gamma\mu - 2\nu\beta + \bar{\gamma}\mu + \mu^2 + \lambda\bar{\lambda}, \quad (10i)$$

$$\delta\gamma - \Delta\beta = \tau\mu - \sigma\nu + (\mu - \gamma + \bar{\gamma})\beta + \bar{\lambda}\alpha, \quad (10j)$$

$$\delta\tau - \Delta\sigma = 2\tau\beta + (\bar{\gamma} + \mu - 3\gamma)\sigma + \bar{\lambda}\rho, \quad (10k)$$

$$\Delta\rho - \bar{\delta}\tau = (\gamma + \bar{\gamma} - \bar{\mu})\rho - 2\alpha\tau - \lambda\sigma - \Psi_2 \quad (10l)$$

$$\Delta\alpha - \bar{\delta}\gamma = \rho\nu - \tau\lambda - \lambda\beta + (\bar{\gamma} - \gamma - \bar{\mu})\alpha - \Psi_3, \quad (10m)$$

The third group, derived from the Bianchi identities, determines the propagation of the tetrad components of the Weyl tensor in the u-direction, from null surface to null surface.

III. The u-derivative equations:

$$\Delta\Psi_0 - \delta\Psi_1 = [4\gamma - \mu]\Psi_0 - [4\tau + 2\beta]\Psi_1 + 3\sigma\Psi_2 \quad (11a)$$

$$\Delta\Psi_1 - \delta\Psi_2 = \nu\Psi_0 + [2\gamma - 2\mu]\Psi_1 - 3\tau\Psi_2 + 2\sigma\Psi_3 \quad (11b)$$

$$\Delta\Psi_2 - \delta\Psi_3 = 2\nu\Psi_1 - 3\mu\Psi_2 + [-2\tau + 2\beta]\Psi_3 + \sigma\Psi_4 \quad (11c)$$

$$\Delta\Psi_3 - \delta\Psi_4 = 3\nu\Psi_2 - [2\gamma + 4\mu]\Psi_3 + [-\tau + 4\beta]\Psi_4 \quad (11d)$$

The complex functions ρ, σ, α, $\cdots$ are defined either in terms of the Ricci rotation coefficients, or the spinor affine connection. In NP they are named "spin-coefficients." Their definitions in terms of the tetrad are:

$$\rho = l_{\mu;\nu}m^\mu\bar{m}^\nu, \quad (12a)$$

$$\sigma = l_{\mu;\nu}m^\mu m^\nu, \quad (12b)$$

$$\tau = l_{\mu;\nu}m^\mu n^\nu, \quad (12c)$$

$$\alpha = \tfrac{1}{2}(l_{\mu;\nu}n^\mu\bar{m}^\nu - m_{\mu;\nu}\bar{m}^\mu\bar{m}^\nu), \quad (12d)$$

$$\beta = \tfrac{1}{2}(l_{\mu;\nu}n^\mu m^\nu - m_{\mu;\nu}\bar{m}^\mu m^\nu), \quad (12e)$$

$$\gamma = \tfrac{1}{2}(l_{\mu;\nu}n^\mu n^\nu - m_{\mu;\nu}\bar{m}^\mu n^\nu), \quad (12f)$$

$$\lambda = -n_{\mu;\nu}\bar{m}^\mu\bar{m}^\nu, \quad (12g)$$

$$\mu = n_{\mu;\nu}\bar{m}^\mu m^\nu, \quad (12h)$$

$$\nu = -n_{\mu;\nu}\bar{m}^\mu n^\nu. \quad (12i)$$

In NP it is shown that, in this coordinate system, $\tau = \bar{\alpha} + \beta$.

As mentioned above, many of the spin-coefficients have simple geometric meanings, such as the shear and divergence of the null geodesic congruence, etc. (See reference 7 and NP, Sec. IV).

The Ψ_A are the tetrad components of the Weyl tensor (or empty space Riemann tensor);

$$\Psi_0 = -C_{\alpha\beta\gamma\delta}l^\alpha m^\beta l^\gamma m^\delta, \quad (12j)$$

$$\Psi_1 = -C_{\alpha\beta\gamma\delta}l^\alpha n^\beta l^\gamma m^\delta, \quad (12k)$$

$$\Psi_2 = -\tfrac{1}{2}C_{\alpha\beta\gamma\delta}(l^\alpha n^\beta l^\gamma n^\delta + l^\alpha n^\beta m^\gamma\bar{m}^\delta), \quad (12l)$$

$$\Psi_3 = C_{\alpha\beta\gamma\delta}l^\alpha n^\beta n^\gamma\bar{m}^\delta, \quad (12m)$$

$$\Psi_4 = -C_{\alpha\beta\gamma\delta}n^\alpha\bar{m}^\beta n^\gamma\bar{m}^\delta. \quad (12n)$$

We wish to solve the field equations with the condition that all the Ψ_A approach zero as r approaches infinity. It was shown in NP that an extremely weak (if not the weakest) assumption to ensure this boundary condition is $\Psi_0 = O(r^{-5})$. Here we will adopt the slightly stronger condition[9]

$$\Psi_0 = \Psi_0^\circ r^{-5} + O(r^{-6}). \quad (13)$$

(In the remainder of the paper, a zero superscript indicates the variable is independent of r). It will also be necessary to make assumptions on the derivatives of Ψ_0;

$$D\Psi_0 = \partial\Psi_0/\partial r = -5\Psi_0^\circ r^{-6} + O(r^{-7}), \quad (14)$$

and "uniform smoothness,"

$$\begin{aligned} &d_i\Psi_0 = (d_i\Psi_0^\circ)r^{-5} + O(r^{-6}), \cdots, \\ &\quad d_id_jd_md_n\Psi_0 = (d_id_jd_md_n\Psi_0^\circ)r^{-5} + O(r^{-6}), \\ &d_iD\Psi_0 = -5(d_i\Psi_0^\circ)r^{-6} + O(r^{-7}), \cdots, \\ &\quad d_id_jd_kD\Psi_0 = -5(d_id_jd_k\Psi_0^\circ)r^{-6} + O(r^{-7}), \\ &\text{where } (i, j, k, m, n = 3, 4) \qquad d_i \equiv \partial/\partial x^i. \end{aligned} \quad (15)$$

Needed in the next section is the asymptotic behavior of each of the spin-coefficients, tetrad components, and remaining Weyl tensor components. This information was derived in NP, using some powerful theorems of Levinson and Cottington.[10]

$$\begin{aligned} &\rho = -r^{-1} + O(r^{-2}) \qquad \xi^k = O(r^{-1}) \\ &\sigma = O(r^{-2}) \qquad X^k, \omega = O(1) \\ &\alpha, \beta, \lambda, \mu, \tau = O(r^{-1}) \qquad U = O(r) \\ &\nu, \gamma = O(1) \end{aligned} \quad (16)$$

$$\begin{aligned} \Psi_1 &= O(r^{-4}) \\ \Psi_2 &= O(r^{-3}) \\ \Psi_3 &= O(r^{-2}) \\ \Psi_4 &= O(r^{-1}) \end{aligned} \quad (17)$$

[9] We wish to study the propagation of Ψ° in the u "direction." Such a study would be more difficult under the general assumption $\Psi_0 = O(r^{-5})$.

[10] E. Coddington and N. Levinson, *Theory of Ordinary Differential Equations* (McGraw-Hill Book Company, Inc., New York, 1955), p. 103.

III. SOLUTIONS

This section outlines the method of solving the empty space field equations listed in Sec. II. The first-order radial differential equations [all of which are linear except Eqs. (9e, f), and these can be made linear by increasing the order of differentiation] fall naturally into certain subgroups, which are solved simultaneously. All the theorems on linear differential equations, existence, uniqueness, etc., thus apply. By integrating these equations, the r dependence of all the variables is easily obtained. In general there is a "constant" of integration, a function of the remaining coordinates, u, x^3, x^4, in the solution of each of these differential equations. Relations between these integration "constants" are obtained from the equations having compound derivatives $\delta = \omega(\partial/\partial r) + \xi^k(\partial/\partial x^k)$, $\Delta = U(\partial/\partial r) + (\partial/\partial u) + X^k(\partial/\partial x^k)$, by setting the coefficients of powers of r^{-1} independently equal to zero.

The pair of equations (9e, f)

$$D\rho = \rho^2 + \sigma\bar{\sigma},$$
$$D\sigma = 2\rho\sigma + \Psi_0$$

with the asymptotic behavior, Eqs. (16), $\rho = -r^{-1} + O(r^{-2})$, $\sigma = O(r^{-2})$, can be integrated for the complete r dependence of ρ and σ, up to $O(r^{-5})$. The method of solution uses formal integration of the order of magnitude symbols (which is permissable), but not differentiation of them.[11]

In the following, $g(r)$ and $h(r)$ will be used generically. We have

$$\rho = -r^{-1} + g(r),$$
$$\sigma = h(r), \quad \text{where} \quad g, h = O(r^{-2}).$$

By Eqs. (9e, f),

$$Dg + 2r^{-1}g = g^2 + h\bar{h} = O(r^{-4})$$
$$Dh + 2r^{-1}h = 2hg + \Psi_0 = O(r^{-4}).$$

The solution will be determined in steps. The first step is to find ρ and σ up to $O(r^{-3})$. Hence, the equations to be solved at this stage are

$$Dg + 2r^{-1}g = O(r^{-4})$$
$$Dh + 2r^{-1}h = O(r^{-4})$$

with the immediate solution,[12]

$$g = e^{-\int 2dr/r}\left\{\int e^{\int 2dr/r} O(r^{-4})\, dr + \rho^0\right\},$$
$$= r^{-2}\left\{\int O(r^{-2})\, dr + \rho^0\right\} = \rho^0 r^{-2} + O(r^{-3}),$$

and with a similar solution for $h(r)$. Therefore

$$\rho = -r^{-1} + \rho^0 r^{-2} + O(r^{-3}), \qquad \rho^0 = \rho^0(u, x^k)$$
$$\sigma = \sigma^0 r^{-2} + O(r^{-3}), \qquad \sigma^0 = \sigma^0(u, x^k).$$

By a coordinate transformation $r' = r - \rho^0(u, x^k)$, the ρ^0/r'^2 term is eliminated. Dropping primes, we have

$$\rho = -r^{-1} + O(r^{-3})$$
$$\sigma = \sigma^0 r^{-2} + O(r^{-3}).$$

Again letting

$$\rho = -r^{-1} + g(r)$$
$$\sigma = \sigma^0 r^{-2} + h(r),$$

where this time

$$g(r), h(r) = O(r^{-3}), \tag{18}$$

and grouping all terms of $O(r^{-5})$ under the order symbol, we have,

$$Dg + 2r^{-1}g = \sigma^0\bar{\sigma}^0 r^{-4} + O(r^{-5})$$
$$Dh + 2r^{-1}h = O(r^{-5})$$

with solutions

$$g = r^{-2}\left\{\int r^2[\sigma^0\bar{\sigma}^0 r^{-4} + O(r^{-5})]\, dr + C_1\right\}$$
$$h = r^{-2}\left\{\int r^2 O(r^{-5})\, dr + C_2\right\},$$
$$g = C_1 r^{-2} - \sigma^0\bar{\sigma}^0 r^{-3} + O(r^{-4})$$
$$h = C_2 r^{-2} + O(r^{-4}).$$

The conditions (18), $g, h = O(r^{-3})$, imply that $C_1 = C_2 = 0$. Hence

$$\rho = -r^{-1} - \sigma^0\bar{\sigma}^0 r^{-3} + O(r^{-4})$$
$$\sigma = \sigma^0 r^{-2} + O(r^{-4}). \tag{19}$$

Repeating the process, it is found that

$$\rho = -r^{-1} - \sigma^0\bar{\sigma}^0 r^{-3} + O(r^{-5})$$
$$\sigma = \sigma^0 r^{-2} + (\bar{\sigma}^0\sigma^{0\prime} - \tfrac{1}{2}\Psi_0^0)r^{-4} + O(r^{-5}), \tag{20}$$

which are the desired asymptotic expressions for ρ, σ. At this point it should be observed that, had the more general assumption $\Psi_0 = O(r^{-5})$ been made, we would have obtained explicitly the forms (19) but not the extra term in Eq. (20). It is evident

---

[11] A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 7. The theorem is

$$\int_r^\infty 0(f(r'))\, dr' = 0\left(\int_r^\infty |f(r')|\, dr'\right), \quad \text{as} \quad r \to \infty.$$

[12] H. Margenau and G. Murphy, *The Mathematics of Physics and Chemistry* (D. Van Nostrand Company, Inc., Princeton, New Jersey, 1943), p. 42, Eq. (2-6).

that this extra term results from the specialization $\Psi_0 = \Psi_0^\circ r^{-5} + O(r^{-6})$. On the other hand, the general term "$O(r^{-6})$" of Ψ_0 prevents us from obtaining forms more explicit than Eqs. (20), as follows: In accordance with Eqs. (20) let

$$\rho = -r^{-1} - \sigma^\circ\bar{\sigma}^\circ r^{-3} + g(r)$$
$$\sigma = \sigma^\circ r^{-2} + Qr^{-4} + h(r),$$

where

$$g(r), h(r) = O(r^{-5}),$$

and

$$Q \equiv \bar{\sigma}^\circ\sigma^{\circ 2} - \tfrac{1}{2}\Psi_0^\circ.$$

By Eq. (9f)

$$Dh + \frac{2}{r}h = \frac{2\sigma^\circ g}{r^2} - \frac{2\sigma^\circ\bar{\sigma}^\circ Q}{r^7} - \frac{2\sigma^\circ\bar{\sigma}^\circ h}{r^3} + \frac{2gQ}{r^4} + 2gh + O(r^{-6}),$$

where the final term $O(r^{-6})$ comes from Ψ_0. However, the five preceding terms are seen to be absorbed into $O(r^{-6})$, so that

$$Dh + (2/r)h = O(r^{-6}).$$

Integration gives (the integration constant is zero),

$$h = r^{-2}\int r^2 O(r^{-6})\,dr = r^{-2}O(r^{-3}) = O(r^{-5}),$$

and no new information is obtained. Similarly Eq. (9e) fails to produce further explicit terms.

Equation (9a) is solved in the same manner as Eqs. (9e, f). The condition (16), $\xi^k = O(r^{-1})$ is translated into

$$\xi^k = g(r), \qquad g(r) = O(r^{-1}).$$

Putting this into Eq. (9a) yields

$$Dg + r^{-1}g = O(r^{-3})$$
$$\xi^k = g = r^{-1}\left\{\xi^{\circ k} + \int rO(r^{-3})\,dr\right\} = \xi^{\circ k}r^{-1} + O(r^{-2}).$$

The higher order terms of ξ^k are obtained similarly.

The next group of related equations involve α, β, ω, and Ψ_1. They are Eqs. (9h, i), Eq. (9b), and Eq. (9n). Ψ_1 is known to be $O(r^{-4})$ by Eqs. (17), which allows the solutions of α and β from Eqs. (9h, i) up to $O(r^{-3})$. Upon performing these integrations using the techniques shown above, it is found that

$$\alpha = \alpha^\circ r^{-1} + \bar{\sigma}^\circ\bar{\alpha}^\circ r^{-2} + O(r^{-3})$$
$$\beta = \beta^\circ r^{-1} + \sigma^\circ\bar{\beta}^\circ r^{-2} + O(r^{-3}).$$

By use of the null rotation Eq. (8), we may set

$$\tau^\circ = \bar{\alpha}^\circ + \beta^\circ = 0. \tag{21}$$

Therefore, $\tau = \bar{\alpha} + \beta = O(r^{-3})$, which is used in the integration of

$$D\omega = \rho\omega + \sigma\bar{\omega} - \tau.$$

Here the r dependence is found by considering the coupled equations,

$$D\omega = \rho\omega + \sigma\bar{\omega} - \tau$$
$$D\bar{\omega} = \rho\bar{\omega} + \bar{\sigma}\omega - \bar{\tau}.$$

Inserting orders of magnitude we get

$$D\omega + r^{-1}\omega = \sigma^\circ r^{-2}\bar{\omega} + O(r^{-3}),$$
$$D\bar{\omega} + r^{-1}\bar{\omega} = \bar{\sigma}^\circ r^{-2}\omega + O(r^{-3}).$$

Formal integrations give

$$\omega = r^{-1}\left\{\int \sigma^\circ r^{-1}\bar{\omega}\,dr + O(r^{-1}) + K_1\right\}$$
$$\bar{\omega} = r^{-1}\left\{\int \bar{\sigma}^\circ r^{-1}\omega\,dr + O(r^{-1}) + \bar{K}_1\right\}. \tag{22}$$

Using the information $\omega = O(1)$, $\bar{\omega} = O(1)$, it is seen that

$$\omega = K_1 r^{-1} + O(r^{-2}) + \sigma^\circ r^{-1}\int O(r^{-1})\,dr$$
$$= K_1 r^{-1} + O(r^{-2}) + O(\ln r/r) = O(\ln r/r)$$

and

$$\bar{\omega} = O(\ln r/r).$$

When these orders are again inserted in Eq. (22), the results are

$$\omega = K_1 r^{-1} + O(r^{-2}) + O(\ln r/r^2) = O(r^{-1})$$
$$\bar{\omega} = O(r^{-1}).$$

Finally, when this information is fed into Eq. (22) for the third time, we arrive at the desired result,

$$\omega = \omega^\circ r^{-1} + O(r^{-2}).$$

This is sufficient to derive the r dependence of Ψ_1 from Eq. (9n). We have

$$\bar{\delta}\Psi_0 = \bar{\omega}\Psi_{0,2} + \bar{\xi}^k\Psi_{0,k} = \bar{\xi}^{\circ k}\Psi^\circ_{0,k}r^{-6} + O(r^{-7})$$

by the initial assumptions on Ψ_0, Eqs. (14) and (15). Equation (9n) becomes

$$D\Psi_1 + 4\Psi_1 r^{-1} = \bar{\xi}^{\circ k}\Psi^\circ_{0,k}r^{-6} - 4\alpha^\circ\Psi_0^\circ r^{-6} + O(r^{-7})$$

where $\Psi_1 = O(r^{-4})$ has been used. Integration immediately gives

$$\Psi_1 = r^{-4}\left\{\int r^4[(-4\alpha^o\Psi_0^o + \bar{\xi}^{ok}\Psi_{0,k}^o)r^{-6} + O(r^{-7})]\,dr + \Psi_1^o\right\}$$

$$\Psi_1 = \frac{\Psi_1^o}{r^4} + \frac{4\alpha^o\Psi_0^o - \bar{\xi}^{ok}\Psi_{0,k}^o}{r^5} + O(r^{-6}).$$

This information is sufficient to complete the integrations of Eqs. (9h, i) and Eq. (9b).

The asymptotic r dependence of all the variables may be calculated by this method. The results are as follows:

$$\Psi_0 = \Psi_0^o r^{-5} + O(r^{-6}), \tag{23a}$$

$$\Psi_1 = \Psi_1^o r^{-4} + (4\alpha^o\Psi_0^o - \bar{\xi}^{ok}\Psi_{0,k}^o)r^{-5} + O(r^{-6}), \tag{23b}$$

$$\Psi_2 = \Psi_2^0 r^{-3} + (2\alpha^o\Psi_1^o - \bar{\xi}^{ok}\Psi_{1,k}^o)r^{-4} + O(r^{-5}), \tag{23c}$$

$$\Psi_3 = \Psi_3^o r^{-2} - \bar{\xi}^{ok}\Psi_{2,k}^o r^{-3} + O(r^{-4}), \tag{23d}$$

$$\Psi_4 = \Psi_4^o r^{-1} - (2\alpha^o\Psi_3^o + \bar{\xi}^{ok}\Psi_{3,k}^o)r^{-2} + O(r^{-3}). \tag{23e}$$

$$\rho = -r^{-1} - \sigma^o\bar{\sigma}^o r^{-3} + O(r^{-5}), \tag{24a}$$

$$\sigma = \sigma^o r^{-2} + (\bar{\sigma}^o\sigma^{o\cdot} - \tfrac{1}{2}\Psi_0^o)r^{-4} + O(r^{-5}), \tag{24b}$$

$$\alpha = \alpha^o r^{-1} + \bar{\sigma}^o\bar{\alpha}^o r^{-2} + \sigma^o\bar{\sigma}^o\alpha^o r^{-3} + O(r^{-4}), \tag{24c}$$

$$\beta = -\bar{\alpha}^o r^{-1} - \sigma^o\alpha^o r^{-2} - (\sigma^o\bar{\sigma}^o\bar{\alpha}^o + \tfrac{1}{2}\Psi_1^o)r^{-3} + O(r^{-4}), \tag{24d}$$

$$\tau = -(1/2r^3)\Psi_1^o + (1/6r^4)(2\bar{\xi}^{ok}\Psi_{0,k}^o - 8\alpha^o\Psi_0^o + \sigma^o\bar{\Psi}_1^o) + O(r^{-5}), \tag{24e}$$

$$\lambda = \lambda^o r^{-1} - \bar{\sigma}^o\mu^o r^{-2} + (1/r^3)(\sigma^o\bar{\sigma}^o\lambda^o + \tfrac{1}{2}\bar{\sigma}^o\Psi_2^o) + O(r^{-4}), \tag{24f}$$

$$\mu = \mu^o r^{-1} - (\sigma^o\lambda^o + \Psi_2^o)r^{-2} + (\sigma^o\bar{\sigma}^o\mu^o - \alpha^o\Psi_1^o + \tfrac{1}{2}\bar{\xi}^{ok}\Psi_{1,k}^o)r^{-3} + O(r^{-4}), \tag{24g}$$

$$\gamma = \gamma^o - (1/2r^2)\Psi_2^o + (1/r^3)(\tfrac{1}{3}\bar{\xi}^{ok}\Psi_{1,k}^o - \tfrac{1}{6}\bar{\alpha}^o\bar{\Psi}_1^o - \tfrac{1}{2}\alpha^o\Psi_1^o) + O(r^{-4}), \tag{24h}$$

$$\nu = \nu^o - (1/r)\Psi_3^o + (1/2r^2)\bar{\xi}^{ok}\Psi_{2,k}^o + O(r^{-3}). \tag{24i}$$

$$U = -(\gamma^o + \bar{\gamma}^o)r + U^o - (1/2r)(\Psi_2^o + \bar{\Psi}_2^o) + (1/6r^2)[(\bar{\xi}^{ok}\Psi_{1,k}^o + \xi^{ok}\bar{\Psi}_{1,k}^o) - 2(\alpha^o\Psi_1^o + \bar{\alpha}^o\bar{\Psi}_1^o)] + O(r^{-3}), \tag{25a}$$

$$X^k = (1/6r^3)(\Psi_1^o\bar{\xi}^{ok} + \bar{\Psi}_1^o\xi^{ok}) + O(r^{-4}), \tag{25b}$$

$$\xi^k = \xi^{ok}r^{-1} - \sigma^o\bar{\xi}^{ok}r^{-2} + \sigma^o\bar{\sigma}^o\xi^{ok}r^{-3} + O(r^{-4}), \tag{25c}$$

$$\omega = \omega^o r^{-1} - (1/r^2)(\sigma^o\bar{\omega}^o + \tfrac{1}{2}\Psi_1^o) + O(r^{-3}). \tag{25d}$$

Further terms of most of these functions can be easily calculated.

In the above we have tacitly used a coordinate transformation (which will be discussed) to eliminate the "constant" X^{ok}.

Another coordinate transformation can be made at this point to simplify the remaining calculations. The metric is in the form Eq. (4), with

$$g^{mn} = -(\xi^m\bar{\xi}^n + \bar{\xi}^m\xi^n) = -(\xi^{om}\bar{\xi}^{on} + \bar{\xi}^{om}\xi^{on})r^{-2} + \cdots .$$

Under the remaining coordinate transformations, the leading term of g^{mn} transforms as a 2×2 metric, hence it may be reduced to a conformally flat metric.[13] (Robinson and Trautman[8] make g^{mn} conformally flat. We can only do this for the leading term of g^{mn}). Up to $O(r^{-3})$, $g^{33} = g^{44}$, and $g^{34} = g^{43} = 0$. However, since

$$g^{33} = -2\xi^{o3}\bar{\xi}^{o3}r^{-2} + O(r^{-3}),$$

$$g^{34} = (\xi^{o3}\bar{\xi}^{o4} + \bar{\xi}^{o3}\xi^{o4})r^{-2} + O(r^{-3}),$$

$$g^{44} = -2\xi^{o4}\bar{\xi}^{o4}r^{-2} + O(r^{-3}),$$

we have $\xi^{o3} = -i\xi^{o4} = P(u, x^k)$.

At this point the remaining coordinate freedom for the variables x^3 and x^4 is,[14]

$$x^{3\prime} + ix^{4\prime} = f(x^3 + ix^k, u). \tag{26}$$

The next step is to solve the nonradial equations, which allow us to express most of the "constants" of integration α^o, ω^o, λ^o, $\cdots$ as functions of σ^o and P.

Equation (10 l) will be done as an example of the procedure. Explicitly it takes the form,

$$\rho_{,1} + U\rho_{,2} + X^k\rho_{,k} - \bar{\omega}\tau_{,2} - \bar{\xi}^k\tau_{,k} + (\bar{\mu} - \gamma - \bar{\gamma})\rho + 2\alpha\tau + \lambda\sigma + \Psi_2 = 0.$$

When the differentiations with respect to r are performed and the coefficients of the various powers of $1/r$ set equal to zero,[15] we find

(1) the coefficient of $1/r$ is identically zero,
(2) the coefficient of $1/r^2$ is $(U^o - \bar{\mu}^o)$, implying

$$U^o = \bar{\mu}^o,$$

(3) the coefficient of $1/r^3$ is

$$(\sigma^o\bar{\sigma}^o)_{,1} + 2\sigma^o\bar{\sigma}^o(\gamma^o + \bar{\gamma}^o) - (\sigma^o\lambda^o + \bar{\sigma}^o\bar{\lambda}^o) = 0.$$

This defines either γ^o or λ^o if the other is known.

---

[13] L. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1960), p. 90.

[14] L. Eisenhart, *A Treatise on the Differential Geometry of Curves and Surfaces* (Ginn and Company, Boston, Massachusetts 1949), Chap. 2, Sec. 40.

[15] For example, if given the asymptotic expression $Ar^{-1} + Br^{-2} + Cr^{-3} + O(r^{-4}) = 0$ (A, B, C independent of r), multiplying through by r and taking the limit as $r \to \infty$ implies $A = 0$; similarly multiplying by r^2 implies $B = 0$, etc.

It turns out that both γ° and λ° can be obtained more easily from the other nonradial equations.

(4) the coefficient of $1/r^4$ is identically zero by virtue of information from the other nonradial equations.

In general it was found the lowest nontrivial powers of $1/r$ yielded all necessary information. Coefficients of the higher powers duplicated this information, but in a rather complicated form; for example a coefficient of $1/r^4$ would generally involve the sums and products of several $1/r$ and $1/r^2$ coefficients. A few of the higher powers of $1/r$ gave the same information as group III, the u-derivative equations. Due to the tediousness of the calculations, we did not carry the computation of the coefficients of some higher powers in $1/r$ to completion. Partial checks were made however, where undefined variables appeared. When the coefficients of these undefined variables (namely, Ψ_0°, Ψ_1°, and $\Psi_2^\circ + \bar{\Psi}_2^\circ$) were collected, they were found to vanish identically. For this reason it appeared that no new information was forthcoming.

The nonradial equations yield the following relations[16]; let $\nabla \equiv \partial/\partial x^3 + i\,\partial/\partial x^4$, then

$$\begin{aligned}
\gamma^\circ &= -\tfrac{1}{2}(\ln \bar{P})_{,1}, \\
\alpha^\circ &= \tfrac{1}{2}\bar{P}\overline{\nabla}(\ln P), \\
\nu^\circ &= -\tfrac{1}{2}\bar{P}\overline{\nabla}(\ln P\bar{P})_{,1}, \\
\omega^\circ &= \bar{P}[\overline{\nabla}\sigma^\circ - 2\sigma^\circ\overline{\nabla}(\ln P)], \\
\lambda^\circ &= \bar{\sigma}^\circ[\ln(\bar{\sigma}^\circ P^{1/2}/\bar{P}^{3/2})]_{,1}, \\
\mu^\circ &= U^\circ = -\tfrac{1}{2}P\bar{P}\,\nabla\overline{\nabla}\ln(P\bar{P}).
\end{aligned} \tag{27}$$

$$\begin{aligned}
\Psi_2^\circ - \bar{\Psi}_2^\circ &= (\bar{P}\overline{\nabla}\omega^\circ + 2\bar{\alpha}^\circ\bar{\omega}^\circ + \bar{\sigma}^\circ\bar{\lambda}^\circ) \\
&\quad - (P\nabla\bar{\omega}^\circ + 2\alpha^\circ\omega^\circ + \sigma^\circ\lambda^\circ),
\end{aligned} \tag{28}$$

$$\Psi_3^\circ = \bar{P}\overline{\nabla}\mu^\circ - P\nabla\lambda^\circ + 4\bar{\alpha}^\circ\lambda^\circ,$$

$$\Psi_4^\circ = \bar{P}\overline{\nabla}\nu^\circ + 2\alpha^\circ\nu^\circ - \lambda^\circ_{,1} - 4\gamma^\circ\lambda^\circ.$$

The only functions left undefined are σ° and P, chosen to be the basic functions, and Ψ_0°, Ψ_1°, and $(\Psi_2^\circ + \bar{\Psi}_2^\circ)$. (The significance of these quantities is explained in Sec. V). However, the propagation of Ψ_0°, Ψ_1°, and Ψ_2° in the u direction is determined by group III, the u-derivative equations. Equation (11a) is worked out as as an example.

[16] Actually the integrations of group I give a more complicated set of expressions than Eqs. (23), (24), and (25), involving the integration "constant" X^{ok} in several places. However, one of the nonradial equations, Eq. (10a), indicates that $X^{o3} + iX^{o4}$ is an analytic function of $x^3 + ix^4$, and may be transformed away by Eq. (26).

$$\begin{aligned}
&\frac{\Psi^\circ_{0,1}}{r^5} + \frac{5(\gamma^\circ + \bar{\gamma}^\circ)\Psi_0^\circ}{r^5} - \frac{P\nabla\Psi_1^\circ}{r^5} \\
&\qquad - \frac{4\gamma^\circ\Psi_0^\circ}{r^5} - \frac{2\bar{\alpha}^\circ\Psi_1^\circ}{r^5} - \frac{3\sigma^\circ\Psi_2^\circ}{r^5} + O(r^{-6}) = 0.
\end{aligned}$$

Here again the explicit powers of $1/r$ are limited by assumption (13). Hence

$$\begin{aligned}
&\Psi^\circ_{0,1} - P\nabla\Psi_1^\circ + (\gamma^\circ + 5\bar{\gamma}^\circ)\Psi_0^\circ \\
&\qquad - 2\bar{\alpha}^\circ\Psi_1^\circ - 3\sigma^\circ\Psi_2^\circ = 0.
\end{aligned} \tag{29}$$

Similarly the next two equations of group III, Eqs. (11b, c), yield

$$\begin{aligned}
&\Psi^\circ_{1,1} - P\nabla\Psi_2^\circ + 2(\gamma^\circ + 2\bar{\gamma}^\circ)\Psi_1^\circ - 2\sigma^\circ\Psi_3^\circ = 0 \\
&\Psi^\circ_{2,1} + 3(\gamma^\circ + \bar{\gamma}^\circ)\Psi_2^\circ \\
&\qquad - P\nabla\Psi_3^\circ + 2\bar{\alpha}^\circ\Psi_3^\circ - \sigma^\circ\Psi_4^\circ = 0.
\end{aligned} \tag{30}$$

The final equation of group III, Eq. (11d), yielded identities, as far as we had carried it.

IV. COORDINATE TRANSFORMATIONS

At this point the explicit r dependence of the metric, Eqs. (4) and (6), is

$$\begin{aligned}
g^{12} &= 1, \qquad g^{11} = g^{1k} = 0, \\
g^{22} &= a_{-1}r + a_0 + a_1r^{-1} + a_2r^{-2} + \cdots, \\
g^{2k} &= b_2^k r^{-2} + b_3^k r^{-3} + \cdots, \\
g^{mn} &= -2P\bar{P}\,\delta^{mn}r^{-2} + d_3^{mn}r^{-3} \\
&\qquad - 6\sigma^\circ\bar{\sigma}^\circ P\bar{P}\,\delta^{mn}r^{-4} + \cdots,
\end{aligned} \tag{31}$$

where

$$\begin{aligned}
a_{-1} &\equiv (\ln P\bar{P})_{,1}, \qquad a_0 \equiv 2\mu^\circ, \\
a_1 &\equiv -(\Psi_2^\circ + \bar{\Psi}_2^\circ), \cdots, \\
b_2^k &\equiv -(\xi^{ok}\bar{\omega}^\circ + \bar{\xi}^{ok}\omega^\circ), \cdots, \\
d_3^{mn} &\equiv 2(\bar{\sigma}^\circ\xi^{om}\xi^{on} + \sigma^\circ\bar{\xi}^{om}\bar{\xi}^{on}), \cdots.
\end{aligned} \tag{32}$$

It is of interest to find the most general coordinate transformation which preserves every relation determined up to this point. We can introduce a new system of hypersurfaces u' and a new tetrad exactly as before, and arrive at all the previous results. Moreover, this coordinate freedom can be used to specify the u dependence of P, and in particular can be used to set $P_{,1} = 0$.

These coordinate transformations may be arrived at more simply by first considering the analogous infinitesimal transformations, $x^{\mu'} = x^\mu + \zeta^\mu(x^\nu)$, and using

$$\begin{aligned}
\tilde{\delta}g^{\mu\nu} &\equiv g^{\mu\nu\prime}(x^\alpha) - g^{\mu\nu}(x^\alpha) \\
&= g^{\mu\alpha}\zeta^\nu_{,\alpha} + g^{\alpha\nu}\zeta^\mu_{,\alpha} - g^{\mu\nu}_{,\alpha}\zeta^\alpha.
\end{aligned}$$

In order to preserve the form of the metric Eq. (31), it is necessary that

$$\delta g^{11} = \delta g^{12} = \delta g^{1k} = 0,$$

$$\delta g^{2k} = (b_2^{k'} - b_2^k)r^{-2} + (b_3^{k'} - b_3^k)r^{-3} + \cdots,$$

$$\delta g^{33} = -2(P'\bar{P}' - P\bar{P})r^{-2} + (d_3^{33'} - d_3^{33})r^{-3} - 6(\sigma^{o'}\bar{\sigma}^{o'}P'\bar{P}' - \sigma^o\bar{\sigma}^oP\bar{P})r^{-4} + \cdots,$$

$$\delta g^{34} = (d_3^{34'} - d_3^{34})r^{-3} + \cdots,$$

$$\delta g^{44} = -2(P'\bar{P}' - P\bar{P})r^{-2} - (d_3^{33'} - d_3^{33})r^{-3} - 6(\sigma^{o'}\bar{\sigma}^{o'}P'\bar{P}' - \sigma^o\bar{\sigma}^oP\bar{P})r^{-4} + \cdots.$$

Under the above restrictions, the infinitesimal transformations reduce to

$$\zeta^1 = \zeta^{o1}(u, x^k),$$

$$\zeta^2 = -\zeta^{o1}_{,1}r + \zeta^{o2}(u, x^k) - \zeta^{o1}_{,k}[b_2^k r^{-1} + \tfrac{1}{2}b_3^k r^{-2} + \cdots],$$

$$\zeta^k = \zeta^{ok}(x^n) + \zeta^{o1}_{,m}[-2P\bar{P}\,\delta^{mk}r^{-1} + \tfrac{1}{2}\,d_3^{mk}r^{-2} - 2\sigma^o\bar{\sigma}^oP\bar{P}\,\delta^{mk}r^{-3} + \cdots],$$

where

$$\zeta^{o3}_{,3} = \zeta^{o4}_{,4},$$

$$\zeta^{o3}_{,4} = -\zeta^{o4}_{,3},$$

$$\zeta^{o2} = P\bar{P}(\zeta^{o1}_{,33} + \zeta^{o1}_{,44}).$$

The restrictions on ζ^{ok} result from the conformal form of the leading term of $g^{mn'}$, whereas the last equation reflects the fact that when the hypersurfaces u' are chosen, there is no freedom in the choice of r'. As yet no restrictions exist on $\zeta^{o1}(u, x^k)$. If this is chosen to satisfy a certain differential equation, the coefficient of r' in $g^{22'}$, namely, $(\ln P'\bar{P}')_{,1}$, vanishes.

The finite transformations built up from the infinitesimal ones take the form,

$$u' = V_0 + V_{-1}r^{-1} + V_{-2}r^{-2} + \cdots, \tag{33a}$$

$$r' = R_1 r + R_0 + R_{-1}r^{-1} + \cdots, \tag{33b}$$

$$x^{m'} = Y_0^m + Y_{-1}^m r^{-1} + \cdots, \tag{33c}$$

where the upper case letters are functions of u, x^k.

The condition $g^{11'} = 0$ leaves V_0 unspecified but determines the remaining V's;

$$V_{-1} = -P\bar{P}(V_{0,1})^{-1}[(V_{0,3})^2 + (V_{0,4})^2], \tag{34}$$

and V_{-2} etc. as functions of the V_0, P, and σ^o.

The condition $g^{12'} = 0$ leaves R_0 unspecified but yields the remaining R's;

$$R_1 = (V_{0,1})^{-1}, \tag{35}$$

the R_{-1}, R_{-2} etc. being functions of V_0, R_0, P, and σ^o.

Setting $g^{1k'} = 0$ determines Y_{-1}^k as

$$Y_{-1}^k = -2P\bar{P}(V_{0,1})^{-1}[V_{0,3}Y_{0,3}^k + V_{0,4}Y_{0,4}^k], \tag{36}$$

the Y_{-2}^k, Y_{-3}^k, etc. also being determined.

Prescribing the form $g^{2k'}$ and $g^{mn'}$ as in Eqs. (31) determines R_0 as a function of the remaining free variables, and imposes relations on Y_0^k,

$$Y_0^k = Y_0^k(x^m)$$

$$Y_{0,3}^3 = \pm Y_{0,4}^4 \tag{37}$$

$$Y_{0,4}^3 = \mp Y_{0,3}^4.$$

Finally, choosing

$$V_0 = J(x^3, x^4)\int (P\bar{P})^{1/2}\,du + K(x^3, x^4), \tag{38}$$

where J and K are arbitrary functions of x^3 and x^4, eliminates the coefficient of r' in $g^{22'}$

$$(\ln P'\bar{P}')_{,1} = 0.$$

The tetrad transformation Eq. (7) can be used to make $P \equiv |P|e^{i\phi}$ real as follows,

$$m^{\mu'} = m^\mu e^{iC}.$$

Since $m^k = \xi^{ok}r^{-1}\delta_k^\mu + O(r^{-2})$, $\xi^{o3} = -i\xi^{o4} = P$, by choosing $C = -\phi$, we have

$$P' = |P|.$$

This, together with $(\ln P'\bar{P}')_{,1} = 0$, yields

$$P' = P'(x^{m'}). \tag{39}$$

This has the effect of immensely simplifying most of the previous relations. Equations (27) and (28) become (dropping primes)

$$\gamma^o = 0, \tag{40a}$$

$$\alpha^o = \tfrac{1}{2}\overline{\nabla}P, \tag{40b}$$

$$\nu^o = 0, \tag{40c}$$

$$\omega^o = P^3\overline{\nabla}(\sigma^o/P^2), \tag{40d}$$

$$\lambda^o = \bar{\sigma}^o_{,1} \tag{40e}$$

$$\mu^o = -P^2\nabla\overline{\nabla}\ln P, \tag{40f}$$

$$(\Psi_2^o - \bar{\Psi}_2^o) = P^2[\overline{\nabla}(\omega^o/P) - \nabla(\bar{\omega}^o/P)] + \bar{\sigma}^o\sigma^o_{,1} - \sigma^o\bar{\sigma}^o_{,1}, \tag{40g}$$

$$\Psi_3^o = -P\overline{\nabla}(P^2\nabla\overline{\nabla}\ln P) - P^3(\bar{\sigma}^o_{,1}/P^2), \tag{40h}$$

$$\Psi_4^o = -\bar{\sigma}^o_{,11}. \tag{40i}$$

The u-derivative equations (29) and (30) become

$$\Psi^o_{0,1} - \nabla(P\Psi^o_1) - 3\sigma^o\Psi^o_2 = 0 \tag{40j}$$

$$\Psi^o_{1,1} - P\nabla\Psi^o_2 - 2\sigma^o\Psi^o_3 = 0 \tag{40k}$$

$$\Psi^o_{2,1} - P^2\nabla(\Psi^o_3/P) + \sigma^o\bar{\sigma}^o_{,11} = 0. \tag{40l}$$

V. RESULTS AND DISCUSSION

The metric has the form

$$g^{\mu\nu} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & g^{22} & g^{2k} & \\ 0 & g^{2k} & g^{mn} & \\ 0 & & & \end{vmatrix} \tag{41a}$$

with the coordinate dependence,

$$g^{22} = -2P^2\left(\frac{\partial^2}{\partial x^{3^2}} + \frac{\partial^2}{\partial x^{4^2}}\right)\ln P - (\Psi^o_2 + \bar{\Psi}^o_2)r^{-1} + \tfrac{1}{3}P^2\left[\nabla\left(\frac{\bar{\Psi}^o_1}{P}\right) + \overline{\nabla}\left(\frac{\Psi^o_1}{P}\right) - 6P^4\nabla\left(\frac{\bar{\sigma}^o}{P^2}\right)\overline{\nabla}\left(\frac{\sigma^o}{P^2}\right)\right]r^{-2} + O(r^{-3}), \tag{41b}$$

$$g^{23} = -r^{-2}\,\mathrm{Re}\,\{f\} + r^{-3}\,\mathrm{Re}\,\{h\} + O(r^{-4}), \tag{41c}$$

$$g^{24} = r^{-2}\,\mathrm{Im}\,\{f\} + r^{-3}\,\mathrm{Im}\,\{h\} + O(r^{-4}), \tag{41d}$$

where

$$P = P(x^3, x^4),$$

$$f \equiv 2P^4\nabla(\bar{\sigma}^o/P^2), \tag{41e}$$

$$h \equiv 4P[\tfrac{1}{3}\Psi^o_1 + P^3\sigma^o\nabla(\bar{\sigma}^o/P^2)],$$

$$g^{33} = -2P^2r^{-2} + 2P^2(\sigma^o + \bar{\sigma}^o)r^{-3} - 6\sigma^o\bar{\sigma}^oP^2r^{-4} + O(r^{-5}), \tag{41f}$$

$$g^{34} = -2iP^2(\sigma^o - \bar{\sigma}^o)r^{-3} + O(r^{-5}), \tag{41g}$$

$$g^{44} = -2P^2r^{-2} - 2P^2(\sigma^o + \bar{\sigma}^o)r^{-3} - 6\sigma^o\bar{\sigma}^oP^2r^{-4} + O(r^{-5}). \tag{41h}$$

More terms could easily have been calculated; g^{22} to $O(r^{-4})$ and g^{2k} to $O(r^{-5})$.

The u-derivative equations (40j,k,l), derived from the Bianchi identities, give the propagation of the tetrad components of the Weyl tensor as follows,

$$\Psi^o_{0,1} - \nabla(P\Psi^o_1) - 3\sigma^o\Psi^o_2 = 0 \tag{42a}$$

$$\Psi^o_{1,1} - P\nabla\Psi^o_2 - 2\sigma^o\Psi^o_3 = 0 \tag{42b}$$

$$\Psi^o_{2,1} + \sigma^o\bar{\sigma}^o_{,11} - P^2\nabla(\Psi^o_3/P) = 0 \tag{42c}$$

where

$$\Psi^o_3 = -P^3\nabla\left(\frac{\bar{\sigma}^o_{,1} + P\overline{\nabla}^2P}{P^2}\right).$$

With these results (also considering the higher order terms not explicitly written out) the solution of the initial value problem can be stated simply.

The first piece of data that is chosen is Ψ_0 on an initial null-surface denoted by u_0 subject to the condition that $\lim_{r\to\infty}(r^5\Psi_0) < \infty$. For convenience, we use a slightly stronger condition in the form $\Psi_0 = \Psi^o_0r^{-5} + O(r^{-6})$. It should be noted that in this paper we do not specify Ψ_0 in complete detail, and hence, the data are correspondingly not completely specified. The second piece of data is given on the timelike world-tube taken at spatial infinity. On this tube we choose $\sigma^o \equiv \lim_{r\to\infty}(r^2\sigma)$, as an arbitrary function of u, x^3, and x^4. The last data are given on the two-dimensional surface at infinity which is defined by the intersection of the null surface u_0 and the world-tube. On this two-surface we give

$$\Psi^o_1 \equiv \lim_{r\to\infty} r^4\Psi_1, \qquad \Psi^o_2 + \bar{\Psi}^o_2 \equiv \lim_{r\to\infty} r^3(\Psi_2 + \bar{\Psi}_2)$$

and

$$P^2\,\delta^{ij} \equiv \lim_{r\to\infty}(r^2g^{ij})$$

as functions of x^3 and x^4.

This information, used with the three differential equations (42), completes the solution for the metric tensor and Riemann tensor up to the completeness of the data on u_0. (If, for example, we had taken

$$\Psi_0 = \sum_{n=0}^{\infty}\frac{\Psi^n_0}{r^{n+5}}$$

on u_0, there would have been an additional equation for the u derivative of each Ψ^n_0.) However, aside from the incompleteness of the u_0 data and the consequent incompleteness of the solution (we could give additional data and carry the solution as far as desired) the problem of finding the metric tensor for empty space, flat at infinity, appears to be solved. It should be pointed out that if Ψ_0 is not an analytic function of $(1/r)$, it has not been proven that the solution of Eq. (11a), which yields the propagation of Ψ_0 off the hypersurface, is unique, (although it is generally believed that the solution is unique).

An interesting special case which we will consider is σ^o initially zero, then varying in an arbitrary manner, and eventually becoming zero again. By multiplying Eq. (42c) by P^{-2} and integrating over x^3 and x^4, we obtain after a few operations

$$\left[\int\frac{1}{P^2}(\Psi^o_2 + \sigma^o\bar{\sigma}^o_{,1})\,dx^3\,dx^4\right]_{,1} = \int\frac{1}{P^2}\sigma^o_{,1}\bar{\sigma}^o_{,1}\,dx^3\,dx^4 + \int\nabla\left(\frac{\Psi^o_3}{P}\right)dx^3\,dx^4. \tag{43}$$

Remembering that the integral is over the closed two-surface with metric

$$g^{o\,ij} = \lim_{r\to\infty}(r^2 g^{ij}) = P^2\,\delta^{ij}$$

(the element of area is $P^{-2}\,dx^3\,dx^4$), it is easy to see that the last term on the right is zero. We now integrate over u, and using the fact that $\int P^{-2}\sigma^o_{,1}\bar{\sigma}^o_{,1}\,dx^3\,dx^4$ is positive and that σ^o vanishes at the two end points, we have

$$\int P^{-2}\Psi_2^o\,dx^3\,dx^4\Big|_{u_0}^{u_1} > 0; \tag{44}$$

in other words the quantity $M \equiv -\int P^{-2}\Psi_2^o\,dx^3\,dx^4$ decreases. We can rewrite this as

$$M = -\int \mathrm{Re}\,\Psi_2^o\,dS, \tag{45}$$

because from Eqs. (40g) and (40d), Im $\Psi_2^o = 0$ when $\sigma^o = 0$. In the case of the Schwartzchild solution M is proportional to the mass, from which we generalize and say M is always proportional to the mass.[8]

The remaining coordinate freedom is given by transformations of the form Eqs. (33). As $r \to \infty$, the leading terms of these transformations are

$$\begin{aligned} u' &= J(x^3, x^4)u + K(x^3, x^4), \\ r' &= [1/J(x^3, x^4)]r, \\ x^{k\prime} &= Y^k(x^3, x^4), \end{aligned} \tag{46}$$

where J and K are arbitrary functions of x^3, x^4. The Y^k must satisfy

$$\begin{aligned} Y^3_{,3} &= \pm Y^4_{,4} \\ Y^3_{,4} &= \mp Y^4_{,3} \end{aligned} \tag{47}$$

and the Y^k are otherwise arbitrary functions of x^3, x^4.[17] The transformation law of the basic variable $P(x^k)$ is easily obtained by considering the transformation of the $1/r^2$ term in g^{33} or g^{44};

$$P'^2 = P^2 J^{-2}\,\det|Y^m_{,n}|. \tag{48}$$

The transformation

$$P'^2 = P^2\,\det|Y^m_{,n}|$$

represents a two-dimensional conformal coordinate transformation of a two-surface, with metric tensor $g^{o\,ij} = P^2\delta^{ij}$, onto itself.[14] Neglecting this and considering only $P' = PJ^{-1}$, we see that Eq. (48) represents the transformation of one surface into another. Equation (46) indicates that $J(x^3, x^4)$ may have neither zeros nor infinities, hence from Eq. (48) we can conclude that P' must have the same number of zeros and infinities as P. (We are assuming that P, which is part of the initial data, is given not in patches but over the entire surface.) It thus appears that many topological properties of the surface are retained under the transformation Eq. (48). This leads to several interesting topological questions. For example: What is the simplest surface that may be obtained by Eq. (48) from a given surface?

Such difficulties can be avoided by taking as part of the initial data a relatively simple surface such as a sphere,[17] or possibly a torus, and restricting Eq. (48) to the identity $P' = P$; in other words, by requiring $J^2 = \det|Y^m_{,n}|$.

As a closing remark, it may be pointed out that Newman and Tamburino[18] have found exact solutions corresponding to a special case ($\Psi_0 = 0$) of the asymptotic solutions found here.

ACKNOWLEDGMENTS

The authors wish to thank Dr. A. Janis, Dr. R. Penrose, Dr. I. Robinson, Dr. J. Stachel, L. Tamburino, and Dr. A. Trautman for valuable suggestions and discussions. We are particularly indebted to Dr. R. Sachs for a detailed criticism of the manuscript.

[17] In a recent preprint of R. Sachs, there is a detailed discussion of the transformations, Eq. (46).

[18] E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962), following paper.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Empty Space Metrics Containing Hypersurface Orthogonal Geodesic Rays*

EZRA T. NEWMAN AND LOUIS A. TAMBURINO
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received April 27, 1962)

In this paper we obtain all empty space metrics which possess hypersurface orthogonal geodesic rays with nonvanishing shear and divergence. By straightforward integration of the Newman-Penrose equations, which are equivalent to the Einstein equations, all solutions are found in closed form and are unique up to a few arbitrary constants. The method of integration is illustrated in detail for the Robinson-Trautman solutions.

I. INTRODUCTION

IN 1960, Robinson and Trautman published the general solutions for the class of metrics containing geodesic rays with nonvanishing divergence and vanishing shear.[1] Geodesic rays are principal null directions of the Riemann tensor, sometimes called Ruse, Debever, or Penrose vectors which are tangent to a congruence of null geodesics. Geometrically, geodesic rays, discussed by Sachs,[2] can be thought of as propagation vectors for outgoing gravitational fields. Because the Robinson-Trautman solutions contained the Schwarzchild metric, the algebraic generalization of their class (the removal of the condition of vanishing shear) was expected to be physically important.

In this paper we obtain this generalized solution for the shearing class of hypersurface orthogonal geodesic rays with nonvanishing divergence. There are two general subcases which we shall designate as cyclindrical and spherical. By a straightforward integration of the Newman-Penrose equations,[3] which are equivalent to the Einstein equation, all solutions are found in closed form and are unique up to a few arbitrary constants. The method of integration is illustrated in detail for the Robinson-Trautman solutions (see Sec. III). The assumption of nonvanishing shear then leads to additional equations that do not appear in the nonshearing case. Subsequently these additional equations severely restrict our class of solutions and prohibit the limit of vanishing shear, hence the Robinson-Trautman metrics are not limiting cases of our solutions.

II. PRELIMINARY REMARKS

A vacuum metric contains a geodesic ray if there exists a principal null direction of the curvature tensor, defined by

$$l_{[\mu}R_{\alpha]\beta\gamma[\delta}l_{\nu]}l^{\beta}l^{\gamma} = 0, \tag{II.1}$$

that is tangent to a congruence of null geodesics[2,4]

$$l_{\mu;\nu}l^{\nu} = 0, \qquad l^{\mu}l_{\mu} = 0. \tag{II.2}$$

Throughout this paper we shall further demand that the geodesic ray l^{μ} be hypersurface orthogonal,

$$l_{\mu} = u_{,\mu}. \tag{II.3}$$

From this condition we shall adopt the same coordinate system that was constructed by Robinson and Trautman[1]:

$$\begin{aligned} x^1 &= u, \\ x^2 &= r, \quad r \text{ is the affine parameter along the null geodesics.} \end{aligned} \tag{II.4}$$

x^i, $i = 3, 4$, x^i label the geodesics on each hypersurface, $u = \text{const}$.

Associated with this coordinate system is the following tetrad[5,6]:

$$\begin{aligned} l^{\mu} &= \delta^{\mu}_2, \qquad l_{\mu} = \delta^1_{\mu}, \\ n^{\mu} &= \delta^{\mu}_1 + U\,\delta^{\mu}_2 + X^i\,\delta^{\mu}_i, && \text{real null vectors} \\ m^{\mu} &= \omega\,\delta^{\mu}_2 + \xi^i\,\delta^{\mu}_i, \\ \bar{m}^{\mu} &= \bar{\omega}\,\delta^{\mu}_2 + \bar{\xi}^i\,\delta^{\mu}_i. && \text{complex vectors} \end{aligned}$$

These vectors satisfy the following orthogonality conditions

* This work has been supported by the Office of Aerospace Research, U.S. Air Force.

[1] I. Robinson and A. Trautman, Phys. Rev. Letters **4**, 431 (1960).

[2] R. Sachs, Proc. Roy. Soc. (London) **264**, 309 (1961).

[3] E. Newman and R. Penrose, J. Math. Phys. **3**, 565 (1962).

[4] Square brackets on indices denote antisymmetrization.

[5] See reference 3. Sec. VI, special coordinates.

[6] E. Newman and T. Unti, preceding paper J. Math. Phys., **3**, 891 (1962).

$$l_\mu n^\mu = -m_\mu \bar{m}^\mu = 1. \tag{II.6}$$

All other products vanish; $l_\mu m^\mu = m_\mu m^\mu = \cdots = 0$.

It follows that

$$g^{\mu\nu} = l^\mu n^\nu + n^\mu l^\nu - m^\mu \bar{m}^\nu - \bar{m}^\mu m^\nu, \tag{II.7}$$

$$g^{\mu\nu} = \begin{vmatrix} 0 & 1 & 0 & 0 \\ 1 & g^{22} & g^{23} & g^{24} \\ 0 & g^{23} & g^{33} & g^{34} \\ 0 & g^{24} & g^{34} & g^{44} \end{vmatrix}. \tag{II.8}$$

The form (II.8) is invariant under the following coordinate transformations:

$$r' = r + R(1, 3, 4), \qquad u' = u, \qquad x^{i'} = x^i \tag{II.9a}$$

shifts the r origin,

$$r' = r/\dot{\gamma}, \qquad u' = \gamma(u), \qquad x^{i'} = x^i \tag{II.9b}$$

relabels hypersurfaces,

$$r' = r \qquad u' = u, \qquad x^{i'} = x^{i'}(1, 3, 4), \tag{II.9c}$$

relabels geodesics.

The following tetrad transformations do not effect l^μ or the orthogonality conditions:

$$l^{\mu'} = l^\mu$$
$$n^{\mu'} = n^\mu + \bar{B}m^\mu + B\bar{m}^\mu + B\bar{B}l^\mu \tag{II.10a}$$
$$m^{\mu'} = m^\mu + Bl^\mu$$

'null rotations,'

where B is a complex scalar independent of r,

$$l^{\mu'} = l^\mu \qquad \text{'spatial rotation'}$$
$$n^{\mu'} = n^\mu \tag{II.10b}$$
$$m^{\mu'} = m^\mu e^{iC}$$

where C is real and independent of r.

Associated with a null geodesic field l^μ are three scalars, the divergence, the curl, and the complex shear, whose geometrical significance has been discussed by Sachs.[2] In the Newman-Penrose paper NP[3] these scalars are related to two of the spin coefficients (linear combinations of the Ricci rotation coefficients) ρ and σ in the following manner:

$$\rho = l_{\mu;\nu} m^\mu \bar{m}^\nu = \tfrac{1}{2}[-l^\mu{}_{;\mu} + i \text{ curl } l^\mu]$$
$$\text{where curl } l^\mu = (l_{[\mu;\nu]} l^{\mu;\nu})^{1/2} \tag{II.11}$$

$$\sigma = l_{\mu;\nu} m^\mu m^\nu \quad \text{complex shear}$$
$$\text{where } \sigma\bar{\sigma} = \tfrac{1}{2}[l_{(\mu;\nu)} l^{\mu;\nu} - \tfrac{1}{2}(l^\mu{}_{;\mu})^2]. \tag{II.12}$$

We now state two pertinent theorems from NP.

Theorem I. If $\Psi_0 \equiv -R_{\mu\nu\rho\sigma} l^\mu n^\nu l^\rho m^\sigma = 0$ then l^μ is a geodesic ray; conversely if l^μ is a geodesic ray, then $\Psi_0 = 0$.

Theorem II (Goldberg-Sachs' theorem). If $\Psi_0 = 0$ and $\Psi_1 \equiv -R_{\mu\nu\rho\sigma} l^\mu n^\nu l^\rho n^\sigma = 0$, then there exists a geodesic ray with $\sigma = 0$; conversely if there exists a shear-free null geodesic, then $\Psi_0 = \Psi_1 = 0$.

The existence of a hypersurface orthogonal geodesic ray is characterized by the following mathematical statements:

$$\Psi_0 = 0 \tag{II.13}$$

$$\rho = \bar{\rho}; \qquad \text{curl } l^\mu = 0. \tag{II.14}$$

All the general solutions for the class of metrics containing hypersurface orthogonal geodesic rays, outlined in Table I, have, with our solutions, been exhausted.

The field equations and coordinate system used in our calculations are formally the same as those found in the preceding Newman-Unti paper NU[6] and originally in NP.[3] One can utilize everything in Sec. II of (NU) up to the paragraph containing Eq. (II.13), keeping in mind one important reservation; the arbitrary null hypersurfaces employed by (NU) are uniquely determined in this paper by the curvature tensor from Eqs. (II.1) and (II.3). Since our l^μ is a geodesic ray, Theorem I implies that Ψ_0 be set equal to zero everywhere.

III. PROCEDURE: ROBINSON-TRAUTMAN SOLUTION AS ILLUSTRATION

The general procedure used in our lengthy calculations will be illustrated by obtaining the Robinson-Trautman solution which is a concise example demonstrating the inherent power of the Newman-Penrose formalism. One can, in general, construct a tetrad in which l^μ (here l^μ is not necessarily a geodesic ray) has the components $l^\mu = \delta^\mu_2$ and is

TABLE I.

Case I. Spherical Rays	$\rho^2 \neq \sigma\bar{\sigma}$, $\rho \neq 0$	
A. $\sigma = 0$		I. Robinson–A. Trautman (1960) (reference 1)
B. $\sigma \neq 0$		Sec. IV
Case II. Cylindrical Rays	$\rho^2 = \sigma\bar{\sigma} \neq 0$	Sec. V
Case III. Plane Rays	$\rho = \sigma = 0$	W. Kundt (1961) (reference 7)

[7] W. Kundt, Z. Physik **163**, 77 (1961).

tangent to a congruence of null geodesics and in which n^μ and m^μ are parallel propagated along the geodesics. With such a tetrad, the Newman-Penrose field equations split evenly into two groups, the radial equations and the nonradial equations.

The radial equations express the partial derivatives, $D \equiv (\partial/\partial r)(r \equiv x^2)$, of unknown functions in terms of known functions of r. Since Ψ_0 is the only variable whose partial derivative $D\Psi_0$ does not appear anywhere in the equations, Ψ_0 must be given as initial data. If there exists a geodesic ray then Ψ_0 is identically zero. A step by step integration of the radial equations is straightforward, the sequence of integration being more or less dictated by the equations. Each step yields the r dependence of an unknown and a "constant" of integration depending only on u and x^i.

The nonradial equations involve more complicated differential operators,

$$\Delta\phi = \phi_{,\mu}n^\mu \qquad \delta\phi^\mu = \phi_{,\mu}m^\mu.$$

The r dependence obtained from the radial equations is substituted into the nonradial equations and relationships involving the constants of integration are obtained essentially by comparing the coefficients of like powers of r. These relationships, together with the available coordinate freedom, enable one to determine the constants of integration in terms of the initial data.

The Robinson-Trautman solution, which contains a hypersurface orthogonal geodesic ray with vanishing shear and nonvanishing divergence, is characterized in the (NP) formalism by the following:

$$\Psi_0 = \Psi_1 = \sigma = \rho - \bar{\rho} = 0, \qquad \rho \neq 0. \tag{III.1}$$

References to the preceding paper, (NU), will be placed to the left of the appropriate equations. A degree sign superscript indicates a function is independent of r; *I.S.* will mean an equation is identically satisfied or void of any new information. Initially the (NU) equations (9f), (9n) and (11a) are identically satisfied by Eq. (III.1).

$$\text{(9e)} \qquad D\rho = \rho^2 \rightarrow \rho = -(r + \rho^\circ)^{-1}$$

One can set $\rho^\circ = 0$ by means of the coordinate transformation (II.9a), hence

$$\rho = -1/r \tag{III.2}$$

$$\text{(9h)} \qquad D\alpha = \rho\alpha \rightarrow \alpha = \alpha^\circ/r \tag{III.3}$$

$$\text{(9i)} \qquad D\beta = \rho\beta \rightarrow \beta = \beta^\circ/r \tag{III.4}$$

$$\text{(9g)} \qquad D\tau = \rho\tau \rightarrow \tau = \tau^\circ/r. \tag{III.5}$$

In (NP) it is shown that $\tau = \bar{\alpha} + \beta$ in this coordinate system, therefore

$$\tau^\circ = \bar{\alpha}^\circ + \beta^\circ. \tag{III.6}$$

τ° may be set equal to zero by the transformation (II 10), hence

$$\alpha = -\bar{\beta}, \qquad \tau = 0. \tag{III.7}$$

$$\text{(10k)} \qquad \bar{\lambda}\rho = 0 \rightarrow \lambda = 0 \tag{III.8}$$

$$\text{(9k)} \qquad D\lambda = \rho\lambda, \qquad I.S. \tag{III.9}$$

$$\text{(9o)} \qquad D\Psi_2 = 3\rho\Psi_2 \rightarrow \Psi_2 = \Psi_2^\circ/r^3 \tag{III.10}$$

$$\text{(9j)} \qquad D\gamma = \Psi_2 \rightarrow \gamma = \gamma^\circ - \Psi_2^\circ/2r^2 \tag{III.11}$$

$$\text{(9l)} \qquad D\mu = \mu\rho + \Psi_2 \rightarrow \mu = \mu^\circ r^{-1} - \Psi_2^\circ r^{-2} \tag{III.12}$$

$$\text{(10f)} \qquad \delta\rho \equiv \omega\, D\rho = 0 \rightarrow \omega = 0 \tag{III.13}$$

$$\text{(9b)} \qquad D\omega = \rho\omega \quad I.S. \tag{III.14}$$

$$\text{(9c)} \qquad DX^i = 0 \rightarrow X^i = X^{\circ i} \tag{III.15}$$

$$\text{(9d)} \qquad DU = -(\gamma + \bar{\gamma}) \rightarrow U = U^\circ - (\gamma^\circ + \bar{\gamma}^\circ)r - \tfrac{1}{2}(\Psi_2^\circ + \bar{\Psi}_2^\circ)r^{-1} \tag{III.16}$$

$$\text{(9a)} \qquad D\xi^i = \rho\xi^i \rightarrow \xi^i = \xi^{\circ i}r^{-1}. \tag{III.17}$$

Using (II 9c), we may set

$$\xi^{\circ 3} = P, \qquad \xi^{\circ 4} = iP. \tag{III.18}$$

The remaining coordinate freedom is

$$r' = r/\dot{\gamma}, \qquad u' = \gamma(u), \qquad \zeta' = x^{3\prime} + ix^{4\prime} = f(\zeta, u)$$

$$\text{where} \qquad \zeta = x^3 + ix^4, \tag{III.19}$$

$$\text{(10c)} \qquad \mu - \bar{\mu} = 0 \rightarrow \mu^\circ \text{ and } \Psi_2^\circ \text{ are real.} \tag{III.20}$$

Substituting the above results into

$$\text{(10a)} \qquad \delta X^i - \Delta\xi^i = (\mu + \bar{\gamma} - \gamma)\xi^i, \tag{III.26}$$

$$\xi^{\circ i}X^\circ_{,i}/r + \xi^{\circ i}/r^2[U^\circ - (\gamma^\circ + \bar{\gamma}^\circ)r - \Psi_2^\circ/r] - (1/r)[\xi^{\circ i}_{,1} + \xi^{\circ i}_{,i}X^{\circ i}] = (\xi^{\circ i}/r)\{\mu^\circ/r - \Psi_2^\circ/r^2 + (\bar{\gamma}^\circ - \gamma^\circ)\}. \tag{III.27}$$

Equating coefficients of like powers in $1/r$ yields

$$1/r \rightarrow \xi^{\circ i}X^\circ_{,i} - \xi^{\circ i}_{,1} - \xi^{\circ i}_{,i}X^{\circ i} = 2\bar{\gamma}^\circ\xi^{\circ i} \tag{III.28}$$

$$1/r^2 \rightarrow \xi^{\circ i}U^\circ = \xi^{\circ i}\mu^\circ \quad \text{or} \quad \mu^\circ = U^\circ. \tag{III.29}$$

From (III 28), one obtains

$$\xi^{\circ i}(X^\circ_{,i}{}^3 + iX^\circ_{,i}{}^4) \equiv P\nabla X^\circ = 0 \tag{III.30}$$

where

$$\nabla \equiv \partial/\partial x^3 + i(\partial/\partial x^4) \equiv 2(\partial/\partial\zeta),\ X^\circ = X^{\circ 3} + iX^{\circ 4}.$$

Because X° is analytic, $X^\circ = X^\circ(\zeta, u)$, it can be shown that by using the coordinate transformation (III 19) one can set

$$X^{\circ i} = 0. \tag{III.31}$$

The remaining coordinate freedom is now

$$r' = r/\dot{\gamma}, \qquad u' = \gamma(u), \qquad \zeta' = f(\zeta). \tag{III.32}$$

Equations (III 28) and (III 31) yield

$$2\dot{\gamma}^\circ = -(\ln P)_{,1}. \tag{III.33}$$

(11b) $$\delta\Psi_2 \equiv P\nabla\Psi_2^\circ/r^3 = 0. \tag{III.34}$$

From (34) and (20) we obtain

$$\nabla\Psi_2^\circ = \overline{\nabla}\Psi_2^\circ = 0$$

$$\therefore\ \Psi_2^\circ = \Psi_2^\circ(u). \tag{III.35}$$

(9p) $$D\Psi_3 = 2\rho\Psi_3 \to \Psi_3 = \Psi_3^\circ/r \tag{III.36}$$

(9q) $$D\Psi_4 = \rho\Psi_4 + \bar{\delta}\Psi_3 - 2\alpha\Psi_3 \to \Psi_4$$

$$= \Psi_4^\circ r^{-1} - (\bar{P}\overline{\nabla}\Psi_3^\circ + 2\alpha^\circ\Psi_3^\circ)r^{-2} \tag{III.37}$$

(9m) $$D\nu = \Psi_3 \to \nu = \nu^\circ - \Psi_3^\circ/r \tag{III.38}$$

The remaining nonradial equations may be worked out in the same manner as (26).

(10b) $$\delta\bar{\xi}^i - \bar{\delta}\xi^i = -2\alpha\xi^i + 2\bar{\alpha}\bar{\xi}^i \to 2\alpha^\circ = \bar{P}\overline{\nabla}\ln P \tag{III.39}$$

(10d) $$\delta U = -\bar{\nu} \to \bar{\nu}^\circ = P\nabla(\gamma^\circ + \bar{\gamma}^\circ) = -\tfrac{1}{2}P\nabla(\ln P\bar{P})_{,1} \tag{III.40}$$

(10e) $$\bar{\delta}\nu = -2\alpha\nu + \Psi_4 \to \Psi_4^\circ = \bar{P}\overline{\nabla}\nu^\circ + 2\alpha^\circ\nu^\circ \tag{III.41}$$

(10g) $$\delta\alpha + \bar{\delta}\bar{\alpha} = \mu\rho + 4\alpha\bar{\alpha} - \Psi_2 \to \mu^\circ = U^\circ = -\tfrac{1}{2}P\bar{P}\nabla\overline{\nabla}\ln(P\bar{P}) \tag{III.42}$$

(10h) $$\bar{\delta}\mu = \Psi_3 \to \Psi_3^\circ = \bar{P}\overline{\nabla}\mu^\circ \tag{III.43}$$

Every function is now expressed in terms of P and Ψ_2°. The final equation relates P and Ψ_2°.

(11c) $$\Delta\Psi_2 - \delta\Psi_3 = -3\mu\Psi_2 - 2\bar{\alpha}\Psi_3 \to \{\partial/\partial u - 3P_{,1}/P\}\Psi_2^\circ = +P\nabla\Psi_3^\circ - 2\bar{\alpha}^\circ\Psi_3^\circ \tag{III.44}$$

(10i), (10j), (10l), (10m), and (11d), *I.S.* Using (II 10b), we can set

$$P = \bar{P}. \tag{III.45}$$

From (II 7) and (II 5), the metric becomes

$$g^{22} = -g_{11} = 2U^\circ - 4\dot{\gamma}^\circ r - 2\Psi_2^\circ/r$$

$$g^{21} = g_{12} = 1, \quad g^{1i} = g^{2i} = g_{1i} = g_{2i} = 0 \tag{III.46}$$

$$g^{ij} = (g_{ij})^{-1} = -2(P^2/r^2)\,\delta^{ij}$$

where $U^\circ = -P^2\nabla^2\ln P, \gamma^\circ = -\tfrac{1}{2}(\ln P)_{,1}, \nabla^2 \equiv \nabla\overline{\nabla}$. If $\Psi_2 = 0$, then (44) becomes

$$\nabla^2(P^2\nabla^2\ln P) = 0. \tag{III.47}$$

If $\Psi_2 \neq 0$, then by using (II 9b) one can set $\Psi_2^\circ = 1$. Hence Eq. (III 44) becomes

$$P_{,1} = (P^3/3)\nabla^2(P^2\nabla^2\ln P). \tag{III.48}$$

The only remaining coordinate freedom, (III 32), is

$$P' = P\,|df(\zeta)/d\zeta|. \tag{III.49}$$

This completes the Robinson-Trautman solution where their p, C, K, m, H are, respectively, our $\sqrt{2}\,P$, $-2U$, $-2U^\circ$, $-\Psi^\circ$, -2γ.

IV. SPHERICAL RAYS

The spherical class of metrics containing hypersurface orthogonal geodesic rays with nonvanishing shear and divergence is characterized in the (NP) formalism by the following

$$\Psi_0 = \rho - \bar{\rho} = 0, \quad \rho^2 \neq \sigma\bar{\sigma}, \quad \rho \neq 0, \quad \sigma \neq 0. \tag{IV.1}$$

Only the initial radial integrations will be given in detail in order to illustrate the technique employed and to introduce the pertinent "constants" of integration. The pair of equations

(9e) $$D\rho = \rho^2 + \sigma\bar{\sigma} \tag{IV.2}$$

(9f) $$D\sigma = 2\sigma\rho$$

may be expressed as

$$D\mathbf{M} = \mathbf{M}^2, \quad \text{or} \quad D\mathbf{M}^{-1} = \mathbf{I}, \tag{IV.3}$$

where

$$\mathbf{M} = \begin{Vmatrix} \rho & \sigma \\ \bar{\sigma} & \rho \end{Vmatrix} \quad \text{is a nonsingular matrix.} \tag{IV.4}$$

Equation (IV.3) is readily integrated to give

$$\mathbf{M}^{-1} = \begin{Vmatrix} \rho^\circ - r & -\sigma^\circ \\ \bar{\sigma}^\circ & \rho^\circ - r \end{Vmatrix}, \tag{IV.5}$$

$$\mathbf{M} = \begin{Vmatrix} \dfrac{\rho^\circ - r}{R^2} & \dfrac{\sigma^\circ}{R^2} \\ \dfrac{\bar{\sigma}^\circ}{R^2} & \dfrac{\rho^\circ - r}{R^2} \end{Vmatrix} \tag{IV.6}$$

where

$$R^2 \equiv \det(\mathbf{M}^{-1}) = (\rho^{\circ 2} - r^2) - \sigma^\circ\bar{\sigma}^\circ \neq 0. \quad \text{(IV.7)}$$

Using coordinate transformation (II 9a), one can can set $\rho^\circ = 0$, hence

$$\rho = -r/R^2 \quad \text{(divergence)} \quad \text{(IV.8)}$$

$$\sigma = \sigma^\circ/R \quad \text{(complex shear)} \quad \text{(IV.9)}$$

where

$$R^2 = r^2 - \sigma^\circ\bar{\sigma}^\circ, \qquad a = |\sigma^\circ|. \quad \text{(IV.10)}$$

(9n) $$D\Psi_1 = 4\rho\Psi_1 \rightarrow \Psi_1 = \Psi_1^\circ/R^4 \quad \text{(IV.11)}$$

(9a) $$D\xi^i = \rho\xi^i + \sigma\bar{\xi}^i \quad \text{(IV.12)}$$

(12) is equivalent to

$$D\xi^i = \mathbf{M}\xi^i, \quad \text{where} \quad \xi^i = \begin{vmatrix}\xi^i\\ \bar{\xi}^i\end{vmatrix}. \quad \text{(IV.13)}$$

If we let $\xi^i = \mathbf{M}\mathbf{V}^i$, then from (3) and (13) we obtain

$$D\mathbf{V}^i = 0, \qquad \mathbf{V} = \begin{vmatrix}\xi^{\circ i}\\ \bar{\xi}^{\circ i}\end{vmatrix}.$$

Hence

$$\xi^i = (r\xi^{\circ i} - \sigma^\circ\bar{\xi}^{\circ i})/R^2. \quad \text{(IV.14)}$$

Using (II.9c), we can set

$$\xi^{\circ 3} = P^\circ \quad \text{and} \quad \xi^{\circ 4} = iP^\circ. \quad \text{(IV.15)}$$

The main features of succeeding calculations have been illustrated in Sec. II, however, in this case the scalars become increasingly complex and the manipulation of them extremely tedious, though straightforward. The nonradial equations are complicated by having r dependence appear in varied products of R and a logrithmic function $L = \frac{1}{2}\ln(r + a)/(r - a)$, for example L/R^2, Lr/R^2, L/R^4, etc. These products were shown to be linearly independent and thus the matching of coefficients of like products becomes equivalent to matching powers of r. With the exception of ν, Ψ_3, and Ψ_4, all scalars, including the metric components, were obtained in closed form. To avoid lengthy expressions in the final phases of the calculation, we resorted to using power series expansions in r for all equations containing ν, Ψ_3, and Ψ_4. We believe that our use of a power series yields all obtainable information in the lowest powers and that a more exhaustive treatment would be fruitless. Up to the equations

(10a) $$\delta X^i - \Delta\xi^i = (\mu + \bar{\gamma} - \gamma)\xi^i + \bar{\lambda}\bar{\xi}^i, \quad \text{(IV.16)}$$

the Robinson-Trautman solution is obtainable by taking suitable limits as Ψ_1° and $\sigma \rightarrow 0$. However, at this point one is confronted with the following system of equations:

$$\nabla \ln(\Psi_1^\circ\bar{P}\,|\sigma^\circ|^{-3}) = 0, \quad \overline{\nabla} \ln(\Psi_1^\circ\sigma^\circ P^{-3}) = 0 \quad \text{(IV.17)}$$

$$\Psi_1^\circ = |\sigma^\circ|^2 P\nabla \ln(\bar{P}^4/|\sigma^\circ|^2\,\bar{\sigma}^{\circ 2}),$$

where $\nabla \equiv 2(\partial/\partial\zeta)$.

In the Robinson-Trautman limit, Ψ_1° and $\sigma \rightarrow 0$, (IV.17) are identities. On the other hand, if one proceeds to solve these equations under the assumption that $\sigma \neq 0$ and $\Psi_1^\circ \neq 0$, then these equations will eventually restrict the solutions to the following exact metrics which do not include the Robinson-Trautman solution as a limit.

$$\begin{aligned}
g^{22} &= -\frac{2r^2(\zeta\bar{\zeta})^{1/2}}{R^2} + \frac{2rL}{A} + \frac{2r^3A(\zeta^2 + \bar{\zeta}^2)}{R^4} - \frac{4r^2A^2(\zeta\bar{\zeta})^{3/2}}{R^4},\\
g^{23} &= 4A^2(\zeta\bar{\zeta})^{3/2}x^3\left\{\frac{L}{2a^3} - \frac{1}{2a^2R^2}(r - 2a) - \frac{(r - a)}{R^4}\right\},\\
g^{24} &= 4A^2(\zeta\bar{\zeta})^{3/2}x^4\left\{\frac{L}{2a^3} - \frac{1}{2a^2R^2}(r + 2a) - \frac{(r + a)}{R^4}\right\},\\
g^{33} &= -\frac{2(\zeta\bar{\zeta})^{3/2}}{(r + a)^2}, \qquad g^{44} = -\frac{2(\zeta\bar{\zeta})^{3/2}}{(r - a)^2},\\
g^{12} &= 1, \qquad g^{34} = g^{13} = g^{14} = 0
\end{aligned} \quad \text{(IV.18)}$$

where

$$\frac{\partial A}{\partial x^i} = \frac{d^2A}{du^2} = 0 \rightarrow A = \begin{cases}Bu\\ B\end{cases} \quad B \text{ is a real constant}$$

$$a = A(\zeta\bar{\zeta})^{1/2}, \quad x^1 \equiv u, \quad x^2 \equiv r, \quad \zeta = x^3 + ix^4$$

$$R^2 = r^2 - a^2, \qquad L = \tfrac{1}{2}\ln((r + a)/(r - a)).$$

In the above solutions (IV 18), the limit as $A \rightarrow 0$ yields flat space. Some other pertinent variables are

$$\rho = -r/R^2$$

$$\sigma = a/R^2$$

$$\xi^i = (r\xi^{\circ i} - a\bar{\xi}^{\circ i})/R^2 \qquad \xi^{\circ 3} = -i\xi^{\circ 4} = (\zeta\bar{\zeta})^{3/4}$$

$$\psi_1 = 2A^2(\zeta\bar{\zeta})^{3/4}\zeta/R^4$$

$$\psi_2 = -[4A^2(\zeta\bar{\zeta})^{1/2}\zeta^2 L + 2A\zeta^2 r + 4A^2(\zeta\bar{\zeta})^{3/2}]/R^4$$
$$\psi_3 = (1/r^2)[2A_{,1}(\zeta\bar{\zeta})^{1/4}\zeta] + (1/r^3)8A\zeta + \cdots$$
$$\psi_4 = -(1/r^2)[8A_{,1}\zeta\bar{\zeta}] + \cdots .$$

V. CYLINDRICAL RAYS

The cylindrical class of metrics containing hypersurface orthogonal geodesic rays with nonvanishing shear and divergence is characterized in the (NP) formalism by the following

$$\Psi_0 = \rho - \bar{\rho} = 0 \qquad \rho^2 = \sigma\bar{\sigma} \neq 0. \tag{V.1}$$

Initially we have

$$\text{(9e)} \quad D\rho = \rho^2 + \sigma\bar{\sigma} \rightarrow \rho = (R^\circ - 2r)^{-1} \tag{V.2}$$

$$\text{(9f)} \quad D\sigma = 2\rho\sigma \qquad \rightarrow \sigma = \sigma^\circ(R^\circ - 2r)^{-1}. \tag{V.3}$$

One can, using transformations (II.9a) and (II.10b), readily set

$$\rho = \sigma = -1/2r. \tag{V.4}$$

All the remaining scalars were obtained in closed form in a manner analogous to that used in Secs. III and IV.

The solution does not depend on x^4 and contains two arbitrary constants, a and b.

$$\begin{aligned}
g_{11} &= -4a^2(cn^2 ay)(\ln r)^2 \\
&\quad - [b + a^2 \ln (r^2 cn^4 ay)]/cn^2 ay, \\
g_{12} &= 1, \qquad g_{23} = g_{24} = 0 \\
g_{13} &= -4Y[r + 4a^2 u(cn^2 ay) \ln r] \\
g_{14} &= -2(cn^2 ay) \ln r \\
g_{33} &= -r^2/2 - 4a^2u^2Y^2(cn^2 ay) \ln r \\
g_{34} &= -8uYcn^2(ay) \\
g_{44} &= -cn^2(ay)/a^2
\end{aligned} \tag{V.5}$$

where

$$x^1 \equiv u, \qquad x^2 \equiv r, \qquad x^3 \equiv y,$$

$cn(ay)$ is an elliptic function with modulus $k = 1/\sqrt{2}$,

$$Y = \pm\frac{a(1 - cn^4 ay)^{\frac{1}{2}}}{2\sqrt{2}\, cn(ay)} \qquad \begin{matrix} + \text{ if } y > 0 \\ - \text{ if } y < 0. \end{matrix}$$

One can obtain Sachs' metric[2] by shifting the y origin by K/a, $y \rightarrow y' - K/a$, where $4K$ is the period of $cn(ay, 1/\sqrt{2})$, and then taking the limit $a \rightarrow 0$.

$$\begin{aligned}
&\lim_{a\to 0} \frac{cn(ay' - K)}{a} = y'/\sqrt{2} \\
&\lim_{a\to 0} Y = -(2y')^{-1}.
\end{aligned} \tag{V.6}$$

Dropping the primes and rescaling r and y, one obtains the Sachs' metric directly

$$\begin{aligned}
g_{11} &= -[b + \ln r^2 y^4] \\
g_{13} &= 2r/y, \qquad g_{12} = 1 \\
g_{14} &= g_{23} = g_{24} = g_{34} = 0 \\
g_{33} &= -r^2 \\
g_{44} &= -y^2.
\end{aligned} \tag{V.7}$$

VI. DISCUSSION-CONCLUSION

The metrics obtained in Secs. IV and V (Petrov type I nondegenerate) together with those of Robinson-Trautman and Kundt (degenerate) exhause all metrics that contain hypersurface orthogonal geodesic rays.

Since the Robinson-Trautman solution contains the Schwarzschild metric as a special case, it was anticipated that the solutions to the generalized class containing shearing geodesic rays would enable one to derive physically interesting relations concerning gravitational radiation and loss of mass. To our disappointment we found that this algebraic generalization leads to a very restricted class of metrics (the solutions do not depend on any arbitrary functions and hence are not suited to describe a radiation field). Our spherical solutions do not yield the Robinson-Trautman metrics in the limit of vanishing shear, $\sigma \rightarrow 0$, but in fact as $A \rightarrow 0$ the metric degenerates into flat space.

The application of the Newman-Penrose formulism to the class of metrics containing nonshearing geodesic rays with nonvanishing curl will be given in another paper.

Note added in proof. It now appears as if the nonexistence of physically interesting solutions (in the class described in this paper) is related to the inability of free particles with no structure to undergo self-acceleration.

ACKNOWLEDGMENT

We are indebted to Dr. I. Robinson for having suggested this problem.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

On the Characteristic Initial Value Problem in Gravitational Theory

R. K. SACHS
U. S. Army Signal Research and Development Laboratory, Fort Monmouth, New Jersey
(Received March 14, 1962)

The problem of determining a solution of the Einstein field equations for the gravitational field from data set on a pair of intersecting characteristic (that is, null) hypersurfaces and on their intersection Σ is considered. It is shown that by giving the conformal inner metric of each hypersurface, the inner geometry of Σ, the two mean extrinsic curvatures of Σ, and one additional extrinsic quantity for Σ, one completely determines a solution in a space-time region. In a suitable coordinate system giving the above data amounts to giving four functions of three variables, four functions of two variables, and one function of one variable; all these can be given without any constraints. In a given space time with given Σ the functions are almost, but not entirely, known functions of their arguments. All the derivations and discussions are applicable only in a sufficiently small four-dimensional region.

INTRODUCTION

THE Einstein field equations for the gravitational field have various pathological properties similar to, but more complicated than the pathologies that arise in electrodynamics as a result of gauge invariance.[1] In particular, if one tries to specify some gravitational field by giving information about the field and its first time derivatives at one time (that is, on a space-like three-dimensional hypersurface), then one runs into constraint equations analogous to the constraint equation div $\mathbf{E} = 0$ of electrodynamics. These constraint equations are not easy to handle, and although various methods have been developed for dealing with them,[2] it would be over optimistic to claim that completely satisfactory results have been obtained. A closely related problem is the following: By judicious choice of the initial surface or of the (otherwise arbitrary) coordinates within this initial surface, it may be possible to change or even set to zero much of the initial data without in any way altering the physical meaning of the corresponding gravitational field. It is generally agreed that the physically meaningful data on an initial hypersurface consist of only four functions of the three coordinates within the initial hypersurface. But unless one goes through very complicated elimination processes, one is forced to work instead with 12 functions—the metric of the initial hypersurface and the extrinsic curvature[3] (second fundamental form)—subject to four differential constraints of the type discussed above and subject to change by various coordinate transformations.

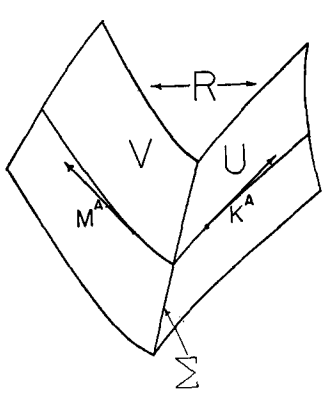


FIG. 1. The hypersurfaces U and V. One spatial dimension is suppressed.

Now, recent work by Bondi, Penrose, and others[4] has suggested that in some arguments an improvement is obtained by setting some or all of the initial data on three-dimensional null hypersurfaces (that is, hypersurfaces everywhere tangent to the light cone). In this paper we consider the situation where data are set on a pair of null hypersurfaces U and V, and on their two-dimensional space-like intersection Σ (see Fig. 1). One can then show that the

[1] P. G. Bergmann, Revs. Modern Phys. **33**, 510 (1961), summarizes the main difficulties and gives further references.

[2] Among the more extensive treatments are those given in A Lichnerowicz, *Theories relativistes de gravitation et de l'electromagnitsme* (Masson et Cie, Paris, 1955); P. A. M. Dirac, Phys. Rev. **114**, 924 (1959); and the recent series of papers in the Physical Review by R. Arnowitt, S. Deser, and C. W. Misner, for example, Phys. Rev. **118**, 1100 (1960). A promising approach has very recently been suggested by A. Peres (report to the Stevens conference, December, 1961, unpublished).

[3] Reference books for the various geometrical concepts discussed in this paper are L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, New Jersey, 1949); and J. A. Schouten, *Ricci Calculus* (Springer-Verlag, Berlin 1954).

[4] Some, though not all, of R. Penrose's results are contained in the preprint "Null Hypersurface Initial Data for Classical Fields of Arbitrary Spin and for General Relativity" (1961). H. Bondi, M. van der Burg, and A. Metzner, Proc. Roy. Soc. (London), (to be published); R. K. Sachs, *ibid.* (to be published). E. Newman and T. Unti (unpublished); A. Peres (unpublished).

data needed to determine a field from the field equations are quite simple, have direct geometrical interpretations, and are not subject to differential or algebraic constraints of any kind. Moreover, in a given gravitational field the initial data are comparatively (though not completely) unique; their values in a given field can be computed by comparatively, though not completely, straightforward manipulations.

For the two null hypersurfaces U and V one must specify the "conformal" inner metric. In a conformal geometry one can assign a meaning to the angle between two directions at a point and also compare two lengths at the same point, but one cannot make a meaningful comparison between lengths at different points.[3] In analytic terms, to determine the conformal geometry of any space one gives its metric tensor, but only up to a factor that is arbitrary and may be position dependent. It turns out that to give the inner conformal metric of U and V is equivalent, in suitable coordinate systems, to giving two functions of three variables on each of the hypersurfaces U and V—four functions of three variables in all. In addition to the above data one must specify the entire inner geometry of the intersection surface Σ, the two mean extrinsic curvatures[3] that Σ has by virtue of being imbedded in a four-dimensional space, and one other extrinsic quantity; in a suitably chosen specialization of the above mentioned coordinate systems, giving these additional data amounts to giving four functions of two variables and one function of one variable.

A comparison between the gravitational field and a string vibrating in one dimension may help clarify the above numerology. To specify the particular vibration which the string is carrying out one can either (a) give the amplitude A of the vibration and $\partial A/\partial t$ at time $t = 0$, or (b) give A on each of two lines U, V tangent to the sound cone (in an x, t diagram) and also give a single number on the intersection point Σ of U and V.[5] One sees from the above discussion that gravitational fields behave similarly but there are the following differences: (a) one works in three spatial dimensions rather than one spatial dimension, (b) the gravitational field has two degrees of freedom rather than one, (c) in the gravitational case the characteristic initial value problem is in some ways simpler than the usual initial value problem. In addition, one would have expected that the gravitational field differs from a rest-mass nonzero Lorentz covariant field with two degrees of freedom by the fact that the gravitational field also contains "longitudinal" modes analogous to the longitudinal modes of the electromagnetic field. Perhaps this difference manifests itself in the scheme that we are here discussing by the fact that one has to set four pieces of data on Σ, but this point is not entirely clear.

Two limitations on the results of this paper are worth emphasizing. First, all considerations are carried through only locally and all assertions in the following development should be qualified by the statement that they hold only for a sufficiently small four-dimensional region. Second, very complicated integration processes for various field variables will be discussed without paying any attention to the question of where, if anywhere, the solutions are finite.

The author would like to emphasize also that he does not at present believe that consideration of the characteristic initial value problem is more fundamental than the consideration of the usual initial value problem. The two seem to him to be mutually complementary alternatives.

The actual calculations and the details of the results presented here are believed to be new. However, many of the main ideas were originally given in the pioneering work of Darmois[6] and in a recent paper by Penrose.[7]

COORDINATE CONVENTIONS[8]

In order to simplify the (in any case clumsy) calculations, a special coordinate system will be introduced. The final results stated in the introduction can be formulated in purely geometrical terms, so the introduction of special coordinate frames is a matter of convenience, not of principle.

Suppose one is given an arbitrary space-time,

[5] Discussions of this problem and of characteristic initial value problems in general can be found in R. Courant and D. Hilbert, *Methoden der Mathematischen Physik* (Verlag Julius Springer, Berlin, 1937), Vol. 2, or in I. G. Petrovsky, *Partial Differential Equations* (Cambridge University Press, New York, 1954).

[6] G. Darmois, *Les equations de la gravitation Einsteinienne* (Mem. Soc. Math., Paris, 1927).

[7] R. Penrose, reference 4, particularly pp. 34 and 35.

[8] The following conventions will be used throughout: Capital Latin letters A, B, etc. run from 2 to 3; small Latin letters a, b, etc. run from 0 to 3. The signature of the metric tensor g_{ab} is taken as $+2$ (so that the metric of a space-like surface is positive definite). Commas denote ordinary derivatives; semicolons denote covariant differentiation with respect to the metric of the four-space; colons denote covariant differentiation of a geometrical object within Σ with respect to the metric of Σ. R_{ab} is the Ricci tensor of the four-space; R is the curvature scalar of the four-space, $u = x^0$ and $v = x^1$ once we have gone over to the specially adapted coordinate systems discussed in this section.

g^{ab}, g^{AB}, h^{AB}, and C_A are defined by the equations

$$g^{ab}\,g_{bc} = \delta^a{}_c,\ g^{AB}\,g_{BC} = \delta^A{}_C,\ h^{AB}\,h_{BC} = \delta^A{}_C,\ C_A = g_{AB}\,C^B.$$

specified by its metric tensor as a function of any coordinates, and in this space-time a pair of null hypersurfaces U and V intersecting in a space-like surface Σ. Within Σ let us choose a pair of coordinates $x^A (A = 2, 3)$ for which the inner metric of Σ takes the form

$$g_{AB} = \exp [2h(x^C)]\ \delta_{AB}, \tag{1}$$

where δ_{AB} is the two-dimensional Kronecker delta. Such coordinates can always be introduced locally.[9] Next, consider the two linearly independent null vectors defined at each point of Σ by the property that they are orthogonal to every direction in Σ. One of these, say k^a (with $a = 0 \cdots 3$, $k^a \neq 0$, $k^a k_a = 0$), must lie within the hypersurface U; at the same time k^a is the normal to the hypersurface U since it is orthogonal to itself and to every other direction in U. The other null vector, say m^a, bears the same relation to the hypersurface V. These two vectors are not defined uniquely at Σ; they are still subject to the scale transformations

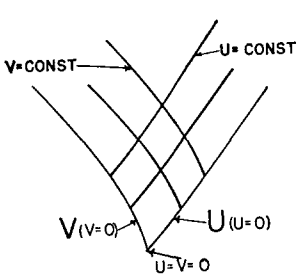
$$k'^a = rk^a, \qquad m'^a = sm^a, \tag{2}$$

where r and s are any functions of the coordinates x^A in Σ. Therefore, we can and shall demand

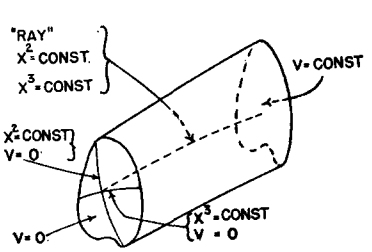
$$m^a k_a = -1 \quad \text{on} \quad \Sigma \tag{3}$$

without loss of generality.

Next consider the lines within U that are orthogonal to U; these will be called rays. They are automatically geodesics[10]; therefore, one can introduce along each of them a preferred parameter,[10] which will be called v. One can then obtain everywhere in U a normal vector field k'^a by the requirement $dx^a/dv = k'^a$ along any ray in any coordinate system x^a. If one now demands that $k'^a = k^a$ and $v = 0$ on Σ, one has fixed the preferred parameters v completely for a given choice of k^a on Σ; it will be supposed that these conventions have been made. Consider now the two-dimensional space-like surfaces $v = \text{const}$ within the hypersurface U. Except for U itself, there is one and only one null hypersurface which passes through any one of these two-dimensional surfaces. Let us call this null hypersurface in four-space the hypersurface $v = \text{const}$. In this way one extends the definition of the quantity v to a four-dimensional region once v is determined on U. The reader will note that the hypersurface V is now characterized by the equation $v = 0$, and this fact will be used very often in the following arguments.



(a)



(b)

FIG. 2. The coordinate system adapted to the hypersurfaces U and V. 2 (a) A surface $x^2 = \text{const}$, $x^3 = \text{const}$. 2 (b) A hypersurface $u = \text{const}$.

One can now repeat all the above manipulations on the hypersurface V instead of the hypersurface U. One obtains another quantity, say u, which has the following properties: (a) within V, u is a preferred parameter along each of the geodesics that are both normal to V and lie in V; (b) the hypersurfaces $u = \text{const}$ in four-space are null hypersurfaces and the hypersurface U is characterized by the equation $u = 0$; (c) $dx^a/du = m^a$ at Σ. Since both u and v are now defined in a four-dimensional region one can now define m^a and k^a in a four-dimensional region by the equations[8]

$$k_a = u_{,a} \quad , \quad m_a = v_{,a}. \tag{4}$$

Finally, we extend the definition of x^A to a four-dimensional region by the requirements

$$x^A_{,a} m^a = 0 \quad \text{in} \quad V \qquad x^A_{,a} k^a = 0 \quad \text{everywhere.} \tag{5}$$

The geometrical picture of the four quantities u, v, and x^A is given in Fig. 2.

Since we have defined four scalar functions, one can use these as coordinates and it is easy to show that the metric in terms of the new coordinates takes the form

$$ds^2 = -e^{-2q}\, du\, dv + g_{AB}(dx^A + C^A\, du)(dx^B + C^B\, du), \tag{6}$$

where q, g_{AB}, and C^A are any six functions of the four coordinates. They obey the following restrictions:

$$\begin{aligned} q &= 0 \quad \text{on} \quad U \quad \text{and on} \quad V \\ C^A &= 0 \quad \text{on} \quad V \\ g_{AB} &= e^{+2h}\, \delta_{AB} \quad \text{in} \quad \Sigma. \end{aligned} \tag{7}$$

[9] L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton University Press, Princeton, New Jersey, 1940), p. 203.

[10] P. Jordan, J. Ehlers, and R. Sachs, Akad. Wiss. Mainz 1 (1961).

Conversely, it is easy to show that whenever one writes a metric in the form (6) and (7), the coordinates u, v, and x^A automatically have all the geometric properties discussed above.

THE FIELD EQUATIONS

Next let us consider what restrictions the Einstein field equations $R_{ab} = 0$ place on the six unknowns in Eq. (6). By standard manipulations one finds for the Christoffel symbols the values given in the Appendix. Using the Christoffel symbols one can work out the Ricci tensor. It will be convenient to divide the ten field equations into four groups as follows:

hypersurface equations;	$R_{11} = R_{1A} = 0$
propagating equations;	$R_{AB} = 0$
subsidiary equations;	$R_{00} = R_{0A} = 0$
trivial equation;	$R_{01} = 0.$

The terminology is similar to that previously used by Bondi and by the author. From the Bianchi identities $(R^{ab} - 1/2g^{ab}R)_{;b} = 0$ one infers by the usual arguments[11] the following lemma[12]: If the three hypersurface equations and the three propagating equations hold everywhere in a sufficiently small four-dimensional region R, bounded from below by U and V, and if the three subsidiary conditions hold on the hypersurface V, then the three subsidiary equations and the trivial equation automatically hold everywhere in R.

It will be convenient to start by discussing how the integration of the subsidiary equations hold on the hypersurface V can be performed. Let us split the quantity g_{AB} into two parts

$$g_{AB} = e^{2h}h_{AB} \tag{8}$$

by the requirement determinant $(h_{AB}) = 1$. Define quantities g^{AB}, h^{AB}, and C_A as follows:

$$g^{AB}g_{BC} = \delta^A_C \quad h^{AB}h_{BC} = \delta^A_C \quad C_A = g_{AB}C^B. \tag{9}$$

One finds that the first subsidiary condition $R_{00} = 0$ takes on V the form[13]

$$-2h_{,00} - 2(h_{,0})^2 + h^{AE}_{,0}h_{AE,0} = 0$$
$$(h_{,0} \equiv \partial h/\partial u \text{ etc}). \tag{10}$$

One sees from Eq. (9) that if h_{AB} is given on V, h is given on Σ and $h_{,0}$ is given on Σ then the value of h throughout V is completely determined; in fact, Eq. (10) enables us to integrate one ordinary differential equation along each line $x_A = \text{const}$ within V to get h.

The remaining two subsidiary conditions have on V the form

$$C_{A,01} + JC_{A,1} + J' = 0,$$
$$C_{A,1} \equiv \partial C_A/\partial v \quad \text{etc.} \tag{11}$$

Here J and J' are "junk factors"—that is, quantities which can be considered as known functions at any one stage of integrating the field equations. In the case of Eq. (11) the junk factors have the following property: If $g_{AB} \equiv e^{2h}h_{AB}$ is known on the hypersurface V, then J and J' are also known on this hypersurface. Since h_{AB} on V has already been assumed as part of the initial data, and since Eq. (10) has already been integrated to obtain h on V, one may consider J and J' known. One can now integrate Eq. (11) to obtain the value of $C_{A,1}$ everywhere on V, provided one gives as additional data the values of $C_{A,1}$ on Σ. In the following discussion it will be supposed that the integrations in Eqs. (10) and (11) have been carried out explicitly. Thus at this stage one has the following data everywhere on V (including Σ):

$$q = C_A = 0 \quad \text{on} \quad V, \tag{12}$$

with h, h_{AB}, and $C_{A,1}$ known functions on V.

Having satisfied the subsidiary conditions on V, one turns next to the hypersurface equations. The equation $R_{11} = 0$ has the form

$$-2h_{,11} - 4h_{,1}q_{,1} - 2(h_{,1})^2 + h^{AE}_{,1}h_{AE,1} = 0. \tag{13}$$

Since $q = 0$ on U, one can use Eq. (13) to integrate for h in terms of h_{AE} on U and $h_{,1}$ on Σ; it will be supposed that this integration has been done. Moreover, on any other hypersurface $u = \text{const} \neq 0$, one can use Eq. (11) to integrate for q provided the expansion $k^a_{;a} = e^{2q}h_{,1}$ is different from zero (compare footnote 12) and g_{AB} is known on the hypersurface $u = \text{const} \neq 0$. The necessary initial value is the value $q = 0$ on V. The next two hypersurface equations have the form

$$C_{A,11} + JC_{A,1} + J' = 0. \tag{14}$$

This time the junk is characterized by the fact that if h, h_{AB}, and q are known on any one hypersurface $u = \text{const}$, then J and J' are known on this hypersurface. Thus, one can find C_A on such a hypersurface provided one knows in addition the initial values of

---

[11] Compare Lichnerowicz,[2] and the references given in footnote 4.

[12] The lemma holds only if the quantity $k^a{}_{;a}$, usually called the expansion, is different from zero. One can always arrange for $k^a{}_{;a} \neq 0$ to hold in some four-dimensional region by suitable choice of Σ. An extensive discussion of spaces in which $k^a{}_{;a} = 0$ for some k^a has been given, for example, by P. Jordan, J. Ehlers, and R. Sachs.[10]

[13] The reader should not try to evaluate the form of any of the subsidiary conditions without first setting $v = 0$ so that $C_A = 0$. The form of the subsidiary conditions when $v \neq 0$ is very complicated; this form is not needed in the following argument because of the lemma stated above.

C_A and $C_{A,1}$ on V. But $C_A = 0$ on V and $C_{A,1}$ has already been obtained there by integrating Eq. (11).

One now proceeds as follows: Imagine that Eqs. (13) and (14) have been solved to yield q and C_A as some complicated functionals of h and h_{AB}. Consider now the three propagating equations, which have the form

$$g_{AB,01} + Jg_{AB,0} + J' = 0. \tag{15}$$

In Eq. (15) the junk terms J and J' are characterized as follows: If all six metric components are known on a hypersurface $u = \text{const}$, then J and J' are known there. According to the remarks made about Eqs. (13) and (14), this fact means that when h and h_{AB} are known on a hypersurface $u = \text{const}$, then J and J' are known there—admittedly, only as quite complicated and nonlocal functionals of h and h_{AB}. One therefore sees that Eq. (15) determines the time derivatives $g_{AB,0}$ at $u = \text{const}$ in terms of g_{AB} at $u = \text{const}$ and in terms of the value $g_{AB,0}$ on the surface $u = \text{const}$, $v = 0$. Since, as one sees from Eq. (12), the latter has been previously obtained, the propagating equations enable one to determine the values of g_{AB} throughout some four-dimensional region R in terms of all the initial data that have been set. Then q and C_A are also determined in R by inserting the values of g_{AB} into the proper functionals.

From the lemma and the discussion of the field equations one sees that at this stage all the components of the Ricci tensor vanish in R and all the metric components are known in R. Collecting results one has the main theorem:

Given h_{AB} on U and on V and given h, $C_{A,1}$, $h_{,0}$ and $h_{,1}$ on Σ, there is precisely one gravitational field (in a sufficiently small four-dimensional region R bounded from below by U and by V) which has these initial data as boundary values and obeys the field equations.

The data can be chosen quite arbitrarily and are not subject to constraints. As stated in the introduction, the question of how large R is (and in particular if it is different from zero!) has not been carefully examined.

GEOMETRICAL INTERPRETATION OF THE INITIAL DATA

After the ugly and formal manipulations of the previous section it is a pleasure to turn to the next question—what do the data that have to be given mean geometrically? The interpretation of h_{AB} is straightforward. Note first that the entire inner metric of either U or V is known if and only if g_{AB} is known. The reason one only needs a 2×2 metric tensor to specify the inner metric is that the hypersurfaces U and V are null hypersurfaces. Thus, if one were to adopt a different system of coordinates and obtain some 3×3 metric in, say, U with the values $g_{\alpha\beta}(\alpha, \beta = 1 \cdots 3)$ one would immediately find that this metric possesses a null eigenvector. That is, one would find a direction $k^\alpha \neq 0$ in U for which $g_{\alpha\beta}k^\beta = 0$. Thus, of the six components of $g_{\alpha\beta}$, only three would be algebraically independent, and these three would be essentially g_{AB}. Stated somewhat differently, g_{AB} gives the distance not between two neighboring points in U but between two neighboring rays at a point in U and the latter is the only meaningful distance that can be assigned. Now the data given consist not of g_{AB} but only of h_{AB}, so that one initially knows g_{AB} only up to the factor $e^{2\lambda}$ which must be obtained by integrating the field equations $R_{00} = 0$ and $R_{11} = 0$ on V and on U, respectively. Thus one sees that giving h_{AB} on U and on V is completely equivalent to specifying the inner conformal geometry[3] of these two hypersurfaces. Note that the conformal metric must be given as a function of a preferred affine parameter[3] along the rays.

To give h on Σ is fully equivalent to giving the inner geometry of Σ, as one sees from Eq. (8). The meaning of $h_{,1}$ and $h_{,0}$ is obtained by introducing for Σ the two extrinsic curvatures E_{AB} and F_{AB} that any two-space imbedded in a four-space has.[3] Usually, in a positive definite space, these forms are not uniquely defined; they are still subject to orthogonal rotations of the normal directions which define each.[3] In the case of a normal hyperbolic metric it is convenient to choose the two null directions k^a and m^a as the defining normal directions. Then (in the special coordinate systems that have been introduced)[14]

$$E_{AB} = k_{A;B} \qquad F_{AB} = m_{A;B}. \tag{16}$$

the abovementioned arbitrariness of an orthogonal rotation now manifests itself by the fact that one can still perform the transformations

$$k'^a = rk^a, \qquad m'^a = r^{-1}m^a \Rightarrow E'_{AB} = rE_{AB},$$
$$F'_{AB} = r^{-1}F_{AB} \quad \text{on} \quad \Sigma. \tag{17}$$

This arbitrariness will be discussed in more detail later. For the moment we note from the Christoffel symbols of the Appendix that to give $h_{,1}$ and $h_{,0}$

[14] If one introduces an arbitrary coordinate system, E_{AB} and F_{AB} retain their meaning but are no longer related to k_a and m_a in the simple way given by Eq. (16); the covariant derivatives appearing in Eq. (16) are those of the four-space.

on Σ, merely amounts to giving the trace (with respect to the inner metric of Σ) of each of the two second fundamental forms:

$$E_{AB}g^{AB} = h_{,1} \qquad F_{AB}g^{AB} = h_{,0} \quad \text{on} \quad \Sigma. \tag{18}$$

These two traces are called the mean extrinsic curvatures; they give information about the extent to which a geodesic in Σ deviates from being a geodesic in the full four-space.[3]

Finally, one can consider the two quantities $C_{A,1}$. One sees from the Christoffel symbols that in the special coordinates

$$C_{A,1} = k_{a;A}m^a \quad \text{on} \quad \Sigma. \tag{19}$$

It follows from this equation and Eq. (17) that one can still alter $C_{A,1}$ on Σ by adding an arbitrary gradient:

$$C'_{A,1} = C_{A,1} - (\ln r)_{,A} \tag{20}$$

Thus the geometrically meaningful quantity is

$$C_{A,1B} - C_{B,1A} = E_{AC}F^C_B - E_{BC}F^C_A + R^{ab}{}_{\cdot\cdot AB}k_a m_b \quad \text{on} \quad \Sigma. \tag{21}$$

Here R_{abcd} is the full Riemann tensor of the four-space. There is only one independent quantity defined by Eq. (21) since both sides are antisymmetric in A and B. It is easy to show that one can set $C^A_{,1\,;A} = 0$ on Σ, where a colon denotes the covariant derivative with respect to the metric of Σ. Then r is fixed up to a solution of the homogeneous Laplace equation on Σ and $C_{A,1}$ is fixed up to the gradient of this solution by giving the geometric quantity (21). To specify a particular solution of the homogeneous Laplace equation, one must give one additional function of a single variable (namely, the relevant boundary values); this function and the quantity given in Eq. (21) constitute the remaining independent data that must be set in place of $C_{A,1}$.

Finally, one can ask the question: To what extent do the initial data consist of "true observables?" In other words, suppose someone were to give us a metric expressed in terms of some arbitrary coordinate system. To what extent could we work out the values of all the above quantities as functions of their arguments. From the discussion given above it is easy to see that one must make the following arbitrary choices: one must choose, *ad hoc*, some particular two-dimensional surface Σ; in addition one must choose, *ad hoc*, some particular coordinates in Σ which express the conformal flatness of Σ in the form given by Eq. (1) (as is well known, this choice amounts to choosing a particular solution of the Laplace equation in Σ).[3] Then the numerical values of all the functions given as data are in principle determined.

To summarize the results of this section: The data that have to be set have rather straightforward geometric interpretations; the functions constituting the data are not true observables but the amount of arbitrariness in them is relatively small.

CONCLUSIONS

A discussion of the local characteristic initial value problem in gravitational theory has been given. One sees that apart from questions of mathematical rigor, a complete scheme can be obtained. There are two very difficult and interesting questions that have not been discussed: First, is it easier to quantize the theory of gravitation by using commutation rules set on null hypersurfaces rather than space-like ones, or are there two essentially equivalent schemes, or is it easier to use space-like hypersurfaces? Second, is it possible to relate in a neat way the local methods discussed here to the more difficult and physically more interesting global methods that various authors[15] have recently introduced. Until one can answer the first of these questions, one cannot say whether the use of null initial data is simply a matter of taste, as in the classical theory, or is in principle different from the use of standard initial data. Unless one can answer the second question positively one will probably not be able to draw many useful conclusions about the results of physically conceivable experiments from the results presented here.

ACKNOWLEDGMENTS

The author thanks J. Ehler, W. Kundt, and E. Schucking of Hamburg University; H. Bondi and F. A. E. Pirani of London University; A. Komar of Syracuse University; E. Newman of Pittsburgh University; and A. Schild of Texas University for discussions. This work was strongly influenced by several very pretty ideas and results stated by R. Penrose of London University in private conversations. The advice and encouragement of W. Mc Afee, U. S. Army Signal Research and Development Laboratory were invaluable.

APPENDIX

The Christoffel symbols

$$\Gamma^a_{bc} = \tfrac{1}{2}g^{ad}(g_{bd,c} + g_{cd,b} - g_{bc,d}) \tag{A1}$$

[15] V. Fock, *Theory of Space, Time, and Gravitation* (Moscow 1955); R. Arnowitt, S. Deser, and C. W. Misner[2]; Bondi and his co-workers[4]; R. K. Sachs.[4]

are needed to calculate the Ricci tensor

$$R_{ab} = \Gamma^c_{ab,c} - \Gamma^c_{ac,b} + \Gamma^d_{dc}\Gamma^c_{ab} - \Gamma^c_{da}\Gamma^d_{cb}. \qquad \text{(A2)}$$

In the coordinate system (6) they are:

$$\Gamma^0_{a1} = 0$$

$$\Gamma^0_{00} = +\tfrac{1}{2}e^{2q}[2(e^{-2q})_{,0} + (C^A C_A)_{,1}]$$

$$\Gamma^0_{0A} = \tfrac{1}{2}e^{2q}[(e^{-2q})_{,A} + C_{A,1}]$$

$$\Gamma^0_{AB} = \tfrac{1}{2}e^{2q}g_{AB,1}$$

$$\Gamma^A_{11} = 0 \qquad \Gamma^A_{1B} = \tfrac{1}{2}g^{AD}g_{DB,1}$$

$$\Gamma^A_{10} = \tfrac{1}{2}g^{AD}[C_{D,1} + (e^{-2q})_{,D}]$$

$$\Gamma^A_{00} = \tfrac{1}{2}g^{AD}[2C_{D,0} - (C^A C_A)_{,D}] - \tfrac{1}{2}e^{2q}C^A[2(e^{-2q})_{,0} + (C^A C_A)_{,1}]$$

$$\Gamma^A_{0B} = \tfrac{1}{2}g^{AD}[g_{DB,0} + C_{D,B} - C_{B,D}] + \tfrac{1}{2}e^{2q}C^A[-(e^{-2q})_{,B} - C_{B,1}]$$

$$\Gamma^A_{BC} = \overset{(2)}{\Gamma}{}^A_{BC} - \tfrac{1}{2}e^{2q}C^A g_{BC,1}$$

$$\Gamma^1_{11} = 2q_{,1}$$

$$\Gamma^1_{10} = -\tfrac{1}{2}e^{2q}C^A_{,1}C_A - q_{,A}C^A$$

$$\Gamma^1_{A1} = -\tfrac{1}{2}e^{2q}[C_{A,1} - (e^{-2q})_{,A}] + \tfrac{1}{2}e^{2q}C^D g_{DA,1}$$

$$\Gamma^1_{00} = -\tfrac{1}{2}e^{2q}[(C^A C_A)_{,0} - 2C^D C_{D,0} + C^D(C_A C^A)_{,D}]$$

$$\Gamma^1_{0A} = -\tfrac{1}{2}e^{2q}[(C^B C_B)_{,A} - C^D\{g_{AD,0} + C_{D,A} - C_{A,D}\}]$$

$$\Gamma^1_{AB} = -\tfrac{1}{2}e^{2q}[C_{A,B} + C_{B,A} - g_{AB,0} - 2C_D\overset{(2)}{\Gamma}{}^D_{AB}]$$

$$\Gamma^b_{ab} = (-2q + 2h)_{,a}$$

Here $\overset{(2)}{\Gamma}{}^A_{BC}$ are the Christoffel symbols of the two dimensional spaces $u = \text{const}$, $v = \text{const}$.

Note added in proof. The author has learned that a paper with similar title has been submitted to the L. Witten volume on general relativity by Professor Bruhat; in case there is any overlapping of subject material it should be pointed out that Professor Bruhat's work began as early as 1958 and was submitted long before the submission of the present paper, even though publication of the present paper may happen to take place first. The essential results of this paper were obtained simultaneously and independently by G. Dautcourt in his Ph.D. thesis (Humboldt University, Berlin, 1961–62; to be published, 1963), which also contains an exhaustive discussion of the special case $k^a_{;a} = 0$ mentioned in footnote 12.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

A Note on the Scattering of Boson Fields*

DAVID SHALE
Department of Mathematics, University of California, Berkeley, California
(Received February 2, 1962)

A physically operational and mathematically simple definition is given for the convergence of states for a boson field. The relevant formulation of the scattering matrix is discussed. As an example, the scattering of a free Boson Field by a given linear time-independent unquantized source is treated. It is shown that cases usually called "infrared" and "ultraviolet" catastrophes are actually convergent.

1. INTRODUCTION

FIELD theories which attempt to derive asymtotic behavior from the dynamics specified for all finite times consider limits of the kind $w_\pm = \lim (t \to \pm\infty) w(t)$, where

$$w(t) = \exp [iH_{\text{total}}\, t] \exp [-iH_{\text{free}}\, t].$$

Two difficulties are:

(A.) Supposing H_{total} can be formulated at all, H_{free} and H_{total} still cannot be formulated as operators on the same Hilbert space in physically interesting cases.[1]

(B.) Even in the mathematically most tractable case[2] when H_{free} and H_{total} can be formulated as self-adjoint operators on the same Hilbert space, taking $w_\pm = \lim (t \to \pm\infty) w(t)$, with convergence in the weak operator topology, still requires at least a finite renormalization to obtain a unitary Møller wave operator.

The work of Segal[3-6] has, among other things, gone a considerable way towards resolving (A), which is clearly the more important difficulty. This paper arose from considering (B). Using the work of Segal we formulate the notion of asymptotic convergence of states in a physically reasonable and mathematically simple way which avoids questions of renormalization and unitarity. Consideration is limited to boson fields whose formal total Hamiltonian is at most quadratic. As an application we extend the work of Cook to treat scattering by a wide class of given linear sources. In particular we give a mathematically rigorous treatment of scattering by a point source. Physically the examples are uninteresting since the scattering matrix is 1. Mathematically, however, they offer further hope that the usual formalism may be altered so as to provide a convergent theory.

2. BASIC DEFINITIONS

We reproduce some basic definitions from reference 4 and 5. Let H, to be called the space of wave functions, be a complex pre-Hilbert space with inner product $(\cdot\ ,\ \cdot)_c$. The single particle structure (or classical field) over H, $\Sigma(H)$, is the pair $\{K, B(\cdot\ ,\ \cdot)\}$ where K is H restricted to real operations and with inner product $(\cdot\ ,\ \cdot) = \text{Re}\ (\cdot\ ,\ \cdot)_c$ and $B(\cdot\ ,\ \cdot) = \text{Im}\ (\cdot\ ,\ \cdot)_c$. The real operator in K corresponding to iI is written Λ.

Our basic example is the case of the Klein-Gordon equation of mass $m \geq 0$. H can be taken as the space of complex wave functions $\psi(R)$ on the hyperboloid $(k, k) + m^2 = 0$ in momentum space, which are square integrable with regard to the Lorentz invariant measure $d^3k/|k_0|$ and satisfy the "reality" condition $\psi(-k) = \psi(k)^*$. (* = complex conjugation.) If $\epsilon(k) = \pm 1$ as $k_0 > 0, < 0$. The operator Λ is given by $\Lambda\psi(k) = i\epsilon(k)\psi(k)$.

A quantization over $\Sigma(H)$ is a map $z \to V(z)$ from K to unitary operators on a complex Hilbert space $\mathfrak{H}$ such that:

$$V(z)V(z_1) = \exp [\tfrac{1}{2}iB(z, z_1)]V(z + z_1), \qquad (1)$$

and $z \to V(z)$ is weakly continuous on every finite-dimensional subspace of H. [In a formal way $V(\psi(x)) = \exp [i \int \Phi(x)\psi(x)\, dx]$, where $\psi(x)$ is a classical wave function and $\Phi(x)$ is the "field at x."]

The algebra of field observables $\mathfrak{A}$ is the uniform closure of $U\mathfrak{A}_M$ where M ranges over the finite-dimensional subspaces of H, and $\mathfrak{A}_M$ is the weakly closed ring of operators generated by $\{V(z) : z \in M\}$.

A regular state E is a positive linear functional on $\mathfrak{A}$ normalized so that $E(I) = 1$, and such that for some quantization $V(\cdot)$ there is a vector x such that $E(A) = (Ax, x)$ for all $A \in \mathfrak{A}$. Pure regular

* Prepared with partial support from the National Science Foundation, Contract G19136.

[1] L. Van Hove, Physica **18**, (1952).
[2] J. M. Cook, J. Math. Phys. **2**, 33 (1961).
[3] I. E. Segal, Ann. Math. **48**, 930 (1942).
[4] I. E. Segal, Kgl. Danske Videnskab. Selskab, Mat-fys Medd. **31**, No. 12.
[5] I. E. Segal, Can. J. Math. **13**, 1 (1961).
[6] I. E. Segal, "The Mathematical Characterization of the Physical Vacuum" (to be published).

states E and F are relatively normalizable if they are determined by vectors x and y in the same irreducible quantization space.

3. OPERATIONAL CONVERGENCE OF STATES

Definition 1. A one-parameter family of regular states E_t converges operationally to a regular state E if for each field observable X, $E_t(X) \to E(X)$.

Mathematically this amounts to topologizing the regular states as the weak dual of $\mathfrak{A}$. Let $p(E, F)$ denote the transition probability between states E, F. It follows readily that when H has finite dimension $E_t \to E$ operationally if and only if $p(E_t, F) \to p(E, F)$ for every fixed state F. Now in the infinite dimensional case, since quantization is not unique, the appropriate definition of $p(\cdot\ ,\ \cdot)$ is not entirely clear. It appears reasonable, however, to put $p(E, F) = 0$ if, E and F are pure and not relatively normalizable. Granting this the example given elsewhere[7] shows that there exists a one-parameter family E_λ of regular states, which is continuous by Theorem 1 below, but with $p(E_\lambda, E_\mu) = 0$ if $\lambda \neq \mu$.

Again as an example below (Sec 8) shows, the physical vacuum state cannot be expected, in general, to be relatively normalizable with regard to the free field vacuum. [This is the difficulty of 1. (A).] Consequently if finite times are to be considered, it cannot be expected that a pure "in" state which is relatively normalizable with regard to the free field vacuum will be normalizable with regard to E_t for finite times t. Hence for a fixed state F the transition probability $p(E_t, F)$ will not be a continuous function of time. We claim further that transition probabilities between states do not play any operational role and could be omitted from the theory entirely. We conclude the present section with a brief discussion of this claim.

For systems of finitely many degrees of freedom where irreducible quantization is unique up to unitary equivalence a certain amount of semantic confusion between "state" and "observable" is permissable. Let A be a non-negative self-adjoint operator of trace class and trace 1, then A represents both an observable and the state $E : E(X) = \mathrm{Tr}\ (XA)$. Then if F is any other state, $p(F, E) = F(A)$. For a system with infinitely many degrees of freedom let A, as above, act on the space of an irreducible quantization. Then E is still a state, but A is not a field observable; for it may be seen that A has no representation in any quantization arising from a pure regular state, not relatively normalizable with respect to those arising from vectors in the Hilbert space on which A acts. Thus for fields, "states" and "observables" must be sharply distinguished.

Now it may be argued that what is measured in a sequence of identical scattering experiments is the expected values, or rather the distribution of values, of certain observables in the "in" and "out" states, i.e., given a self-adjoint field observable A and a state E what is actually determined experimentally is a certain borel measure $\mu_{A,E}$ on the real line and concentrated on the spectrum of A.[8-10] If the system has finitely many degrees of freedom and A has pure point spectrum $\{\lambda_i\}$ of multiplicity 1, with x_i the eigenvector corresponding to λ_i, then $\mu_{A,E}(\lambda_i) = p(E, F_i)$ where F_i is the state determined by x_i. This accounts for the historical importance of transition probabilities. The method cannot be applied in the cases usual in field theory where A has continuous spectrum or parts of infinite multiplicity. For a self-adjoint observable A it is simpler to compute $\psi_{A,E}$ the characteristic function (Fourier transform) corresponding to the measure $\mu_{A,E}$. This is given by $\psi_{A,E}(\lambda) = E(\exp [iA\lambda])$.

Now let E_t be a one-parameter family of regular states continuous in the sense of definition 1. For fixed A denote the characteristic functions by $\psi(t, \lambda)$. By definition 1, for each fixed λ, $\psi(t, \lambda)$ is continuous in t, i.e., definition 1 implies continuity of the measures μ_{A,E_t} in a sense common in probability theory.

4. CONVERGENCE THEOREM FOR STATES

The characteristic function $\rho(\cdot)$ of a regular state E is defined by $\rho(z) = E(V(z))$ for all z in the space of wave functions H.[11,12]

It is proved[11] that $\rho(\cdot)$ is a characteristic function if and only if $\rho(z)$ is continuous on every finite-dimensional subspace M of H, $\rho(0) = 1$ and

$$\sum_{ij} a_i a_j^* \exp [(i/2)B(z_i, z_j)]\rho(z_i - z_j) \geq 0 \cdots , \quad (2)$$

for complex $a_1, \cdots, a_n$ and $z_1, \cdots, z_n$ in H. The property (2) will be referred to as quasi-positive definiteness.

The following analog of a well-known theorem of

[7] D. Shale, Trans. Amer. Math. Soc. **103**, 149 (1962), Remark 6.1.

[8] G. W. Mackey, "Lecture Notes on the Mathematical Foundations of Quantum Mechanics," Harvard, 1960 (unpublished).

[9] See reference 3.

[10] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, (Princeton University Press, Princeton, New Jersey, 1955).

[11] See reference 5.

[12] H. Araki, J. Math. Phys. **1**, 492 (1960).

Levy[13] provides a tool for checking whether states converge operationally.

Theorem 1. Let E_t be a one-parameter family of regular states with characteristic functions $\rho_t(\cdot)$. Then there exists a regular state E such that $\lim (t \to t')E_t = E$ (operational convergence), if and only if $\rho_t(z)$ converges pointwise to a function $\rho(z)$ which is continuous at the origin on every finite-dimensional subspace M of the space of wave functions H and then $\rho(z)$ is the characteristic function of E.

Proof. If $E_t \to E$ then the conclusion is trivial. We prove the converse first in the case when dim H is finite. Suppose that the $\rho_t(z)$ converge pointwise to $\rho(z)$ continuous at the origin.

A function $\eta(z)$ on K is integrally quasi-positive definite (q.p.d.) if

$$\int \exp [\tfrac{1}{2}iB(z, z')]\eta(z - z')f(z)f(z')^* \, dz \, dz' \geq 0$$

where $f(z)$ is any Lebesgue integral function on K and the integral is over $K \times K$ with regard to Lebesgue measure.

Now the functions $\rho_t(z)$ being q.p.d. are integrally q.p.d. Hence, $\rho(z)$ is measurable, continuous at the origin and, by the Lebesgue dominated convergence theorem, integrally q.p.d. Hence,[14] there exists a continuous q.p.d. function $\nu(z)$ with $\nu(0) = 1$ such that $\nu(z) = \rho(z)$ a.e. Let E be the regular state determined by $\nu(z)$.

Now let $V(z)$ be the irreducible quantization. Let C_0 be the complex vector space generated by

$$\left\{V(f) = \int V(z)f(z) \, dz \colon f(z) \in L_1(K)\right\}.$$

Then C_0 is an algebra whose uniform closure is C the algebra of completely continuous operators.[14] $E_t(V(f))$ may be evaluated as $\int \rho_t(z)f(z) \, dz$. It follows that for every X in C_0, (and hence every X in C), $E_t(X) \to E(X)$.

Let A_t(respectively A) be the non-negative self-adjoint operator of trace class such that $E_t(X) = \mathrm{Tr}\,(A_tX)$ [respectively, $E(X) = \mathrm{Tr}\,(AX)$]. We wish to show that $E_t(X) \to E(X)$ for all bounded operators X. It is sufficient to establish this when X is self-adjoint. Given X there exists a completely continuous operator Y such that $X + Y$ is self-adjoint with pure point spectrum.[15] Consequently we can suppose that X has pure point spectrum. Let $\{e_n\}$ be an orthonormal basis so that $Xe_n = \lambda_n e_n$, for all n. Let P_n be the projection with range e_n. Then for all n,

$$(A_te_n, e_n) = \mathrm{Tr}(A_tP_n) = E_t(P_n) \to E(P_n) = (Ae_n, e_n).$$

Also $(A_te_n, e_n) \geq 0$, $(Ae_n, e_n) \geq 0$ and

$$\sum_n (A_te_n, e_n) = \sum_n (Ae_n, e_n) = 1.$$

It follows readily that

$$E_t(X) = \sum_n \lambda_n(A_te_n, e_n) \to \sum_n \lambda_n(Ae_n, e_n) = E(X).$$

We have established that $E_t \to E$. It follows that $\nu(z) = \rho(z)$ everywhere.

In the general case, dim H infinite, the above shows that $\rho(z)$ is the characteristic functional of a regular state E and that $E_t(X) \to E(X)$ for all tame field observables X (i.e., $X \in \mathfrak{A}_M$ for some finite subspace M of H). The theorem follows since tame observables are uniformly dense in $\mathfrak{A}$.

5. FORMULATION OF THE SCATTERING AUTOMORPHISM

We now attempt to give some minimal requirements under which a system may be said to undergo convergent scattering. Our treatment is partly adapted from the work of Segal.[16]

From the present point of view, given a fixed frame of reference, it is the states which evolve with time. The evolution will be given by a one-parameter group $\mathfrak{D}(t)$ mapping the regular states into themselves and such that:

(a) For every pair of regular states E, F and $a \geq 0$, $b \geq 0$ and $a + b = 1$, $\mathfrak{D}(t)(aE + bF) = a\mathfrak{D}(t)E + b\mathfrak{D}(t)F$.

(b) $\mathfrak{D}(t)$ is continuous in the sense that for every regular state E, $\mathfrak{D}(t)E$ is operationally continuous.

Any *-automorphism φ of the algebra $\mathfrak{A}$ of field observables determines a contragredient map φ^* of the regular states given by: $\varphi^*(E)(X) = E((\varphi^{-1}X))$ for all X in $\mathfrak{A}$. A one-parameter family $\varphi(t)$ of *-automorphisms is continuous if $\varphi^*(t)$ is continuous in the above sense.

We now assume that the total (respectively free) field dynamics arise from a continuous one-parameter group of *-automorphisms $\varphi(t)$ [respectively, $\varphi_0(t)$|. Let $w(t) = \varphi(t)\varphi_0(-t)$. The dynamics of the total field is asymptotic to that of the free field if the Møller wave automorphisms $w_\pm = \lim (t \to \pm\infty)w(t)$ exist. The scattering automorphism is then $s = w_+^{-1} \cdot w_-$.

[13] M. Loeve, *Probability Theory* (D. Van Nostrand Company, Princeton, New Jersey, 1960), 2nd ed., p. 191.

[14] See reference 5, proof of Theorem 2.

[15] J. von Neumann, "Charackterisierung des Spectrums eines Integral-operatoren," *Actualités Scientifiques et industrielles* (Hermann & Cie., Paris, 1935) p. 229.

[16] See references 4–6.

Now a free field posesses a vacuum state E_0 and hence has a quantization associated with it. Considerations of positivity of the free field energy[17] have made the Fock quantization[18] the most likely candidate to describe the free field. We shall assume this if the final condition on s is that $s^*(E_0) = E_0$. It follows that there then exists a unitary operator S, the "s matrix," such that $s(X) = S.X.S^{-1}$.

6. ADMISSABLE WAVE FUNCTIONS

We now consider H the pre-Hilbert space of normalizable wave functions introduced in Sec. 2. To emphasize its dependence on H we write $\mathfrak{A}(H)$ for the algebra of field observables. For a given differential equation specifying the free field let $\bar{H}$ be the corresponding Hilbert space of normalizable wave functions. Usually H is taken to be $\bar{H}$. If instead H is a proper dense submanifold, then $\mathfrak{A}(H)$ is a proper subalgebra of $\mathfrak{A}(\bar{H})$ (see Sec. 7). However the space of symmetric tensors $S(H)$[19] on which the free field quantization acts, is identical with, $S(\bar{H})$. Consequently, if the dynamics is given as an automorphism of $\mathfrak{A}(H)$ instead of $\mathfrak{A}(\bar{H})$, the s matrix S is unchanged.

Associated with the free field quantization there are certain observables (e.g., free field energy, total number of particles), which are not field observables and need not have any meaning at finite times, but which can presumably be measured in the "in" and "out" states. Now $\bar{H}$ is the subspace of one particle states in $S(H)$. However, only states which assign finite expectation values to observables like the "free field energy" can possibly be attained in practice. Accordingly it seems appropriate to discard the others from the space of admissable wave functions, and hence, restrict the observables for which the dynamics need be specified, should it be convenient to do so. We extend this idea in the following:

Definition 2. Let G denote the proper inhomogeneous Lorentz group and $\mathfrak{g}$ its lie algebra. Let $U(\cdot)$ [respectively, $dU(\cdot)$] be the canonical representation of G (respectively, $\mathfrak{g}$) on $\bar{H}$. A linear submanifold H will be called admissable if there is a submanifold H_0 with $H_0 \subseteq H$, H_0 dense in $\bar{H}$, invariant under $U(\cdot)$ and $dU(\cdot)$, and such that $dU(X)$ is essentially skew adjoint on H_0 for all X in $\mathfrak{g}$.

The existence of H_0 has been established by Nelson.[20]

[17] See reference 6.
[18] J. M. Cook, Trans. Amer. Math. Soc., **74**, (1953).
[19] See reference 18.
[20] E. Nelson, Ann. Math. **70**, 591 (1959).

7. AUGMENTED SYMPLECTIC GROUP

In order to construct examples we give an extension of the symplectic group over the single particle structure $\Sigma(H)$ introduced by Segal.[21] We begin by noting that the commutation relations [Eq. (1)], can be regarded as arising from a unitary representation $V^\sim(\cdot)$ of the group $K^\sim = \{z^\sim = (a, z)$: a is real and $z \in K\}$ with multiplication law:

$$(a, z)(a_1, z_1) = (a + a_1 + \tfrac{1}{2}B(z, z_1), z + z_1) \cdots \quad (3)$$

(cf. reference 5), where

$$V^\sim(z^\sim) = e^{ia}V(z) \cdots . \quad (4)$$

For any finite dimensional subspace M of K let $M^\sim$ be the subgroup $\{z^\sim : z \in M\}$. $M^\sim$ is to be given the Euclidean space topology.

The proper augmented symplectic group given in the following definition describes the group of dynamical transformations arising from Hamiltonians which are at most quadratic.

Definition 3. The augmented symplectic group $Sp(K)^\sim$ is the group of automorphisms $T^\sim$ of $K^\sim$ such that $T^\sim : M^\sim \to T^\sim(M^\sim)$ is continuous for each finite dimensional subspace M. $C = \{(a, 0)\}$ is the center of K. The proper augmented symplectic group $ASp(K)$ is the subgroup which leaves C pointwise fixed.

Since each $T^\sim \in Sp(K)^\sim$ leaves C invariant, the continuity condition implies $T^\sim(a, 0) = (\alpha a, 0)$ with α fixed. Let $\varphi(T^\sim)V^\sim(z^\sim) = V^\sim(T^\sim z^\sim)$. By considering elements of $\mathfrak{A}(H)$ of the form cI, c complex, it becomes clear that $\varphi(T^\sim)$ determines an automorphism of $\mathfrak{A}(H)$ if and only if $T^\sim$ leaves C pointwise fixed, i.e., $T^\sim \in ASp(K)$.

Theorem 2. For $T^\sim \in ASp(K)$ there exists a linear functional λ on K and a linear transformation T on K which preserves $B(\cdot\ ,\ \cdot)$, such that $T^\sim(z^\sim) = (a + \lambda(z), Tz)$. Conversely, every pair (λ, T) determines a unique $T^\sim \in ASp(K)$.

Proof. Writing $T^\sim(0, z) = (\lambda(z), T(z))$ the identity:

$$T^\sim((a, z)(a_1 z_1)) \equiv T^\sim(a, z) \cdot T^\sim(a_1, z_1)$$

gives

$$T(z + z_1) = T(z) + T(z_1) \cdots \quad (5a)$$

$$\tfrac{1}{2}B(T(z), T(z_1)) - \tfrac{1}{2}B(z, z_1) = \lambda(z + z_1) - \lambda(z) - \lambda(z_1) \cdots . \quad (5b)$$

Since the right side of (5b) (left side) is symmetric (antisymmetric) in z, z_1 both must equal zero. It

[21] See reference 4.

follows from Eqs. (5) that $T(\alpha z) = \alpha T(z)$ and $\lambda(\alpha z) = \alpha\lambda(z)$ for all rational α and hence all real α by continuity. This establishes the direct part. The converse is trivial.

For $T^\sim : T^\sim(a, z) = (a + \lambda(z), Tz)$ we write $T^\sim = (\lambda, T)$. If E is a regular state with characteristic functional $\rho(z)$ then $\varphi^*(T^\sim)E$ has characteristic functional

$$\rho'(z) = \exp(-i\lambda(T^{-1}z))\rho(T^{-1}z) \cdots . \tag{6}$$

For

$$\begin{aligned}\rho'(z) = \varphi^*(T^\sim)E[V(z)] &= E[\varphi(T^{\sim -1})V^\sim(0, z)] \\ &= E\{V^\sim[-\lambda(T^{-1}z), T^{-1}z]\} \\ &= \exp[-i\lambda(T^{-1}z)]\rho(T^{-1}z).\end{aligned}$$

Let $V_0(\cdot)$ be the free field quantization and E_0 the free field vacuum state. We take H (and hence K) to be a Hilbert space.

Corollary. For $T^\sim$ in $ASp(K)$ there is a unitary operator Y such that $\varphi(T^\sim)X = Y.X.Y^{-1}$, or equivalently E_0 and $\varphi^*(T^\sim)E_0$ are relatively normalizable if and only if $T^\sim = (\lambda, T)$ with λ continuous on K in the Hilbert space topology and $(T^*T)^{1/2} - I$ is Hilbert-Schmidt. Further if $T^\sim = (\lambda, I)$, Y can be taken to be $V(z_1)$ where z_1 is the unique vector in $\bar{K}$ such that $\lambda(z) = -B(z, z_1)$.[22]

Proof. $T^\sim = (\lambda, T) = (0, T)(\lambda, I)$. Since the part concerning $(0, T)$ has been proved elsewhere[23] we consider only (λ, I). By Eq. (6) the characteristic functional of $E_{T\sim} = \varphi^*(T^\sim)E_0$ is $\rho'(z) = \exp[-i\lambda(z)]\rho(z)$. If $E_{T\sim}$ and E_0 are relatively normalizable then $E_{T\sim}$ arises from a vector x in the Hilbert space of the free field quantization and $\rho'(z) = (V(z)x, x)$. Since $V(\cdot)$ here is continuous from the Hilbert-space topology on K to the weak operator topology it follows that $\lambda(\cdot)$ is continuous. It is straightforward to show that $\lambda(z)$ is continuous on K if and only if $\lambda(z) = -B(z, z_1)$ for a unique z_1 in $\bar{K}$. Finally the commutation relations [Eq. (1)] give

$$\begin{aligned}V(z_1)\cdot V(z)\cdot V(-z_1) &= \exp[-iB(z, z_1)]V(z) \\ &= \varphi(T^\sim)V(z).\end{aligned}$$

To see, as remarked in Sec. 6, that $\mathfrak{A}(H_1)$ is not equal to $\mathfrak{A}(H)$ for H_1 a proper linear submanifold of H, it is only necessary to note that there exists a nonzero linear functional λ on H (not necessarily continuous) which is zero on H_1. Putting $T^\sim = (\lambda, I)$ we have $\varphi(T^\sim) = I$ on $\mathfrak{A}(H_1)$ but not on $\mathfrak{A}(H)$.

[22] Cf. I. E. Segal, Trans. Amer. Math. Soc. **88**, 12 (1958), example 2, p. 33.

[23] See reference 7.

8. SCATTERING OF A FREE BOSON FIELD BY A TIME INDEPENDENT LINEAR UNQUANTIZED SOURCE.

As an example we reformulate the treatment of scattering with a given source due to Cook,[24] and extend it to give a convergent treatment of cases usually labeled as "infrared" and "ultraviolet" catastrophes.

Let $\mathfrak{H}$ denote the space on which the free field quantization $V_0(z)$ acts, and $\bar{H}$ be the space of normalizable wave functions. For $z \in \bar{H}$ let $iR_0(z)$ be the infinitesimal generator of $V_0(tz) - \infty < t < \infty$. The canonical action $\Gamma(\cdot)$ of unitary group $U(\bar{H})$ on $\mathfrak{H}$ will be given explicitly in Sec. 9. If A is a self-adjoint operator on $\bar{H}$ then $d\Gamma(A)$ is the infinitesimal generator of $\Gamma(\exp(iAt))$, $-\infty < t < \infty$. If A is the one-particle Hamiltonian, $d\Gamma(A)$ is the free field Hamiltonian. Recall that Λ is the real form of iI. The basic result due to Cook[24] is:

Theorem 3. If A is a self-adjoint but not necessarily bounded on $\bar{H}$ and $z \in \mathfrak{D}(A)$ then

$$\begin{aligned}[d\Gamma(A) + R_0(\Lambda Az) + \tfrac{1}{2}(Az, z)]^\sim \\ = V_0(z)\, d\Gamma(A)\, V_0(z)^{-1}.\end{aligned}$$

We give a new proof in Sec. 9.

Consideration will be limited to a Klein-Gordon field of mass $m \geq 0$. The free particle Hamiltonian A is multiplication by k_0 on the space of normalizable wave functions on $(k, k) + m^2 = 0$ in momentum space. The following lemma is a simple consequence of the Fubini Theorem and the Riemann-Lebesgue Lemma.

Lemma A. If $f(k)$ is an integrable complex-valued function on $(k, k) + m^2 = 0$, $m \geq 0$ with regard to the measure $dm = d^3\mathbf{k}/|k_0|$, then $F(t) = \int \exp(ik_0t)\cdot f(k)\, dm \to 0$ as $t \to \pm\infty$.

Now let $\varphi(x)$ be a time independent source function. Let $\hat{\varphi}(k) = \hat{\varphi}(k_0, \mathbf{k})$ be the Fourier transform $\int \varphi(x) \exp(-i(k, x)\, d^3x$, regarded as a function on $(k, k) + m^2 = 0$.

Case I (Cook). $\hat{\varphi}(k)$ is normalizable and in $\mathfrak{D}(A^{-1})$. The total Hamiltonian is.

$$\begin{aligned}[d\Gamma(A) + R(\Lambda\hat{\varphi}) + \tfrac{1}{2}(A^{-1}\hat{\varphi}, \hat{\varphi})]^\sim \\ = V_0(A^{-1}\hat{\varphi})\cdot d\Gamma(A)\cdot V_0(A^{-1}\hat{\varphi})^{-1} \cdots ,\end{aligned} \tag{7}$$

using Theorem 3.

Reformulating in the present language we take the space of admissible wave functions to be $\bar{H}$. The field dynamics is given by the automorphism $\varphi_0(t) =$

[24] See reference 2, Theorem 1.

$\varphi(T(t))$ with $T(t)$ the symplectic transformation exp (iAt). Exponentiating Eq. (7) and applying the corollary to Theorem 2 we get the interacting dynamics given by the automorphism $\varphi(t) = \varphi(\lambda, I)\varphi_0(t)\varphi(\lambda, I)^{-1}$ where (λ, I) is augmented symplectic transformation determined by the linear function $\lambda(z) = -B(z, A^{-1}\hat{\varphi})$. Now the Møller wave automorphism $w(t) = \varphi(t)\varphi_0(-t) = \varphi(T_0(t))$ where $T_0(t) = (\lambda(\cdot) - \lambda(T(-t)(\cdot)), I)$. Applying Eq. (6), if $\rho(z)$ is the characteristic functional of the regular state E and $\rho'(z)$ that of $w^*(t)E$, we have

$$\rho'(z) = \rho(z) \exp [-i\lambda(z) + i\lambda(T(-t)z)]. \quad (8)$$

It follows that

$$w_{\pm} = \varphi(\lambda, I), \quad (9)$$

provided $\lambda(T(-t)z) \to 0$ as $t \to \pm\infty$ and then the scattering automorphism $s = w_+^{-1}w_- = I$.

Now

$$\lambda(T(-t)z) = -B(\exp(-iAt)z, A^{-1}\hat{\varphi})$$

$$= -\operatorname{Im} \int \exp(-ik_0t)[z(k)\hat{\varphi}(k)^*/k_0]\, dm. \quad (10)$$

By Lemma A this $\to 0$ as $t \to \pm\infty$.

Case II. Suppose the Fourier transform $\hat{\varphi}(k)$ of the source is normalizable but $\hat{\varphi} \notin \mathfrak{D}(A^{-1})$. This can happen if $m = 0$ when A^{-1} is unbounded. The difficulty with the above formulation is that $\lambda(z) = -\operatorname{Im} \int [z(k)\hat{\varphi}(k)^*/k_0]\, dm$ will not be defined for all normalizable wave functions $z(k)$. Accordingly, we apply the doctrine enunciated in Sec. 6 and take H to be the subspace of $\bar{H}$ consisting of infinitely differentiable functions with compact support on $(k, k) + m^2 = 0$ which does not contain 0. Then H may be seen to satisfy the general requirements of definition 2. In addition, $z(k)/k_0$ is normalizable for all $z \in H$ and the remainder of the reformulated discussion is unchanged. We remark that in this case it follows from the corollary to Theorem 2 that the clothed vacuum $w_{\pm}^{\#}E_0$ is not relatively normalizable with regard to the free field vacuum state.

Case III. The Fourier transform $\hat{\varphi}(k)$ is not a normalizable wave function. For example let $\varphi(x)$ be the δ function at x_0 corresponding to a point source. This is the so-called "ultraviolet" catastrophe. Then $\hat{\varphi}(k) = \exp [-i(k_1x_1 + k_2x_2 + k_3x_3)]$. We take the space of admissible wave functions to be as in case II. Since $x(k)\hat{\varphi}(k)^*/k_0$ is then integrable for $z \in H$ the reformulated treatment is unchanged.

Our aim in treating the above examples has been to establish the viability of the mathematical framework given for scattering. It may be observed that the maneuver used to treat the "divergences" amounts to requiring that when the source is highly concentrated the wave functions used (on space time) must be spread out. There is no difficulty in extending the treatment to a wide class of interactions for which the interaction Hamiltonian is at most quadratic. In particular the infrared, catastrophe which arises for the Maxwell field with a given time dependent unquantized source can be treated by adapting the maneuver given by Cook.[2]

ACKNOWLEDGMENTS

The author wishes to acknowledge his debt to Professor I. E. Segal and to Dr. J. M. Cook for a number of stimulating conversations on field theory as well as his debt to their publications referred to in the text. He wishes also to thank Professor J. Feldman for many helpful conversations.

APPENDIX. THE HOLOMORPHIC QUANTIZATION AND A NEW PROOF OF COOK'S THEOREM

The holomorphic quantization of Segal[25] has appeared since the work of Cook.[26] It has the advantage that field quantities may be computed very simply. We use it to give a new proof of Cook's theorem (Theorem 3 above).

We begin by reproducing the basic definitions. Let H be a complex Hilbert space and $\Sigma(H) = \{K, B(\cdot\ ,\ \cdot)\}$ the single particle structure over H. Let $L_2(K)$ be the Hilbert space of square integrable functions over K with regard to the normal distribution with variance 1. Let $f(t_1, \cdots, t_n)$ be a complex polynomial in n variables and $z_1, \cdots, z_n \in H$. The antiholomorphic polynomial (a.h.p.) determined by f and $z_1 \cdots z_n$ is

$$f(z) = f(t_1, \cdots, t_n), \quad \text{with} \quad t_i = (z_i, z), \quad (11)$$

where $(\cdot\ ,\ \cdot)$ is now the inner product in H. Let $\mathfrak{H}$ be the subspace (closed) of $L_2(K)$ determined by the a.h.p. The quantization on $\mathfrak{H}$ is determined by

$$V_0(z')f(z) = \exp [-(z', z')/4 - \tfrac{1}{2}(z', z)]f(z + z') \quad (12)$$

where f ranges over the a.h.p.

Let $\mathfrak{E} = \{g(t_1 \cdots t_n) = f(t_1, \cdots, t_{n-1}) \exp(-\alpha t_n)\}$, where g is a complex polynomial, $t_i = (z_i, z)$, $i = 1 \cdots n$ (arbitrary $n > 0$), and $z_1, \cdots, z_n \in H$ are arbitrary. It may be seen that $\mathfrak{E}$ is a dense subspace of $\mathfrak{H}$.

The canonical action $\Gamma(\cdot)$ of the group $U(H)$ of unitary operators on H is given by $\Gamma(u)f(z) =$

[25] See reference 6.
[26] See reference 2.

$f(u^{-1}z)$ where f ranges over the a.h.p. If $U(H)$ is given the weak operator topology, $\Gamma(\cdot)$ is weakly continuous. For $g \in \mathfrak{E}$, $\Gamma(u)g(z) = g(t_1, \cdots, t_n$ with $t_i = (uz_i, z)$. Finally recall that if A is self adjoint on H, $id\Gamma(A)$ is the infinitesimal generator of $\Gamma[\exp(iAt)]$.

Lemma 1. (a) On $\mathfrak{E}$, $V_0(z')$ is given by

$$V_0(z')g(z) = g(t_1, \cdots, t_n) \times \exp(-(z', z')/4 - (z', z)/2)$$

with $t_i = (z_i, z) + (z_i, z')$, $i = 1 \cdots n$.

(b) If $iR_0(z')$ is the infinitesimal generator of $V_0(tz')$, on $\mathfrak{E}$, $R_0(z')$ is given by

$$R_0(z')g(z) = \sum_i -i\frac{\partial g(t_1, \cdots, t_n)}{\partial t_i}\cdot(z_i, z') + (i/2)(z', z)g(t_1, \cdots, t_n)$$

with $t_i = (z_i, z')$ $i = 1 \cdots n$.

(c) Let A be self adjoint on H. Let $\mathfrak{D}$ be the set of functionals in $\mathfrak{E}$ for which the z_i's lie in $\mathfrak{D}(A)$ the domain of A. Then $\mathfrak{D}$ lies in the domain of $d\Gamma(A)$ and

$$d\Gamma(A)g(z) = \sum_i \frac{\partial g(t_1, \cdots, t_n)}{\partial t_i}\cdot(Az_i, z)$$

with $t_i = (z_i, z)$.

Proof. Part (a) follows readily from the definition. Proof is omitted.

(b) For g in $\mathfrak{E}$ consider

$$h_t = (1/t)[V_0(tz')g - g] - iR_0(z')g.$$

As $t \to 0$ $h_t \to 0$ pointwise. Now suppose g depends on $z_1 \cdots z_m$. Let H' be the subspace generated by $z', z_1, \cdots, z_n$ with $e_1, \cdots e_r$ an orthonormal basis for H'. Then for $t \neq 0$, h_t is an antiholomorphic function based on H'. It may be seen that for suitable $M > 0$, $|h_t| < M \exp(\beta\sum_i|(z, e_i)|^2)$ for all $\beta > 0$. If K' is the real subspace determined by H', it follows by the Lebesgue dominated theorem that $h_t \to 0$, as $t \to 0$, in $L_2(K')$ (integral with regard to Gaussian measure with variance 1). Therefore $h_t \to 0$ in $L_2(K)$.

(c) It is convenient, though inessential, to establish part (c) using the duality transform $D^\sim$ between $S(H)$ a space of symmetric tensors over H and $\mathfrak{H}$. We define $D^\sim$ by

$$D^\sim(z_1 \otimes \cdots \otimes z_n)_s = (n!)^{-1}\prod_i(z_i, z).$$

If $\Gamma'(\cdot)$ is the canonical action of $U(H)$ on $S(H)$ then clearly $D^\sim$ transforms Γ' and $d\Gamma'$ into Γ and $d\Gamma$, respectively. For $z_1, \cdots, z_n$ in $\mathfrak{D}(A)$ it is a straightforward matter to check that

$$d\Gamma(A)(z_1 \otimes \cdots \otimes z_n)_s = (Az_1 \otimes z_2 \otimes \cdots \otimes z_n)_s + (z_1 \otimes Az_2 \otimes z_3 \otimes \cdots \otimes z_n)_s + \cdots.$$

It follows that for monomials (and hence for polynomials) based on $\mathfrak{D}(A)$, $d\Gamma(A)$ has the form stated. To treat the general case suppose $g \in \mathfrak{D}$, with $g(t_1, \cdots, t_n) = f(t_1, \cdots t_{n-1})\exp(t_n)$. Put

$$g_m = f\left(\sum_{r=0}^{m} t_n^r/r!\right).$$

Now using the Lebesgue dominated convergence theorem and arguing as in the proof of (b) we obtain $g_m(z) \to g(z)$ and $d\Gamma(A)g_m(z) \to d\Gamma(A)g(z)$ as $m \to 0$, with convergence in $L_2(K)$. The result follows.

The following lemma,[27] which is probably not new, proves to be a convenient tool for establishing that certain operators are essentially self-adjoint.

Lemma 2. Let A be a self-adjoint unbounded operator on a complex Hilbert space H and $U(t) = \exp(iAt)$. Let $\mathfrak{D}$ be a dense submanifold of H such that (1) $\mathfrak{D} \subseteq \mathfrak{D}(A)$ (the domain of A), (2) $\mathfrak{D}$ is invariant under $U(t)$ for $-\infty < t < \infty$; then the restriction of A to $\mathfrak{D}$ is essentially self-adjoint.

Proof. Let $\mathfrak{D}_\pm = \{Ax \pm ix : x \in \mathfrak{D}\}$. It is sufficient to show that the closure $\mathfrak{D}_\pm^\sim$ of $\mathfrak{D}_\pm$ is H. We establish this for $\mathfrak{D}_+$. Suppose $\mathfrak{D}_+^\sim \neq H$. Then $\mathfrak{D}_+^\sim$ is invariant under $U(t)$, $-\infty < t < \infty$, and hence reduces A. Hence $(A + iI)^{-1}$ maps $\mathfrak{D}_+$ into $\mathfrak{D}_+^\sim$. Therefore, $\mathfrak{D} = (A + iI)^{-1}\mathfrak{D}_+ \subseteq \mathfrak{D}_+^\sim$. It follows that $\mathfrak{D}$ is not dense in H. Contradiction.

Proof of Theorem 3. In order to establish the theorem in the form stated it is enough to find a dense domain $\mathfrak{D}$ on which

$$d\Gamma(A) + R_0(\Lambda Az') + \tfrac{1}{2}(Az', z') = V_0(z')\, d\Gamma(A)\, V_0(-z'),$$

and such that the right side is essentially self-adjoint. Now take $\mathfrak{D}$ as in Lemma 1, part (c). Then $\mathfrak{D}$ satisfies the requirements of Lemma 2 (with $\mathfrak{H}$ instead of H and $d\Gamma(A)$ instead of A). Then $d\Gamma(A)$ is essentially self-adjoint on $\mathfrak{D}$. Since $V(z')$ and $V(-z')$ leave $\mathfrak{D}$ invariant, $V(z')$ maps $\mathfrak{D}$ onto $\mathfrak{D}$. It follows that $V(z')\, d\Gamma(A)V(-z')$ is essentially self-adjoint on $\mathfrak{D}$. The stated equality on $\mathfrak{D}$ now follows by computation using Lemma 1.

---

[27] *Note added in proof.* Lemma 2 is equivalent to a result of J. L. B. Cooper, Proc. London Math. Soc. **50**, 11 (1948).

Complex Angular Momenta in Potential Scattering

MARCEL FROISSART*†
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received March 5, 1962)

The S matrix associated with a central potential is shown to be meromorphic in the energy and angular momentum variables under very broad conditions. The domain of meromorphy contains the product of a domain in the energy variable by a domain in the angular variable. The former (latter) domain has a very simple connection with the domain of meromorphy of the Laplace (Mellin) transform of the potential with respect to the radius.

I. INTRODUCTION

THIS paper is devoted to an extension of the results obtained by Bottino, Longoni, and Regge,[1] concerning the analyticity properties of the scattering matrix associated with the Schrödinger equation with a central potential. The S matrix is first expressed in Sec. 2 in terms of the Jost functions, which are formally expanded in powers of the potential. This expansion is studied in detail in Sec. 3, and a partial summation of terms corresponding to the Green's function is shown to be possible under very weak conditions upon the potential. This proves the desired properties in a small region. To enlarge this region, it is convenient to study in more detail (Sec. 4) the analytic continuation of the Laplace and Mellin transforms of the Green's function.

This analytic continuation then makes it possible to derive in Sec. 5 the desired properties, by separating out the divergent parts of the Jost functions and expressing them as Laplace and/or Mellin transforms of the Green's function. In the conclusion we gather together all conditions on the potential and the corresponding analyticity properties.

II. FORMAL EXPANSION OF THE S MATRIX

Let the Schrödinger equation read

$$\left[\frac{\partial^2}{\partial r^2} + k^2 - \frac{\lambda^2 - 1/4}{r^2} - V(r)\right]\phi(k, \lambda, r) = 0,$$

where the solution ϕ is the "regular" solution at the origin, or, more precisely, the solution of the integral equation

$$\phi(k, \lambda, r) = j_\lambda(kr) + \frac{1}{k}\int_0^r G_\lambda(kr, kr')\,V(r')\phi(r', k, \lambda)\,dr'. \quad (1)$$

As usual,

$$j_\lambda(kr) = (\tfrac{1}{2}\pi kr)^{1/2}J_\lambda(kr); \qquad n_\lambda(kr) = (\tfrac{1}{2}\pi kr)^{1/2}N_\lambda(kr)$$

$$h_\lambda^{(1)}(kr) = j_\lambda(kr) + in_\lambda(kr); \quad h_\lambda^{(2)}(kr) = j_\lambda(kr) - in_\lambda(kr)$$

$$G_\lambda(kr, kr') = n_\lambda(kr)j_\lambda(kr') - j_\lambda(kr)n_\lambda(kr').$$

The Jost function $f(\lambda, k)$ is defined by the asymptotic behavior for large r of $\phi(k, \lambda, r)$:

$$\phi(k, \lambda, r) \approx (1/2ik)[f(\lambda, k)\,e^{(i\pi/2)(\lambda+1/2)}h_\lambda^{(1)}(kr) - f(\lambda, -k)\,e^{-(i\pi/2)(\lambda+1/2)}h_\lambda^{(2)}(kr)]$$

and the S-matrix element:

$$S(\lambda, k) = e^{2i\delta} = e^{i\pi(\lambda-1/2)}\frac{f(\lambda, k)}{f(\lambda, -k)} = \frac{W_2(\lambda, k)}{W_1(\lambda, k)}, \quad (2)$$

where W_1 stands for:

$$W_1(\lambda, k) = \lim_{r\to\infty} W[h_\lambda^{(1)}(kr), \phi(k, \lambda, r)]; \quad (3)$$

W is the Wronskian, and

$$W_2(\lambda, k) = -[W_1(\lambda^*, k^*)]^*.$$

The solution of Eq. (1) is formally given by

$$\phi(k, \lambda, r) = j_\lambda(kr) + \sum_{n=1}^{\infty}\frac{1}{k^n}\int_{0<r_1<r_2<\cdots<r_n<r_{n+1}=r}\prod_{m=1}^{n}[G_\lambda(kr_{m+1}, kr_m) \times V(r_m)\,dr_m]\,j_\lambda(kr_1)$$

and likewise formally, the expression (3) for W_1 becomes:

$$W_1 = -ik + \sum_{n=1}^{\infty}\frac{1}{k^{n-1}}\int_{0<r_1<\cdots<r_n<\infty} h_\lambda^{(1)}(kr_n) \times \prod_{m=1}^{n-1}G_\lambda(kr_{m+1}, kr_m)\prod_{m=1}^{n}[V(r_m)\,dr_m]j_\lambda(kr_1) \quad (4)$$

---

* On leave of absence from C. E. N. Saclay, France.
† Supported in part by the Air Force Office of Scientific Research, Air Research and Development Command.
[1] Bottino, Longoni, Regge, *Potential Scattering for Complex Energy and Angular Momentum*, (Istituto Di Fisica dell' Università Torino, Torino, Italy, to be published).

by virtue of the following relation for the Wronskian:

$$W[\alpha j_\lambda(kr) + \beta n_\lambda(kr), \gamma j_\lambda(kr) + \delta n_\lambda(kr)] = k(\alpha\delta - \beta\gamma).$$

The fact that Eq. (3) contains the symbol $\lim_{r\to\infty}$ is reflected in the fact that the integral of Eq. (4) extends to infinity.

III. STUDY OF THE CONVERGENCE OF THE FORMAL SERIES

A convenient way to study the series (4) is to carry out all integrations, except those going to zero or to infinity (over r_1 and r_n), which give rise to divergence, then to sum the series, and finally to study the last two integrations. Accordingly, define for $n > 2$

$$I_n(r_1, r_n) = \frac{1}{k^{n-1}} \int_{r_1<\cdots<r_n}^{(n-2)} \prod_{m=1}^{n-1} G_\lambda(kr_{m+1}, kr_m) \times \prod_{m=2}^{n-1} V(r_m)\, dr_2 \cdots dr_{n-1}. \quad (5)$$

An upper bound on this quantity may be put in the following way:

$$I_n(r_1, r_n) = \int_{r_1<s_1<r_2<\cdots<r_{n-1}<s_{n-1}<r_n}^{(2n-3)} \prod_{m=1}^{n-1} G'_\lambda(ks_m, kr_m) \times \prod_{m=2}^{n-1} V(r_m)\, ds_1 \cdots ds_{n-1}\, dr_2 \cdots dr_{n-1} \quad (6)$$

where $G'_\lambda(x, y) \equiv (\partial/\partial x)G_\lambda(x, y)$ satisfies

$$|G'_\lambda(ks, kr)| < M(k, \lambda) \frac{F_{k,\lambda}(s)}{F_{k,\lambda}(r)} \quad \text{for} \quad s > r$$

and:

$$F_{k,\lambda}(s) = s^{|\mathrm{Re}\,\lambda|-1/2} \qquad s \leq 1 \quad |\mathrm{Re}\,\lambda| \geq 1/2 \quad (7a)$$

$$F_{k,\lambda}(s) = 1 \qquad s \leq 1 \quad |\mathrm{Re}\,\lambda| \leq 1/2 \quad (7b)$$

$$F_{k,\lambda}(s) = \exp|\mathrm{Im}\, ks| \qquad s > 1. \quad (7c)$$

(See Appendix for the derivation of this bound.) $F_{k,\lambda}(s)$ is monotonic, therefore:

$$\frac{F_{k,\lambda}(s_m)}{F_{k,\lambda}(r_m)} \leq \frac{F_{k,\lambda}(r_{m+1})}{F_{k,\lambda}(r_m)}.$$

Furthermore, let $U(r) = \max_{s>r} |V(s)|$. $U(r)$ is a monotonically decreasing function, so that $|V(r_m)| < U^{1/2}(r_m)U^{1/2}(s_{m-1})$. By substitution of these different upper bounds into the expression for $I_n(r_1, r_n)$:

$$|I_n(r_1, r_n)| < M_{k,\lambda}^{n-1} \frac{F_{k,\lambda}(r_n)}{F_{k,\lambda}(r_1)} \times \int_{r_1<s_1<r_2<\cdots<r_{n-1}<s_{n-1}<r_n} \prod_{m=2}^{n-1} [U^{1/2}(r_m)U^{1/2}(s_{m-1})] \times dr_2 \cdots dr_{n-1}\, ds_1 \cdots ds_{n-1}.$$

This last integral is equal to:

$$\frac{1}{(2n-4)!} \int_{r_1}^{r_n} \left\{\int_{r_1}^{s_{n-1}} U^{1/2}(r)\, dr\right\}^{2n-4} ds_{n-1} < \frac{1}{(2n-4)!}(r_n - r_1)\left\{\int_0^\infty U^{1/2}(r)\, dr\right\}^{2n-4},$$

if $\int_0^\infty U^{1/2}(r)\, dr < \infty$, which shall be considered to hold throughout the rest of this paper.

Thus, it is possible to sum the series:

$$I(r, r') = \sum_{n=2}^{\infty} I_n(r, r')$$

and it converges uniformly in r, r', k, λ on every compact set. In particular, it is an entire function of k and λ as every term I_n. It may be useful to remark that the function $I(r, r')$ is a Green's function of the total Hamiltonian. The preceding computation gives an upper bound for $I(r, r')$ of the form:

$$|I(r, r')| < N_{k,\lambda} \frac{F_{k,\lambda}(r')}{F_{k,\lambda}(r)}(r' - r). \quad (8)$$

This allows to determine a region where the integral (4) converges uniformly in k and λ, and therefore to establish analyticity in k and λ in some domain for $W_{1(2)}$. However this domain is fairly small, as compared with the domain obtained by Regge *et al.* Therefore it is necessary to devise some analytic continuation procedure, which is developed in the next section.

IV. ANALYTIC PROPERTIES OF THE MELLIN AND LAPLACE TRANSFORMS OF THE GREEN'S FUNCTION

The divergence of the integral (4) is due to the wrong power behavior for small values of r_1 and to the wrong exponential behavior for big values of r_n. It is therefore natural to try and isolate the singularities in λ, k associated with these divergences by studying the analytic properties of Mellin or Laplace transforms of $I(r, r')$. Three cases are to be considered:

a) $r < r' < 1$ $A(s, s')$

$$= \iint_{0<r<r'<1} r^{s-1}r'^{s'-1} I(r, r')\, dr\, dr'$$

b) $r < 1 < r'$ $B(s, p')$

$$= \iint_{0<r<1<r'<\infty} r^{s-1} I(r, r')\, e^{-p'r'}\, dr\, dr'$$

c) $1 < r < r'$ $C(p, p')$

$$= \iint_{1<r<r'<\infty} I(r, r')\, e^{-pr-p'r'}\, dr\, dr'.$$

The limiting value 1 is taken arbitrarily here, and it does not even need to be the same for all three functions, as for any finite range of values of r, r' the integrals converge. As they stand, these functions, according to (8), are analytic in the following regions:

$$\begin{aligned} &\text{A)} \quad \operatorname{Re} s > \max\,[|\operatorname{Re}\lambda| - 1/2, 0] \qquad \operatorname{Re} s' > 1 - \operatorname{Re} s \\ &\text{B)} \quad \operatorname{Re} s > \max\,[|\operatorname{Re}\lambda| - 1/2, 0] \qquad \operatorname{Re} p' > |\operatorname{Im} k| \qquad (9) \\ &\text{C)} \quad \operatorname{Re}(p + p') > 0 \qquad \operatorname{Re} p' > |\operatorname{Im} k|. \end{aligned}$$

A possible way to continue these functions in their arguments is to use the fact that $I(r, r')$ satisfies two differential equations, and therefore, A, B, C satisfy the transformed equations, which may allow to analytically continue them outside of (9).

The differential equations satisfied by $I(r, r')$ are the same with respect to r and with respect to r':

$$\left[\frac{\partial^2}{\partial r^2} + k^2 - \frac{\lambda^2 - 1/4}{r^2} - V(r)\right] I(r, r') = \delta(r - r').$$

The transformed equations read, with the Mellin transform:

$$\begin{aligned} &[(s-1)(s-2) - (\lambda^2 - 1/4)] \\ &\quad \times A(s-2, s') + k^2 A(s, s') - \frac{1}{2\pi i} \\ &\quad \times \int_{-i\infty}^{+i\infty} A(s - \sigma, s') u(\sigma)\, d\sigma = \frac{1}{s + s' - 1}, \qquad (10a) \\ &[(s'-1)(s'-2) - (\lambda^2 - 1/4)] \\ &\quad \times A(s, s'-2) + k^2 A(s, s') - \frac{1}{2\pi i} \\ &\quad \times \int_{-i\infty}^{+i\infty} A(s, s' - \sigma) u(\sigma)\, d\sigma \\ &\quad = \frac{1}{s + s' - 1} + \alpha(s, s'), \qquad (10b) \\ &[(s-1)(s-2) - (\lambda^2 - 1/4)] \\ &\quad \times B(s-2, p') + k^2 B(s, p') - \frac{1}{2\pi i} \\ &\quad \times \int_{-i\infty}^{+i\infty} A(s - \sigma, p') u(\sigma)\, d\sigma = \beta(s, p'), \qquad (10c) \end{aligned}$$

where

$$\begin{aligned} \alpha(s, s') &= -\int_0^1 r^{s-1}\left[\frac{\partial}{\partial r'} I(r, 1) + (1 - s') I(r, 1)\right] dr \\ \beta(s, p') &= -\int_1^\infty e^{-p'r'}\left[\frac{\partial}{\partial r} I(1, r') + (1 - s) I(1, r')\right] dr' \\ u(\sigma) &= \int_0^1 V(r) r^{\sigma-1}\, dr, \end{aligned}$$

and the integral over σ is as yet taken according to the general theory of Mellin transforms, which is possible by virtue of (8) and $\int_0^1 U^{1/2}(r)\, dr < \infty$. Similarly, the Laplace transforms B and C satisfy:

$$\begin{aligned} &(\partial^2/\partial p'^2)[(p'^2 + k^2) B(s, p')] - (\lambda^2 - 1/4) B(s, p') \\ &\quad - \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} B(s, p' - q) \tilde{u}''(q)\, dq = \gamma(s, p') \qquad (11a) \\ &(\partial^2/\partial p^2)[(p^2 + k^2) C(p, p')] - (\lambda^2 - 1/4) C(p, p') \\ &\quad - \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} C(p - q, p') \tilde{u}''(q)\, dq \\ &\quad = \delta(p, p') + \frac{e^{-(p+p')}}{p + p'} \qquad (11b) \\ &(\partial^2/\partial p'^2)[(p'^2 + k^2) C(p, p')] - (\lambda^2 - 1/4) C(p, p') \\ &\quad - \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} C(p, p' - q) \tilde{u}''(q)\, dq = \frac{e^{-(p+p')}}{p + p'} \qquad (11c) \end{aligned}$$

where again:

$$\begin{aligned} \gamma(s, p') &= e^{-p'} \int_0^1 r^{s-1}\left[\frac{\partial}{\partial r'} I(r, 1) + p' I(r, 1)\right] dr \\ \delta(p, p') &= e^{-p'} \int_1^\infty \left[\frac{\partial}{\partial r} I(1, r') + p I(1, r')\right] e^{-p'r'}\, dr' \\ \tilde{u}(q) &= \int_1^\infty e^{-qr} V(r)\, dr. \end{aligned}$$

The integral over q is to be taken along a path satisfying, say, $\operatorname{Re} q > 0$.

It is possible to do the same transformations and to prove that α and γ (β and δ) satisfy equations of type Eq. (10) [Eq. (11)] with an inhomogeneous term that is now an entire function.

The analytic continuation procedure developed here cannot proceed unless some additional conditions on the potential are assumed. The condition $\int_0^\infty U^{1/2}(r)\, dr < \infty$ implies that $r^2 V(r)$ goes to zero with r. It is therefore possible to take $\operatorname{Re}\sigma$ greater

than 2 and arbitrarily close to it in Eq. (10). The present method requires a little more; that is, that there exists a small positive number η such that $r^{2-\eta}V(r)$ goes to zero with r. It is then possible to take the path of the σ integration in Eq. (10) along a line $2 - \eta < \text{Re } \sigma < 2$. This allows to express the first term of Eq. (10) as a function of the others, and therefore allows one to extend domain (9) by a strip of width η in the s or s' plane. This procedure may then be repeated any number of times unless one of the following circumstances occurs [the language adopted is that suitable for Eq. (10a), but it is clearly irrelevant]:

i) The inhomogeneous term is infinite: this generates a pole in $A(s, s')$ at $s = -s' - 1$.
ii) $(s - 1)(s - 2) = \lambda^2 - \frac{1}{4}$: this generates a pole in $A(s, s')$ at $s = \pm\lambda - \frac{1}{2}$.
iii) The second term may have a singularity: this tells us that every singularity in $A(s, s')$ is associated to an infinite chain of singularities of similar type at regular intervals of 2.
iv) The integral over σ may be pinched between a singularity of $u(\sigma)$ and a singularity of $A(s - \sigma, s')$. This shows that if s_0 is a singularity of $A(s, s')$, then $s_0 + \sigma - 2$ is a singularity of $A(s, s')$ whenever σ is a singularity of u. This particular set of points we shall denote by $s_0 + \mathcal{S} - 2$, where $\mathcal{S}$ denotes the set of singular points of $u(\sigma)$. It is important to note that in the case where the pinching is between two poles, the resulting singularity is again a pole, as may be seen by decomposing the contour into a loop around one pole and a contour which is not pinched.
v) The last possibility for the analytic continuation procedure to fail is that the integral over σ starts diverging. This however only occurs for fairly pathological potentials, where the asymptotic behavior of $u(\sigma)$ for $\sigma \to \pm i\infty$ depends upon Re σ. This may be treated together with the preceding case by saying that there is a singularity of $u(\sigma)$ at Re $\sigma + i\infty$. An example of such a potential is:

$$V(r) = e^{-\mu r} \cos (ln^2 r)$$

All these possibilities are then taken together and the singularities of $A(s, s')$ in s are given by the formulas:

$$\text{or} \quad \begin{aligned} s &= -1 - s' - 2n + m(\mathcal{S} - 2) \\ s &= \pm\lambda - 1/2 - 2n + m(\mathcal{S} - 2) \end{aligned} \tag{12}$$

where m, n take all possible non-negative integer values, the symbol $m(\mathcal{S} - 2)$ denoting the set: $\{\sigma_1 + \cdots + \sigma_m - 2m\}$, where $\sigma_1, \cdots, \sigma_m$ are singular points for $u(\sigma)$; it is understood that $(\sigma_0 + i\infty) + (\sigma_1 + i\infty)$ is any number whose real part is the same as that of $(\sigma_0 + \sigma_1)$.

The set of essential singularities of $A(s, s')$ is given by:

$$\begin{aligned} &s = -1 - s' - 2n + m(\mathcal{S} - 2) + (\mathcal{E} - 2) \\ &\text{or} \qquad\qquad (13) \\ &s = \pm\lambda - 1/2 - 2n + m(\mathcal{S} - 2) + (\mathcal{E} - 2) \end{aligned}$$

where $\mathcal{E}$ is the set $\mathcal{S}$ minus the set of all poles of u; in fact, it has been noticed that the only mechanism to produce an essential singularity in $A(s, s')$ is to have an essential singularity in u.

The very same procedure may be used to find the singularities in s' of $A(s, s')$. It is however impossible that there be poles with a fixed location in s', because their residue would then obey an equation of type (10) without right-hand side, and therefore they would appear in the region (9). Therefore, the only singularities in $A(s, s')$ are exactly those already found in (12).

For B, α, and γ the reasoning is similar, but the inhomogeneous term is entire in s. Therefore, the only singularities are those independent of s' in (12).

A completely analogous study may be made to continue in p and p', with the proviso that $V(r)$ decays exponentially, or equivalently, that the integral on q in (11) might be taken inside $-\mu < \text{Re } q < 0$.

In this case, the singularities are created either by the Fuchsian type singularity $p^2 + k^2 = 0$ in Eq. (11), or by the inhomogeneous term, or by the integral on the potential. The reasoning is made along the same lines with the only difference that all singularities appearing are cuts instead of poles. The result is: $C(p, p')$ is analytic except for cuts starting at

$$p' = \pm ik + m\mathcal{P} \quad \text{or} \quad p + p' = m\mathcal{P}, \tag{14}$$

where $\mathcal{P}$ is the set of singularities of $u(q)$; and $B(s, p')$ is analytic in p' except for cuts starting at $p' = \pm ik + m\mathcal{P}$.

A final step in the reasoning is to prove that in the domains determined by Eqs. (12) and (14), the functions A, B, C are analytic in all four variables, as they were in the domains (9). This is done by a straightforward repeated application of a generalized form of Hartog's lemma (See Bochner and Martin,[2] p. 141). It is worth noticing that this theorem is in fact very deep, and that it is perhaps

[2] S. Bochner and W. T. Martin, *Several Complex Variables* (Princeton University Press, Princeton, New Jersey, 1948).

the most powerful tool used in this proof, although it looks almost trivial.

V. SEPARATION OF THE DIVERGENT TERMS

The summation of the series in (4) leads to the formula:

$$W_1(\lambda, k) = -ik + \int_0^\infty h_\lambda^{(1)}(kr)\,V(r)\,j_\lambda(kr)\,dr + \iint_{0<r<r'<\infty} h_\lambda^{(1)}(kr')\,V(r)\,V(r')I(r,r')j_\lambda(kr)\,dr\,dr'.$$

These integrals diverge outside (9), but it is possible to separate explicitly the divergent part in the following way.

For $r < 1$ use:

$$j_\lambda(kr) = \sqrt{\pi}\sum_{n=0}^{N}\frac{(-1)^n}{n!\,n+\lambda!}\left(\frac{kr}{2}\right)^{2n+\lambda+1/2} + \tilde{j}_\lambda(kr), \quad (16)$$

where $\tilde{j}_\lambda(kr)$ is of the order of $r^{2N+\lambda+5/2}$ for small r. For $r > 1$, use the following integral representation for $h_\lambda^{(1)}(kr)$, which can be proved by solving an equation analogous to (11):

$$h_\lambda^{(1)}(kr) = e^{-i(\lambda+1/2)\pi/2} \times \left[e^{ikr} + \frac{1}{ik}\int_{ik}^{-\infty} e^{pr}P'_{\lambda-1/2}\left(\frac{p}{ik}\right)dp\right]$$

where P' is the derivative of the Legendre function of the first kind. Hence,

$$h_\lambda^{(1)}(kr) = e^{-i(\lambda+1/2)\pi/2} \times \left[e^{ikr} + \frac{1}{ik}\int_{ik}^{-A} e^{pr}P'_{\lambda-1/2}\left(\frac{p}{ik}\right)dp\right] + \tilde{h}_\lambda(kr) \quad (17)$$

where $\tilde{h}_\lambda(kr)$ is of the order of e^{-Ar} for large r. A separation of the first integral in (15) into two parts $r \gtrless 1$, followed by the use of (16) and (17) through a suitable expression of $j_\lambda(kr)$ in terms of $h_\lambda^{(1)}(kr)$ and $h_\lambda^{(1)}(-kr)$ for $r > 1$ or of $h_\lambda^{(1)}(kr)$ in terms of $j_\lambda(kr)$ and $j_{-\lambda}(kr)$ for $r < 1$, leads to the evaluation of the following divergent integrals:

$$\int_0^1 r^{2\lambda+1+2n}V(r)\,dr = u(2\lambda + 2n + 2)$$

$$\int_1^\infty e^{(p+p')r}V(r)\,dr = \tilde{u}(-p-p'),$$

with p and p' varying between $-A$ and ik. Therefore, the singularities due to the divergences of the simple integrals are:

$$\lambda = \tfrac{1}{2}(\mathcal{S}-2) - n, \qquad -2ik = \mathcal{P}. \quad (18)$$

Both terms involving the remainders of formulas (16) or (17) are convergent for A and N large enough, therefore the singularities in (18) are the only ones due to the divergence of the simple integral in (4). Of course there is also the branch point for $k = 0$ due to the nonanalyticity of the integrand at that point. The double integral is evaluated along the same lines, dividing it into three parts, according to whether r and r' are $\lessgtr 1$. The separation (16) and/or (17) introduces explicitly the following types of diverging integrals:

(i) $$\iint_{0<r<r'<1} r^{2n+\lambda+1/2}r'^{2n'\pm\lambda+1/2}V(r)\,V(r')I(r,r')\,dr\,dr' = \frac{-1}{4\pi^2}\iint_{-i\infty}^{+i\infty} A(2n+\lambda+3/2-\sigma, 2n' \pm \lambda + 3/2 - \sigma')u(\sigma)u(\sigma')\,d\sigma\,d\sigma'.$$

This integral may be analytically continued as long as [Eq. (12)]:

$$2n + \lambda + 3/2 - \mathcal{S} \neq \pm\lambda - 1/2 - 2n + m(\mathcal{S}-2)$$

and:

$$2n + \lambda + 3/2 - \mathcal{S} + 2n' \pm \lambda + 3/2 - \mathcal{S} \neq -1 - 2n + m(\mathcal{S}-2).$$

It therefore has singularities at

$$\lambda = -n + \tfrac{1}{2}(m+1)(\mathcal{S}-2) \qquad m, n = 0, 1, 2\cdots$$

(ii) $$\iint_{0<r<1<r'<\infty} r^{2n+\lambda+1/2}V(r)\,V(r')e^{-p'r'}I(r,r')\,dr\,dr' = \frac{-1}{4\pi^2}\iint_{-i\infty}^{+i\infty} B(2n+\lambda+3/2-\sigma, p'-q)u(\sigma)\tilde{u}(q)\,d\sigma\,dq.$$

The singularities of this integral are given by:

$$2n + \lambda + 3/2 - \mathcal{S} = \pm\lambda - 1/2 - 2n + m(\mathcal{S}-2)$$
$$p' - \mathcal{P} = \pm ik + m\mathcal{P},$$

p' describes a path going from A to $-ik$. Therefore this integral has the singularities:

$$\lambda = -n + \tfrac{1}{2}(m+1)(\mathcal{S}-2); \quad 2ik = -(m+1)\mathcal{P}$$

(iii) $$\iint_{1<r<r'<\infty} e^{-pr-p'r'}V(r)\,V(r')I(r,r')\,dr\,dr' = -\frac{1}{4\pi^2}\iint_{-i\infty}^{+i\infty} C(p-q, p'-q')\tilde{u}(q)\tilde{\mu}(q')\,dq\,dq'$$

which has singularities at:

$$p' - \mathcal{P} = \pm ik + m\mathcal{P} \quad \text{and} \quad p' + p - 2\mathcal{P} = m\mathcal{P}$$
$$2ik = -(m+1)\mathcal{P}.$$

It is easy to see that, by taking N and A large enough, the remainders of formulas (16) and (17) do not contribute more singularities. The singularities of W_2 are obtained from those of W_1 by complex conjugation. The preceding reasoning also allows us to keep track of the nature of the singularity around $k = 0$ by writing down at each step the circuit relations around $k = 0$. It is easily seen that these circuit relations are exactly those of the formal series (4) and therefore, the formula obtained by Bottino, Longoni, and Regge[1] is proved by this method:

$$S(\lambda, ke^{i\pi}) = \frac{-e^{-2i\pi\lambda}}{S(\lambda, k) - 2\cos\pi\lambda e^{-i\pi\lambda}}. \tag{19}$$

CONCLUSION

We have reached the conclusion that, subject to the conditions $r^{2-\eta}e^{\mu r}V(r)$ bounded from zero to infinity, η and μ arbitrary, positive, defining $u(\sigma) = \int_0^1 r^{\sigma-1}V(r)\,dr$ and $\tilde{u}(q) = \int_1^\infty e^{-qr}V(r)\,dr$ then, the S-matrix element is meromorphic in the (λ, k) plane except for the following points:

$$\begin{aligned} &\lambda = -n + \tfrac{1}{2}m(\mathcal{S} - 2) + \tfrac{1}{2}(\mathcal{E} - 2) \\ &k = \pm[(m+1)/2i]\mathcal{P} \\ &k = 0. \quad [\text{See Eq. (19)}] \end{aligned} \tag{20}$$

where $\mathcal{S}$ is the set of singularities of $u(\sigma)$, $\mathcal{E}$, the set of essential singularities of $u(\sigma)$ [see before Eq. (12) for the pathological case of singularities at $i\infty$] and $\mathcal{P}$ is the set of singularities of $\tilde{u}(q)$.

Of course this is independent of the reality of the potential $V(r)$, except for the symmetry of the domain. Furthermore, it shows clearly that, as long as the potential stays bounded, it has no need to be analytic except near zero and infinity. This of course will correspond to essential singularities at infinity in the (λ, k) space, but our reasoning applies as well to the square-well potential, for example, for which the S-matrix element is meromorphic throughout the (λ, k) space, according to formula (20).

As a last remark it should be noted that the domain defined by (20) has not been shown to be the largest possible, and in particular it may happen that some of the singularities of (20) disappear, by some peculiar cancellation.

APPENDIX

The upper bound upon the derivative of the free Green's function [cf. Eq. (7)] may be obtained very easily by the following identities:

$$\begin{aligned} &n_\lambda'(ks)j_\lambda(kr) - j_\lambda'(ks)n_\lambda(kr) \\ &\quad = (-1/\sin\pi\lambda)[j_{-\lambda}'(ks)j_\lambda(kr) - j_\lambda'(ks)j_{-\lambda}(kr)] \\ &\quad = (1/2i)[h_\lambda^{(1)\prime}(ks)h_\lambda^{(2)}(kr) - h_\lambda^{(2)\prime}(ks)h_\lambda^{(1)}(kr)]. \end{aligned}$$

Let λ be noninteger and k nonzero. Then

$$|j_\lambda(kr)| < C_1 r^{\operatorname{Re}\lambda+1/2} \quad |j_\lambda'(kr)| < C_2 r^{\operatorname{Re}\lambda-1/2} \qquad r < 1.$$

This immediately justifies the upper bound (7) for $r < s < 1$, and the form (7a) of F. It is possible to use (7a) for $|\operatorname{Re}\lambda| < \frac{1}{2}$, but then F is no longer monotonic, a property which was used later. Hence the introduction of (7b). Also

$$|h_\lambda^{(1)}(kr)| < C_3 e^{-\operatorname{Im} kr} \quad |h_\lambda^{(1)\prime}(kr)| < C_4 e^{-\operatorname{Im} kr} \qquad r > 1.$$

This justifies (7) with the definition (7c), for $1 < r < s$. Then for $r < 1 < s$, (7) is trivial whatever representation of the Green's function one chooses. The case λ integer or k zero does not satisfy (7) rigorously, but may be treated by using the theorem on the maximum of an analytic function, as the Green's function involved in (7) is analytic in k and λ. In fact, it is enough to study the questions of convergence on the boundary of a domain to know the result inside the domain, as we only study absolute convergence, and it is always possible to keep this boundary off the points λ integer, $k = 0$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Exact Eigenfunctions of Angular Momentum by Rotational Projection*

J. K. PERCUS AND A. ROTENBERG
Courant Institute of Mathematical Sciences, New York University, New York, New York
(Received April 2, 1962)

By an appropriate rotational procedure, a projection operator can be constructed to give exact eigenfunctions of angular momentum. Applications to cases involving spin and orbital angular momentum are given and, in particular, all spin multiplets of the alternant molecular structure are obtained. The connection with the group theoretical approach is discussed.

1. INTRODUCTION

THERE are at least two commonly used techniques for extracting exact eigenfunctions of spin and orbital angular momentum, and their z components, from an arbitrary coupling of n particles in a quantum mechanical system. The first method is described in essence in Condon and Shortley[1] and involves the repeated use of the vector coupling (Clebsch-Gordan) coefficients. The second method uses a projection operator[2] to obtain a pure state function by projecting out all unwanted components from a mixed or impure wavefunction. In this paper a third approach is described. It is also a projection method but the projection is accomplished indirectly rather than directly. By averaging the rotations of an impure state, with a suitable weight factor, in the space of rotations, the unwanted components are reduced to zero, leaving the desired exact eigenfunction. A similar procedure has been used in recent years in nuclear physics[3] where the impure state corresponds to an intrinsic state of the nucleus and its components are the eigenfunctions associated with collective excitations of the nucleons.

The derivation of the basic formula is presented in the next section in a straightforward way, without explicit use of group representation theory. In Sec. 3, it will be used to solve, almost trivially, two spin projection problems which have heretofore required quite extended and cumbersome procedures. Finally, in Sec. 4, interpretation will be made from the viewpoint of group theory and the unification thus achieved will permit the description of a simple algorithm for the general angular momentum projection.

2. PROJECTION BY ROTATION

Let us, in full generality, include both space and spin coordinates and denote them by $\mathbf{x}_i$ and $\mathbf{u}_i$, respectively. Then if $R_{\phi\theta\chi}$ denotes a space rotation with Euler angles ϕ, θ, χ, and $S_{\phi\theta\chi}$ the corresponding spin rotation, we define

$$\begin{aligned} \mathbf{x}_i' &= R_{\phi\theta\chi}\mathbf{x}_i \\ \mathbf{u}_i' &= S_{\phi\theta\chi}\mathbf{u}_i, \end{aligned} \tag{1}$$

where

$$R_{\phi\theta\chi} = \begin{pmatrix} \cos\phi\cos\theta\cos\chi - \sin\phi\sin\chi & \sin\phi\cos\theta\cos\chi + \cos\phi\sin\chi & -\sin\theta\cos\chi \\ -\cos\phi\cos\theta\sin\chi - \sin\phi\cos\chi & -\sin\phi\cos\theta\sin\chi + \cos\phi\cos\chi & \sin\theta\sin\chi \\ \cos\phi\sin\theta & \sin\phi\sin\theta & \cos\theta \end{pmatrix} \tag{2}$$

and

$$S_{\phi\theta\chi} = \begin{pmatrix} e^{-(i/2)(\phi+\chi)}\cos(\theta/2) & -e^{-(i/2)(\phi-\chi)}\sin(\theta/2) \\ e^{(i/2)(\phi-\chi)}\sin(\theta/2) & e^{(i/2)(\phi+\chi)}\cos(\theta/2) \end{pmatrix} \tag{3}$$

We must determine a weight function $F(\phi, \theta, \chi)$ which, on averaging over both space and spin rotation, will project the impure state function $\psi = \psi(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots)$ onto a wavefunction $\psi_{JM}(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots)$ of definite angular momentum J and its z component M. That is, we require

$$\begin{aligned} \psi_{JM}(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) &= \iiint F(\phi, \theta, \chi)\psi'(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) \\ &\quad \times \sin\theta\, d\theta\, d\phi\, d\chi, \end{aligned} \tag{4}$$

* Work performed at the AEC Computing and Applied Mathematics Center under auspices of the U. S. Atomic Energy Commission.

[1] E. U. Condon and G. H. Shortley, *The Theory of Atomic Spectra*, (Cambridge University Press, New York, 1959).
[2] P. O. Löwdin, Phys. Rev. **97**, 1507 (1955).
[3] D. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).

where

$$\psi' = R_{\phi\theta\chi}S_{\phi\theta\chi}\psi(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) \equiv \psi(\cdots \mathbf{x}_i', \mathbf{u}_i' \cdots).$$

It is readily verified that

$$J_z = -i \sum_i [x_i\, \partial/\partial y_i - y_i\, \partial/\partial x_i + i\sigma_{zi}/2],$$

operating on $\psi(\cdots \mathbf{x}_i', \mathbf{u}_i' \cdots)$, is identical with

$$\bar{J}_z = i\, \partial/\partial\phi, \tag{5}$$

where the σ's are the usual Pauli spin matrices and units have been chosen so that $\hbar = 1$. Correspondingly, we have

$$\bar{J}_x = -i[-(\cos\phi/\sin\theta)\cos\theta\, \partial/\partial\phi - \sin\phi\, \partial/\partial\theta + (\cos\phi/\sin\theta)\, \partial/\partial\chi] \tag{6}$$

$$\bar{J}_y = -i[(\sin\phi/\sin\theta)\cos\theta\, \partial/\partial\phi - \cos\phi\, \partial/\partial\theta - (\sin\phi/\sin\theta)\, \partial/\partial\chi] \tag{7}$$

and

$$\bar{J}^2 = (-1/\sin^2\theta)[\partial^2/\partial\phi^2 + \sin\theta\, (\partial/\partial\theta)\sin\theta\, \partial/\partial\theta + \partial^2/\partial\chi^2 - 2\cos\theta\, (\partial/\partial\phi)\, \partial/\partial\chi]. \tag{8}$$

Hence, in order that $J_z\psi_{JM} = M\psi_{JM}$ we need, from Eq. (4),

$$J_z\psi_{JM} = \iiint F(\phi, \theta, \chi)\bar{J}_z\psi'(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) \times \sin\theta\, d\theta\, d\phi\, d\chi. \tag{9}$$

Inserting the expression (5) for $\bar{J}_z$ in Eq. (9) and performing a partial integration with respect to ϕ, we find

$$J_z\psi_{JM} = -i \iiint (\partial/\partial\phi)\, F(\phi, \theta, \chi) \times \psi'(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots)\sin\theta\, d\theta\, d\phi\, d\chi. \tag{10}$$

But this must coincide with $M\psi_{JM}$ and thus,

$$\bar{J}_zF(\phi, \theta, \chi) \equiv i(\partial/\partial\phi)\, F(\phi, \theta, \chi) = -MF(\phi, \theta, \chi). \tag{11}$$

In a similar way we determine the condition on F that ψ_{JM} be an eigenfunction of J^2 with eigenvalue $J(J+1)$. The calculation is the same as the one just given except, of course, that the partial integrations are performed over all the Euler angles and we use the property that $\bar{J}^2$ is self-adjoint with respect to the volume element in rotational space. The result is

$$\bar{J}^2F(\phi, \theta, \chi) = J(J+1)F(\phi, \theta, \chi). \tag{12}$$

Since Eqs. (11) and (12) are precisely the equations for a rigor rotator, we can write down the solution[4] immediately,

$$F_{JM}(\phi, \theta, \chi) = \sum_K a_{JMK}F_{JMK}(\phi, \theta, \chi), \tag{13}$$

where the a_{JMK} are arbitrary and

$$F_{JMK}(\phi, \theta, \chi) = e^{iM\phi}e^{iK\chi}\cos^{|K+M|}(\theta/2)\sin^{|K-M|}(\theta/2) \times F(G - J, 1 + G + J; \times 1 + |K - M|; \sin^2(\theta/2)) \tag{14}$$

where $G = \frac{1}{2}|K+M| + \frac{1}{2}|K - M|$ and $F(a, b; c; z)$ is the hypergeometric function.

The expressions (4) and (13) constitute a generating function: given a state ψ, a sequence ψ_{JM} of eigenstates of J^2 and J_z may be constructed. The construction consists literally of a z-axis rotation to extract an internal J_z of value K, an x-axis rotation to obtain a J^2 of $J(J+1)$ and a final z-axis rotation to produce an external J_z of M. If we demand in addition a real decomposition of ψ, a state originally diagonal in J_z can give rise only to states of the same J_z. This requires $K = M$ and, of course, appropriate normalization of the a_{JMK}. We write

$$\psi_{JM} = \mathcal{O}_{JM}\psi(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) = a_{JMM} \iiint F_{JMM}(\phi, \theta, \chi)\psi'(\cdots \mathbf{x}_i, \mathbf{u}_i \cdots) \times \sin\theta\, d\theta\, d\phi\, d\chi \tag{15}$$

as the defining equation for the projection $\mathcal{O}_{JM}$, and by direct calculation

$$a_{JMM} = (\langle F_{JMM} \mid F_{JMM}\rangle)^{-1} = (2J+1)/8\pi^2. \tag{16}$$

Let us verify that the ψ_{JM} constitute a complete decomposition of the mixed state ψ. Since the F_{JMK} are a complete set of symmetric top eigenfunctions, then

$$\sum a_{JMK}F_{JMK}(\phi, \theta, \chi)F^*_{JMK}(\phi', \theta', \chi') = \delta(\phi - \phi', \theta - \theta', \chi - \chi')$$

where the δ function is defined with respect to the volume element $\sin\theta\, d\theta\, d\phi\, d\chi$. Choosing $\phi' = \theta' = \chi' = 0$, then since $F_{JMK}(0, 0, 0) = \delta_{MK}$, we have $\sum a_{JMM}F_{JMM}(\phi, \theta, \chi) = \delta(\phi, \theta, \chi)$, producing no rotation in (15), and hence $\sum \mathcal{O}_{JM} = 1$ as desired. Further, with $K = M$, the reciprocal rotation $(-\chi, -\theta, -\phi)$ is seen to convert F_{JMM} to F^*_{JMM}

[4] L. Pauling and E. B. Wilson, *Quantum Mechanics*, (McGraw-Hill Book Company, Inc., New York, 1935), p. 278.

and it readily follows that $\mathcal{O}_{JM}$ is Hermitian. Finally since $J_z\mathcal{O}_{JM} = \mathcal{O}_{JM}J_z = M\mathcal{O}_{JM}$ and $J^2\mathcal{O}_{JM} = \mathcal{O}_{JM}J^2 = J(J+1)\mathcal{O}_{JM}$, the $\mathcal{O}_{JM}$ are obviously orthogonal, e.g.,

$$M'\mathcal{O}_{J'M'}\mathcal{O}_{JM} = \mathcal{O}_{J'M'}J_z\mathcal{O}_{JM} = \mathcal{O}_{J'M'}\mathcal{O}_{JM}M$$

so $\mathcal{O}_{J'M'}\mathcal{O}_{JM} = 0$ unless $M = M'$.

We conclude that

$$\mathcal{O}_{JM}\psi = \frac{2J+1}{8\pi^2}\iiint e^{iM(\phi+\chi)}\cos^{2M}(\theta/2) \times F(M-J, 1+M+J; 1; \sin^2(\theta/2)) \times R_{\phi\theta\chi}S_{\phi\theta\chi}\psi \sin\theta\, d\theta\, d\phi\, d\chi. \quad (17)$$

3. SPIN PROJECTION OPERATOR

Consider the problem of obtaining an eigenfunction of total spin S with eigenvalue S^2 equal to $S(S+1)$ and of S_z equal to M, in the case of n spin 1/2 particles. Let μ be the number of particles with spin z component up and $\nu = n - \mu$ the number with spin z component down. Then we denote the mixed state by

$$\psi = [a^\mu \mid b^\nu], \quad (18)$$

as an abbreviation for a Slater determinant built up from the μ orbitals $a_1, a_2, \cdots, a_\mu$ in order and the ν orbitals $b_1, b_2, \cdots, b_\nu$ in order. The spin part of each term in the determinant is the product $\prod_{i=1}^{\mu} u(j) \prod_{l=1}^{\nu} v(l)$. Operating with the spin rotation operator given in Eq. (3), we have

$$S_{\phi\theta\chi}\prod_{j=1}^{\mu}u(j)\prod_{l=1}^{\nu}v(l) = \prod_{j=1}^{\mu}\{S_{11}u(j)+S_{21}v(j)\} \times \prod_{l=1}^{\nu}\{S_{12}u(l)+S_{22}v(l)\},$$

where S_{ij} is the ijth element of the spin rotation operator. After carrying out the indicated multiplications we obtain

$$S_{\phi\theta\chi}[a^\mu \mid b^\nu] = \sum_{j=0}^{\mu} S_{11}^{\mu-j}S_{21}^{j}[a^{\mu-j} \mid a^j] \times \sum_{l=0}^{\nu} S_{12}^{l}S_{22}^{\nu-l}[b^l \mid b^{\nu-l}] \quad (19)$$

where now the notation [|] includes the sum of all possible permutations of the individual orbitals between spin types but maintaining the order within each type.

From an examination of Eqs. (3) and (14), it is seen that the ϕ and χ integrations in Eq. (17) are trivial. The χ integration gives

$$2\pi\,\delta\left(M - \frac{\mu}{2} + \frac{j}{2} - \frac{j}{2} + \frac{l}{2} + \frac{\nu}{2} - \frac{l}{2}\right) = 2\pi\,\delta\left(M - \frac{\mu}{2} + \frac{\nu}{2}\right) \quad (20)$$

where δ denotes the Kronecker δ function. But $M = (1/2)(\mu - \nu)$ so the integral is 2π. For ϕ, we have

$$2\pi\,\delta\left(M - \frac{\mu}{2} + \frac{j}{2} - \frac{j}{2} - \frac{l}{2} + \frac{\nu}{2} - \frac{l}{2}\right) = 2\pi\,\delta(l-j) \quad (21)$$

and a nonzero contribution is obtained only for $l = j$. Thus Eq. (19) yields

$$S_{\phi\theta\chi}\psi = \sum_{j=0}^{\nu}(-1)^j S_{11}^{\mu-j}S_{21}^{j}S_{12}^{j}S_{22}^{\nu-j} \times [a^{\mu-j}b^j \mid a^j b^{\nu-j}], \quad (22)$$

where $[a^{\mu-j}b^j \mid a^j b^{\nu-j}]$ is the sum of $\binom{\mu}{j} \times \binom{\nu}{j}$ determinants obtained from $[a^\mu \mid b^\nu]$ by interchanging j orbitals from the a group, in order, in all possible ways with j orbitals of the b group, in order. The factor $(-1)^j$ arises because, in each determinant, j columns of a orbitals are interchanged with j columns of b orbitals in writing Eq. (19) in the above form

Next we substitute Eq. (22) with the values of the spin matrix elements inserted into the expression for the projection, Eq. (17), obtaining

$$\mathcal{O}_{SM}[a^\mu \mid b^\nu] = \frac{2S+1}{2}\int_0^\pi \cos^{2M}\frac{\theta}{2} \times F(M-S, 1+M+S; 1; \sin^2\theta/2). \times \sum_{j=0}^{\nu}(-1)^j(-1)^j(\cos\theta/2)^{\mu-j+\nu-j} \times (\sin\theta/2)^{2j}[a^{\mu-j}b^j \mid a^j b^{\nu-j}]\sin\theta\, d\theta. \quad (23)$$

The integral is easily evaluated for two special values of M. In the first case $M = S$, and then the hypergeometric function $F(0, 1+2S; 1; \sin^2\theta/2) = 1$. On interchanging the summation and integration and substituttng $\cos\theta = 2y - 1$, we get

$$\mathcal{O}_{SS}[a^\mu \mid b^\nu] = \frac{2S+1}{\mu+1}\sum_{j=0}^{\nu}\binom{\mu}{j}^{-1}[a^{\mu-j}b^j \mid a^j b^{\mu-j}]. \quad (24)$$

This result has been obtained by Löwdin[5] by a rather lengthy derivation.

The second case is $M = 0$ and this implies $\mu = \nu$.

[5] R. Pauncz (private communication).

Then the hypergeometric function reduces to a Legendre function,[6]

$$F(-S, S+1; 1; \sin^2 \theta/2) = P_S(\cos \theta) \qquad (25)$$

and the Legendre function can be expanded[6] to give (on setting $\cos \theta = x$),

$$P_S(x) = \frac{(-1)^S}{2^S S!} \left(\frac{d}{dx}\right)^S [(1-x)(1+x)]^S$$

$$= \frac{(-1)^S}{2^S S!} \sum_{l=0}^{S} (-1)^l \frac{[S!]^3 (1+x)^l (1-x)^{S-l}}{[(S-l)!\, l!]^2}. \qquad (26)$$

The second part of Eq. (26) is obtained by carrying out the differentiation explicitly. We set this expression into Eq. (23) and substitute the values of elements of the spin rotation matrix from Eq. (3). Making the substitution $x = 2y - 1$, the integral is readily evaluated, and after some reduction, we find

$$\Theta_{S0}[a^\mu \mid b^\mu] = (-1)^S \frac{2S+1}{S+\mu+1} \sum_{j=0}^{\mu} \sum_{l=0}^{S} (-1)^{j+l} \times \binom{S}{l}^2 \binom{\mu+S}{S+j-l}^{-1} [a^{\mu-j} b^j \mid a^j b^{\mu-j}]. \qquad (27)$$

This expression has application to alternant molecular systems which possess closed shell structure. Such systems have been studied by Pauncz, de Heer, and Löwdin[7] who obtained the projected spin eigenfunctions for some special cases.

4. CONNECTION WITH GROUP REPRESENTATION THEORY

The basic expression (17) is easily obtained by an application of the theory of group representations. Suppose the $\mathfrak{D}^i(R) = (\mathfrak{D}^i_{m\mu}(R))$ to be a complete set of irreducible representations, of dimension n_i, of the transformation group $G = \{R\}$ of order h. From the general orthogonality theorem[8]

$$\sum_R \mathfrak{D}^i_{m\mu}(R)\, \mathfrak{D}^{i'}_{\mu' m'}(R^{-1}) = (h/n_i)\, \delta_{mm'}\, \delta_{\mu\mu'}\, \delta_{ii'}, \qquad (28)$$

and the fact that $\mathfrak{D}^i$ is a unitary matrix representation, it readily follows that

$$\left(\sum_R \mathfrak{D}^{i*}_{mm'}(R) R\right)\left(\sum_{R'} \mathfrak{D}^{i'*}_{m'm'}(R') R'\right) = (h/n_i)\, \delta_{mm'}\, \delta_{ii'} \sum_R \mathfrak{D}^{i*}_{mm}(R) R. \qquad (29)$$

In other words, the

$$\Theta_{im} = n_i h^{-1} \sum_R \mathfrak{D}^{i*}_{mm}(R) R \qquad (30)$$

are a set of orthogonal projections. Not only does $\Theta_{im}\psi$ transform according to the jth representation, but, as a consequence of

$$R \sum_{R'} \mathfrak{D}^{i*}_{mm}(R') R' = \sum_{\mu, R'} \mathfrak{D}^i_{\mu m}(R)\, \mathfrak{D}^{i*}_{\mu m}(R') R', \qquad (31)$$

it is seen that $\Theta_{im}\psi$ is an eigenfunction of any R on which $\mathfrak{D}^i$ is diagonal. Finally the Θ_{im} are complete, a result of the second orthogonality theorem[8]

$$\sum_l \chi^l(C_\alpha)^* \chi^l(C_\beta) = (h/h_\alpha)\, \delta_{\alpha\beta}, \qquad (32)$$

where $\chi^l(R) = \mathrm{Tr}\ \mathfrak{D}^l(R)$ and C_α is any element of the α class; this, coupled with $\mathfrak{D}^i_{mm}(I) = 1$, shows that $\sum \Theta_{im} = I$.

The representations of the rotation group in three dimensions may be generated by choosing the complete set $\{Y_{lm}\}$ of eigenfunctions of L^2 and L_z as a basis (including integer and half-integer l):

$$R Y_{lm} = \sum_\mu \mathfrak{D}^l_{m\mu}(R) Y_{l\mu}. \qquad (33)$$

The representations generated this way are, in the notation of (13) and (14), given explicitly by

$$\mathfrak{D}^l_{m\mu}(\phi, \theta, \chi) = e^{-im\phi} e^{-i\mu\chi} \cos^{|m+\mu|}(\theta/2) \sin^{|m-\mu|} \theta/2 \times F(g-l, 1+g-l; 1+|m-\mu|; \sin^2 \theta/2) \qquad (34)$$

with $g = \frac{1}{2}|m-\mu| + \frac{1}{2}|m+\mu|$, where R is specified by its Euler angles. Since the dimension of $\mathfrak{D}^l$ is $2l+1$ and the volume in the space of rotations is $8\pi^2$, Eq. (30) indeed reproduces Eq. (17).

Let us, as an example, consider the case of three electrons coupled to give a resultant orbital angular momentum L with z component M. In the usual way the wave functions will be taken as products of single-particle wave functions and for the mixed state we write

$$\psi(m_1 m_2 m_3) = \Phi_{l_1 m_1}(1) \Phi_{l_2 m_2}(2) \Phi_{l_3 m_3}(3). \qquad (35)$$

It is assumed that the angular part of the single-particle function is Y_{lm} so that the effect of operating with the space rotation operator is given by (33). Since the rotation operator operates on all coordinates we get a product of three such terms. The rotation matrices can be combined using the Clebsch-Gordan[9] series, namely,

[6] W. Magnus and F. Oberhettinger, *Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1954), p. 50.

[7] R. Pauncz, J. de Heer, and P. O. Löwdin, J. Chem. Phys. **36**, 2247 (1962).

[8] V. Heine, *Group Theory in Quantum Mechanics* (Pergamon Press, New York, 1960), Sec. 14.

[9] M. E. Rose, *Elementary Theory of Angular Momentum* (John Wiley & Sons, Inc., New York, 1957), Chap. IV.

$$\mathfrak{D}^{l_1}_{m_1\mu_1}\,\mathfrak{D}^{l_2}_{m_2\mu_2} = \sum_j C(l_1, l_2, j; m_1, m_2) \times C(l_1, l_2, j; \mu_1, \mu_2)\,\mathfrak{D}^{j}_{m_1+m_2,\mu_1+\mu_2}, \quad (36)$$

where the coefficients are the Clebsch-Gordan coefficients and the limits on j are $|l_1 - l_2|$ and $(l_1 + l_2)$. Using Eq. (36) again, $\mathfrak{D}^{l_3}_{m_3\mu_3}$ can be combined with $\mathfrak{D}^{j}_{m_1+m_2,\mu_1+\mu_2}$ to give a sum of terms involving $\mathfrak{D}^{j'}_{m,\mu}$ where $m = m_1 + m_2 + m_3$, $\mu = \mu_1 + \mu_2 + \mu_3$, and j' is the new summation index. From the orthogonality of the rotation matrices we find that $j' = L$ and $m = \mu = M$. Thus two of the summations are eliminated and we have

$$\Theta_{LM}\psi(m_1, m_2, m_3) = \sum_{\mu_1}\sum_{\mu_2}\sum_j C(l_1, l_2, j; m_1, m_2) \times C(l_1, l_2, j; \mu_1, \mu_2) \times C(l_3, j, L; M - m_1 - m_2, m_1 + m_2) \times C(l_3, j, L; M - \mu_1 - \mu_2; \mu_1 + \mu_2) \times \psi(\mu_1, \mu_2, M - \mu_1 - \mu_2). \quad (37)$$

Wigner[10] has written the formula in the case of two particles. The extension of Eq. (37) to more particles simply involves repeated use of Eqs. (33) and (36), there being a μ and j summation added for each extra particle. If two or more particles are in the same shell, there are further limitations on the μ summation arising from the Pauli principle, but these need not be introduced explicitly as the superfluous terms are eliminated upon antisymmetrization. By using Eq. (37) to construct eigenfunctions of L and M the concept of fractional parentage is by-passed; this is a feature of projection operator techniques[11] in general. Other methods of constructing eigenfunctions usually solve for the principal case ($M = L$) and find the eigenfunctions for other values of M by repeated use of the step-down operation. Equation (37) gives the functions for all values of M directly.

Since the rotation matrices for half-integral order obey all the relations given here, with Eq. (33) generalized to the form

$$R_{\phi\theta\chi}S_{\phi\theta\chi}\Phi_{JM} = \sum_\mu \mathfrak{D}^{j}_{M\mu}(\phi, \theta, \chi)\Phi_{J\mu}, \quad (33')$$

the result given in Eq. (37) is valid, with appropriate changes in notation, for j-j coupling.

We have added this discussion of orbital angular momentum because there are advantages of using projection operators in the form given here that are not as widely recognized as they might be. Equation (37) is probably the simplest way to compute the eigenfunctions and is certainly the easiest method for doing hand calculations, assuming, of course, the availability of a table of Clebsch-Gordan coefficients. More important is the fact that for many problems it is not necessary to carry out all the summations indicated in Eq. (37); indeed the very form of this equation makes the application of selection rules very easy.

[10] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959), p. 192.

[11] J.-L. Calais, Preprint No. 7, Quantum Theory Project, University of Florida, Gainesville, Florida, 1960 (unpublished).

On a Paper of Green and Lanford*

S. T. KURODA†
Department of Mathematics, University of California, Berkeley, California
(Received April 27, 1962)

The phase shift formulas for the wave and scattering operators in the potential scattering of a single nonrelativistic particle are proved under weaker assumptions on the potential than the one used previously by Green and Lanford.

CONCERNING the potential scattering of a single nonrelativistic particle by a spherically symmetric potential $V(r)$, Green and Lanford,[1] proved, among other things, that, if the function $V(r)$, $0 < r < \infty$, satisfies

$$\int_0^R r\,|V(r)|\,dr < \infty \quad \text{and} \quad \int_R^\infty |V(r)|\,dr < \infty, \quad \text{for} \quad 0 < R < \infty, \tag{1}$$

and either

$$\int_r^\infty V(s)\,ds \quad \text{belongs to} \quad L^2(R, \infty) \tag{2}$$

or for some $\epsilon > 0$,

$$V(r) = O(r^{-1-\epsilon}), \qquad r \to \infty,$$

then the wave and scattering operators exist and the phase shift formulas for them hold true. The purpose of the present paper is to supplement their results by showing that the assumption (2) can be made redundant. This will be done by a limiting procedure starting from a case of the potential vanishing outside of a certain sphere, to which the result of I can be applied, and using a kind of continuity theorem for wave and scattering operators proved in a previous work of the writer.[2]

By the separation of the angular and radial variables, our three-dimensional problem is essentially reduced to the one-dimensional problem of radial equations. Therefore, in order to avoid the complication of the notations, we shall state our result in the one-dimensional form as in the theorem given below. Then, the proof of the three-dimensional version of the theorem will be sketched in the remark after the theorem. Hence, except in that remark, we shall fix a partial wave space with arbitrarily specified quantum numbers l and m, and treat the problem in the Hilbert space $X = L^2(0, \infty)$ of all radial wave functions of that partial wave space. In X we consider the following two ordinary differential operators as the free and total Hamiltonians:

$$H_0 = -d^2/dr^2 + l(l+1)/r^2, \qquad H = -d^2/dr^2 + l(l+1)/r^2 + V(r), \tag{3}$$

where $0 < r < \infty$, and l is a non-negative integer. The exact definition of H_0 and H as self-adjoint operators in the Hilbert space $X = L^2(0, \infty)$ is given in I under the assumption (1), using the theory of eigenfunction expansion associated with the ordinary differential operators (3). We follow this definition and denote the generalized Fourier transforms associated with H_0 and H by F_0 and F, respectively.[3] As is well known, F_0 is the isometric integral operator from X onto[4] $\hat{X} = L^2(0, \infty)$ given by the kernel

$$\psi_0(r, k) = \varphi(kr), \qquad \varphi(r) = r^{1/2} J_{l+1/2}(r), \tag{4}$$

where J denotes the Bessel function. The restriction F_c of F on the continuum subspace X_c of H (i.e., the subspace of X consisting of all functions orthogonal to all the bound states of H) is the isometric integral operator from X_c on[5] $\hat{X}$ whose kernel $\psi(r, k)$ is a solution of the differential equation

$$-d^2\psi/dr^2 + \{l(l+1)/r^2 + V(r)\}\psi = k^2\psi,$$

satisfying the boundary and asymptotic conditions

$$\lim_{r\to 0} \psi(r, k) = 0, \qquad \psi(r, k) \sim (2/\pi)^{1/2} \sin[kr - l\pi/2 + \delta(k)], \quad r \to \infty, \tag{5}$$

* Work on this paper was partly supported by National Science Foundation Grant G-19136.

† On leave of absence from the University of Tokyo.

[1] T. A. Green and O. E. Lanford, III, J. Math. Phys. **1**, 139 (1960). Hereafter, we shall refer to this paper as I.

[2] S. T. Kuroda, J. Math. Soc. Japan **12**, 243 (1960), hereafter referred to as II.

[3] The transforms F_0 and F are the restrictions on the partial wave space under consideration of the transforms F_0 and F given in I.

[4] Although $\hat{X}$ may be regarded as identical with X, we make a distinction mainly because X and $\hat{X}$ correspond to the configuration and momentum spaces, respectively.

[5] We regard the momentum spaces of the free and actual particles as identical. This is for the convenience in handling the phase shift formula without introducing the canonical mapping between two momentum spaces.

where we assume $0 \leq \delta(k) < 2\pi$. By these requirements, $\psi(r, k)$ and $\delta(k)$ are determined uniquely.

In what follows we shall prove the following theorem.

Theorem. Let H_0 and H be as above with $V(r)$ satisfying (1). Then the wave operators

$$\Omega_{\pm} = \underset{t \to \pm\infty}{\text{strong-lim}} \exp(itH) \exp(-itH_0)$$

and the scattering operator $S = \Omega_-^*\Omega_+$ exists, and the latter is unitary. Furthermore, we have the phase shift formulas

$$\Omega_{\pm} = F_c^{-1} \exp\{\pm i\,\delta(k)\}F_0, \tag{6}$$

$$S = F_0^{-1} \exp\{2i\,\delta(k)\}F_0. \tag{7}$$

Remark. The similar phase shift formulas for the three-dimensional Hamiltonian as given in I are readily obtained from this theorem. [See (4.32) and (4.34) of I. We note that, in I, F_0 and F stand for the three-dimensional version of the generalized Fourier transforms.] Since the three-dimensional Hamiltonians are considered as the direct sums of Hamiltonians in each partial wave spaces,[6] the proof is readily obtained as follows:

Let a Hilbert space X be a direct sum of Hilbert spaces X_k, $k = 1, 2, \cdots : X = \sum_{k=1}^{\infty} \oplus X_k$, and let self-adjoint operators H_0 and H in X be direct sums of self-adjoint operators $H_{0,k}$ and H_k in X_k, respectively. Then, the wave operator

$$\Omega_{\pm} = \underset{t \to \pm\infty}{\text{strong-lim}} \exp(itH) \exp(-itH_0)$$

exists if and only if

$$\Omega_{\pm,k} = \underset{t \to \pm\infty}{\text{strong-lim}} \exp(itH_k) \exp(-itH_{0,k})$$

exists for each k. Furthermore, $\Omega_{\pm}$ is the direct sum of $\Omega_{\pm,k}$: $\Omega_{\pm} = \sum_{k=1}^{\infty} \oplus \Omega_{\pm,k}$.

Proof. Put $\Omega(t) = \exp(itH) \exp(-itH_0)$ and define $\Omega_k(t)$ similarly with H_k and $H_{0,k}$. Each $u \in X$ is expressible as $u = \sum \oplus u_k \in X_k$, and the following formula holds true:

$$||(\Omega(t) - \Omega(t'))u||^2 = \sum_{k=1}^{\infty} ||(\Omega_k(t) - \Omega_k(t'))u_k||^2.$$

The necessity of the condition follows from this at once. For the proof of the sufficiency, take an arbitrary $\epsilon > 0$. Since the series $\sum ||u_k||^2 = ||u||^2$ is convergent, there exists N such that $\sum_{k=N}^{\infty} ||u_k||^2 < \epsilon$. Hence, by the above formula, we have

$$||(\Omega(t) - \Omega(t'))u||^2 \leq \sum_{k=1}^{N-1} ||(\Omega_k(t) - \Omega_k(t'))u_k||^2 + 2\epsilon.$$

By taking t and t' sufficiently large, the first term on the right-hand side can also be made less than ϵ. This proves the sufficiency of the condition. The formula $\Omega_{\pm} = \sum_{k=1}^{\infty} \oplus \Omega_{\pm,k}$ is proved in a similar way.

Proof of the theorem. In virtue of the theory of eigenfunction expansions, we have formally

$$[(H_0 + 1)^{-1/2}u](r) = \int_0^{\infty} \varphi(kr)(k^2 + 1)^{-1/2}\,dk \int_0^{\infty} \varphi(ks)u(s)\,ds.$$

The operator $(H_0 + 1)^{-1/2}$ is therefore an integral operator with the kernel

$$\chi(r, s) = \int_0^{\infty} \varphi(kr)\varphi(ks)(k^2 + 1)^{-1/2}\,dk.$$

Rigorously speaking, these integrals may not be convergent and appropriate definitions are required. In particular, $\chi(r, s)$ is actually defined as the transform F_0^{-1} of the function $\varphi(kr)(k^2 + 1)^{-1/2}$, namely,

$$\chi(r, s) = \underset{L \to \infty}{\text{strong-lim}} \int_0^{L} \varphi(kr)(k^2 + 1)^{-1/2}\varphi(ks)\,dk. \tag{8}$$

The verification of this fact is straightforward. Then, Parseval's formula for F_0^{-1} applied to (8) gives, for any fixed r,

$$\int_0^{\infty} |\chi(r, s)|^2\,ds = \int_0^{\infty} |\varphi(kr)|^2\,(k^2 + 1)^{-1}\,dk = r\int_0^{\infty} |\varphi(x)|^2\,(x^2 + r^2)^{-1}\,dx. \tag{9}$$

Since $\varphi(x)$ is bounded and $\varphi(x) \sim x^{l+1}$, $x \to 0$, by (4), we see from (9) that there exists a positive constant K such that

$$\int_0^{\infty} |\chi(r, s)|^2\,ds \leq \min(Kr, K).$$

By virtue of our assumption (1) this estimate readily yields

$$\int_0^{\infty}\int_0^{\infty} |V(r)|\,|\chi(r, s)|^2\,dr\,ds \leq K\int_0^{R} r\,|V(r)|\,dr + K\int_R^{\infty} |V(r)|\,dr < \infty. \tag{10}$$

In other words, the operator $|V|^{1/2}(H_0 + 1)^{-1/2}$ is an integral operator of Hilbert-Schmidt type. According to the corollary to Theorem 1 of II, this in turn establishes the existence of $\Omega_{\pm}$ and the unitarity of S. (In II, H is defined in terms of the theory of

[6] See Secs. II and III of reference 1.

closed forms. However, that definition is easily shown to be equivalent to the present one.[7])

In order to prove (6) and (7) we approximate $V(r)$ by a sequence of functions $V_n(r)$, $n = 1, 2, \cdots$, defined to be $V_n(r) = V(r)$, if $r < n$, and 0, otherwise. Let $\Omega_{\pm}^{(n)}$ denote the wave operators corresponding to the scattering potential $V_n(r)$. Then we have the following four statements[8]:

Equation (6) holds for $\Omega_{\pm}^{(n)}$

$$\text{(with } \delta \text{ replaced by } \delta_n); \tag{11}$$

$$\operatorname*{strong\text{-}lim}_{n\to\infty} \Omega_{\pm}^{(n)} = \Omega_{\pm}; \tag{12}$$

$$\lim_{n\to\infty} \delta_n(k) = \delta(k); \tag{13}$$

$$\operatorname*{strong\text{-}lim}_{n\to\infty} F_c^{(n)-1} = F_c^{-1}; \tag{14}$$

where $\delta_n(k)$ and $F_c^{(n)}$ are the phase shift and the generalized Fourier transform determined by H, with V replaced by V_n. The statement (11) is due to Green and Lanford, for $V_n(r)$ clearly satisfies the condition (2).

The formula (12) is a consequence of Theorem 2 of II, for it follows from (10) that

$$\lim_{n\to\infty} \int_0^\infty \int_0^\infty |V(r) - V_n(r)|\, |\chi(r, s)|^2\, dr\, ds = 0.$$

The proof of (13) and (14) will be sketched later.

By (13), the sequence of the multiplicative operators given by $\exp\{i\delta_n(k)\}$ converges strongly to the one given by $\exp\{i\delta(k)\}$. Therefore, starting from the formula (6) for $\Omega_{\pm}^{(n)}$ and taking limit as $n \to \infty$ with the aid of (12), (13), and (14), we finally obtain the formula (6) for $\Omega_{\pm}$. The formula (7) follows immediately from (6).

Sketch of the proof of (13) and (14). Let us consider the two solutions $\psi_n(r, k)$ and $\tilde{\psi}_n(r, k)$ of the differential equation

$$-d^2\psi/dr^2 + \{l(l+1)/r^2 + V_n(r)\}\psi = k^2\psi,$$

where $\psi_n(r, k)$ satisfies the requirement (5) with $\delta(k)$ replaced by $\delta_n(k)$, and $\tilde{\psi}_n(r, k) = \psi(r, k)$, if $0 < r < n$. Since $V_n(r) = V(r)$ if $0 < r < n$, we readily obtain

$$\tilde{\psi}_n(r, k) = c_{n,k}\psi_n(r, k), \tag{15}$$

$$(\tilde{\psi}_n - \psi)(r, k) = -k^{-1} \int_n^r \sin k(r - s) \times (V_n - V)(s)\psi(s, k)\, ds, \qquad r > n, \tag{16}$$

where $c_{n,k}$ is a constant. Since $\psi(s, k)$ is known to be bounded in $\{(s, k) \mid 0 \le s < \infty,\ \eta < k < \infty\}$ for each $\eta > 0$, it follows from (16) that

$$\lim_{n\to\infty} (\tilde{\psi}_n - \psi)(k, r) = 0, \tag{17}$$

uniformly for all r and $k \in (0, \infty)$.[9] Hence, by virtue of the asymptotic forms of $\tilde{\psi}_n$ and ψ, and the periodic property of the sine function, we get

$$\lim_{n\to\infty} \{\sin(kr - \delta(k)) - c_{n,k} \sin(kr - \delta_n(k))\} = 0,$$

uniformly for all r and $k \in (0, \infty)$. This readily implies

$$\lim_{n\to\infty} \delta_n(k) = \delta(k), \qquad \lim_{n\to\infty} c_{n,k} = 1, \tag{18}$$

both uniformly for $k \in (0, \infty)$. To prove (14), we note that (15), (17), and (18) give that

$$\lim_{n\to\infty} \psi_n(r, k) = \psi(r, k)$$

uniformly for all r and $k \in (0, \infty)$. From this we get (14) by a standard argument, using the fact that ψ_n and ψ are kernels of the isometric integral operators $F_c^{(n)-1}$ and F_c^{-1}, respectively.

[7] Incidentally, we remark that, in II, $\Omega_{\pm}$ is denoted by $W_{\pm}$.

[8] Although (12) is established here only for the one-dimensional case, the same argument as in the above remark shows that (12) holds also for three-dimensional wave operators.

[9] "Uniformly for $k \in (0, \infty)$" means that the convergence is uniform with respect to k in any closed interval of $(0, \infty)$ of k.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Introduction to Momentum Space Integration Techniques in Perturbation Theory

M. FOWLER
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England
(Received March 14, 1962)

Generalized "pinch" techniques are developed for analyzing singularity configurations of any Feynman integral in the product space of the loop momenta. The Landau equations follow immediately, and the dual diagrams arise naturally as geometric singularity criteria. Cutkosky's formula for the discontinuity is derived by an elementary method, and its structure is clearly exhibited by this approach. The basic differences between Landau and non-Landau singularities for single loop diagrams are discussed, and it is shown why the presence of non-Landau singularities, in contrast to those in the Landau scheme, depends on the dimensionality of space.

1. INTRODUCTION

THE aim of this paper is to establish a new approach to the study of singularities in perturbation theory, based on direct examination of the mass shell singularities in k space, the product space of the loop momenta. Previous work in perturbation theory has been based on a study of singularity configurations either after effecting a transformation to invariants formed from the loop momenta or by introducing Feynman parameters and eliminating the loop momenta by symmetric integration.[1,2]

The present method has not been used before because it seems at first more difficult conceptually, and singularity criteria are more subtle than those derived previously. These difficulties are in fact only initial, and the method turns out to be no harder than alternative techniques, and for some problems it is more helpful towards a deeper understanding of the structure of Feynman integrals.

In Sec. 2 a simple generalization of the pinch criteria for singularities is derived,[3] and it is shown that this is equivalent to the Landau conditions. The dual diagrams appear as an integral part of the geometric structure, and the method gives a vivid picture of why the Landau conditions are equivalent to mechanical tautening in the dual diagram. In Sec. 3, a new but fairly simple technique is introduced by which the cuts beginning at singular points can be examined. This is applied to give an elementary derivation of Cutkosky's formula for the discontinuity. In Sec. 4 the non-Landau singularities[4,5] for single loop graphs are discussed. The structural connection between Cutkosky's formula and the general pinch criterion is exhibited and used to define an equivalence relationship among the sets of singularities. Under this relationship the non-Landau singularities form a group on their own, and this explains their dimensional dependence.

It appears that k-space techniques provide a powerful tool for examining the structure of perturbation theory. They are being used at the moment in an investigation of "mixed" singularities,[5] and it is hoped that they may be useful in solving other outstanding problems in perturbation theory, such as the appearance of isolated real points on Landau curves.[6,7]

2. DERIVATION OF THE LANDAU EQUATIONS

2.1. Notation

Spin and isotopic spin are consistently ignored. The standard notation for Feynman diagrams is adopted, p denoting external, q internal, and k loop momenta with masses m. The Feynman integral has the form

$$\int \frac{\prod_i d^4k_i}{\prod_i (q_i^2 - m_i^2)}$$

[1] The possibility of deriving the dual diagrams by k-space methods is suggested by the work of J. C. Taylor, Phys. Rev. **117**, 261 (1960).

[2] A list of references to most recent work in perturbation theory is given by R. J. Eden, Maryland lecture notes, (1961), (unpublished). R. J. Eden and J. C. Polkinghorne, *Brandeis University, Summer School of Theoretical Physics, 1961* (W. A. Benjamin Publishing Company, New York, 1962).

[3] This generalization has been derived independently by P. V. Landshoff and J. C. Polkinghorne (unpublished). See J. C. Polkinghorne, Nuovo cimento **23**, 360 (1962).

[4] R. E. Cutkosky, J. Math. Phys. **1**, 429 (1960).

[5] D. B. Fairlie, P. V. Landshoff, J. Nuttall, and J. C. Polkinghorne, J. Math. Phys. **3**, 594 (1962). In this paper non-Landau singularities are referred to as "second type."

[6] R. J. Eden, P. V. Landshoff, J. C. Polkinghorne, and J. C. Taylor, J. Math. Phys. **2**, 656 (1961).

[7] S. Mandelstam, Phys. Rev. **112**, 1344 (1958); **115**, 1741, 1752 (1959).

where i goes over the loop momenta, j over all internal lines.

2.2. Generalized Pinch Criteria

The singularity criteria used by previous authors[2] are valid only when there is one surface of singularity. They correspond essentially to this surface having a locally conelike degeneracy. An example of more general pinch conditions arises in considering the Feynman integral for the self-energy part with two-dimensional momenta.

$$\int \frac{dk_1\, dk_2}{(k_1^2 + k_2^2 - m_1^2 - i\epsilon)[(k_1 - p)^2 + k_2^2 - m_2^2 - i\epsilon]}.$$

In the real (k_1, k_2) plane there are two circles of pole with attached complex singular surfaces.

A plane contour is trapped at a point if, and only if, every line lying in the plane and passing through the point is pinched at the point. For if one is not, by continuity lines through the point lying in some segment including this line are not pinched, so the plane can be distorted away from the singularity, into the four-dimensional space. It follows from this that a single circle of finite radius does not cause singular behavior although all lines tangential to it are pinched (see Fig. 1). "Line" is used sometimes to mean the real line and sometimes to include the complex surface attached to the real line. A line is pinched if the complex points where it meets singular surfaces pinch the real subsection. If a contour has to be distorted away from the real subsection to avoid oncoming singularities, a pinch can occur by singularities coming together at a complex point.

If the radius of a circle shrinks to zero, a pinch develops on every line through the center, so $m_1 = 0$ gives rise to a singularity (Fig. 2). Note that the attached surface is conelike. To derive conditions for the two circles acting together to produce a pinch, we examine lines through the common point. It is clear that if both circles have the same sign

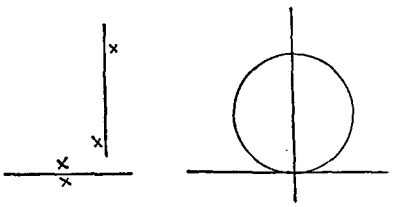


FIG. 1. Intersections of $k_1^2 + k_2^2 = m^2 + i\epsilon$ with lines.

for $i\epsilon$, a is pinched and b is not pinched at X (see Fig. 3). Generally, any line through X in the sector p between the tangents to the circles is pinched.

To identify the sectors p, n we attach a direction to every line through X, then denote p by $(0, i)$ meaning that at X every l in p is going, e.g., out of C_1 and into C_2. The condition for a singularity is that every line through X is pinched, so that the region n, or (i, i) goes to zero. This means that the two circles touch externally. Touching occurs also if $(0, i)$ shrinks to zero, but in this case one circle lies inside the other and no line except the common tangent is pinched. However, if the $i\epsilon$ on one of the circles is reversed $(0, i)$ going to zero gives a singular situation, so it corresponds essentially to a singularity on another sheet. (The $i\epsilon$'s having the same sign corresponds to the physical sheet.)

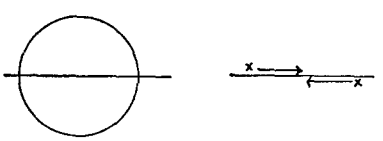


FIG. 2. Pinch arising as $m \to 0$.

Hence for two circles in a plane, singularities arise when (i) the radius of either circle goes to zero or (ii) the circles touch *externally*. This last condition is of course analogous to the criterion that both Feynman parameters be positive for the physical sheet singularity.

2.3. Three- and Four-Dimensional Analogs

For the vertex part in three dimensions (Fig. 4), the Feynman integral is

$$\int \frac{d^3k}{(k^2 - m_1^2 - i\epsilon)[(k + p')^2 - m_2^2 - i\epsilon][(k - p)^2 - m_3^2 - i\epsilon]}$$

The denominator corresponds to three spheres of pole in k space. We consider the analytic behavior near a point X where the three spheres intersect. The solid angle about X can be divided into sets (i, i, i), $(i, i, 0)$, $(i, 0, i)$, $(0, i, i)$ for the two circles, making the obvious generalization. A pinch occurs at X if (i, i, i) goes to zero. The condition for this is that the three tangent planes at 0 to the sphere

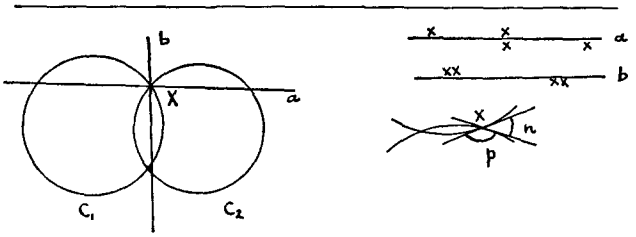


FIG. 3. The self-energy part in two dimensions. Singularity configurations on lines through X.

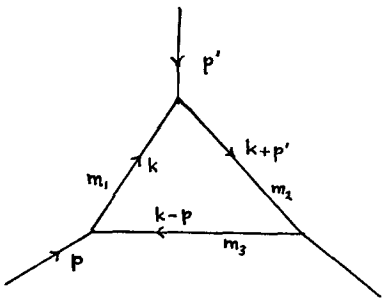


FIG. 4. The vertex part.

have a common line, that is the two points X, X' of intersection of the spheres coincide. This means that X lies in the plane of centers, immediately giving the well-known dual diagram condition for singularity (Fig. 5).

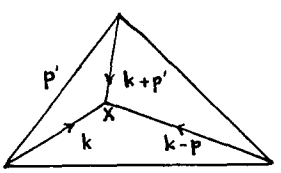


FIG. 5. Dual diagram for the vertex part.

If X lies in the plane of centers, but outside the triangle of centers, one of the regions $(i, i, 0)$, etc., has gone to zero (i, i, i) still being nonzero, so there is no singularity. This is because our criterion conveniently includes the implicit condition that the singularity be on the physical sheet. As before, changing one $i\epsilon$ would swap the corresponding i, 0 and hence change sheet.

If the points of intersection of the spheres move off to ∞ (as the centers become collinear) the tangent planes become parallel, and all the regions except (i, i, i) collapse. This is best seen in the (Re k_1, Re k_2, Im k_3) section, where the "spheres" are hyperboloids. These non-Landau configurations will be discussed in detail in Sec. 4.[8]

It is clear that another possible singularity in the three-dimensional case is given by two spheres touching. The third sphere is of course irrelevant for this singularity. Taking coordinates in a plane through the line of centers of S_1, S_2 with an angle variable ϕ to position the plane, the (i, i) region in the plane is zero, and independent of ϕ, so the solid (i, i) region must also be zero.

By the same argument, three spheres in four dimensions have the same singularity configurations as three spheres in three dimensions because coordinates can be taken in a solid through the plane of centers, with a ϕ variable to position the solid, and the configurations are independent of ϕ.

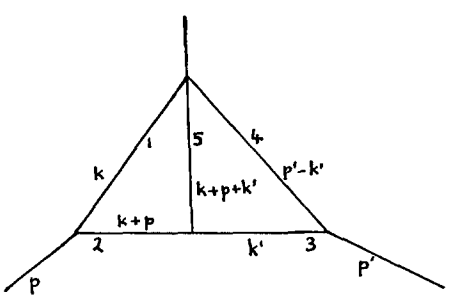


FIG. 6. The wigwam diagram.

2.4. *N*-Space Criteria and the Landau Conditions

The above conditions for singularity can be expressed more succinctly as follows: There exists a point X in the space such that m of the surfaces of singularity pass through X and the m tangent primes to these surfaces form a linearly dependent set. The arguments used above are not really dimensionality dependent, and the generalization to this criterion is fairly trivial. It is given in Appendix 1.

We now give a proof that the above condition is equivalent to the Landau equations for a singularity. This is done most easily by taking a specific example, for instance the wigwam diagram in Fig. 6 for which the Feynman integral is

$$\int \frac{d^4k\, d^4k'}{(k^2 - m_1^2)[(k+p)^2 - m_2^2](k'^2 - m_3^2)[(p'-k')^2 - m_4^2][(k'+p+k)^2 - m_5^2]}.$$

TABLE I. Direction ratios of normals to singularity surfaces for wigwam diagram.

	k	k'
1	k	0
2	$k+p$	0
3	0	k'
4	0	$p'-k'$
5	$k'+p+k'$	$k'+p+k'$

The direction ratios for the normals at points on a surface have eight components (working in k, k' space). The components are listed for the five surfaces of singularity in the integrand. The important point is that the normal to 5 does not lie in one of the subspaces, as the surface is symmetric in k, k'.

[8] In (5) it is shown that non-Landau singularities are associated with infinite internal momenta.

It is clear from Table I that when the normals are linearly dependent at some common point of the five surfaces we have the complete set of Landau equations, the α's being just the coefficients in the dependency equation. Any other diagram has the same formal solution.

For the single loops, the dual diagrams appear immediately as part of the structure. For higher diagrams, the situation is a little more subtle. The dual diagrams do not appear complete because for the n-loop case the piece of diagram corresponding to each loop appears in a four-dimensional subspace of the $4n$-dimensional space. However, the full diagram can be constructed in a single four-space from these bits without difficulty, because where the same line occurs in two or more subspaces, it occurs in "parallel" sets (see Table I above). Of course, it may be impossible to put the bits together to form a dual diagram. This corresponds to the "improper"[9] case and for this the best that can be done is to duplicate lines (bearing in mind that such lines must be parallel).

2.5. Tautening

As the procedure for constructing dual diagrams outlined above defines directions as well as lengths, the tautening conditions are automatically satisfied (as indeed they must be since the procedure is equivalent to the Landau equations). A more vivid way of seeing how the tautening conditions arise is given by the following example.

For the wigwam diagram in three dimensions, the five surfaces of singularity meet in a one-dimensional curve, corresponding to the one degree of freedom in the dual diagram. If the external variables are such that the dual diagram is nearly in a plane, the curve approximates to an ellipse in the (h, h') variables (see Fig. 7) with only second-order variations in the other four variables [the six-dimensional space is the (k, k') space]. As the ellipse shrinks to a point, all five surfaces of singularity have the (h, h') plane as a common tangent plane, so that their five tangent primes at the point are linearly dependent (they always have a line in common). Also, when the ellipse shrinks to a point there is of course no freedom of movement left in the dual diagram, that is, it is mechanically taut. This perhaps illuminates the close connection between tautening and effective pinches in n dimensions.

A less trivial tautening condition, that for the Mercedes diagram, (Fig. 8) is given by exactly the same argument. Six surfaces in 9-space meet in a 3-curve, and at the point of tautening this lies in a 4-plane tangential to all six surfaces.

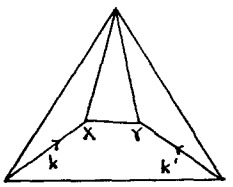
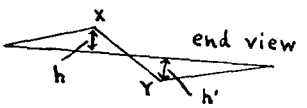


FIG. 7. Dual of wigwam diagram.



3. DERIVATION OF CUTKOSKY'S FORMULA

3.1. Preliminaries

Cutskosky[4] has shown that the discontinuity of the Feynman integral across a branch cut starting from a singularity defined by Landau's conditions for which $q_i^2 = m_i^2$ for $i \leq m$ is

$$(2\pi i)^m \int \frac{\pi\, d^4k \prod_{i \leq m} \delta^{(+)}(q_i^2 - m_i^2)}{\prod_{i \geq m} (q_i^2 - m_i^2)}.$$

The method of proof is to transform to the q_i^2's with extra angle variables and use a pinch analysis

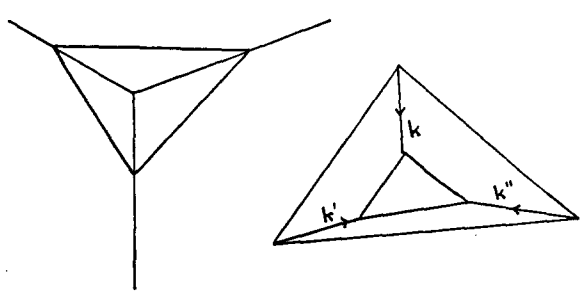


FIG. 8. The Mercedes diagram.

on the q_i^2 contours. The formula is derived below by k-space methods. All the essential features are exhibited in the following example.

3.2. Self-Energy Part in Two Dimensions

The normal threshold singularity is given by

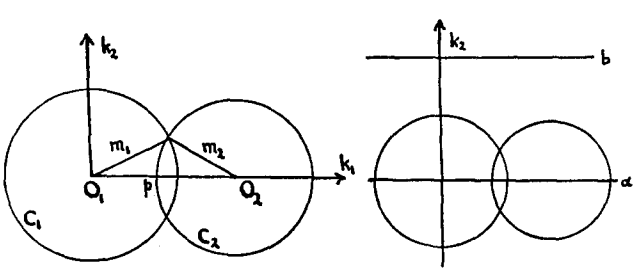


FIG. 9. The self-energy part in two dimensions.

[9] P. V. Landshoff, Nuclear Phys. 20, 129 (1960).

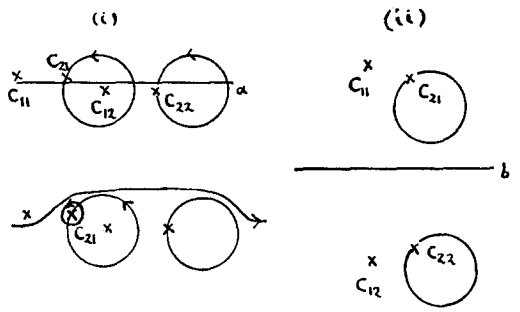


FIG. 10. Movement of singularities in the complex planes a and b (see Fig. 9) on moving around the normal threshold in the complex p-plane.

$p = m_1 + m_2$ (see Fig. 12). Keeping m_1, m_2 fixed, the center 0_2 of C_2 is moved in the complex plane $k_2 = 0$ in a path such that p goes around $m_1 + m_2$ and returns to its original position.

To see what is happening in k space, we take various sections $k_2 = \text{const}$. In the section $k_2 = 0$ there are four poles as shown in Fig. 10. If the contour is distorted upwards, leaving a bubble contour about C_{21} (Fig. 10), it is easy to see that moving 0_2 in a circle about the point $m_1 + m_2$ gives no change in the integral over k_1 for this k_2. Similarly for the section b (Figs. 9, 10) there is no change in the integral over k_1 on varying 0_2.

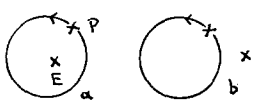


FIG. 11. The essential difference between a and b.

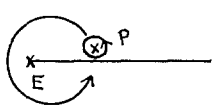


FIG. 12. The k_2 integration is equivalent to allowing the pole to trace out a cut.

There is, however, an important difference. For a, C_{21} describes a path about C_{12}, but for b, C_{12} lies outside the path of C_{21}. This means, by continuity, that at some intermediate stage C_{12}, C_{21} actually collide, and as C_{21} will be surrounded by a contour this gives rise to a singular situation. [The bubble about C_{21} in i (which could equally well have been put about C_{12}) is necessary simply because C_{21} goes around C_{12}.] Distorting the path traced by C_{21} merely changes the point at which the collision occurs. So does distorting the $\int dk_2$ contour away from the real axis. This collision, unlike the normal pinch situation, cannot be avoided by making suitable changes in parameters. This is essentially because performing the circuit about the singular point $m_1 + m_2$, at some stage the cut from this singularity must be encountered, and the equivalent of this in the above picture is the pole collision situation, as is shown in the next section.

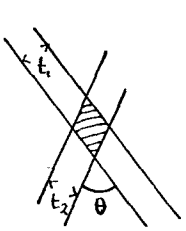


FIG. 13. Limiting procedure for defining δ functions.

3.3. Equivalence of "Colliding Poles" and a Cut

Performing the integral over k_2 is equivalent to summing over a continuous series of situations between a and b and so on to ∞ (see Fig. 11). The position is not altered if the path P is fixed and the integral represented by allowing the other pole to trace out a cut, the discontinuity or strength corresponding essentially to the relative velocities of the two poles k_2 is increased.

It now becomes apparent why the collision cannot be avoided—P is a path about an end point E of an infinite cut. (It is clear that the "line of pole" must be a cut, because moving the k_2 integration contour moves it about, giving an analytic continuation which would not apply if, for instance, the line of pole corresponded to a series of cuts having different end points along its length.) Hence the discontinuity in the original integral, measured by the small contour round P, is equal to the product of the discontinuity across the cut and the strength of P (Fig. 12).

It is easy to see that if the k_2 contour is undistorted, for the two circles the collision occurs at the point of intersection X (Fig. 9) and the discontinuity will be greater if the circles cut at a smaller angle, depending on the "relative velocity" of the two poles, as k_2 is increased at a uniform rate. A more detailed discussion of this point of view is given in Appendix 2.

3.4. Integrating Over δ Functions

Perhaps the easiest method at this stage is to anticipate the answer, and consider evaluating the integral

$$I = \int \delta(k^2 - m_1^2)\ \delta[(k - p)^2 - m_2^2]\ dk_1\ dk_2.$$

The standard method is to change to variables k^2, $(k - p)^2$ in which case the value of the integral is just that of the Jacobian at the point of intersection of the circles, $\propto A^{-1}$ where A is the area of the triangle $0_1 0_2 X$ (Fig. 9). It is clearer that this is the answer if the delta functions are defined by a limiting process,

$$\delta = \lim_{t\to 0} \delta_t$$

where $\delta_t = 0$ outside $(-t/2, t/2)$ $\delta_t = t^{-1}$ inside this interval. Replacing the two δ functions in J by δ_t functions the circles become ring-like and I gives their common area, (see Fig. 13) that is, $t_1 t_2/\sin\theta$, ($t_1 = t/m_1$, $t_2 = t/m_2$) multipled by the weighting factor t^{-2}, hence

$$I = \tfrac{1}{4}A^{-1}.$$

3.5. Limiting Processes for Colliding Pole Situations

It has been shown in the last section that in evaluating integrals over products of δ functions it is sometimes helpful to define the δ function by a limiting method, because this elucidates the structure of the integral at the vital points where the curves defined by the δ functions overlap, $\delta(C_1)$ $\delta(C_2)$, in the obvious notation.

It is plausible that the amount of overlap $\delta(C_1)$ $\delta(C_2)$ is a measure of the intensity of singularity caused by the colliding poles when one circle crashes through the other one as its center executes a path in the complex plane.

A more acceptable picture of this is given by replacing one of the circles, which corresponds to a distribution of pole with measure $\delta(C_1)$, by a distribution with measure $\delta_t(C_1)$ and take the limit as $t \to 0$. This means the circle is replaced by an annular cut, and the pole collision (which occurred for a discrete value of k_2) is replaced by a pole passing through a cut of strength $\sim 1/t$ over a range $\sim t$ of values of k_2.

On a contour C in this range the singularity configuration is as shown in Fig. 14 (ii), and the discontinuity caused by the collision is measured by the bubble contour enclosing $\delta(C_2)$ on opposite sides of the cut $\delta_t(C_1)$. It is clear that the total discontinuity on integrating over k_2 is given by the total amount of overlap of the pole $\delta(C_2)$ and the cut $\delta_t(C_2)$, that is,

$$(2\pi i)^2 \lim_{t\to 0} \int \delta(C_2)\ \delta_t(C_1)\ dk$$
$$= (2\pi i)^2 \int \delta(C_1)\ \delta(C_2)\ dk.$$

This is, of course, exactly Cutkosky's formula for the discontinuity.

3.6. Summary

At this point we review what has been established thus far. The k integral is over a two-dimensional (Re k_1, Re k_2) contour in a four-dimensional space.

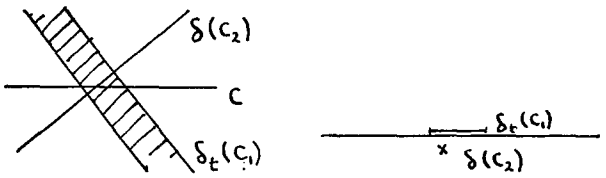


FIG. 14. Discontinuity across normal threshold for self-energy part.

We consider the singularity configuration in the three-dimensional subspace (Re k_1, Re k_2, Im k_1) that is, the space formed by the stack of complex planes with k_2 a real constant. On moving the external variable P in its complex plane in a path around the normal threshold singularity, at some stage the k contour, viewed in the three-dimensional subspace, is trapped in a collision. This corresponds to encountering the cut in complex p space. Varying the three-dimensional space by replacing Re k_2 with a path off the k_2 real axis in the complex k_2 plane merely moves the cut around and alters the point at which the collision occurs. The discontinuity across the cut in p space, corresponding to the "strength" of the collision, can be found by replacing the k-space singularities by small cuts, and taking the limit as the length of the cut goes to zero, keeping its total strength constant. This leads immediately to Cutkosky's formula.

3.7. Vertex Part in Three Dimensions

The method used is essentially the same as that for the two circles in a plane, but is a little more difficult to visualize. The singularity configurations are considered in the four-dimensional space (Re k_1, Re k_2, Re k_3, Im k_1) by fixing k_2, k_3 at real values and examining the situation in the k_1 complex plane, then integrating over k_2, k_3.

We replace two of the three spheres of pole by shells of finite thickness, giving a total pole distribution $\delta(S_1)$, $\delta_t(S_2)$, $\delta_t(S_3)$. Nothing is gained by taking different t's for S_2, S_3.

For a line (k_2, k_3) through the area $\delta(S_1)$ $\delta_t(S_2)$ $\delta_t(S_3)$ the singularity configuration in the complex k_1 plane is as shown in Fig. 15. and on varying the external parameters about the anomalous threshold singularity, the pole $\delta(S_1)$ will trace a path P to a point on the other side of *both* the cuts, hence the difference in the integral round the bubble contour will be proportional to the product of the strengths of the two cuts.

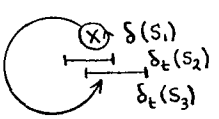


FIG. 15. Discontinuity across leading singularity for vertex part.

To find the total discontinuity on going round the anomalous threshold, the integral is taken as before over all values of k_2, k_3 giving a contribution, that is, the values in the region $\delta(S_1)\ \delta_\iota(S_2)\ \delta_\iota(S_3)$, giving

$$(2\pi i)^3 \lim_{\iota\to 0} \int \delta(S_1)\ \delta_\iota(S_2)\ \delta_\iota(S_3)\ dk$$

$$= (2\pi i)^3 \int \delta(S_1)\ \delta(S_2)\ \delta(S_3)\ dk.$$

It should perhaps be emphasized that there is no need to vary the external parameters in such a way that only S_1 moves with S_2 and S_3 remaining stationary. The only configurations of importance are the initial and final ones, corresponding to moving from one side to the other of a cut in an external variable.

For completeness, we must take into account the fact that a colliding-poles situation occurs also on any line corresponding to the intersection of two of the spheres, $\delta(S_1)\ \delta(S_2)$. This corresponds to the normal threshold, and on going around the anomalous threshold in the above example, if the k_2 and k_3 integration paths are left undistorted along the real axes, the "normal threshold" collision occurs twice, corresponding to going through the normal threshold cut and coming back. Suitable distortion of the k_2, k_3 contours would move the normal threshold cut away from the anomalous threshold (as indeed it is in the usual representation adopted, with both going in a positive direction).

3.8. Generalization to any Diagram

For definiteness the example of the wigwam diagram (Fig. 7) in three dimensions is taken. There are five surfaces of singularity in a six-dimensional space, so in general, they meet in a one-dimensional curve. This corresponds to the fact that the dual diagram has one degree of freedom. By analogy with the simpler case, four of the five surfaces are replaced by δ_ι surfaces, then sections are taken corresponding to fixing five of the six complex variables at real values, these five then being integrated over. The leading singularity arises from the pole going from one side to the other of all four small cuts. It is clear that the discontinuity given is the integral over the product of the five delta functions, and that the method generalizes immediately to any diagram.

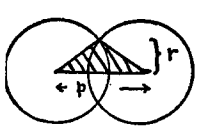


FIG. 16. Non-Landau term for self-energy part.

4. NON-LANDAUIAN SINGULARITIES FOR SINGLE LOOP DIAGRAMS

4.1. Introduction

The presence of singularities apparently not included in the Landauian scheme was first noticed by Cutkosky[4] in some elementary applications of his formula. For instance, the discontinuity across the normal threshold for the self-energy part in four dimensions is $1/s$ (see Sec. 3) so the function itself must have a pole at $s = 0$ on some sheet reached through the normal threshold. In general, the forward scattering curve is singular on some sheet, if the dimensionality of the space is greater than the number of sides of the loop. A general discussion is given in reference 5.

In this section a discussion of how non-Landau singularities arise in the k-space approach is given. The single loop self-energy part is considered in some detail. The dimensional dependence of the singularity and its relation to touching of surfaces at infinity are examined. The techniques developed in Sec. 3 are applied to give the discontinuity, and it is shown that in p space the function has a cut along the whole length of the imaginary axis. Essentially, this corresponds to the function depending only on $S = p^2$. This cut is closely related to the non-Landau singularity in S. It is shown that these singularities do not occur on the physical sheet.

Finally, the structural connection between Cutkosky's formula and the general pinch criterion is used to define an equivalence relationship which separates out the non-Landau singularities and explains why their existence (unlike that of Landau singularities) depends on the dimensionality of the space.

4.2. Single Loop Self-Energy Part

We have seen that the non-Landauian "forward scattering" term appearing in the discontinuity $\sim 1/A$ in the two-dimensional case, where A is the area of the shaded triangle (Fig. 16). As $p \to 0$, $rp \sim$ const, so A does not go to zero at $p = 0$. Hence for the two-dimensional case there is, in fact, no non-Landauian singularity. For the n-dimensional case, $n > 2$, there is a weighting factor corresponding to the area of the hypersphere of radius r, that is, $\propto r^{n-2}$. Since $rp \sim$ const as p goes to zero, if $p^2 = S$ the form of the non-Landauian singularity in S is $S^{1-n/2}$ at the origin.

Hence, in general, a singularity arises when the two spheres of singularity become concentric (or have their centers separated by a zero length vector). It is a well-known result in projective geometry that under these conditions the surfaces touch at infinity, so in this sense the singularities are a natural extension of the Landauian scheme. Of course, the same remarks apply equally to the two-dimensional case, for which there is no singularity, and the cases are topologically equivalent, but since the critical situation occurs at infinity the r^n weighting factor is all important in determining whether, or not there is a singularity, and of what type. This is discussed further in Sec. 4.5. The touching arises at infinity as the circles become concentric and is not equivalent to a common tangent situation at some point X where X goes to infinity. Hence the surface of singularity cannot be merely part of the Landau surface. The configuration is best seen in Re k_1, Im k_2) space where there are two rectangular hyperbolas becoming concentric at $p = 0$.

4.3. Discontinuity Across the Singularity

Examining the singularity as it appears in the discontinuity across the normal threshold, its distinuity is $2/p$ in three dimensions, zero in two and four dimensions. It is interesting to study the singularity more directly by applying the collision techniques developed in Sec. 3, as this gives a clearer picture of why it arises. If 0_2, the center of the circle C_2, follows a circular path P in the complex plane about 0, (to avoid unnecessary complications, sufficiently small for the circles never to cut at real points, see Fig. 17), then the intersect of C_2 with the complex plane $y = y_0$ (Re) goes around the intersect of C_1 if y_0 is sufficiently large. This gives rise to the familiar situation that $P(y_0 = 0)$ does not go round $C_1(y_0 = 0)$ but P (y_0 sufficiently large) includes C_1. Thus at some intermediate point there is a collision, and it must correspond to p being pure imaginary. A similar phenomenon occurs in the lower half-plane, so the entire imaginary axis is cut, as is best seen in the (Re, Im) plane where two rectangular hyperbolas crash through each other at the point where p is pure imaginary. Applying Cutkosky's formula the "common area" of the two rectangular hyperbolas is identical to that of two circles with the same centers and radii, so this provides a check with the other method.

The cut along the imaginary axis corresponds to the function depending only on $S = p^2$. Transforming to S and mapping the Re $p > 0$ half-plane

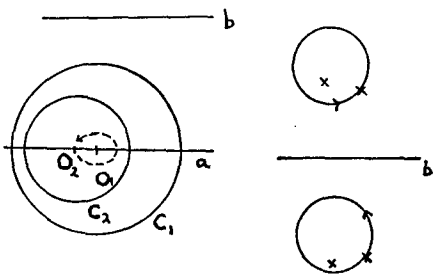


FIG. 17. Path around non-Landau singularity.

into the entire plane, the imaginary axis cut runs twice along the negative real axis in S. It is easy to prove that in this mapping the cut alternately disappears and has two sheets as the number of dimensions is increased. In Fig. 18, $2f(A) = f(A) - f(B)$ is $2r^{n-2}/A$ and so (r pure imaginary) is alternately pure real and pure imaginary. As the function is real, $f(B) = (\pm)^n f(-B)$, so there is alternation between single and two-sheeted behavior. None of the single-loop forward-scattering singularities is on the

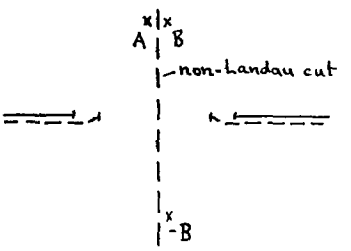


FIG. 18. Singularities in p-complex plane for self-energy part.

physical sheet. For the self-energy part, continuing towards $S = 0$ on the physical sheet gives the configuration in Fig. 19, for $y = 0$. Integrating over all real y, the contour is never distorted as the singularities remain in their own half planes. At infinity, C_{11}, C_{21} coincide and this gives rise to possible singularity only on those sheets where the contour is caught between them. Similar considerations hold for higher single loop graphs.

4.4. Connection Between Cutkosky's Formula and General Pinch Criteria

Theorem. Higher-order singularities appear in the discontinuity across Landau cuts.

Proof. If an integration is over a surface δ in a space D, with surfaces of singularity S_i, a singularity occurs when $S_i \cap \delta$ meet at a point where the normals (in δ) are linearly dependent. As the normal to each S_i at the point can be written in components

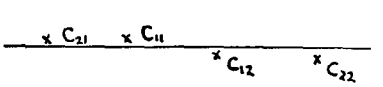


FIG. 19. Singularities on a (Fig. 17) for $p = 0$.

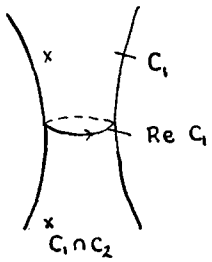


FIG. 20. Singularity configuration in the discontinuity across the normal threshold for the vertex part.

on δ and perpendicular to δ, in D the normals to the set (S_i, δ) are linearly dependent, so there is singularity in the integral over D with surfaces (S_i, δ) singular.

This theorem can be used to define an *equivalence relationship* among singular configurations. For instance, if three spheres have two coincident common points, and hence pinch the real three-space, the two circles $S_1 \cap S_2$ and $S_1 \cap S_3$ pinch the real two-space S_1. Further, if two circles touch and thus pinch the real two-space, the two points $C_1 \cap C_2$ pinch the real subspace of the surface C_1. (see Fig. 20). This configuration corresponds to two-sheeted inverse square root behavior, so we have established an induction procedure for reducing complicated configurations to very simple ones whose behavior is well known. This is equivalent to examining the singularity as it appears across successively lower-order thresholds.

4.5. Differences between Landauian and non-Landauian Singularities

Using this procedure it is easy to see that all Landauian singularities are equivalent to two poles pinching a line or an n sphere pinching real n space when its radius goes to zero. The order of magnitude of such a singularity is given by

$$\lim_{a\to 0} \int \frac{dx_1 \cdots dx_n}{x_1^2 + \cdots + x_n^2 - a^2} = \lim_{a\to 0} \int \frac{r^{n-1}\, dr}{r^2 - a^2} \sim a^{n-2}$$

If the order is a^0 the singularity is logarithmic—it never disappears. For a given singularity, the order depends on the dimensionality of space.

The vital point is that the non-Landauian singularities are not equivalent to the singularities listed above, and it is because of this that their existence is dimensional dependent, as we now demonstrate.

For example, consider the vertex part in three dimensions, taking the integral around the real circle $C = S_2 \cap S_3$ with two singularities S, $S'(S_1 \cap S_2 \cap S_3)$ on the complex surface of C. This configuration is as in Fig. 20, and corresponds to the discontinuity across the normal threshold. If on varying the external parameters, S', S exchange places the contour C has two extra parts, bubbles around S and S'. If now S_1, S_2, S_3, are moved so that their centers become collinear; S, S' move to infinity on C. This does not give rise to a singularity, because the residues at S, S' remain finite throughout. In four-dimensional space, S, S' are replaced by a circle on the complex part of a sphere. This can be adequately represented by multiplying the residues at S, S' by a weighting factor r corresponding to the radius of the circle. In this case, as S, S' go to infinity their residues become infinite and there is a singularity.

This can be regarded as a pole of constant residue approaching a singularity r^{n-3} at infinity, and in this way gives an easy derivation of the alternate one- and two-sheeted behavior as the dimensionality of the space is increased. Also, for $n = 3$ the pole of "constant residue" approaches a place where the function is analytic, so no singularity can arise.

From this viewpoint it is easy to see why the leading curve (corresponding to S, S' pinching C) switches the non-Landau singularity on and off, because on going around it S, S' pick up or lose the bubble contours measuring their residues. No singularity can be generated by a pinch between S and S' at ∞, as this would require only one of them to have a bubble contour.

Thus the non-Landau singularity arises through a pinch at infinity between a pole and a singularity r^{n-3} which actually disappears for $n = 3$. (For $n < 3$ there is no singularity for various obvious reasons.) This is quite unlike the Landau situation where varying the dimensionality only varies the type of pinch, as explained above.

ACKNOWLEDGMENTS

I am grateful to Dr. J. C. Polkinghorne for posing the problems that initiated this investigation, and for many valuable suggestions and criticisms throughout. I am also indebted to I. Drummond for many interesting conversations, and Dr. R. J. Eden for helpful comments during the preparation of this work for publication.

APPENDIX 1. N-SPACE SINGULARITY CRITERIA

We give here a generalization of the results of 2.3. If m surfaces of singularity in n-dimensional space meet at a point X, the lines in n-space through X can be divided into sets labeled $(i, i, \cdots, i)$, $(0, i, \cdots)$, etc., generalizing the previous notation. The normals to the surfaces at X span an m-dimensional subspace. It is clear that when one of these

sets collapses the normals lie in a space of dimension $m - 1$, and if the surfaces are quadratic with the same sign $i\epsilon$ on each, the collapse of $(i, i, \cdots, i)$ gives rise to a singularity. Thus if the surfaces are all spherical, the condition for singularity [on the $(+i\epsilon, \cdots, +i\epsilon)$ sheet] is that X lies inside the simplex of centers.

APPENDIX 2. AN ALTERNATIVE DERIVATION OF CUTKOSKY'S FORMULA FOR THE S.E. PART

It is possible to derive Cutkosky's formula without resorting to limiting processes by the following rather graphic method, which is admittedly not very rigorous in its present form. Consider the singularity configuration in the k_1 complex plane as k_2 is varied at constant speed along the real axis. We are only interested in behavior near the critical point, where the singularities can be taken as lines of strength t_1, t_2 at angles θ_1, θ_2 (see Fig. 21). Then in the complex k_1 plane given by $k_2 = \text{const}$ there are two poles, of strengths $t_1/\sin\theta_1$, $t_2/\sin\theta_2$ with velocity of approach v. One of these can be regarded as stationary, and varying k_2 is equivalent to letting the other one paint out a cut of strength (pole strength/velocity of approach), so the Cutkosky discontinuity, measured by a bubble contour around the stationary pole, is given by the product of the pole strengths divided by the velocity. In terms of θ this is

$$\frac{t_1}{\sin\theta_1}\frac{t_2}{\sin\theta_2}\left(\frac{1}{\tan\theta_1} - \frac{1}{\tan\theta_2}\right) = \frac{t_1 t_2}{\sin\theta}, \quad \theta = \theta_2 - \theta_1$$

It is clear from the usual limiting arguments that this is the integral over the product of delta functions, but it is interesting to show how this approach ties more directly with the delta integral.

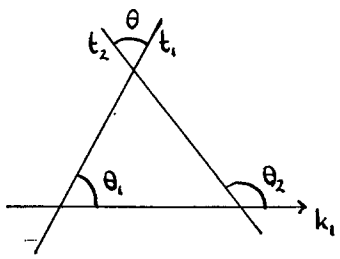


FIG. 21. Singularities near point of intersection for self-energy part.

$$\int \delta^{(+)}(k_1^2 + k_2^2 - m_1^2) \times \delta^{(-)}[(k_1 - p)^2 + k_2^2 - m_2^2]\, dk_1\, dk_2$$

$$= \int \delta^{(+)}(k_1^2 - f^2) \times \delta^{(-)}[(k_1 - p)^2 - g^2]\, dk_1\, dk_2, \quad \text{say,}$$

$$= \frac{1}{fg}\int \delta(k_1 - f)\, \delta(k_1 - p - g)\, dk_1\, dk_2$$

$$= \int \frac{dk_2}{fg}\, \delta(f - g - p)$$

$$= \frac{1}{fg(f' - g')}, \quad \text{where} \quad f - g - p = 0.$$

It is easy to show that the poles have strengths $1/f$, $1/g$ in the k_1 complex plane, and the velocity of approach is manifestly $f' - g'$. Doubtless this method can also be generalized, but becomes rather more difficult conceptually than the alternative approach developed in the text.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Asymptotic Properties of the Wave Function for a Bound Nonrelativistic Three-Body System*

E. Leo Slaggie†
Department of Physics, Columbia University, New York, New York
AND
Eyvind H. Wichmann‡
Department of Physics, University of California, Berkeley, California
(Received January 17, 1962)

The asymptotic properties of a Schrödinger wave function which represents the bound ground state of a system of three interacting particles are examined. It is assumed that the interaction can be described by a static potential which is the sum of three two-body potentials and one three-body potential, where the potentials have the property that if any one of the particles is separated from the other two by a distance which tends to infinity, then the part of the potential energy which depends on the position of that particle tends to zero. The problem is treated nonrelativistically. Decreasing exponential bounds on the ground-state three-body wave function are established in configuration space. It is shown that these bounds depend only on the masses of the three particles, on the ground-state energy of the three-body system, and on the lower bounds on the spectra of the Hamiltonians for the three two-body systems arising when one of the three particles is removed and the remaining two interact through the corresponding two-body potential.

It is furthermore shown that the three-body wave function in momentum space admits of analytic continuation into a tube-region in complex momentum space, which tube-region is the product of real momentum space and a convex bounded region in imaginary momentum space. The tube-region is explicitly determined.

The implications of the results for the theory of the vertex function in quantum field theory are discussed. The relevancy of the results obtained in this paper to the variational computation of energy levels and wave functions of three-body systems is also briefly discussed.

In the course of the derivation some generalizations are pointed out; in particular, the final results remain valid for a class of more general interactions, which cannot be described by static potentials. Likewise, the results remain valid for a class of spin-dependent interactions.

I. INTRODUCTION

We consider three spinless particles, numbered 1, 2, and 3. Let, for $u = 1, 2$, or 3, m_u be the mass of particle u, and let ϱ_u be the position vector of particle u with respect to some origin. The vector ϱ_{uv}, defined by

$$\varrho_{uv} = \varrho_u - \varrho_v, \tag{1}$$

is the vector joining particles u and v, and pointing towards particle u.

We consider the motion of these three particles, within the framework of the nonrelativistic Schrödinger equation, under the assumption that the interparticle interaction is described by the potential

$$U = U_{12}(\varrho_{12}) + U_{23}(\varrho_{23}) + U_{31}(\varrho_{31}) + U_{123}(\varrho_{12}, \varrho_{23}). \tag{2}$$

Since the Hamiltonian is translationally invariant the motion of the center of mass is independent of the motion relative to the center of mass. We are only interested in the relative motion and shall, therefore, introduce the constraint that the center of mass is at rest at the origin in three-space. This constraint is expressed by

$$m_1\varrho_1 + m_2\varrho_2 + m_3\varrho_3 = 0. \tag{3}$$

In the following we shall let the letters r, s, t, u, v stand for the indices which label the particles, with the convention that $u = 1, 2$, or 3 and $v = 1, 2$, or 3, whereas the triplet (r, s, t) always stands for some *cyclic* permutation of the triplet (1, 2, 3).

We shall assume that the potentials satisfy the bounds

$$|U_{rs}(\varrho_{rs})| < Q(|\varrho_{rs}|), \quad \text{for} \quad \varrho_{rs} \neq 0, \tag{4a}$$

$$|U_{123}(\varrho_{12}, \varrho_{23})| < Q(|\varrho_{12}| + |\varrho_{23}|), \quad \text{for} \quad |\varrho_{12}| + |\varrho_{23}| > 0, \tag{4b}$$

where the function $Q(q)$, defined for all $q > 0$, is a continuous positive function such that

$$\lim_{q\to\infty} Q(q) = 0, \quad \text{steadily}, \tag{5a}$$

$$\lim_{q\to 0} qQ(q) < \infty. \tag{5b}$$

It is not our purpose here to investigate under what conditions on the potentials the Schrödinger equations have meaningful solutions, nor to investigate under what conditions a three-body bound ground state does exist. We shall, therefore, assume that the potentials are sufficiently well behaved

* Research supported in part by the National Science Foundation and by the U. S. Atomic Energy Commission.
† Present address: Department of Physics, University of California, Santa Barbara, California.
‡ Alfred P. Sloan Foundation Fellow.

such that the various Schrödinger equations and corresponding integral equations which we shall consider in the following have physically meaningful solutions. In particular, we assume the potentials to be integrable over any finite region.

About the three-body system, we shall assume that it has a bound ground state of negative energy $-B_0$. We furthermore assume that the greatest lower bound of the spectrum of the Hamiltonian operator for the residual system obtained when particle r is removed and particles s and t interact through the potential U_{st} is given by the non-positive number $-B_r$, where, for $r = 1, 2$, or 3,

$$B_0 > B_r \geqq 0. \tag{6}$$

The reader will thus note that we do not assume that a bound ground state exists for the residual system obtained with the removal of particle r. If a bound ground state does not exist we have $B_r = 0$. If, on the other hand, a bound ground state does exist the assumption expressed by the inequality (6) states that the binding energy of any one of the two-body systems is less than the binding energy of the ground state of the three-body system.

The three-body ground-state wave function satisfies a Schrödinger equation on a six-dimensional configuration space. We may guess that when a point in this space tends towards infinity along some direction then the wavefunction at the point tends to zero exponentially. It is the principal problem of this paper to show that this is indeed the case, and to determine exponential bounds on the wave function in each direction in configuration space.

II. DIFFERENTIAL AND INTEGRAL EQUATIONS SATISFIED BY THE THREE-BODY WAVE FUNCTION

We first consider the question of selection of coordinates. For $r = 1, 2$, or 3 we get three different choices of two *independent* vectors which completely specify the configuration of the system; namely, the vectors $\boldsymbol{\varrho}_{st}$ and $\boldsymbol{\xi}_r$, where

$$\boldsymbol{\xi}_r = \boldsymbol{\varrho}_r(m_1 + m_2 + m_3)(m_s + m_t)^{-1}. \tag{7}$$

The vector $\boldsymbol{\xi}_r$ thus joins particle r with the center of mass of the t-s system. The corresponding reduced masses are given by

$$M_r = m_r(m_t + m_s)(m_1 + m_2 + m_3)^{-1}, \tag{8a}$$

$$\mu_r = m_t m_s(m_t + m_s)^{-1}. \tag{8b}$$

Because of their simple geometric interpretation the coordinates $\boldsymbol{\varrho}_{st}$ and $\boldsymbol{\xi}_r$ are useful for visualizing the configuration of the system. However, for our purposes the following three alternative sets of coordinates will be even more convenient: Let

$$\mathbf{x}_r = (2M_rB_0)^{1/2}\boldsymbol{\xi}_r$$

$$= [2B_0m_r(m_1 + m_2 + m_3)(m_s + m_t)^{-1}]^{1/2}\boldsymbol{\varrho}_r, \tag{9a}$$

$$\mathbf{y}_r = (2\mu_rB_0)^{1/2}\boldsymbol{\varrho}_{st}$$

$$= [2B_0m_tm_s(m_t + m_s)^{-1}]^{1/2}(\boldsymbol{\varrho}_s - \boldsymbol{\varrho}_t), \tag{9b}$$

where (r, s, t) is any cyclic permutation of $(1, 2, 3)$.

Equations (3) and (9) can be used to express $\boldsymbol{\varrho}_1$, $\boldsymbol{\varrho}_2$, and $\boldsymbol{\varrho}_3$ in terms of any one of the systems of coordinates $(\mathbf{x}_u, \mathbf{y}_u)$, $u = 1, 2$, or 3. Substituting the resulting expressions into the relations (9) which define $(\mathbf{x}_v, \mathbf{y}_v)$, $v = 1, 2$, or 3, we may thus explicitly determine the transformation from the $(\mathbf{x}_u, \mathbf{y}_u)$ system to the $(\mathbf{x}_v, \mathbf{y}_v)$ system. This transformation is orthogonal and can be expressed as follows:

$$\begin{aligned} \mathbf{x}_u &= \mathbf{x}_v \cos\theta_{uv} + \mathbf{y}_v \sin\theta_{uv}, \\ \mathbf{y}_u &= -\mathbf{x}_v \sin\theta_{uv} + \mathbf{y}_v \cos\theta_{uv}, \end{aligned} \tag{10a}$$

where

$$\theta_{uv} = -\theta_{vu},$$

$$\cos\theta_{rs} = -[m_rm_s(m_r + m_t)^{-1}(m_s + m_t)^{-1}]^{1/2}, \tag{10b}$$

$$\sin\theta_{rs} = -[m_t(m_1 + m_2 + m_3) \times (m_r + m_t)^{-1}(m_s + m_t)^{-1}]^{1/2}.$$

Thus, for $(r, s) = (1, 2)$, $(2, 3)$, or $(3, 1)$, we have

$$\tfrac{3}{2}\pi > \theta_{rs} > \pi. \tag{10c}$$

We regard the six Cartesian components of the vectors $\mathbf{x}_1$ and $\mathbf{y}_1$ as the Cartesian components of vector z in a six-dimensional Euclidean space $\mathcal{E}$. To express the fact that z can be specified by either one of the three pairs $(\mathbf{x}_r, \mathbf{y}_r)$ of three vectors, we write

$$z = (\mathbf{x}_1, \mathbf{y}_1)_1 = (\mathbf{x}_2, \mathbf{y}_2)_2 = (\mathbf{x}_3, \mathbf{y}_3)_3 \tag{11}$$

when the vectors $(\mathbf{x}_r, \mathbf{y}_r)$ satisfy Eqs. (10a). The subscripts after the parenthesis in Eq. (11) are meant to indicate particular decompositions of $\mathcal{E}$ into a sum of two mutually perpendicular three-dimensional Euclidean subspaces. The vectors $\mathbf{x}_r$ and $\mathbf{y}_r$ are, thus, the projections of z into the three-dimensional subspaces $\mathcal{E}_{xr}$ and $\mathcal{E}_{yr}$, respectively. The consistency of this interpretation follows from the fact that if the three pairs of vectors $(\mathbf{x}_r, \mathbf{y}_r)$, $r = 1, 2$, or 3, satisfy the relations (10a), then

$$z \cdot z = |\mathbf{x}_1|^2 + |\mathbf{y}_1|^2 = |\mathbf{x}_2|^2 + |\mathbf{y}_2|^2 = |\mathbf{x}_3|^2 + |\mathbf{y}_3|^2. \tag{12}$$

In terms of the coordinates which we have introduced the wave equation which the three-body ground-state wave function $\psi(z)$ satisfies assumes the form

$$[-\nabla_6^2 + V(z)]\psi(z) = -\psi(z). \tag{13}$$

The potential $V(z)$ is a sum of four terms:

$$V(z) = V_{12}(z) + V_{23}(z) + V_{31}(z) + V_{123}(z), \tag{14a}$$

where

$$\begin{aligned} V_{rs}(z) &= V_{rs}(\mathbf{y}_t) = U_{rs}(\boldsymbol{\varrho}_{rs})/B_0, \\ V_{123}(z) &= U_{123}(\boldsymbol{\varrho}_{12}, \boldsymbol{\varrho}_{23})/B_0. \end{aligned} \tag{14b}$$

We emphasize that the two-body potential $V_{rs}(z)$ depends only on the projection $\mathbf{y}_t$ of z into $\mathcal{E}_{yt}$ and we, therefore, sometimes write $V_{rs}(\mathbf{y}_t)$ interchangeably with $V_{rs}(z)$ to bring out this fact explicitly.

The symbol ∇_6^2 stands for the Laplacian operator in the six-dimensional space $\mathcal{E}$. It may be expressed in terms of the Laplacians ∇_{xr}^2 and ∇_{yr}^2 in the three-dimensional subspaces $\mathcal{E}_{xr}$ and $\mathcal{E}_{yr}$, respectively, as follows:

$$\nabla_6^2 = \nabla_{x1}^2 + \nabla_{y1}^2 = \nabla_{x2}^2 + \nabla_{y2}^2 = \nabla_{x3}^2 + \nabla_{y3}^2, \tag{15}$$

which relations hold when the various coordinates satisfy the relations (10a).

Let us restate the assumptions which we made in the introduction regarding the potentials as follows:

There exists a positive continuous function $Q(q)$, defined for all positive values of its argument q, and satisfying the conditions (5), such that:

$$\begin{aligned} |V_{rs}(z)| = |V_{rs}(\mathbf{y}_t)| < Q(|\mathbf{y}_t|), \quad \text{for} \quad \mathbf{y}_t \neq 0, \\ |V_{123}(z)| < Q(|z|), \quad \text{for} \quad z \neq 0, \end{aligned} \tag{16}$$

where

$$|z| = (z \cdot z)^{1/2}.$$

Let us likewise restate the conditions (6) as well as the assumptions contained in the paragraph preceding the inequalities (6):

The greatest lower bound on the spectrum of the operator

$$H'_{rs} = -\nabla_{yt}^2 + V_{rs}(\mathbf{y}_t) \tag{17}$$

is the nonpositive number $-b_t$, where

$$b_t = B_t/B_0, \qquad 1 > b_t \geqq 0. \tag{18}$$

We conclude this section by stating four integral equations which the three-body ground-state wave function $\psi(z)$ satisfies. Let

$$H_0 = -\nabla_6^2, \qquad H''_{rs} = -\nabla_{xt}^2, \tag{19a}$$

$$H_{rs} = H'_{rs} + H''_{rs} = -\nabla_6^2 + V_{rs}(z). \tag{19b}$$

The regions associated with these operators are taken to be the six- and three-dimensional Euclidean spaces, respectively. Then the spectrum of H_0 consists of all non-negative real numbers; the same being true for each one of the operators H''_{rs}. Since the operators H'_{rs} and H''_{rs}, where (r, s) is any one of the pairs (1, 2), (2, 3), or (3, 1), act on different Hilbert spaces, it follows that the operator $H_{rs} = H'_{rs} + H''_{rs}$, defined in Eq. (19), has as its spectrum the set of all real numbers $\geqq -b_t$. (The operator H_{rs} acts, of course, on the direct product of the two Hilbert spaces on which H'_{rs}, H''_{rs}, respectively, act.)

We define the resolvent operators (Green's functions)

$$G_0(\eta) = (H_0 - \eta)^{-1}, \tag{20a}$$

$$G_{rs}(\eta) = (H_{rs} - \eta)^{-1}, \tag{20b}$$

where the complex variable η is outside the respective spectra.

In the configuration space representation the Green's functions are, thus, solutions to the differential equations

$$(-\nabla_6^2 - \eta)G_0(z, z'; \eta) = \delta_6(z - z'), \tag{21a}$$

$$[-\nabla_6^2 + V_{rs}(z) - \eta]G_{rs}(z, z'; \eta) = \delta_6(z - z'), \tag{21b}$$

with the appropriate boundary conditions. In the Eqs. (21) the differentiations are with respect to z, and $\delta_6(z)$ is the Dirac delta function in $\mathcal{E}$.

Since the point $\eta = -1$ is outside the spectra of the operators H_0 and H_{rs} we may reformulate the differential equation (13) into four alternative integral equations, namely,

$$\psi(z) = -\int_{(\infty)} d^6(z')G_0(z, z'; -1)V(z')\psi(z'), \tag{22a}$$

$$\begin{aligned} \psi(z) = -\int_{(\infty)} d^6(z')G_{rs}(z, z'; -1) \\ \times [V_{st}(z') + V_{tr}(z') + V_{123}(z')]\psi(z'), \end{aligned} \tag{22b}$$

where (r, s, t) is any cyclic permutation of (1, 2, 3).

III. EXPONENTIAL BOUNDS FOR THE WAVE FUNCTION FOR A SINGLE PARTICLE BOUND IN A POTENTIAL

Before we attack the problem of determining the exponential bounds on the three-body wave function it will be useful to consider the analogous problem for the very simple case of a single particle bound by a static potential $V(\mathbf{x})$. By so doing, we can illustrate simply the main idea employed in this paper, and at the same time we will reach an under-

standing of the important difference between the three-body problem and the problem of the single particle.

We thus consider the Schrödinger equation for a single particle bound in the potential $V(\mathbf{x})$. With a suitable choice of variables the equation takes the form

$$[-\nabla_x^2 + V(\mathbf{x})]\phi(\mathbf{x}) = -\phi(\mathbf{x}), \tag{23}$$

where $\phi(\mathbf{x})$ is the bound ground-state wave function. About the potential we make the assumption that

$$|V(\mathbf{x})| < Q(|\mathbf{x}|), \quad \text{for} \quad \mathbf{x} \neq 0, \tag{24}$$

where $Q(q)$ is a continuous function, defined for all $q > 0$, which satisfies the conditions (5).

We denote by $G(\eta)$ the resolvent (Green's function) of the Laplacian in three-space:

$$G(\eta) = (-\nabla_x^2 - \eta)^{-1}, \tag{25a}$$

where η is any complex number not on the non-negative part of the real axis. Explicitly, the Green's function is given by

$$G(\mathbf{x}, \mathbf{x}'; \eta) = \exp(i\eta^{1/2}|\mathbf{x} - \mathbf{x}'|)/4\pi|\mathbf{x} - \mathbf{x}'|, \tag{25b}$$

where, in the complex η-plane cut along the positive real axis,

$$\text{Im}(\eta^{1/2}) > 0.$$

The wave function $\phi(\mathbf{x})$ then satisfies the integral equation

$$\phi(\mathbf{x}) = -\int_{(\infty)} d^3(\mathbf{x}') \frac{\exp(-|\mathbf{x} - \mathbf{x}'|)}{4\pi|\mathbf{x} - \mathbf{x}'|} V(\mathbf{x}')\phi(\mathbf{x}'). \tag{26}$$

We assert that for every θ such that $1 > \theta > 0$ there exists a constant $K(\theta)$ such that

$$|\phi(\mathbf{x})| < K(\theta) \exp(-\theta|\mathbf{x}|) \tag{27}$$

for all $\mathbf{x}$.

Proof. We select any θ such that $1 > \theta > 0$, and define a function $m(\mathbf{x})$ of $\mathbf{x}$ (and of θ) by

$$m(\mathbf{x}) = \underset{\mathbf{x}'}{\text{l.u.b.}} \{|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\}. \tag{28}$$

Here, and in the following, we employ the notation

$$\underset{z \in R}{\text{l.u.b.}} \{f(z)\}$$

for the least upper bound of the real function $f(z)$ as the variable z varies over the region R. To avoid misunderstandings we always exhibit the variable and its domain as above, except that if the region R is not mentioned the variable varies over all real values. We employ a similar notation for the greatest lower bound, which we abbreviate by g.l.b.

We use Eq. (26) and inequality (24), to derive the inequality

$$|\phi(\mathbf{x})| \leqq \int_{(\infty)} d^3(\mathbf{x}') \frac{\exp(-|\mathbf{x} - \mathbf{x}'|)}{4\pi|\mathbf{x} - \mathbf{x}'|} \times Q(|\mathbf{x}'|)\,|\phi(\mathbf{x}')|. \tag{29}$$

From the definition (28) it follows that

$$|\phi(\mathbf{x}')| \leqq m(\mathbf{x}) \exp(\theta|\mathbf{x} - \mathbf{x}'|).$$

When this estimate is inserted into the inequality (29) we obtain

$$|\phi(\mathbf{x})| \leqq h(\mathbf{x})m(\mathbf{x}), \tag{30}$$

where

$$h(\mathbf{x}) = \int_{(\infty)} d^3(\mathbf{x}') \times \frac{\exp[-(1-\theta)|\mathbf{x} - \mathbf{x}'|]}{4\pi|\mathbf{x} - \mathbf{x}'|} Q(|\mathbf{x}'|). \tag{31}$$

Let R be any region such that $h(\mathbf{x}) < 1$ throughout R, and let $\tilde{R}$ be the complement of R. By the definition (28) we may write

$$m(\mathbf{x}) = \max\{\underset{\mathbf{x}' \in R}{\text{l.u.b.}} [|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)], \underset{\mathbf{x}' \in \tilde{R}}{\text{l.u.b.}} [|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)]\}. \tag{32}$$

By the inequality (30), and by the definition of R, we have

$$\begin{aligned} \underset{\mathbf{x}' \in R}{\text{l.u.b.}} \{|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\} &< \underset{\mathbf{x}' \in R}{\text{l.u.b.}} \{m(\mathbf{x}') \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\} \\ &\leqq \underset{\mathbf{x}'}{\text{l.u.b.}} \{m(\mathbf{x}') \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\}. \end{aligned} \tag{33}$$

Since

$$\underset{\mathbf{x}'}{\text{l.u.b.}} \{\exp(-\theta|\mathbf{x} - \mathbf{x}'| - \theta|\mathbf{x}' - \mathbf{x}''|)\} = \exp(-\theta|\mathbf{x} - \mathbf{x}''|),$$

we obtain, by the definition (28),

$$m(\mathbf{x}) = \underset{\mathbf{x}'}{\text{l.u.b.}} \{m(\mathbf{x}') \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\}, \tag{34}$$

which result, when substituted into (33), gives

$$\underset{\mathbf{x}' \in R}{\text{l.u.b.}} \{|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\} < m(\mathbf{x}).$$

This means that Eq. (32) may, in fact, be written

$$m(\mathbf{x}) = \underset{\mathbf{x}' \in \tilde{R}}{\text{l.u.b.}} \{|\phi(\mathbf{x}')| \exp(-\theta|\mathbf{x} - \mathbf{x}'|)\}. \tag{35}$$

Let us now consider the function $h(\mathbf{x})$, defined by

Eq. (31). We note that $h(\mathbf{x})$ in fact depends only on $|\mathbf{x}|$. Furthermore, we easily see that since $Q(q)$ tends monotonically to zero as q tends to infinity it follows that $h(\mathbf{x})$ tends to zero as $|\mathbf{x}|$ tends to infinity. The region R may therefore be selected as the surface and exterior of some sphere. The region $\tilde{R}$ is consequently the interior of this sphere, and hence a bounded region. It then follows immediately from Eq. (35) that

$$m(\mathbf{x}) < K(\theta) \exp(-\theta |\mathbf{x}|) \tag{36}$$

for some constant $K(\theta)$. From the definition (28) follows that

$$|\phi(\mathbf{x})| \leqq m(\mathbf{x}), \tag{37}$$

which fact, combined with the inequality (36), proves our assertion.[1]

We note that this procedure works for any bound state, not only for the ground state. By considering the case of a Coulomb potential for which explicit solutions are known we can conclude that the exponential bound expressed by the set of inequalities (27) is the best possible in the sense that, in general, an inequality like (27) cannot hold for $\theta \geqq 1$.

We furthermore note that the restrictions on the potential are unnecessarily severe; all that is needed for the success of our method is that for every θ such that $1 > \theta > 0$ the function $h(\mathbf{x})$, as defined in Eq. (31), is less than one outside some bounded region. (The region may, of course, depend on θ.)

Let us now try to generalize this method to the case of the three-body wave function, basing our discussion on the integral equation (22a) of the preceding section. As is well known, the Green's function $G_0(z, z'; -1)$ is a function of $|z - z'|$ only, and has the property that it falls off faster, as $|z - z'|$ tends to infinity, than the function $\exp(-\theta |z - z'|)$ whenever $1 > \theta$. Let us select any θ such that $1 > \theta > 0$, and let us define a function $h_0(z)$ by

$$h_0(z) = \int_{(\infty)} d^6(z') \, |G_0(z, z'; -1)| \times [Q(|\mathbf{y}_1'|) + Q(|\mathbf{y}_2'|) + Q(|\mathbf{y}_3'|) + Q(|z'|)] \exp(\theta |z - z'|). \tag{38}$$

Let R_0 be any region in $\mathcal{E}$ such that $h_0(z) < 1$ throughout the region, and let $\tilde{R}_0$ be the complement of R_0. By a procedure very similar to the one by which we have derived the equality (35), we may derive the result that

$$|\psi(z)| \leqq \text{l.u.b.}_{z' \in \tilde{R}_0} \{|\psi(z')| \exp(-\theta |z - z'|)\}. \tag{39}$$

If we now compare $h_0(z)$, as defined by Eq. (38), to $h(\mathbf{x})$, as defined by Eq. (31), we notice an essential difference; the function $h(\mathbf{x})$ is less than one outside some bounded region in three-space, whereas it cannot be concluded that $h_0(z)$ is less than one everywhere in $\mathcal{E}$ outside some bounded region. The reason, is, of course, that $V(z)$ does not tend to zero as $|z|$ tends to infinity, or differently stated, $Q(|\mathbf{y}_r|)$, for $r = 1$, 2, or 3, does not necessarily tend to zero as $|z|$ tends to infinity.[2]

A moment's reflection tells us that the above state of affairs is entirely in accordance with expectations. We may consider, for instance, the deuterium atom as an example. In this case it is intuitively clear that the probability density of the electron in space at large distances from the center of mass depends primarily on the binding energy of the three-body system with respect to a breakup of the system into an electron and a deuteron, and not on the binding energy with respect to a breakup in which all three particles are completely separated. In fact, it is a good approximation for the purposes of atomic physics to regard the deuteron as a single particle, in which case its binding energy with respect to its breakup into a proton and neutron plays no role. However, the procedure leading to the inequality (39) takes no account of any two-body binding energies, and can, therefore, not give us the final answer desired.

Intuitive physical arguments, thus, suggest that the asymptotic behavior of the three-body wave function must necessarily depend on the constants B_1, B_2, and B_3, as well as on B_0, and the masses m_1, m_2, and m_3. If we now examine the three integral equations (22b) we note two facts: (a) The Green's functions G_{rs} "contain" the solutions to the three two-body problems and we may expect that certain exponential bounds on the functions G_{rs} will depend explicitly on the constants B_1, B_2, and B_3 (or rather on b_1, b_2, and b_3); (b) one of the two-body potentials is absent in each one of the integrands. This latter fact makes it plausible that one may employ the integral equations (22b) to derive three inequalities of a similar nature as the inequality (39), but with $\tilde{R}_0$ replaced by larger regions. We shall show that this is indeed the case, and that the four inequalities which one obtains in this manner are sufficiently stringent for the establishment of exponential bounds on the ground-state wave function $\psi(z)$ in every direction in $\mathcal{E}$. The procedure which we shall

[1] For an alternative demonstration of this assertion, see Appendix III.

[2] See also E. Gerjuoy, Ann. Phys. (New York) 5, 58 (1958)

follow can be regarded as fairly straightforward: First, we establish exponential bounds on the Green's functions $G_{rs}(z, z'; -1)$ as functions of $(z - z')$, and then we employ the integral equations (22) to establish exponential bounds on $\psi(z)$.

Let us return to Eq. (38) in which $h_0(z)$ is defined. By inspection of the integrand we note the following: For every θ such that $1 > \theta > 0$ there exists a λ_θ such that when $|\mathbf{y}_r| > \lambda_\theta$ for $r = 1, 2$, and 3, then $h_0(z) < 1$. This enables us to state:

Lemma I. Let θ be any number such that $1 > \theta > 0$. Then there exists a number λ_θ, such that if R_0 is the region consisting of all points z such that $|\mathbf{y}_r| > \lambda_\theta$ for $r = 1, 2$, and 3, and if $\tilde{R}_0$ is the complement of R_0, then the three-body wave function $\psi(z)$ satisfies the inequality

$$|\psi(z)| \leqq \underset{z' \in \tilde{R}_0}{\text{l.u.b.}} \{|\psi(z')| \exp(-\theta |z - z'|)\} \tag{40}$$

for all z in $\mathcal{E}$.

This result indicates the importance of the region $\tilde{R}_0$ which can be characterized physically as a region such that if the three-body system "breaks up" along a direction within $\tilde{R}_0$, i.e., if z tends to infinity within $\tilde{R}_0$, then the separation between two of the particles remains bounded.

IV. EXPONENTIAL BOUNDS ON THE GREEN'S FUNCTIONS $G_{rs}(z, z'; -1)$

We begin by considering the problem of a single particle in a field of force derivable from the potential $V(\mathbf{y})$, where $V(\mathbf{y})$ satisfies the inequality

$$|V(\mathbf{y})| < Q(|\mathbf{y}|), \quad \text{for} \quad \mathbf{y} \neq 0, \tag{41}$$

in three-dimensional $\mathbf{y}$ space, for some positive continuous function $Q(q)$, defined for all $q > 0$, which satisfies the conditions (5).

Let H_y' be the operator

$$H_y' = -\nabla_y^2 + V(\mathbf{y}), \tag{42}$$

and let the number $-b$ be the greatest lower bound on the spectrum of H_y'. We thus have $b > 0$ if a bound ground state exists; otherwise $b = 0$.

We shall furthermore assume that $1 > b$. The spectrum of H_y' is confined to the real axis, and includes the positive half of the real axis.

Let η be a complex number not in the spectrum of H_y'. Then the Green's function $G'(\eta)$, defined symbolically by

$$G'(\eta) = (H_y' - \eta)^{-1}, \tag{43a}$$

is the solution, with the appropriate boundary conditions, of the differential equation

$$(-\nabla_y^2 + V(\mathbf{y}) - \eta)G'(\mathbf{y}, \mathbf{y}'; \eta) = \delta_3(\mathbf{y} - \mathbf{y}'). \tag{43b}$$

Concerning the Green's function we make the following assertion,

Lemma II. Let

$$\rho = \text{Im}\,(\eta^{1/2}), \tag{44}$$

where $\eta^{1/2}$ is that branch of the function $\eta^{1/2}$ which has a positive imaginary part on the complement of the spectrum of H_y'. Then, for every $r_0 > b$, and every θ such that $1 > \theta > 0$, there exists a constant $K(\theta, r_0)$ such that

$$|G'(\mathbf{y}, \mathbf{y}'; \eta)| < \frac{K(\theta, r_0)}{|\mathbf{y} - \mathbf{y}'|} \exp(-\theta\rho\, |\mathbf{y} - \mathbf{y}'|) \tag{45}$$

for all $\mathbf{y}$, $\mathbf{y}'$, and throughout the half-plane $\text{Re}\,(\eta) \leqq -r_0$.

Since the detailed proof of this lemma is a bit cumbersome, we shall defer it until Appendix I. To give a plausibility argument, we point out that for fixed $\mathbf{y}'$, and for $\mathbf{y} \neq \mathbf{y}'$, the Green's function as a function of $\mathbf{y}$ satisfies a "Schrödinger equation with energy η," and it is therefore reasonable that the method used in the preceding section to establish an exponential bound on a one-particular wave function can be modified to yield the exponential bound stated in inequality (45), but with the constant K replaced by a constant which depends on θ and η. It is perhaps not quite so obvious that the dependence of this constant on η is as stated in the lemma.

Let us next study the resolvent of the operator

$$H = H_x'' + H_y', \tag{46}$$

where

$$H_x'' = -\nabla_x^2.$$

The remarks following Eqs. (19) apply, and the spectrum of H thus consists of all real numbers $\geqq -b$. Since we assumed that $1 > b$, the number -1 is not in the spectrum of H. This fact, as well as the fact that H_x'' and H_y' act on different Hilbert spaces (and hence commute), permits us to write symbolically

$$G(-1) = (H + 1)^{-1}$$

$$= \frac{1}{2\pi i} \int_C d\eta (H_x'' + 1 + \eta)^{-1}(H_y' - \eta)^{-1}, \tag{47a}$$

where C is any contour going from infinity in the lower half of the η plane to infinity in the upper half of the η plane, and in such a manner that the spectrum of H_y' is entirely to the right of C, whereas the spectrum of $-(H_x'' + 1)$ is entirely to the left

of C. Integral representations like the one in Eq. (47a) have been considered quite generally by Friedman.[3] In Appendix II we shall comment briefly on the validity of the representation. Let us now select a straight line parallel to the imaginary axis as the contour of integration C. We then have explicitly

$$G(z, z'; -1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt \frac{G'(\mathbf{y}, \mathbf{y}'; -r + it)}{4\pi |\mathbf{x} - \mathbf{x}'|} \times \exp [i(r - 1 - it)^{1/2} |\mathbf{x} - \mathbf{x}'|], \quad (47b)$$

where $z = (\mathbf{x}, \mathbf{y})$; $z' = (\mathbf{x}', \mathbf{y}')$, and where the point $-r$, at which the contour intersects the real axis, satisfies the condition

$$1 > r > b. \quad (48)$$

Let us now select a number r_0 such that $1 > r_0 > b$, and let us select a number θ' such that $1 > \theta' > 0$. Let r be a fixed number such that $1 > r \geqq r_0$. Employing the estimate on $G'(\mathbf{y}, \mathbf{y}'; -r + it)$ given in Lemma II, we may trivially derive an inequality from Eq. (47b), namely,

$$|G(z, z'; -1)| < K_1(\theta', r_0)[|\mathbf{x} - \mathbf{x}'| \times |\mathbf{y} - \mathbf{y}'| (|\mathbf{x} - \mathbf{x}'| + |\mathbf{y} - \mathbf{y}'|)^2]^{-1} \times \exp [-\theta' r^{1/2} |\mathbf{y} - \mathbf{y}'| - \theta'(1 - r)^{1/2} |\mathbf{x} - \mathbf{x}'|] \quad (49)$$

for some constant K_1 which depends on θ' and r_0 only.

The number r, subject to the restriction $1 > r \geqq r_0$, is at our disposal, and we may select it such that the most stringent bound on $G(z, z'; -1)$ is obtained. Hence,

$$|G(z, z'; -1)| \leqq K_1(\theta', r_0)[|\mathbf{x} - \mathbf{x}'| \times |\mathbf{y} - \mathbf{y}'| (|\mathbf{x} - \mathbf{x}'| + |\mathbf{y} - \mathbf{y}'|)^2]^{-1} \times \underset{1 \geqq r \geqq r_0}{\text{g.l.b.}} \{\exp [-\theta' r^{1/2} |\mathbf{y} - \mathbf{y}'| - \theta'(1 - r)^{1/2} |\mathbf{x} - \mathbf{x}'|]\}. \quad (50)$$

If we now select

$$\theta = \theta'[(1 - r_0)/(1 - b)]^{1/2}, \quad (51)$$

we obtain from the inequality (50), the inequality

$$|G(z, z'; -1)| \leqq K_1(\theta', r_0)[|\mathbf{x} - \mathbf{x}'| \times |\mathbf{y} - \mathbf{y}'| (|\mathbf{x} - \mathbf{x}'| + |\mathbf{y} - \mathbf{y}'|)^2]^{-1} \times \underset{1 \geqq r \geqq b}{\text{g.l.b.}} \{\exp [-\theta r^{1/2} |\mathbf{y} - \mathbf{y}'| - \theta(1 - r)^{1/2} |\mathbf{x} - \mathbf{x}'|]\}. \quad (52)$$

We note that by selecting θ' sufficiently close to 1, and by selecting r_0 sufficiently close to b, we may obtain a θ, as defined by Eq. (51), which is as close to 1 as we wish. This permits us to formulate the following lemma.

Lemma III. For every θ such that $1 > \theta > 0$ there exists a constant $K(\theta)$ such that

$$|G(z, z'; -1)| < K(\theta)[|\mathbf{x} - \mathbf{x}'| \times |\mathbf{y} - \mathbf{y}'| (|\mathbf{x} - \mathbf{x}'| + |\mathbf{y} - \mathbf{y}'|)^2]^{-1} \times E(\theta(\mathbf{x} - \mathbf{x}'), \theta(\mathbf{y} - \mathbf{y}'); b), \quad (53)$$

where the function E is defined by

$$E(\mathbf{x}, \mathbf{y}; b) = \underset{1 \geqq r \geqq b}{\text{g.l.b}} \{\exp [-r^{1/2} |\mathbf{y}| - (1 - r)^{1/2} |\mathbf{x}|]\}. \quad (54)$$

This lemma thus gives the required exponential bounds on the Green's functions $G_{r_0}(z, z'; -1)$, as functions of $(z - z')$. Before we state these bounds in detail, let us first establish some properties of the functions $E(\mathbf{x}, \mathbf{y}; b)$ which we shall later make use of. We thus formulate Lemma IV.

Lemma IV. Let, for $1 > b \geqq 0$, the function $E(\mathbf{x}, \mathbf{y}; b)$ be defined by Eq. (54). Let the function $E_0(z)$ be given by

$$E_0(z) = E_0(\mathbf{x}, \mathbf{y}) = \exp [-(|\mathbf{x}|^2 + |\mathbf{y}|^2)^{1/2}] = \exp (-|z|). \quad (55)$$

Then,

$$\text{(a)} \quad E(\mathbf{x}, \mathbf{y}; b) = \begin{cases} E_0(\mathbf{x}, \mathbf{y}), & \text{if } |\mathbf{y}|^2 \geqq b(|\mathbf{x}|^2 + |\mathbf{y}|^2), \\ \exp [-b^{1/2} |\mathbf{y}| - (1 - b)^{1/2} |\mathbf{x}|], & \text{if } |\mathbf{y}|^2 \leqq b(|\mathbf{x}|^2 + |\mathbf{y}|^2), \end{cases} \quad (56a)$$

$$\text{(b)} \quad E(\mathbf{x}, \mathbf{y}; b) \geqq E_0(\mathbf{x}, \mathbf{y}) = E(\mathbf{x}, \mathbf{y}; 0) > 0, \quad (56b)$$

$$\text{(c)} \quad E(\mathbf{x}, \mathbf{y}; b) = \underset{\mathbf{x}', \mathbf{y}'}{\text{l.u.b.}} \{E(\mathbf{x}', \mathbf{y}'; b) \times E(\mathbf{x} - \mathbf{x}', \mathbf{y} - \mathbf{y}'; b)\}; \quad (56c)$$

$$\text{(d)} \quad E(\mathbf{x}, \mathbf{y}; b) = \underset{\mathbf{x}'}{\text{l.u.b.}} \{E(\mathbf{x} - \mathbf{x}', \mathbf{y}; b) \times \exp [-(1 - b)^{1/2} |\mathbf{x}'|]\}; \quad (56d)$$

[3] B. Friedman, "An Abstract Formulation of the Method of Separation of Variables," Proceedings of the Conference on Differential Equations held at the University of Maryland, March, 1955 (unpublished); See also B. Friedman *Principles and Techniques of Applied Mathematics* (John Wiley & Sons, Inc., New York, 1957), p. 273.

$$\text{(e)}\quad E(\mathbf{x},\mathbf{y};b) = \text{l.u.b.}_{\mathbf{x}'}\{E_0(\mathbf{x}-\mathbf{x}',\mathbf{y}) \times \exp[-(1-b)^{1/2}|\mathbf{x}'|]\}; \tag{56e}$$

$$\text{(f)}\quad E(\mathbf{x},\mathbf{y};b) = \text{l.u.b.}_{\mathbf{x}',\mathbf{y}'}\{E_0(\mathbf{x}-\mathbf{x}',\mathbf{y}-\mathbf{y}') \times E(\mathbf{x}',\mathbf{y}';b)\}. \tag{56f}$$

Proof. The proofs of the statements (a) and (b) are trivial. To prove statement (c) we first note that since $E(0,0;b) = 1$, we have

$$\text{l.u.b.}_{\mathbf{x}',\mathbf{y}'}\{E(\mathbf{x}',\mathbf{y}';b)E(\mathbf{x}-\mathbf{x}',\mathbf{y}-\mathbf{y}';b)\} \geqq E(\mathbf{x},\mathbf{y};b). \tag{57}$$

On the other hand,

$$\begin{aligned} &E(\mathbf{x}',\mathbf{y}';b)E(\mathbf{x}-\mathbf{x}',\mathbf{y}-\mathbf{y}';b) \\ &\quad = \text{g.l.b.}_{\substack{1\geqq r\geqq b\\ 1\geqq r'\geqq b}}\{\exp[-r^{1/2}|\mathbf{y}'| - (1-r)^{1/2}|\mathbf{x}'| \\ &\qquad - r'^{1/2}|\mathbf{y}-\mathbf{y}'| - (1-r')^{1/2}|\mathbf{x}-\mathbf{x}'|]\} \\ &\quad \leqq \text{g.l.b.}_{1\geqq r\geqq b}\{\exp[-r^{1/2}(|\mathbf{y}'| + |\mathbf{y}-\mathbf{y}'|) \\ &\qquad - (1-r)^{1/2}(|\mathbf{x}'| + |\mathbf{x}-\mathbf{x}'|)]\} \leqq E(\mathbf{x},\mathbf{y};b), \end{aligned}$$

which, together with the inequality (57), establishes part (c) of the lemma.

To prove statement (d) we note that

$$E(\mathbf{x},0;b) = \exp[-(1-b)^{1/2}|\mathbf{x}|].$$

Therefore,

$$\begin{aligned} &\text{l.u.b.}_{\mathbf{x}',\mathbf{y}'}\{E(\mathbf{x}-\mathbf{x}',\mathbf{y}-\mathbf{y}';b)E(\mathbf{x}',\mathbf{y}';b)\} \\ &\qquad \geqq \text{l.u.b.}_{\mathbf{x}'}\{E(\mathbf{x}-\mathbf{x}',\mathbf{y};b) \\ &\qquad\qquad \times \exp[-(1-b)^{1/2}|\mathbf{x}'|]\} \geqq E(\mathbf{x},\mathbf{y};b), \end{aligned}$$

which, together with the statement in (c), proves the assertion.

To prove part (e) we first note that it follows from parts (b) and (d) that

$$\text{l.u.b.}_{\mathbf{x}'}\{E_0(\mathbf{x}-\mathbf{x}',\mathbf{y})\exp[-(1-b)^{1/2}|\mathbf{x}'|]\} \leqq E(\mathbf{x},\mathbf{y};b). \tag{58}$$

Furthermore, we trivially have

$$\text{l.u.b.}_{\mathbf{x}'}\{E_0(\mathbf{x}-\mathbf{x}',\mathbf{y})\exp[-(1-b)^{1/2}|\mathbf{x}'|]\} \geqq E_0(\mathbf{x},\mathbf{y}).$$

This result, taken together with the inequality (58), proves the assertion for the case where $|\mathbf{y}|^2 \geqq b(|\mathbf{x}|^2 + |\mathbf{y}|^2)$.

For the case where $b(|\mathbf{x}|^2 + |\mathbf{y}|^2) > |\mathbf{y}|^2$ (which can obtain only if $b > 0$, and $\mathbf{x} \neq 0$), we simply note that for

$$\mathbf{x}' = \mathbf{x}\left[1 - \frac{|\mathbf{y}|}{|\mathbf{x}|}\left(\frac{1-b}{b}\right)^{1/2}\right],$$

we have

$$E_0(\mathbf{x}-\mathbf{x}',\mathbf{y})\exp[-(1-b)^{1/2}|\mathbf{x}'|] = E(\mathbf{x},\mathbf{y};b),$$

which, together with the inequality (58), completes the proof.

The proof of assertion (f) is trivial once assertions (a) to (e) have been proved.

We conclude by stating in a theorem the principal result of this section, namely, the exponential bounds on the Green's function G_{rs}.

Theorem I. Let $G_{rs}(z,z';\eta)$ be the Green's functions defined by Eq. (20b) or (21b), subject to all the assumptions made in Secs. I and II. Let the three functions $E_r(z)$, $r = 1, 2$, or 3, be defined by

$$E_r(z) = E(\mathbf{x}_r,\mathbf{y}_r;b_r), \tag{59}$$

where the function $E(\mathbf{x},\mathbf{y};b)$ is defined by Eq. (54).

Let θ be any number such that $1 > \theta > 0$. Then there exists a constant $K(\theta)$ such that for (r, s, t) any cyclic permutation of (1, 2, 3)

$$|G_{rs}(z,z';-1)| < K(\theta)E_t(\theta(z-z'))[|\mathbf{x}_t - \mathbf{x}_t'| \times |\mathbf{y}_t - \mathbf{y}_t'|(|\mathbf{x}_t - \mathbf{x}_t'| + |\mathbf{y}_t - \mathbf{y}_t'|)^2]^{-1}. \tag{60}$$

V. EXPONENTIAL BOUNDS ON THE THREE-BODY GROUND-STATE WAVE FUNCTION

We shall now make use of the results contained in Theorem I to derive from the integral equations (22b) some inequalities satisfied by the wave function $\psi(z)$. Let θ be a fixed constant such that $1 > \theta > 0$, and let (r, s, t) be a fixed cyclic permutation of (1, 2, 3). By inspection of the integral equations (22b) we obtain the inequality

$$|\psi(z)| \leqq \int_{(\infty)} d^6(z')F_1(z,z')F_2(z,z')F_3(z,z'), \tag{61a}$$

where

$$F_1(z,z') = E_t(\theta(z-z'))\,|\psi(z')|, \tag{61b}$$

$$F_2(z,z') = E_t[\tfrac{1}{2}(1-\theta)(z-z')]\left\{|V_{st}(z')|\frac{|\mathbf{y}_r'|}{1+|\mathbf{y}_r'|} + |V_{tr}(z')|\frac{|\mathbf{y}_s'|}{1+|\mathbf{y}_s'|} + |V_{123}(z')|\frac{|z'|}{1+|z'|}\right\}, \tag{61c}$$

$$F_3(z,z') = |G_{rs}(z,z';-1)|\,\{E_t[\tfrac{1}{2}(1+\theta)(z-z')]\}^{-1} \times \left\{\frac{1+|\mathbf{y}_r'|}{|\mathbf{y}_r'|} + \frac{1+|\mathbf{y}_s'|}{|\mathbf{y}_s'|} + \frac{1+|z'|}{|z'|}\right\}. \tag{61d}$$

It follows from Theorem I that the integral

$$\int_{(\infty)} d^6(z')F_3(z, z') < K_1 \tag{62}$$

for some constant K_1, which depends on θ but *not* on z.

Let

$$h(z) = \text{l.u.b.}_{z'} \left\{K_1 E_t[\tfrac{1}{2}(1-\theta)(z-z')]\left[\frac{Q(|\mathbf{y}_r'|)\ |\mathbf{y}_r'|}{1+|\mathbf{y}_r'|} + \frac{Q(|\mathbf{y}_s'|)\ |\mathbf{y}_s'|}{1+|\mathbf{y}_s'|} + \frac{Q(|z'|)\ |z'|}{1+|z'|}\right]\right\}, \tag{63a}$$

where $Q(q)$ is the function introduced in connection with the inequalities (16). The function $h(z)$ exists for all z, and it follows from Eq. (61c) and the inequalities (16) that

$$K_1 F_2(z, z') \leqq h(z), \quad \text{for all} \quad z'. \tag{63b}$$

Let

$$m(z) = \text{l.u.b.}_{z'} \{E_t(\theta(z-z'))\ |\psi(z')|\}. \tag{64}$$

We then have

$$F_1(z, z') \leqq m(z); \qquad |\psi(z)| \leqq m(z). \tag{65}$$

Combining statements (61) to (65), we then obtain

$$|\psi(z)| < h(z)m(z). \tag{66}$$

It is worthwhile to compare the present study with the investigation in Sec. III. We note that the definition (64) is analogous to the definition (28); Eq. (65) is analogous to Eq. (37), and the inequality (66) is analogous to inequality (30). Equation (34) played an important role in the proof given in Sec. III; the corresponding relation for the present study is

$$\text{l.u.b.}_{z'} \{E_t[\theta(z-z')]m(z')\} = m(z). \tag{67}$$

This relation is an immediate consequence of the definition (64) and Eq. (56c) in Lemma IV.

Now let R be any region such that $h(z) < 1$ throughout R, and let $\tilde{R}$ be the complement of R. By a procedure entirely analogous to the one followed in Sec. III we obtain the analog of Eq. (35), namely,

$$m(z) = \text{l.u.b.}_{z' \in \tilde{R}} \{|\psi(z')| E_t[\theta(z-z')]\}. \tag{68}$$

By inspection of the definition (63a) we see that we may select as the region R the set of all z such that

$$|\mathbf{y}_r| > \lambda, \qquad |\mathbf{y}_s| > \lambda, \tag{69}$$

where λ is a sufficiently large constant (which depends on θ).

We shall now formulate in the form of a lemma the results obtained so far in this section, together with the results expressed by Lemma I.

Lemma V. Let θ be any number such that $1 > \theta > 0$. Then there exists a number λ such that the three-body ground-state wave function $\psi(z)$ satisfies the four inequalities

$$|\psi(z)| \leqq \text{l.u.b.}_{z' \in \tilde{R}_1 \cup \tilde{R}_2 \cup \tilde{R}_3} \{E_0[\theta(z-z')]\ |\psi(z')|\}, \tag{70a}$$

$$|\psi(z)| \leqq \text{l.u.b.}_{z' \in \tilde{R}_r \cup \tilde{R}_s} \{E_t[\theta(z-z')]\ |\psi(z')|\}, \tag{70b}$$

where the functions $E_t(z)$, $t = 1, 2, 3$, are defined in Eq. (59), and where the function $E_0(z)$ is defined in Eq. (55), and where the region R_t, $t = 1, 2$, and 3, is the set of all points z such that $|\mathbf{y}_t| > \lambda$, and where $\tilde{R}_t$ is the complement of R_t.

We shall show that the four inequalities (70) are inconsistent if $|\psi(z)|$ falls off too slowly as z tends to infinity. For this purpose we state:

Lemma VI. Let k be any number such that $1 \geqq k \geqq 0$. Let

$$B(z) = \max\{E_1(z), E_2(z), E_3(z)\}. \tag{71}$$

Then, with the notation of Lemma V,

$$\text{(a)} \quad \text{l.u.b.}_{z' \in \tilde{R}_1 \cup \tilde{R}_2 \cup \tilde{R}_3} \{B(z')E_0(z-z')\} = B(z), \tag{72a}$$

$$\text{(b)} \quad \text{l.u.b.}_{z' \in \tilde{R}_r \cup \tilde{R}_s} \{E_t(z-z')B(z')\} \geqq B(z), \tag{72b}$$

$$\text{(c)} \quad \text{l.u.b.}_{z' \in \tilde{R}_1 \cup \tilde{R}_2 \cup \tilde{R}_3} \{B(kz')E_0(z-z')\} \leqq B(kz). \tag{72c}$$

Finally, for all z in $\tilde{R}_t$,

$$\text{(d)} \quad \text{l.u.b.}_{z' \in \tilde{R}_r \cup \tilde{R}_s} \{E_t(z-z')B(kz')\} < K_0 B\{[k + \delta(1-k)]z\}, \tag{72d}$$

where K_0 and $1 \geqq \delta \geqq 0$ are some fixed constants which *do not* depend on k.

Proof. The statements (a) to (c) follow trivially from the definition (71) and the properties of the functions E listed in Lemma IV, and we may omit the detailed demonstration.

To prove the assertion (d), we define the linear manifolds N_t, $t = 1$, 2, or 3, such that N_t is the set of all z such that $\mathbf{y}_t = 0$. We note that λ is the upper bound on the distance from any point of $\tilde{R}_t$ to N_t. Let z_1 be in $\tilde{R}_t$, and let z_0 be the projection of z_1 into N_t. We then have $|z_1 - z_0| \leqq \lambda$. From the exponential nature of the functions $E_t(z)$ and $B(z)$ follows that

$$B([k + (1-k)\ \delta]z_0) < K_1 B([k + (1-k)\ \delta]z_1),$$

and

$$\text{l.u.b.}_{z'\in\tilde{R}_t\cup\tilde{R}_s}\{E_t(z_1-z')B(kz')\} < K_2\ \text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t(z_0-z')B(kz')\},$$

where K_1 and K_2 are constants independent of k and δ, where k and δ are such that $1 \geqq k \geqq 0$, $1 \geqq \delta \geqq 0$.

To prove assertion (d) it suffices to show that there exists a δ, $1 \geqq \delta > 0$, with the property that for every z in N_t, and for every k such that $1 \geqq k \geqq 0$,

$$\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t(z-z')B(kz')\} \leqq B\{[k+\delta(1-k)]z\}. \tag{73}$$

We show this as follows: For z in N_t we have

$$\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t(z-z')B(kz')\} \leqq [\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t[(1-k)(z-z')]\}] \times [\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t[k(z-z')]B(kz')\}]. \tag{74}$$

But we certainly have

$$\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t[(1-k)(z-z')]\} \leqq B[\delta(1-k)z] \tag{75}$$

for some δ which satisfies $1 \geqq \delta > 0$. Furthermore, by part (e) of Lemma IV,

$$\text{l.u.b.}_{z'\in N_t\cup N_s}\{E_t(z-z')B(z')\} = \text{l.u.b.}_{\substack{z'\in N_t\cup N_s\\ z''\in N_t}}\{E_0(z''-z')B(z')E_t(z-z'')\} \leqq \text{l.u.b.}_{z''\in N_t}\{B(z'')E_t(z-z'')\} = B(z). \tag{76}$$

Inserting the results contained in inequalities (75) and (76) into inequality (74), we obtain the inequality (73), which completes the proof of the lemma.

We are now in a position to state the principal result of this paper as follows:

Theorem II. (a) For every θ such that $1 > \theta > 0$ there exists a constant $C(\theta)$ such that the three-body ground-state wave function $\psi(z)$ satisfies the inequality

$$|\psi(z)| < C(\theta)B(\theta z), \tag{77}$$

where the function $B(z)$ is defined by Eq. (71).

(b) If $|\psi(z)|$ is replaced by $B(\theta z)$ in the inequalities (70) of Lemma V, then these inequalities are satisfied.[4]

[4] The result contained in Theorem II, for the particular case where all the potentials are bounded by decreasing exponentials, had been obtained previously by a different method by one of the present authors; E. Leo Slaggie, Ph.D. thesis, University of California 1960 (unpublished).

Proof. The proof of the theorem is very simple with the aid of Lemma VI. To prove part (a) we select a θ such that $1 > \theta > 0$, and a constant k such that $1 \geqq k \geqq 0$, and a constant C_1 such that

$$|\psi(z)| < C_1B(k\theta z).$$

This is always possible since we may, in particular, select $k = 0$. It then follows from assertions (c) and (d) of Lemma VI that

$$|\psi(z)| < C_1K_0B([k+\delta(1-k)]\theta z),$$

or, repeating the procedure n times,

$$|\psi(z)| < C_1K_0^nB([1-(1-\delta)^n(1-k)]\theta z).$$

Since $\delta > 0$, we, thus, obtain the estimate

$$|\psi(z)| < C_2(\theta, \theta')B(\theta\theta' z),$$

where θ and θ' can be selected as close to 1 as we please. For every such selection $C_2(\theta, \theta')$ exists, and the part (a) of the theorem thus follows.

Part (b) of the theorem follows by inspection from assertions (a) and (b) of Lemma VI.

The reader will note that part (b) of Theorem II states that the inequalities (70) in Lemma V cannot yield a "better" bound on $|\psi(z)|$ than the one given in inequality (77).

In Appendix IV we shall show that there exist potentials which satisfy our assumptions and for which the inequality (77) is the best possible in the sense that for any $\theta > 1$ the inequality

$$|\psi(z)| > B(\theta z)$$

holds for sufficiently large $|z|$. We conjecture, but have been unable to prove, that the same might be true for all potentials which satisfy our assumptions. As it stands, we have thus proved that without further assumptions on the potentials the exponential bounds given cannot be "improved."

Our proof perhaps appears to be quite complicated at first sight. There exists, however, simple geometrical interpretations for all the steps carried out in this section in terms of which most of the discussion becomes intuitively "obvious." For a discussion of this interpretation we refer to Appendix V.

VI. ANALYTICITY PROPERTIES OF THE THREE-BODY GROUND-STATE WAVE FUNCTION IN MOMENTUM SPACE

Let $f(k)$ be the Fourier transform of $\psi(z)$,

$$f(k) = (2\pi)^{-3}\int_{(\infty)} d^6(z)\psi(z)\exp(-ik\cdot z), \tag{78}$$

where k is a real vector in the six-dimensional Euclidean space $\mathcal{E}$.

We denote by $\mathbf{p}_t$ and $\mathbf{q}_t$ the projections of k into the three-dimensional Euclidean spaces $\mathcal{E}_{x_t}$ and $\mathcal{E}_{y_t}$ which we introduced in Sec. II. In analogy with Eq. (11) we, thus, have

$$k = (\mathbf{p}_1, \mathbf{q}_1)_1 = (\mathbf{p}_2, \mathbf{q}_2)_2 = (\mathbf{p}_3, \mathbf{q}_3)_3, \tag{79}$$

and the pair $(\mathbf{p}_u, \mathbf{q}_u)$ is related to the pair $(\mathbf{p}_v, \mathbf{q}_v)$ by the relations (10a). Furthermore,

$$k\cdot z = \mathbf{p}_1\cdot\mathbf{x}_1 + \mathbf{q}_1\cdot\mathbf{y}_1 = \mathbf{p}_2\cdot\mathbf{x}_2 + \mathbf{q}_2\cdot\mathbf{y}_2 = \mathbf{p}_3\cdot\mathbf{x}_3 + \mathbf{q}_3\cdot\mathbf{y}_3. \tag{80}$$

Since the wave function $\psi(z)$ is exponentially bounded, we can extend $f(k)$ by analytic continuation into some region in the complex momentum space $\mathcal{E}_c$. Let, therefore, k be complex, and let us employ the notation

$$\begin{aligned} k' &= \mathrm{Re}\,(k), & k'' &= \mathrm{Im}\,(k), \\ \mathbf{p}_t' &= \mathrm{Re}\,(\mathbf{p}_t), & \mathbf{p}_t'' &= \mathrm{Im}\,(\mathbf{p}_t), \\ \mathbf{q}_t' &= \mathrm{Re}\,(\mathbf{q}_t), & \mathbf{q}_t'' &= \mathrm{Im}\,(\mathbf{q}_t). \end{aligned} \tag{81}$$

To determine the region of analyticity of $f(k)$, we investigate the convergence of the integral

$$\int_{(\infty)} d^6(z)B(z)\,\exp\,(k''\cdot z), \tag{82}$$

where $B(z)$ is the positive function defined in Eq. (71).

From the definition of $B(z)$, it follows that the integral (82) converges if an only if each one of the integrals

$$\int_{(\infty)} d^6(z)E_t(z)\,\exp\,(k''\cdot z), \tag{83}$$

where $t = 1$, 2, or 3, converges. We examine the integrand

$$E_t(z)\,\exp\,(k''\cdot z) = \underset{1\geqq r\geqq b_t}{\text{g.l.b.}}\,\{\exp\,[-r^{1/2}\,|\mathbf{y}_t| - (1-r)^{1/2}\,|\mathbf{x}_t| + \mathbf{p}_t''\cdot\mathbf{x}_t + \mathbf{q}_t''\cdot\mathbf{y}_t]\}.$$

The integrand is, thus, a decreasing exponential function if and only if for some r, $1 \geqq r \geqq b_t$, we have

$$|\mathbf{p}_t''| < (1-r)^{1/2}, \qquad |\mathbf{q}_t''| < r^{1/2}. \tag{84}$$

The condition (84) may be restated in the form

$$|\mathbf{p}_t''| < (1-b_t)^{1/2}, \qquad |\mathbf{p}_t''|^2 + |\mathbf{q}_t''|^2 < 1. \tag{85}$$

From these results and from Theorem II follows Theorem III.

Theorem III. The three-body ground-state wave function $f(k)$ in momentum space, defined in Eq. (78) as the Fourier transform of $\psi(z)$, can be analytically continued into a region T_c in complex momentum space $\mathcal{E}_c$, the region T_c being defined by the four inequalities

$$|k''|^2 = |\mathbf{p}_t''|^2 + |\mathbf{q}_t''|^2 < 1, \qquad |\mathbf{p}_t''| < (1-b_t)^{1/2}, \quad t = 1, 2, \text{ or } 3, \tag{86}$$

where k'', $\mathbf{p}_t''$, and $\mathbf{q}_t''$ are the imaginary parts of the vectors k, $\mathbf{p}_t$, and $\mathbf{q}_t$, respectively. If we interpret the inequalities (86) as confining k'' to a region T in the real space $\mathcal{E}$, then T_c is the product of $\mathcal{E}$ with T, and is, thus, a tube region. The region T is convex, and so is T_c.

The momentum space wave function is, thus, analytic in the region defined by the inequalities (86), which region has a very simple geometrical and physical interpretation. It is quite conceivable that the results contained in Theorem III could be obtained much more easily. It must be noted, however, that whereas Theorem II implies Theorem III, the converse is not in general true. We shall consider this question, and some related topics, in some detail in Appendix V.

It is easy to see that the region T_c is the largest tube of analyticity for $f(k)$ which can be obtained from the estimates in Theorem II, i.e., the largest region of analyticity which is a product of real momentum space and a region in imaginary momentum space. The largest *tube* of analyticity may, of course, be smaller than the largest region of analyticity of $f(k)$, even if the estimates in Theorem II are the best possible.

VII. ON ELECTROMAGNETIC FORM FACTORS

Suppose that particle 1 carries a charge, whereas particles 2 and 3 are neutral. Considering the ground state of the three-body system as a particle, we may inquire into the distribution of charge in this particle, i.e., study the charge form factor of the particle. We are, of course, not in a position to discuss anything but the asymptotic distribution of charge.

To cast the discussion into a form in which the relationship to the theory of the vertex operator in quantum field theory is apparent, we consider the matrix element

$$M(\Delta) = \int_{(\infty)} d^3(\mathbf{r}_1)\,d^3(\mathbf{r}_2)\,d^3(\mathbf{r}_3)\psi_f^*(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3) \times \exp\,(i\Delta\cdot\mathbf{r}_1)\psi_i(\mathbf{r}_1,\mathbf{r}_2,\mathbf{r}_3), \tag{87}$$

where

$$\mathbf{r}_u = \mathbf{r} + \boldsymbol{\varrho}_u, \qquad u = 1, 2, \text{ and } 3,$$

$$\Psi_{i,f}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \psi(\mathbf{x}_1, \mathbf{y}_1) \exp(i\mathbf{K}_{i,f}\cdot\mathbf{r}).$$

The vector $\mathbf{r}$ is thus the position vector of the center of mass, and the vectors $\mathbf{K}_i$ and $\mathbf{K}_f$ denote, respectively, the initial and final total momenta of the three-body system, regarded as an elementary particle on the mass shell.

Taking the variables $\mathbf{r}$, $\mathbf{x}_1$, and $\mathbf{y}_1$ as variables of integration instead of $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$, which variable transformation may be carried out by aid of the relations (9), we may carry out the integration over $\mathbf{r}$ and obtain

$$M(\boldsymbol{\Delta}) = C\,\delta_3(\mathbf{K}_f - \mathbf{K}_i - \boldsymbol{\Delta})M_0(\boldsymbol{\Delta}) \tag{88}$$

for some constant C, and with

$$M_0(\boldsymbol{\Delta}) = \int_{(\infty)} d^3(\mathbf{x}_1)\, d^3(\mathbf{y}_1)\psi^*(\mathbf{x}_1, \mathbf{y}_1)\psi(\mathbf{x}_1, \mathbf{y}_1) \times \exp(i\gamma\, \boldsymbol{\Delta}\cdot\mathbf{x}_1), \tag{89}$$

where the constant γ is given by

$$\gamma = (m_2 + m_3)^{1/2}[2B_0 m_1(m_1 + m_2 + m_3)]^{-1/2}. \tag{90}$$

Because the ground-state wave function $\psi(\mathbf{x}_1, \mathbf{y}_1)$ falls off exponentially, the "vertex operator" $M_0(\boldsymbol{\Delta})$ can be analytically continued, as a function of the momentum transfer $\boldsymbol{\Delta}$, into a tube region in complex three-dimensional space. This tube region may be determined either from the estimates in Theorem II, or directly from the facts stated in Theorem III. The result is:

The function $M_0(\boldsymbol{\Delta})$, as a function of complex momentum transfer $\boldsymbol{\Delta}$, is analytic in the tube region

$$|\boldsymbol{\Delta}''| < 2g_1, \tag{91}$$

where $\boldsymbol{\Delta}'' = \text{Im}\,(\boldsymbol{\Delta})$, and where

$$g_1 = (m_1/M_1)\min\{[2M_1(B_0 - B_1)]^{1/2}, [(m_2 + m_3)/m_2][2M_2(B_0 - B_2)]^{1/2}, [(m_2 + m_3)/m_3][2M_3(B_0 - B_3)]^{1/2}\}. \tag{92}$$

The function $M_0(\boldsymbol{\Delta})$ is essentially the charge form factor. The charge distribution for the particle at rest is given by

$$C(\boldsymbol{\varrho}_1) = c' \int_{(\infty)} d^3(\mathbf{y}_1)\, |\psi(\mathbf{x}_1, \mathbf{y}_1)|^2. \tag{93}$$

This function then satisfies the inequality

$$|C(\boldsymbol{\varrho}_1)| < C''(\theta) \exp(-2\theta g_1 |\boldsymbol{\varrho}_1|) \tag{94}$$

for every θ such that $1 > \theta > 0$, where $C''(\theta)$ is some constant which may depend on θ.

The formula (92), whereby g_1 is determined, has a simple physical interpretation, and is, in the opinion of the authors, entirely in accordance with what one might expect on the basis of intuitive physical considerations. We shall support this assertion by considering some examples in Sec. VIII.

If we assume that the potentials are invariant under rotations, the function $C(\boldsymbol{\varrho}_1)$ will be a function of $|\boldsymbol{\varrho}_1|$ only since the ground state is nondegenerate. If, furthermore, the bound expressed by inequality (94) is indeed the best possible, in the sense that it fails for any $\theta > 1$, then the tube, defined by inequality (91), in which $M_0(\boldsymbol{\Delta})$ is analytic is the largest possible. Actually $M_0(\boldsymbol{\Delta})$ is a function of $\boldsymbol{\Delta}^2$ only, and as a function of this variable it must exhibit a singularity at the point $\boldsymbol{\Delta}^2 = -4g_1^2$. This corresponds, in the theory of the vertex function in quantum field theory, to what has become known as an "anomalous threshold."[5] The authors wish to express the opinion that this singularity is a "nonrelativistic form" of a singularity which will always be present in the vertex function of *any* particle, in the framework of field theory; although the vertex function may, of course, have other singularities which occur for smaller values of $-\boldsymbol{\Delta}^2$. For the study of form factors we believe one is permitted to regard any stable particle as the "bound state" of any set of particles into which it could decay, through interactions which exist in nature, and in conformity with selection rules, given the necessary extra energy. For further discussion of these and related questions we refer the reader to the literature.[6,7]

We wish to emphasize, however, that a knowledge of the asymptotic form of the charge distribution of a particle can be no substitute for the kind of detailed knowledge which only a detailed dynamical theory can provide. The methods of this paper can at most tell us that a slowly falling-off exponential "tail" may be present, but they fail to give any information about the size of the coefficient which multiplies this exponential function.

VIII. CONSIDERATION OF SOME EXAMPLES

In this section we shall consider some specific examples of the function $B(z)$ in physically interest-

[5] This terminology is somewhat unfortunate since nothing could be more "normal" than a form factor computed from a nonrelativistic wave function when a nonrelativistic model seems permissible.

[6] R. Karplus, C. Sommerfield, and E. H. Wichmann, Phys. Rev. **111**, 1187 (1958).

[7] R. Oehme, Nuovo cimento, **13**, 778 (1959).

ing cases. Or rather, we study a function $F(z)$ which has the same asymptotic exponential behavior as $B(z)$, but which has the advantage that it can be represented by a single and simple analytic expression.

Let us thus define the function $F(\mathbf{x}, \mathbf{y}; b)$, where $1 > b \geqq 0$, by

$$F(\mathbf{x}, \mathbf{y}; b) = [\exp(-(1-b)^{1/2}|\mathbf{x}|) + \exp(-|\mathbf{x}|^2/|z|)][\exp(b^{1/2}|\mathbf{y}|) + \exp(|\mathbf{y}|^2/|z|)]^{-1}. \quad (95)$$

It is easily seen that the asymptotic behavior of the function $F(\mathbf{x}, \mathbf{y}; b)$ is similar to the asymptotic behavior of the function $E(\mathbf{x}, \mathbf{y}, b)$, defined in Eq. (54) in Sec. IV. We may therefore regard the function

$$F(z) = F(\mathbf{x}_1, \mathbf{y}_1; b_1) + F(\mathbf{x}_2, \mathbf{y}_2; b_2) + F(\mathbf{x}_3, \mathbf{y}_3; b_3) \quad (96)$$

as a function which characterizes the asymptotic exponential behavior of the wave function, since for every θ such that $1 > \theta > 0$ there exists a constant $K(\theta)$ such that

$$|\psi(z)| < K(\theta)F(\theta z). \quad (97)$$

It should be noted, however, that our discussion is defective in the sense that we have not proved, for the cases which we are going to consider, that there does not exist a "better" estimate than the one given in inequality (97).

Let us now consider the expression (96) qualitatively, employing a physical language which we believe has a considerable intuitive appeal. We first examine the function $F(\mathbf{x}_1, \mathbf{y}_1; b_1)$, and remind the reader that $\mathbf{x}_1$, which we may call the "separation coordinate," is essentially the distance of particle 1 from the center of mass of the 2-3 system, whereas $\mathbf{y}_1$, which we may call the "internal coordinate," is essentially the vector joining particles 2 and 3. The quantity b_1 is the binding energy of the ground state of the 2-3 system, measured in units of B_0, the binding energy of the ground state of the three-body system. The quantity $(1 - b_1)$ is then the separation energy, or the binding energy of particle 1 to the ground state of the 2-3 system. We may therefore expect that when the separation coordinate $\mathbf{x}_1$ is large compared to the internal coordinate $\mathbf{y}_1$ the wavefunction should behave like the product

$$\exp(-b_1^{1/2}|\mathbf{y}_1|)\exp[-(1-b_1)^{1/2}|\mathbf{x}_1|]. \quad (98)$$

This is indeed the asymptotic behavior of the function $F(\mathbf{x}_1, \mathbf{y}_1; b_1)$ whenever $|\mathbf{y}_1|^2 < b_1(|\mathbf{x}_1|^2 + |\mathbf{y}_1|^2)$.

However, as the ratio $|\mathbf{y}_1|/|\mathbf{x}_1|$ grows larger the behavior of the wave function changes, and we have

$$F(\mathbf{x}_1, \mathbf{y}_1; b_1) \sim \exp[-(|\mathbf{x}_1|^2 + |\mathbf{y}_1|^2)^{1/2}] \quad (99)$$

whenever $|\mathbf{y}_1|^2 \geqq b_1(|\mathbf{x}_1|^2 + |\mathbf{y}_1|^2)$.

It is clear that the behavior of the function F must undergo a change of this nature, since for $|\mathbf{y}_1|$ large compared to $|\mathbf{x}_1|$ one can hardly think about the system as a bound state in which particle 1 is bound to the "particle" formed by the ground state of the 2-3 system. This is particularly clear in the case where $b_1 = 0$. We may say that the more loosely bound is the 2-3 system the more easily does it lose its individuality as a "particle."

Suppose now that we wish to study the distribution of particle 1 in space, disregarding the location of particles 2 and 3. This means that we study the asymptotic behavior of $F(z)$ as a function of $\mathbf{x}_1$, but with $\mathbf{y}_1$ arbitrary; in particular $\mathbf{y}_1$ could be a function of $\mathbf{x}_1$. At first sight it appears natural to think that the required asymptotic distribution of particle 1 is given by the "product wave function" in expression (98), in which case the extension of the distribution would be determined by the separation energy $(1 - b_1)$ (and, of course, by the various masses). It may, however, happen that one of the other separation energies, say $(1 - b_2)$ is very small, which means that the binding energy of the 1-3 system is large. In this case we expect the three-body system to have an extended structure, and the particle 1 may have an extended distribution in space as a *component* of the bound 1-3 system. If we now consider Eq. (92) in Sec. VIII we see that the somewhat complicated expression for the constant g_1, which describes the distribution of particle 1, has a very simple and natural physical interpretation.

We may summarize this discussion by saying that to determine the asymptotic exponential behavior of the three-body system one must consider all possible virtual disintegrations of the system into two or three separated particles. The exponential bounds may then be computed in a trivial fashion from the binding energies and the masses.

Before we consider specific examples we may note that the function $F(z)$ is invariant under simultaneous rotations of $\mathbf{x}_1$ and $\mathbf{y}_1$, and is therefore, a function only of the three invariants which can be formed from these vectors. The ground-state wave function for the case of spherically symmetric potentials has the same property. As the three invariants we may select the quantities $|\mathbf{y}_1|^2$, $|\mathbf{y}_2|^2$, and $|\mathbf{y}_3|^2$. The three vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ are, of course, not independent, but satisfy the identity

$$\mathbf{y}_1\mu_1^{-1/2} + \mathbf{y}_2\mu_2^{-1/2} + \mathbf{y}_3\mu_3^{-1/2} = 0, \quad (100)$$

and we may, therefore, express the scalar product of any two of the vectors $\mathbf{y}_u$ in terms of their squares. We quote a number of relations which hold between the various invariants, following the notation in Sec. II.

$$\begin{aligned}|z|^2 &= (m_1 + m_2 + m_3)^{-1}[|\mathbf{y}_1|^2(m_2 + m_3) \\ &\quad + |\mathbf{y}_2|^2(m_3 + m_1) + |\mathbf{y}_3|^2(m_1 + m_2)] \\ &= 2B_0(m_1 + m_2 + m_3)^{-1}[|\varrho_{12}|^2 m_1 m_2 \\ &\quad + |\varrho_{23}|^2 m_2 m_3 + |\varrho_{31}|^2 m_3 m_1]; \quad (101)\end{aligned}$$

$$\begin{aligned}|\mathbf{x}_r|^2 &= (m_1 + m_2 + m_3)^{-1}[|\mathbf{y}_s|^2 (m_r + m_t) \\ &\quad + |\mathbf{y}_t|^2 (m_r + m_s) - |\mathbf{y}_r|^2 m_r] \\ &= 2B_0 m_r(m_1 + m_2 + m_3)^{-1}[|\varrho_{rs}|^2 m_s \\ &\quad + |\varrho_{tr}|^2 m_t - |\varrho_{st}|^2 \mu_r]. \quad (102)\end{aligned}$$

The three invariants $|\varrho_{rs}|^2$, which are simply the squares of the sides of the triangle formed by particles 1, 2, and 3, have the advantage of having a very direct geometric meaning. It is worthwhile to restate, in terms of these invariants, the conditions which determine whether $F(\mathbf{x}_t, \mathbf{y}_t; b_t)$ has the form given in expression (98) or the form given in expression (99).

We thus have two cases:

Case A:

$$\begin{aligned}&|\varrho_{rs}|^2 [(1 - b_t) + m_t(m_r + m_s)^{-1}] \\ &\quad \geqq b_t[|\varrho_{tr}|^2 (m_t/m_s) + |\varrho_{st}|^2 (m_t/m_r)], \quad (103a)\end{aligned}$$

in which case

$$F(\mathbf{x}_t, \mathbf{y}_t; b_t) \sim \exp(-|z|).$$

Case B:

$$\begin{aligned}&|\varrho_{rs}|^2 [(1 - b_t) + m_t(m_r + m_s)^{-1}] \\ &\quad < b_t[|\varrho_{tr}|^2 (m_t/m_s) + |\varrho_{st}|^2 (m_t/m_r)], \quad (103b)\end{aligned}$$

in which case

$$F(\mathbf{x}_t, \mathbf{y}_t; b_t) \sim \exp[-b^{1/2}|\mathbf{y}_t| - (1 - b_t)^{1/2}|\mathbf{x}_t|].$$

Let us now consider the case of the neutral helium atom. We label the two electrons by the indices 1 and 2, and the nucleus by the index 3. We thus have $m_1 = m_2 = m$ = the electronic mass; to simplify the discussion we shall assume that the nucleus is infinitely heavy. In this limit we have

$|\mathbf{x}_1| = |\mathbf{y}_2|$ = the distance of electron 1 from the nucleus;

$|\mathbf{x}_2| = |\mathbf{y}_1|$ = the distance of electron 2 from the nucleus;

The binding energies are: $B_1 = B_2 = 54.3$ eV, which is the approximate ionization potential of the singly ionized ion; $B_3 = 0$, and $B_0 = B_1 + 24.8$ eV, where the number 24.8 eV is the approximate ionization potential of the atom. We thus have, approximately, $b_1 = b_2 = b = 0.69$; $b_3 = 0$. Selecting $0.41 \times a_0$ as the unit of length, where a_0 is the Bohr radius in hydrogen, we thus obtain the following expression for the function $F(z)$:[8]

$$\begin{aligned}F(z) &= \frac{\exp(-(1-b)^{1/2}|\mathbf{x}_1|) + \exp(-|\mathbf{x}_1|^2/|z|)}{\exp(b^{1/2}|\mathbf{x}_2|) + \exp(|\mathbf{x}_2|^2/|z|)} \\ &\quad + \frac{\exp(-(1-b)^{1/2}|\mathbf{x}_2|) + \exp(-|\mathbf{x}_2|^2/|z|)}{\exp(b^{1/2}|\mathbf{x}_1|) + \exp(|\mathbf{x}_1|^2/|z|)} \\ &\quad + \frac{\exp(-|\mathbf{x}_3|) + \exp(-|\mathbf{x}_3|^2/|z|)}{1 + \exp(|\mathbf{y}_3|^2/|z|)}, \quad (104)\end{aligned}$$

where

$$|z| = (|\mathbf{x}_1|^2 + |\mathbf{x}_2|^2)^{1/2}, \qquad |\mathbf{x}_3|^2 = (\mathbf{x}_1 + \mathbf{x}_2)^2/2,$$
$$|\mathbf{y}_3|^2 = (\mathbf{x}_1 - \mathbf{x}_2)^2/2.$$

As a second example we consider the cases of the H^3 and He^3 nuclei.[9] We neglect the proton-neutron mass difference and, thus, have $m_1 = m_2 = m_3 = m$. Let the deuteron binding energy be $B_1 = B_2 = 2.2$ MeV. Two neutrons or two protons do not bind and we thus have $B_3 = 0$. In the case of tritium the proton is indexed by 3, whereas in the case of He^3 the neutron is so indexed. The binding energies are $B_0 = 8.5$ MeV for H^3, (hence, $b_1 = b_2 = b = 0.26$), and $B_0 = 7.7$ MeV for He^3, (hence, $b_1 = b_2 = b = 0.29$).

Letting $c = (2B_0 m/3)^{1/2}$ and employing the notation

$$\begin{aligned}z &= c(|\varrho_{12}|^2 + |\varrho_{23}|^2 + |\varrho_{31}|^2)^{1/2}, \\ x_1 &= c(|\varrho_{12}|^2 + |\varrho_{31}|^2 - \tfrac{1}{2}|\varrho_{23}|^2)^{1/2}, \\ x_2 &= c(|\varrho_{23}|^2 + |\varrho_{12}|^2 - \tfrac{1}{2}|\varrho_{31}|^2)^{1/2}, \\ x_3 &= c(|\varrho_{31}|^2 + |\varrho_{23}|^2 - \tfrac{1}{2}|\varrho_{12}|^2)^{1/2},\end{aligned}$$

$$y_1 = c(3/2)^{1/2}|\varrho_{23}|, \qquad y_2 = c(3/2)^{1/2}|\varrho_{31}|,$$
$$y_3 = c(3/2)^{1/2}|\varrho_{12}|,$$

$$p = (1 - b)^{1/2}, \qquad q = b^{1/2},$$

[8] The last term in the expression for $F(z)$ may be ignored as it is easily seen that it does not affect the asymptotic exponential behavior of $F(z)$.

[9] Wave functions for H^3 have been studied in great detail by G. H. Derrick, Nuclear Phys. **16**, 405 (1959) and references therein. In particular, he finds restrictions on the behavior of the wave function near the in line and equilateral configurations.

we then have[8]

$$F(z) = \frac{\exp(-px_1) + \exp(-x_1^2/z)}{\exp(qy_1) + \exp(y_1^2/z)} + \frac{\exp(-px_2) + \exp(-x_2^2/z)}{\exp(qy_2) + \exp(y_2^2/z)} + \frac{\exp(-x_3) + \exp(-x_3^2/z)}{1 + \exp(y_3^2/z)}.$$

IX. SOME REMARKS ON VARIATIONAL COMPUTATIONS

We wish to comment briefly on the relevancy of our results for the computation of energy levels and wave functions by variational methods.

As far as the authors know, no variational computation involving a three-body system has ever been carried out in such a way that each element in the set of trial functions would satisfy the asymptotic conditions which we have derived. As an example, we may cite the computations of the energy and wave function for the ground state of the helium atom. As we can see from Eq. (104) in Sec. VIII, the asymptotic behavior of the ground-state wave function will not be correctly described by a single exponential like $\exp(-c|z|)$, nor by an exponential function like $\exp(-c|\mathbf{x}_1| - c|\mathbf{x}_2|)$, where $|\mathbf{x}_1|$ and $|\mathbf{x}_2|$ are the distances between the electrons and the nucleus. The latter form is assumed, for instance, in the computation carried out by Pekeris.[10] We wish to emphasize strongly, however, that the fact that the trial functions have the "wrong" asymptotic behavior by no means implies that the variational computation of the energy would be wrong in principle. All that is required for the success of a variational computation of a ground-state energy is that the expectation values of the Hamiltonian operator evaluated over the set of trial functions should have the true energy as the greatest lower bound. The asymptotic behavior is of no great consequence. In fact it is perfectly possible, for Schrödinger equations of the kind we have considered, to construct a sequence of wave functions, each one of which goes asymptotically like, say $1/|z|^5$, but such that the expectation value of the Hamiltonian converges to $-B_0$, the ground-state energy. We believe these facts to be well known, and we will, therefore, not elaborate.

In spite of what has been said about the unimportance in principle of the asymptotic behavior of the trial functions in a variational computation, the possibility remains that an improvement, from a practical computational standpoint, can be achieved by selecting trial functions which have a realistic asymptotic behavior. For instance, in the case of the helium atom referred to earlier the trial functions are polynomials in the position vectors multiplied by the exponential function quoted. It is tempting to say, very loosely, that the polynomials may be "partly wasted in correcting the asymptotic behavior," the implication being that it might be in the interest of computational economy to replace the exponential function actually used by one which better conforms to reality. This may in particular be the case if one desires to compute the ground-state wave function rather than the energy. On the other hand, it is undeniable that a function like the one exhibited in expression (104) in Sec. VIII is more complicated than a simple exponential function, and any advantage gained in the sense of faster convergence may be nullified by the increased difficulty in the evaluation of expectation values of the Hamiltonian. It is clear that in selecting the proper computational procedure it is essential to take the behavior of the potentials into account, since they determine the relative importance of the various regions in configuration space. We have not studied any concrete cases and we therefore do not know whether the above suggestions have any merit. We feel, however, that it is of interest to consider these matters in connection with variational computations. In this connection we wish to refer the reader to a variational computation on the hyper-triton carried out by Downs and Dalitz.[11] These authors discuss the asymptotic behavior of the wave function, and comment upon the relative merits of various kinds of trial functions.

X. POSSIBLE GENERALIZATIONS

In this section we shall comment briefly upon a number of possible generalizations of our investigation.

(a) First of all, we note that in deriving Theorem II we have not really made use of the fact that $\psi(z)$ is the *ground*-state wave function. Theorem II, therefore, holds for any bound state of the three-body system, of energy $-B_0$, where $B_0 > B_t$ for $t = 1, 2$ and 3.

(b) That the conditions which we have assumed the potentials to satisfy are unnecessarily severe is suggested by our discussion of the simple case of a particle bound in a potential, in Sec. III.

Although our theory applies to most of the three-

[10] C. L. Pekeris, Phys. Rev. **112**, 1649 (1959).

[11] B. W. Downs and R. H. Dalitz, Phys. Rev. **114**, 593 (1959).

body problems defined in terms of potentials which seem to be physically meaningful, nevertheless there is a notable exception, namely, the case in which the forces become infinitely repulsive at small distances, "too rapidly" for our conditions to hold. It may, of course, be argued, and we are willing to subscribe to this opinion, that force fields of this kind derivable from a *static potential* do not occur in nature and are therefore meaningless physically. Since, however, potentials of this kind have often been considered in models of interparticle interactions, particularly in nuclear theory, it would nevertheless be interesting to know whether our results, as expressed in Theorem II, remain valid. Our results certainly hold for *finite* repulsive cores, no matter how strongly repulsive. The methods which we have employed cannot, however, without modifications be applied to the case of an infinitely repulsive core described by a potential which becomes infinite at some finite interparticle separation. The authors believe, but have not proved, that the necessary modifications can be made, and that Theorem II continues to hold in the case of an infinitely repulsive core also.

(c) Our procedure may be readily generalized to the case when the particles have spins, and where the interaction is described by a Hermitian matrix acting on spin space, the matrix elements of which are functions of the position vectors which satisfy the conditions which we have imposed on the potentials. In carrying out this generalization we would study, in place of our wave function $\psi(z)$, *that* spinor component which is largest in absolute value. The Green's functions $G_{r_s}(z, z'; -1)$ would be replaced by matrices. The modifications introduced by the spins are not of an essential nature, and they do not affect the validity of Theorem II. The discussion in Appendices III and IV cannot, however, be generalized in such a trivial fashion.

(d) If we consider the method whereby Theorem II was proved, we may also note that this method applies to some problems in which the interactions cannot be described by potentials at all. Let us, for instance, consider the modification arising when a potential $V(\mathbf{y})$ [or rather, the operation of multiplying the wave function by $V(\mathbf{y})$], is replaced by an integral transform on the wave function where the kernel $V(\mathbf{y}, \mathbf{y}')$ satisfies the condition

$$|V(\mathbf{y}, \mathbf{y}')| < Q(|\mathbf{y}| + |\mathbf{y}'|) \exp(-a|\mathbf{y} - \mathbf{y}'|),$$

where $Q(q)$, as before, is a function which satisfies the conditions (5) in Sec. I. We claim, but shall not prove, that if $a > 1$, then Theorem II holds and may be proved by a minor modification of the method which we have used. We are not here claiming that kernel functions of the above kind are of physical interest, but we merely wish to point out that it is not *essential* for the success of our methods that the interaction can be described by a static potential.

(e) Concerning the generalization to a relativistic theory, the situation is somewhat obscure. It does not appear particularly meaningful to try to consider relativistic wave equations describing a finite number of particles, since creation and destruction phenomena play an essential role whenever the particles have energies in the relativistic region. Problems of this nature belong to the realm of quantum field theory. One may, however, expect that some of our results can be looked upon as nonrelativistic limits of some features in field theory, especially since our results can be regarded as being essentially of a kinematical and geometrical nature.

(f) The question arises whether our methods can be readily generalized to the case of more than three particles. The nature of the results for the case of three particles suggests that this can be done, and that the asymptotic behavior of a many-body wave function is always determined, insofar as the exponential falloff is concerned, by the masses and the binding energies appropriate to the different modes in which the particle can be broken up. For instance, in the case of a bound state of four particles we would have to consider the breaking up of the system into two, three, and four fragments, where the fragments themselves may be bound states of a smaller number of particles.

Concerning the generalization of our methods, it seems likely to us that the generalization can be carried out step by step as follows: Suppose that exponential bounds have been determined for n-particle wave functions, where $n \leqq N$ for some N. Then it should also be possible to determine exponential bounds, like the ones we have found, on the n-particle Green's functions, when the complex energy parameter is outside the spectrum of the Hamiltonian for the n-particle system. If this can be done, one may trivially determine exponential bounds on the Green's functions which describe a system of particles consisting of two subsystems between which there is no interaction, and such that the number of particles in each subsystem does not exceed N. These last-mentioned Green's functions occur as factors in the integrands of a number of integral equations satisfied by the wave function (or Green's function) for a system of

$N + 1$ interacting particles, and it seems reasonable that these integral equations can be used to derive exponential bounds on the $(N + 1)$-particle wave functions or Green's functions.

The path to an extension of the procedure to any number of particles thus seems clear. We have not, however, investigated the question in detail, and we, therefore, do not wish to assert that an unforeseen complication cannot arise.

XI. ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. Friedman and Professor C. Schwartz for interesting and illuminating conversations.

APPENDIX I

In this Appendix we shall prove Lemma II, stated in Sec. IV. We follow the notation of Sec. IV.

The Green's function $G'(\eta)$, defined in Eqs. (43), satisfies a well-known integral equation which we may write symbolically as

$$(H'_y - \eta)^{-1} = (-\nabla_y^2 - \eta)^{-1} - (-\nabla_y^2 - \eta)^{-1} V(-\nabla_y^2 - \eta)^{-1} + (-\nabla_y^2 - \eta)^{-1} V(H'_y - \eta)^{-1} V(-\nabla_y^2 - \eta)^{-1}, \tag{105a}$$

where η is a complex number not in the spectrum of H'_y; η will then also be outside the spectrum of $-\nabla_y^2$.

More explicitly we may write the integral equation in the form

$$G'(\mathbf{y}, \mathbf{y}'; \eta) = \frac{\exp(i\eta^{1/2}|\mathbf{y} - \mathbf{y}'|)}{4\pi|\mathbf{y} - \mathbf{y}'|} - \int_{(\infty)} d^3(\mathbf{y}_0) \frac{V(\mathbf{y}_0) \exp[i\eta^{1/2}(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}_0|)]}{(4\pi)^2 |\mathbf{y} - \mathbf{y}_0| |\mathbf{y}' - \mathbf{y}_0|} + \int_{(\infty)} d^3(\mathbf{y}_0)\, d^3(\mathbf{y}'_0) \frac{V(\mathbf{y}_0) V(\mathbf{y}'_0) G'(\mathbf{y}_0, \mathbf{y}'_0; \eta) \exp[i\eta^{1/2}(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}'_0|)]}{(4\pi)^2 |\mathbf{y} - \mathbf{y}_0| |\mathbf{y}' - \mathbf{y}'_0|} \tag{105b}$$

In the above equation we have selected that branch of the function $\eta^{1/2}$ for which

$$\rho = \mathrm{Im}(\eta^{1/2}) > 0. \tag{106}$$

We select a fixed constant θ such that $1 > \theta > 0$, and define two functions, $m(\mathbf{y}, \mathbf{y}'; \theta, \eta)$ and $h(\mathbf{y}, \mathbf{y}'; \theta, \rho)$, by

$$m(\mathbf{y}, \mathbf{y}'; \theta, \eta) = \underset{\mathbf{y}_0, \mathbf{y}_0'}{\text{l.u.b.}} \left\{ \frac{|\mathbf{y}_0 - \mathbf{y}'_0|}{1 + |\mathbf{y}_0 - \mathbf{y}'_0|} \times |G'(\mathbf{y}_0, \mathbf{y}'_0; \eta)| \times \exp[-\theta\rho(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}'_0|)] \right\}; \tag{107}$$

$$h(\mathbf{y}, \mathbf{y}'; \theta, \rho) = \int_{(\infty)} d^3(\mathbf{y}_0)\, d^3(\mathbf{y}'_0) \times \frac{|\mathbf{y} - \mathbf{y}'| (1 + |\mathbf{y}_0 - \mathbf{y}'_0|) Q(|\mathbf{y}_0|) Q(|\mathbf{y}'_0|)}{(4\pi)^2 |\mathbf{y} - \mathbf{y}_0| |\mathbf{y}' - \mathbf{y}'_0| (1 + |\mathbf{y} - \mathbf{y}'|) |\mathbf{y}_0 - \mathbf{y}'_0|} \times \exp[-(1 - \theta)\rho(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}'_0|)], \tag{108}$$

where ρ is given by Eq. (106). We shall take it for granted that the function $m(\mathbf{y}, \mathbf{y}'; \theta, \eta)$ always exists. It is easy to see, by inspection of the integrand in Eq. (108), that the function $h(\mathbf{y}, \mathbf{y}'; \theta, \rho)$ exists.

Making use of the above definitions, as well as of the estimate on the potential given in inequality (41) of Sec. IV. we may derive, from the integral equation (105b), the inequality

$$\frac{|\mathbf{y} - \mathbf{y}'|}{1 + |\mathbf{y} - \mathbf{y}'|} |G'(\mathbf{y}, \mathbf{y}'; \eta)| \leqq h(\mathbf{y}, \mathbf{y}'; \theta, \rho) m(\mathbf{y}, \mathbf{y}'; \theta, \eta) + A(\theta, \rho) \exp(-\theta\rho |\mathbf{y} - \mathbf{y}'|), \tag{109}$$

where

$$A(\theta, \rho) = \frac{1}{4\pi} + \underset{\mathbf{y}, \mathbf{y}'}{\text{l.u.b.}} \left\{ \int_{(\infty)} d^3(\mathbf{y}_0) \times \frac{Q(|\mathbf{y}_0|) |\mathbf{y} - \mathbf{y}'|}{(4\pi)^2 |\mathbf{y} - \mathbf{y}_0| |\mathbf{y}' - \mathbf{y}_0| (1 + |\mathbf{y} - \mathbf{y}'|)} \times \exp[-\rho(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}_0|) + \theta\rho |\mathbf{y} - \mathbf{y}'|] \right\}. \tag{110}$$

It is easy to see that the number $A(\theta, \rho)$ exists for all θ and ρ, such that $1 > \theta > 0$ and $\rho > 0$, and that it furthermore satisfies the condition

$$A(\theta, \rho) < A_1 + A_2/(1 - \theta)\rho, \tag{111}$$

where A_1 and A_2 are constants independent of θ and ρ.

We next derive an estimate on the function

$h(\mathbf{y}, \mathbf{y}'; \theta, \rho)$, defined in Eq. (108). Let δ be a fixed constant such that

$$(1 - \theta)\rho \geqq 2\,\delta > 0, \tag{112}$$

and let the function $g(\mathbf{y}; \delta)$ be defined by

$$g(\mathbf{y}; \delta) = \text{l.u.b.}_{\mathbf{y}_0} \left\{ \frac{Q(|\mathbf{y}_0|)\ |\mathbf{y}_0|}{1 + |\mathbf{y}_0|} \exp\left(-\delta\ |\mathbf{y} - \mathbf{y}_0|\right) \right\}, \tag{113}$$

and let $J(\delta)$ be defined by

$$J(\delta) = \text{l.u.b.}_{\mathbf{y},\mathbf{y}'} \left\{ \int_{(\infty)} d^3(\mathbf{y}_0)\, d^3(\mathbf{y}_0') \exp\left[-\delta(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}_0'|)\right] \times \frac{|\mathbf{y} - \mathbf{y}'|\,(1 + |\mathbf{y}_0 - \mathbf{y}_0'|)(1 + |\mathbf{y}_0|)(1 + |\mathbf{y}_0'|)}{(4\pi)^2\ |\mathbf{y}_0|\ |\mathbf{y}_0'|\ |\mathbf{y} - \mathbf{y}_0|\ |\mathbf{y}' - \mathbf{y}_0'|\ |\mathbf{y}_0 - \mathbf{y}_0'|\ (1 + |\mathbf{y} - \mathbf{y}'|)} \right\}. \tag{114}$$

By inspection of the integral in this definition one sees that the integral is a bounded function of $\mathbf{y}$ and $\mathbf{y}'$, and the quantity $J(\delta)$, therefore, exists.

Using the definitions (113) and (114) we may now derive, from the defining equation (108), the following estimate:

$$h(\mathbf{y}, \mathbf{y}'; \theta, \rho) \leqq g(\mathbf{y}; \delta) g(\mathbf{y}'; \delta) J(\delta). \tag{115}$$

It follows, however, from the conditions (5) on the function $Q(q)$, that the function $g(\mathbf{y}; \delta)$, which depends on $\mathbf{y}$ only through $|\mathbf{y}|$, tends to zero as $|\mathbf{y}|$ tends to infinity. Since $g(\mathbf{y}; \delta)$ is furthermore obviously bounded, we may conclude that for every $\delta > 0$ there exists a number r_δ such that

$$h(\mathbf{y}, \mathbf{y}'; \theta, \rho) < 1/2. \tag{116}$$

whenever

$$|\mathbf{y}| + |\mathbf{y}'| \geqq r_\delta, \quad \text{and} \quad (1 - \theta)\rho \geqq \delta.$$

Taking into account all the facts established so far in this Appendix, we may now derive, by a minor modification of the procedure employed in Sec. III to derive Eq. (35), the following inequality:

$$m(\mathbf{y}, \mathbf{y}'; \theta, \eta) < A(\theta, \rho) \times \exp\left(-\theta\rho\ |\mathbf{y} - \mathbf{y}'|\right) + \tfrac{1}{2} m(\mathbf{y}, \mathbf{y}'; \theta, \eta) + \text{l.u.b.}_{(\mathbf{y}_0, \mathbf{y}_0') \in R_\delta} \left\{ \frac{|\mathbf{y}_0 - \mathbf{y}_0'|}{1 + |\mathbf{y}_0 - \mathbf{y}_0'|} |G'(\mathbf{y}_0, \mathbf{y}_0'; \eta)| \times \exp\left[-\theta\rho(|\mathbf{y} - \mathbf{y}_0| + |\mathbf{y}' - \mathbf{y}_0'|)\right] \right\}, \tag{117}$$

where R_δ is so defined that $(\mathbf{y}_0, \mathbf{y}_0')$ is in R_δ if and only if $|\mathbf{y}_0| + |\mathbf{y}_0'| < r_\delta$.

In the remainder of this Appendix we confine η to the half-plane

$$\text{Re}\,(\eta) \leqq -r < -b, \tag{118}$$

where r is a fixed constant. This half-plane is the union of two disjoint regions R_i and R_e which we define by

$$R_i\text{: } \text{Re}\,(\eta) \leqq -r, \quad |\eta| \leqq r_1; \tag{119a}$$

$$R_e\text{: } \text{Re}\,(\eta) \leqq -r, \quad |\eta| > r_1. \tag{119b}$$

Given any $\rho_0 > 0$, we may always select r_1 so large that

$$\text{Im}\,(\eta^{1/2}) = \rho > \rho_0 \tag{120}$$

for all η in R_e.

Let us now consider the case when η is in R_i. From the inequality (111) follows that the quantity $A(\theta, \rho)$ is bounded (for θ fixed) when η is in R_i. Furthermore, we may select a number δ such that the inequality (112) is satisfied for all η in R_i.

We shall take it for granted that

$$\text{l.u.b.}_{\substack{(\mathbf{y}_0, \mathbf{y}_0') \in R_\delta \\ \eta \in R_i}} \left\{ \frac{|\mathbf{y}_0 - \mathbf{y}_0'|}{1 + |\mathbf{y}_0 - \mathbf{y}_0'|} |G'(\mathbf{y}_0, \mathbf{y}_0'; \eta)| \right\}$$

exists, and we may then conclude that there exists a constant K' such that

$$m(\mathbf{y}, \mathbf{y}'; \theta, \eta) < K' \exp\left(-\theta\rho\ |\mathbf{y} - \mathbf{y}'|\right) \tag{121}$$

for all η in R_i, and for all $\mathbf{y}$ and $\mathbf{y}'$. The constant K' depends, of course, on θ and on R_i, i.e., on r and r_1.

It then follows, from the definition (107) of the function $m(\mathbf{y}, \mathbf{y}'; \theta, \eta)$, that

$$|G'(\mathbf{y}, \mathbf{y}'; \eta)| < K' \frac{1 + |\mathbf{y} - \mathbf{y}'|}{|\mathbf{y} - \mathbf{y}'|} \times \exp\left(-\theta\rho\ |\mathbf{y} - \mathbf{y}'|\right) \tag{122}$$

for all $\mathbf{y}$ and $\mathbf{y}'$ and for all η in R_i.

To complete the proof of Lemma II we must show that an estimate like the one above also holds for η in R_e. To do this we exploit the fact that for a sufficiently large r_1 the Liouville-Neumann expansion for $G'(\eta)$, given symbolically by

$$G'(\eta) = (-\nabla_y^2 - \eta)^{-1} \sum_{n=0}^{\infty} [-V(-\nabla_y^2 - \eta)^{-1}]^n, \tag{123}$$

will converge for all η in R_e.

Let us write

$$G_n'(\eta) = (-\nabla_y^2 - \eta)^{-1}[-V(-\nabla_y^2 - \eta)^{-1}]^n. \tag{124}$$

It is obvious that for any finite N there exists a K_N such that

$$\left|\sum_{n=0}^{N} G_n'(\mathbf{y}, \mathbf{y}'; \eta)\right| < \frac{K_N}{|\mathbf{y} - \mathbf{y}'|} \exp(-\theta\rho\,|\mathbf{y} - \mathbf{y}'|) \quad (125)$$

for all η in R_ϵ. To examine the remainder we select $N = 2$ and $n > N$. We then have the estimate

$$|G_n'(\mathbf{y}, \mathbf{y}'; \eta)| \leqq I_n(\mathbf{y}, \mathbf{y}'; (1-\theta)\rho_0) \times \frac{\exp(-\theta\rho\,|\mathbf{y} - \mathbf{y}'|)}{|\mathbf{y} - \mathbf{y}'|}, \quad (126a)$$

where

$$\begin{aligned} I_n(\mathbf{y}, \mathbf{y}'; (1-\theta)\rho_0) &= |\mathbf{y} - \mathbf{y}'| \\ &\times \int_{(\infty)} d^3(\mathbf{y}_1) \cdots d^3(\mathbf{y}_n)(4\pi)^{-n} Q(|\mathbf{y}_1|) \cdots \\ &\times Q(|\mathbf{y}_n|)[|\mathbf{y} - \mathbf{y}_1|\,|\mathbf{y}_1 - \mathbf{y}_2| \cdots \\ &\times |\mathbf{y}_{n-1} - \mathbf{y}_n|\,|\mathbf{y}_n - \mathbf{y}'|]^{-1} \\ &\times \exp[-(1-\theta)\rho_0(|\mathbf{y} - \mathbf{y}_1| + |\mathbf{y}_1 - \mathbf{y}_2| + \cdots \\ &+ |\mathbf{y}_{n-1} - \mathbf{y}_n| + |\mathbf{y}_n - \mathbf{y}'|)], \end{aligned} \quad (126b)$$

and where we have selected $\rho_0 > 0$ to be such that for all η in R_ϵ we have $\rho \geqq \rho_0$. Now, by selecting r_1 sufficiently large we may select ρ_0 so large (θ is, of course, fixed) that

$$\sum_{n=3}^{\infty} I_n(\mathbf{y}, \mathbf{y}'; (1-\theta)\rho_0) < K_R \quad (127)$$

for some K_R, for all $\mathbf{y}$ and $\mathbf{y}'$.

It follows that

$$|G'(\mathbf{y}, \mathbf{y}'; \eta)| < \frac{K_2 + K_R}{|\mathbf{y} - \mathbf{y}'|} \exp(-\theta\rho\,|\mathbf{y} - \mathbf{y}'|) \quad (128)$$

for all η in R_ϵ. Combining the results expressed in inequalities (122) and (128), we thus obtain, finally, the result that for every θ such that $1 > \theta > 0$, and for every $r > b$, there exists a constant $K(\theta, r)$ such that

$$|G'(\mathbf{y}, \mathbf{y}'; \eta)| < \frac{K(\theta, r)(1 + |\mathbf{y} - \mathbf{y}'|)}{|\mathbf{y} - \mathbf{y}'|} \times \exp(-\theta\rho\,|\mathbf{y} - \mathbf{y}'|) \quad (129)$$

for every η such that $\mathrm{Re}\,(\eta) \leqq -r$, and where $\rho = \mathrm{Im}\,(\eta^{1/2})$.

Lemma II follows trivially from the estimate in inequality (129).

APPENDIX II

In this Appendix we shall discuss, in a partly heuristic fashion, the validity of the integral representation for the function $G(z, z'; -1)$, given in Eqs. (47) in Sec. IV. Let us regard Eq. (47b) as *defining* a function $G(z, z'; -1)$ for the region

$$|\mathbf{x} - \mathbf{x}'| > 0, \qquad |\mathbf{y} - \mathbf{y}'| > 0. \quad (130)$$

Let us define, in the region defined by inequalities (130), two additional functions, $G_R(z, z'; -1, s)$ and $G_L(z, z'; -1, s)$, by

$$G_{R,L}(z, z'; -1, s) = \frac{1}{2\pi i} \times \int_{C_{R,L}(s)} d\eta (H_y' - \eta)^{-1}(H_x'' + 1 + \eta)^{-1}, \quad (131)$$

where $s > 0$ is a constant, and where the contour $C_L(s)$ goes from the point $\eta = -1 - s - i\cdot 0$ to the point $\eta = -1 - s + i\cdot 0$ in such a way that the contour intersects the real axis only at some point $\eta = -r$, where $1 > r > b$. Similarly the contour $C_R(s)$ goes from $\eta = s - i\cdot 0$ to $\eta = s + i\cdot 0$, intersecting the real axis only at the point $\eta = -r$.

Then it is easy to see that for fixed z and z' satisfying the conditions in inequality (130) we have

$$\lim_{s\to\infty} G_R(z, z'; -1, s) = \lim_{s\to\infty} G_L(z, z'; -1, s) = G(z, z'; -1). \quad (132)$$

We may furthermore see that if we multiply the members in Eq. (132) by the factors $|\mathbf{x}-\mathbf{x}'|$, $|\mathbf{y}-\mathbf{y}'|$ we may relax the conditions (130) on z and z', and we then have

$$\lim_{s\to\infty} |\mathbf{x} - \mathbf{x}'|\, G_L(z, z'; -1, s) = |\mathbf{x} - \mathbf{x}'|\, G(z, z'; -1), \quad (133a)$$

for $|\mathbf{x} - \mathbf{x}'| \geqq 0$, $|\mathbf{y} - \mathbf{y}'| > 0$;

$$\lim_{s\to\infty} |\mathbf{y} - \mathbf{y}'|\, G_R(z, z'; -1, s) = |\mathbf{y} - \mathbf{y}'|\, G(z, z'; -1), \quad (133b)$$

for $|\mathbf{x} - \mathbf{x}'| > 0$, $|\mathbf{y} - \mathbf{y}'| \geqq 0$.

We next note that the functions G_R and G_L satisfy inhomogeneous differential equations which we may write in the form

$$(H_y' + H_x'' + 1)G_R(z, z'; -1, s) = \delta_3(\mathbf{x} - \mathbf{x}')\frac{1}{2\pi i}\int_{C_R(s)} d\eta (H_y' - \eta)^{-1}; \quad (134a)$$

$$(H_y' + H_x'' + 1)G_L(z, z'; -1, s) = \delta_3(\mathbf{y} - \mathbf{y}')\frac{1}{2\pi i}\int_{C_L(s)} d\eta (H_x'' + 1 + \eta)^{-1}. \quad (134b)$$

The two integrals define projection operators which in the limit as s tends to infinity become the identity operators on the Hilbert spaces on which H_y' and H_x'' act, respectively. We may thus write, symbolically,

$$\lim_{s\to\infty} G_R(z, z'; -1, s) = \lim_{s\to\infty} G_L(z, z'; -1, s) = (H_y' + H_x'' + 1)^{-1}. \quad (135)$$

The functions $G_R(s)$ and $G_L(s)$ are, of course, to be interpreted as representing transformation on sufficiently well-behaved square integrable functions of $\mathbf{x}$ and $\mathbf{y}$, and the formula (135) thus says that the *transformations* $G_R(s)$ and $G_L(s)$ tend to the *transformation* $(H_y' + H_x'' + 1)^{-1}$ as s tends to infinity. This latter transformation is again representable by the Green's function associated with the differential operator $(H_y' + H_x'')$. We may now conclude that whenever Eqs. (133) apply the function $G(z, z'; -1)$, as defined by Eq. (47b), is indeed the Green's function claimed. This concludes our "demonstration" since we do not regard it as our duty to discuss expansion theorems and related delicate questions.

APPENDIX III

In this Appendix we shall present a very simple alternative proof of the assertion contained in inequality (27) of Sec. III, concerning exponential bounds on the ground-state wave function for a single particle bound in a potential field of force. We follow the notation of Sec. III. The argument goes as follows: The ground-state wave function $\phi(\mathbf{x})$ is real and nodeless, and we may assume it to be positive. Let θ be any number such that $1 > \theta > 0$. Let

$$\psi_M(\mathbf{x}) = \phi(\mathbf{x}) - (M/|\mathbf{x}|) \exp(-\theta\,|\mathbf{x}|),$$

and let

$$\phi_M(\mathbf{x}) = \tfrac{1}{2}(\psi_M(\mathbf{x}) + |\psi_M(\mathbf{x})|),$$

where M is a positive constant. Let R_M be the region in which $\psi_M(\mathbf{x})$ is positive; hence, the region in which $\phi_M(\mathbf{x})$ does not vanish. We shall show that for a sufficiently large M the region R_M must be empty.

Let $\mathbf{x}$ be in R_M. Then

$$[-\nabla_x^2 + V(\mathbf{x})]\phi_M(\mathbf{x}) = -\phi_M(\mathbf{x}) - [1 - \theta^2 + V(\mathbf{x})](M/|\mathbf{x}|) \exp(-\theta\,|\mathbf{x}|). \quad (136)$$

By selecting M sufficiently large, the distance of R_M to the origin can be made as large as we please, and since the potential tends to zero as $\mathbf{x}$ tends to infinity, we may select M such that

$$1 - \theta^2 + V(\mathbf{x}) > 0 \quad (137)$$

throughout R_M, unless R_M is empty.

Now, if R_M is *not* empty, we evaluate the expectation value of the "Hamiltonian" for the trial function $\phi_M(\mathbf{x})$, selecting M so large that the condition (137) holds. Since $\phi_M(\mathbf{x})$ is positive in R_M, it follows from Eq. (136) that

$$\langle\phi_M\,|(-\nabla_x^2 + V(\mathbf{x}))|\,\phi_M\rangle < -1\langle\phi_M\,|\,\phi_M\rangle, \quad (138)$$

which contradicts the assumption that the lowest eigenvalue of the operator in question equals -1. R_M, therefore, must be empty. It follows that for a sufficiently large M

$$|\phi(\mathbf{x})| < (M/|\mathbf{x}|) \exp(-\theta\,|\mathbf{x}|). \quad (139)$$

The constant M depends in general on θ. However, if the potential is positive outside some bounded region, we see that the condition (137) can be satisfied for $\theta = 1$, and an inequality like (139) then holds for $\theta = 1$.

It seems to the authors that it should be possible to treat the case of the three-body ground-state wave function in an analogous fashion. One is, however, then faced with the difficulty of constructing a suitable analog to the function $\phi_M(\mathbf{x})$.

For the case where the potential is spherically symmetric, we have, of course, available a number of simple methods of determining the asymptotic behavior of the wave function, since we may then consider the radial equation which is an ordinary differential equation. Our purpose in presenting methods which are applicable also in the admittedly somewhat academic general case was to illustrate procedures which may be generalized to many-body problems.

We shall finally show that for $\theta > 1$ no inequality like (139) can hold. Let $\theta > 1$. We consider

$$\chi(\mathbf{x}) = (m/|\mathbf{x}|) \exp(-\theta\,|\mathbf{x}|) - \phi(\mathbf{x}).$$

This function is positive in some neighborhood containing the origin. Let R_0 be the largest such open region. Let R_- be the region in which $\chi(\mathbf{x})$ is negative. By selecting m sufficiently small, we can arrange it so that R_- contains the surface of a sphere centered at the origin such that R_0 is in the interior of this sphere. Let R_+ be the complement of the union of R_0 and R_-. By selecting m sufficiently small, we can make the distance of R_+ from the origin as large as we please. We may then select as a "trial function" a function which agrees with $\chi(\mathbf{x})$ on R_+, and is

zero elsewhere. By a procedure analogous to the one leading to inequality (138), we establish that for sufficiently small m the region R_+ must be empty. We thus have the result: For every $\theta > 1$ there exists a constant $m > 0$ such that

$$|\phi(\mathbf{x})| > m \exp(-\theta |\mathbf{x}|) \tag{140}$$

for all $\mathbf{x}$.

APPENDIX IV

In this Appendix we shall show that the exponential bound stated in Theorem II is the best possible that can be obtained *in general*, i.e., without further assumptions on the potentials.

To do this we consider the particular case when all four potentials $V_{rs}(z)$ and $V_{123}(z)$ are nonpositive. We may restrict our discussion to the case where none of the two-body potentials $V_{rs}(z)$ is identically zero.

Since we deal with the ground-state wave function for spinless particles, we may assume that the wave function $\psi(z)$ is positive in $\mathcal{E}$. If we now examine the integral equation (22a), we note that the contribution to the right-hand side from the integration over any region in $\mathcal{E}$ is positive. It follows that for every $\theta > 1$ there exists a constant $m(\theta) > 0$ such that

$$|\psi(z)| > m(\theta) \exp(-\theta |z|). \tag{141}$$

In certain directions in $\mathcal{E}$ the function $B(z)$ equals $\exp(-|z|)$, whereas for certain other directions $B(z)$ is larger, i.e., falls off more slowly, than $\exp(-|z|)$. Let us, therefore, consider a direction in $\mathcal{E}$ for which $B(z)$ is larger than $\exp(-|z|)$; $B(z)$ is then equal to the largest one of the functions $E_t(z)$ along the direction in question. If we now recall definition (54) in Sec. IV of the function $E(\mathbf{x}, \mathbf{y}; b)$, we note that if, in a given direction, $E(\mathbf{x}, \mathbf{y}; b) > \exp(-|z|)$, then we must have $b > 0$, and furthermore $E(\mathbf{x}, \mathbf{y}; b)$ for that direction must be of the form

$$E(\mathbf{x}, \mathbf{y}; b) = \exp[-b^{1/2} |\mathbf{y}| - (1 - b)^{1/2} |\mathbf{x}|]. \tag{142}$$

Let us now consider our derivation in Sec. IV of an exponential bound on the Green's function $G(z, z'; -1)$, and let us consider the estimate in inequality (49) which holds for all r such that $1 > r > b$. Our procedure consisted in selecting the "most favorable" constant r, and we see that for a fixed direction such that $E(\mathbf{x}, \mathbf{y}; b)$ is given by Eq. (142), we get the best bound by letting r tend to b, and we would get a still better bound if we could choose r smaller than b. We are, however, prevented from doing this because that would mean our contour of integration C has been moved past the pole of the function $G'(\mathbf{y}, \mathbf{y}'; \eta)$ occurring at $\eta = -b$, which is the ground-state eigenvalue of the operator H'_y. Suppose, however, that we do shift the contour C past the pole, and that we then select the best possible new point $\eta = -r$ at which C intersects the real axis.[12] The resulting integral will then have a more rapid exponential falloff than the function $E(\mathbf{x}, \mathbf{y}; b)$.

The Green's function $G(z, z'; -1)$ is equal to the sum of the new contour integral and $-2\pi i$ times the residue of the integrand in Eq. (47b) at the point $\eta = -b$. This latter term is, thus, of the form

$$R = \phi(\mathbf{y})\phi^*(\mathbf{y}') \frac{\exp[-(1 - b)^{1/2} |\mathbf{x} - \mathbf{x}'|]}{4\pi |\mathbf{x} - \mathbf{x}'|}, \tag{143}$$

since the ground state of the operator H'_y is nondegenerate. The function $\phi(\mathbf{y})$ is the ground-state wave function.

From our results in Appendix III we can now conclude that the positive term R must dominate the Green's function $G(z, z'; -1)$ in the direction which we are considering in the $(z - z')$ space, if $\mathbf{y}'$ is small. The wave function $\phi(\mathbf{y})$ satisfies, according to the result expressed by inequality (140) in Appendix III, the condition that for any $\theta > 1$ there exists an $m(\theta)$ such that

$$|\phi(\mathbf{y})| > m(\theta) \exp(-\theta b^{1/2} |\mathbf{y}|).$$

We may summarize these results by saying that for any direction in the $(z - z')$ space for which $E_t(z - z') > \exp(-|z - z'|)$ the Green's function $G_{rs}(z, z'; -1)$ must, for z' in some bounded region and for any $\theta > 1$, satisfy the condition

$$|G_{rs}(z, z'; -1)| > m_1(\theta) E_t[\theta(z - z')], \tag{144}$$

as a function of z, where $m_1(\theta)$ is a constant which may depend on θ.

It is then an easy matter to apply the estimate in inequality (144) to derive from the integral equations (22) the following result [taking inequality (141) into account]:

Theorem IV. If the potentials $V_{rs}(z)$ are nonpositive and do not vanish identically, and if $V_{123}(z)$ is nonpositive, then for every $\theta > 1$ there exists a constant $c(\theta) > 0$ such that for all z

$$|\psi(z)| > c(\theta) B(\theta z), \tag{145}$$

where the notation is as in Theorem II.

[12] If there is more than one bound state the new point of intersection must be selected between $\eta = -b$ and the pole closest to $-b$.

APPENDIX V

In this Appendix we shall consider the relationship between the rate of exponential falloff of a function and the tube of analyticity of its Fourier transform. We shall omit all proofs; some of the facts listed are quite trivial and, we believe, well known.

Definitions.

(1) Let R_n be the n-dimensional real Euclidean space. A point in R_n will be denoted by z, where this symbol stands for the position vector of the point with respect to a fixed origin. The absolute magnitude of the vector z is denoted $|z|$.

(2) Let M be the set of all continuous functions $\rho(z)$ satisfying

(a) $\rho(\lambda z) = \lambda\rho(z)$, for $\lambda \geqq 0$;

(b) $\rho(z) \geqq 0$, for all z in R_n.

The functions in M are, thus, nonnegative and homogeneous of degree one. A function in M is determined on any ray in R_n by its value at any point on the ray.

(3) Let $\bar{M}$ be the set of all functions in M which satisfy the further condition that

$$\rho(z) > 0, \quad \text{whenever} \quad z \neq 0.$$

(4) For every $\rho(z)$ in M, let $\mathcal{E}(-\rho(z))$ be the set of all functions $f(z)$ such that for every θ satisfying $1 > \theta \geqq 0$ the function $f(z) \exp[\theta\rho(z)]$ is bounded for all z in R_n.

We have selected this notation for obvious mnemotechnic reasons. Thus, the function $\exp[-\rho(z)]$ in particular belongs to $\mathcal{E}(-\rho(z))$, and so does every function which decays at the same or at a faster rate at infinity.

(5) For any function $\rho(z)$ in $\bar{M}$, let $S(\rho(z))$ be the set of all points in R_n for which $\rho(z) < 1$. This set is clearly a bounded open region containing the origin, and star-like with respect to the origin.

Every function $\rho(z)$ in $\bar{M}$ thus defines a region $S(\rho(z))$, and conversely every bounded open region with a continuous boundary which is star-like with respect to, and contains, the origin, defines a function $\rho(z)$ in $\bar{M}$.

(6) Let $\rho_1(z)$ and $\rho_2(z)$ be two functions in $\bar{M}$, and let $S_1 = S(\rho_1(z))$ and $S_2 = S(\rho_2(z))$ be the corresponding regions in R_n, defined as in definition (5). Let $S = S_1 \otimes S_2$ be the set of all points z in R_n which can be written in the form

$$z = \theta z_1 + (1 - \theta)z_2,$$

where $1 \geqq \theta \geqq 0$, and where z_1 is in S_1 and z_2 is in S_2. We may call S the *joint envelope* of S_1 and S_2. The region S is clearly bounded; it has a continuous boundary, and it contains and is star like with respect to the origin. Therefore, there is associated a function $\rho(z)$ with S, and for this function we may employ the notation

$$\rho(z) = \rho_1(z) \otimes \rho_2(z).$$

We now make the following assertions.

Assertions.

(1) Let $\rho_1(z)$ and $\rho_2(z)$ be any two functions in $\bar{M}$, and let, with the notation in definition (6), $\rho(z) = \rho_1(z) \otimes \rho_2(z)$. Then

$$\rho(z) = \underset{z'}{\text{g.l.b.}} \{\rho_1(z') + \rho_2(z - z')\}.$$

(2) Let $\rho(z)$ be a function in $\bar{M}$. Let the region $T(\rho)$ in R_n be defined by the condition that a point k'' is in $T(\rho)$ if and only if

$$\rho(z) - k'' \cdot z > 0, \quad \text{for all} \quad z \neq 0 \quad \text{in} \quad R_n.$$

Let k be a vector in complex n-dimensional space, and let $k' = \text{Re}\,(k)$ and $k'' = \text{Im}\,(k)$. The vectors k' and k'' are, thus, elements of R_n. Let $T_c(\rho)$ be the tube region in complex n-dimensional space such that k'' is in $T(\rho)$. Then

(a) The region $T(\rho)$ is convex and contains the point $k'' = 0$;

(b) For any function $f(z)$ in $\mathcal{E}(-\rho(z))$ which is integrable over every bounded subregion of R_n the integral

$$\phi(k) = (2\pi)^{-n/2} \int_{(\infty)} d^n(z) f(z) \exp(-ik \cdot z) \qquad (146)$$

exists and defines an analytic function of k whenever k is in $T_c(\rho)$.

(c) For any point k_0 on the boundary of $T_c(\rho)$ there exists a function $f(z)$ in $\mathcal{E}[-\rho(z)]$ such that the corresponding function $\phi(k)$, defined by Eq. (146), fails to be analytic at k_0.

(3) Let $\rho(z)$ be a function in $\bar{M}$, and let the region $S(\rho)$ be defined as in definition (5), and let the region $T(\rho)$ be defined as in the preceding assertion. Then

(a) k'' is in $T(\rho)$ if and only if

$$1/|k''| > \underset{|z|=1}{\text{l.u.b.}} \{k'' \cdot z/\rho(z)\ |k''|\}$$

(b) If S_c is the convex envelope of the region S, and if $\rho_c(z)$ is the function in $\bar{M}$ defined by S_c, and hence $S_c = S(\rho_c(z))$, then

$$T(\rho) = T(\rho_c),$$

and the region $T(\rho)$ is, thus, determined by the convex envelope of $S(\rho)$.

Let us now comment briefly on the relevancy of these definitions and assertions to the discussion in the main body of the paper.

A characteristic problem in our discussion is the following: Given the positive functions $f_1(z)$ and $f_2(z)$ which fall off exponentially at infinity. We desire to place an exponential bound on the function $f(z)$ defined by

$$f(z) = \underset{z'}{\text{l.u.b.}} \{f_1(z')f_2(z - z')\}.$$

We may now reformulate this problem as follows: The functions $f_1(z)$ and $f_2(z)$ belong to the "exponential classes" $\mathcal{E}(-\rho_1(z))$ and $\mathcal{E}(-\rho_2(z))$, respectively, for some functions ρ_1 and ρ_2 in $\tilde{M}$. We desire to find a function $\rho(z)$ in $\tilde{M}$ such that $f(z)$ is in the exponential class $\mathcal{E}(-\rho(z))$. From assertion (1) we get the answer very simply: $\rho(z) = \rho_1(z) \otimes \rho_2(z)$, or, stated differently, the region S which defines $\rho(z)$ is the joint envelope of the regions S_1 and S_2 associated with the functions ρ_1 and ρ_2. With this geometrical interpretation the proofs of Lemmas IV and VI are almost trivial.

The relevance of assertions (2) and (3) in this Appendix to the discussion in Sec. VI is quite obvious. Suppose that the ground-state wave function $\psi(z)$ belongs to the exponential class $\mathcal{E}(-\rho(z))$ for some $\rho(z)$ in $\tilde{M}$. Then the Fourier transform of $\psi(z)$ is analytic in the region $T_c(\rho)$, which region is the intersection of all regions of analyticity of the Fourier transforms of all functions in the class $\mathcal{E}(-\rho(z))$. We have shown that the wave function belongs to the exponential class $\mathcal{E}(\ln [B(z)])$. The region S defined by $-\ln [(B(z)] < 1$ is, however, in general not convex, and is, thus, smaller than the convex envelope S_c of S. The region $T_c(S)$ is, however, the same as the region $T_c(S_c)$, and our Theorem II is, therefore, in general, stronger than our Theorem III.

Studies in Perturbation Theory. IV. Solution of Eigenvalue Problem by Projection Operator Formalism*

PER-OLOV LÖWDIN
Quantum Chemistry Group, University of Uppsala, Uppsala, Sweden
and
Quantum Theory Project, University of Florida, Gainesville, Florida
(Received December 15, 1961)

The partitioning technique for solving secular equations is briefly reviewed. It is then reformulated in terms of an operator language in order to permit a discussion of the various methods of solving the Schrödinger equation. The total space is divided into two parts by means of a self-adjoint projection operator O. Introducing the symbolic inverse $T = (1-O)/(E-H)$, one can show that there exists an operator $\Omega = O + THO$, which is an idempotent eigenoperator to H and satisfies the relations $H\Omega = E\Omega$ and $\Omega^2 = \Omega$. This operator is not normal but has a form which directly corresponds to infinite-order perturbation theory. Both the Brillouin- and Schrödinger-type formulas may be derived by power series expansion of T, even if other forms are perhaps more natural. The concept of the reaction operator is discussed, and upper and lower bounds for the true eigenvalues are finally derived.

1. INTRODUCTION

A FUNDAMENTAL problem in quantum chemistry and solid-state physics is the solution of the Schrödinger equation

$$H\Psi = E\Psi \tag{1}$$

for the stationary states. One of the strongest tools for treating this problem is the so-called partitioning technique, since it contains many of the conventional methods as special cases, particularly the variation principle and the perturbation theory. In pure mathematics, the partitioning technique seems to be a well-known tool in determinant[1] and matrix theory.[2] In wave mechanics, it was early used in Dirac theory[3] for separating the two large components from the two small components in the spinor wave function. In radiation theory, it has been used effectively by Gora,[4] and, during the years 1948–51, several authors[5] became independently interested in the technique as a simple and valuable alternative to the conventional perturbation theory.

This development has continued, and it has turned out that the partitioning technique is highly convenient for the applications to quantum chemistry.[6] It renders an excellent numerical tool for solving secular equations of high order; the eigenvalues E are not explicitly obtained but are implicitly given as solutions to an equation of the type $E = f(E)$, which is conveniently solved by iteration. The various types of iteration procedures and their convergency have also been studied, and the connection with the variation principle as well as with the perturbation theory has been investigated.[7]

The purpose of this paper is to rewrite the basic formulas in the partitioning technique in terms of a projection-operator formalism, which gives a simple and condensed form of the entire approach and shows the connection with the infinite-order perturbation theory and the iteration-variation methods in a highly transparent way. Both the Brillouin-Wigner and the Schrödinger forms of perturbation theory will be studied in this way, and the formalism will further be used to derive an explicit form for the so-called reaction operator, the expectation value of which gives the true energy shift under a perturbation. This operator

* The research reported in this paper has been sponsored in part by the King Gustaf VI Adolf's 70-Years Fund for Swedish Culture, Knut and Alice Wallenberg's Foundation, The Swedish Natural Science Research Council, and in part by the Chemistry Research Branch, ARL, AFRD, of the Air Research and Development Command, U. S. Air Force, through its European Office.

[1] M. Arnaldi, Giornale di Mat. Battaglini **34**, 209 (1896).

[2] See, e.g., E. Bodewig, *Matrix Calculus* (North-Holland Publishing Company, Amsterdam, 1959), 2nd ed.

[3] See W. Pauli in *Handbuch der Physik*, edited by S. Flügge (Springer Verlag, Berlin, 1933), Vol. 24, p. 236.

[4] E. Gora, Z. Physik **120**, 121 (1942-43).

[5] H. A. Kramers, Courant Anniversary Vol., 205 (1948); S. Sueoka, J. Phys. Soc. Japan **4**, 361 (1949); W. Kohn, J. Chem. Phys. **17**, 670(1949); M. H. L. Pryce, Proc. Phys. Soc. (London) **A63**, 25 (1950); S. F. Boys, Proc. Roy. Soc. (London) **A201**, 125 (1950); M. Lax, Phys. Rev. **79**, 200 A (1950); P. O. Löwdin, J. Chem. Phys. **19**, 1396 (1951).

[6] J. O. Hirschfelder, and P. O. Löwdin, Technical Note No. 3, Uppsala Quantum Chemistry Group, 1957 (unpublished); Molecular Physics **2**, 229 (1959); H. Shull and W. T. Simpson, J. Chem. Phys. **28**, 925 (1958).

[7] P. O. Löwdin, Technical Note No. 11, Uppsala Quantum Chemistry Group 1958 (unpublished); *Advances in Chemical Physics* (Interscience Publishers, Inc., New York, 1959), Vol. 2, p. 270 *f*; Technical Note No. 28, Uppsala Quantum Chemistry Group 1959 (unpublished).

will in a following paper be utilized for developing an exact self-consistent-field theory, which bridges the gap between the independent-particle model and the exact many-particle theory.

2. PARTITIONING TECHNIQUE IN SOLVING SECULAR EQUATIONS

In order to familiarize ourselves with an operator formalism which otherwise may seem unnecessarily abstract, we will start out with a brief review of the partitioning technique as a numerical tool for solving the secular equations connected with the eigenvalue problem (1). In using Ritz's method,[8] we will introduce a complete orthonormal basis $\{f_i\}$ and write the eigenfunction in the form $\Psi = \sum_i f_i c_i$, where the coefficients $\{c_i\}$ form a column vector $\mathbf{c}$. Introducing the energy matrix $\mathbf{H}$ having the elements $H_{kl} = \langle f_k | H | f_l \rangle$, one obtains from (1) the matrix equation

$$\mathbf{H}\mathbf{c} = E\mathbf{c}, \tag{2}$$

which is simply the transform of the original Schrödinger equation in the discrete representation chosen.

Let us now divide or "partition" the complete basis $\{f_i\}$ into two subsets (a) and (b). For the sake of simplicity, we choose the subset (a) so that it contains only a finite number of functions. The matrix $\mathbf{H}$ and the vector may now be written in the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{aa} & \mathbf{H}_{ab} \\ \mathbf{H}_{ba} & \mathbf{H}_{bb} \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} \mathbf{c}_a \\ \mathbf{c}_b \end{pmatrix}, \tag{3}$$

and, instead of (2), we obtain two equations:

$$\mathbf{H}_{aa}\mathbf{c}_a + \mathbf{H}_{ab}\mathbf{c}_b = E\mathbf{c}_a, \tag{4}$$

$$\mathbf{H}_{ba}\mathbf{c}_a + \mathbf{H}_{bb}\mathbf{c}_b = E\mathbf{c}_b, \tag{5}$$

which can now be treated in slightly different ways.

Solution of Equation System

We start by reviewing the method used in the numerical applications. In studying both nondegenerate and degenerate levels E, it is convenient to choose the subset (a) to consist of a single function, say f_1, and to put $\mathbf{c}_a = c_1 = 1$. The approach then gives all eigenvalues except those for which accidentally $c_1 = 0$. Starting from a trial value $E = E^{(0)}$, we then determine the corresponding vector $\mathbf{c}_b^{(0)}$ by solving Eq. (5) leading to the system:

$$(E^{(0)}\mathbf{1}_{bb} - \mathbf{H}_{bb})\mathbf{c}_b^{(0)} = \mathbf{H}_{b1}. \tag{6}$$

Substitution of $\mathbf{c}_b^{(0)}$ into (4) gives a new value $E^{(1)}$ defined by the relation

$$E^{(1)} = H_{11} + \mathbf{H}_{1b}\mathbf{c}_b^{(0)}. \tag{7}$$

The right-hand member defines a function $f(E)$ for $E = E^{(0)}$, which has the property that $f'(E) = -\mathbf{c}_b^{\dagger}(E)\cdot\mathbf{c}_b(E)$. Repeating the procedure starting from $E^{(1)}$ instead of $E^{(0)}$, we obtain a new value $E^{(2)}$, etc. The Eqs. (6) and (7) define together a first-order iteration process,[9] leading to a series of numbers $E^{(0)}$, $E^{(1)}$, $E^{(2)}$, $E^{(3)}$, $\cdots$, with the property that any two consecutive numbers $E^{(k)}$ and $E^{(k+1)}$ bracket a true eigenvalue E. The series $E^{(0)}$, $E^{(1)}$, $E^{(2)}$, $\cdots$ is convergent if $\mathbf{c}_b^{\dagger}\mathbf{c}_b < 1$ and divergent if $\mathbf{c}_b^{\dagger}\mathbf{c}_b > 1$. In both cases the first-order process may be replaced by a second-order process based on the formula:

$$E^* = E^{(0)} + \frac{E^{(1)} - E^{(0)}}{1 + \mathbf{c}_b^{(0)\dagger}\mathbf{c}_b^{(0)}}, \tag{8}$$

and we note that the right-hand member is equivalent to the expectation value of $\mathbf{H}$ with respect to the vector $\mathbf{c}^{(0)}$ having the components $c_1 = 1$ and $\mathbf{c}_b^{(0)}$, which gives the connection with the variation principle. The numerical process based on (6)–(8) is very convenient, as each step renders an upper and lower bound $E^{(0)}$ and $E^{(1)}$ as well as a second-order approximation E^* to a true eigenvalue E. The process is very rapidly convergent, since the error in each step is proportional to the *square* of the error in the preceding step.

The numerical procedure does not distinguish between degenerate and nondegenerate eigenvalues E but, in order to obtain *all* eigenvectors $\mathbf{c}$ associated with a degenerate level, one has to solve the equation system $(\mathbf{H} - E\cdot\mathbf{1})\mathbf{c} = 0$, once the E value has been determined.

Solution by Inverse Matrix

One may determine $\mathbf{c}_b$ explicitly from (5) in the form

$$\mathbf{c}_b = (E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{ba}\mathbf{c}_a, \tag{9}$$

provided that the inverse matrix $(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}$ exists for the E value under consideration. Substitution of (9) into (4) leads to the relation

$$\bar{\mathbf{H}}_{aa}\mathbf{c}_a = E\mathbf{c}_a, \tag{10}$$

where

$$\bar{\mathbf{H}}_{aa} = \mathbf{H}_{aa} + \mathbf{H}_{ab}(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{ba}. \tag{11}$$

[8] W. Ritz, Z. reine angew. Math. **135**, 1 (1909).

[9] For the general classification of iteration procedures, see E. Schröder, Math. Ann. **2**, 317 (1870); D. R. Hartree, Proc. Cambridge Phil. Soc. **45**, 230 (1949). Compare also P. O. Löwdin, Technical note No. 11, Uppsala Quantum Chemistry Group, 1958 (unpublished), particularly the Appendix.

Equation (10) has exactly the same form as the original eigenvalue problem (2), but the total matrix $\mathbf{H}$ is now condensed into a finite matrix $\bar{\mathbf{H}}_{aa}$ given by (11). This technique is of importance in many physical and chemical problems, since it enables us to concentrate our interest on a certain part (a) of a system, whereas the influence of the other parts (b) may be considered as a "perturbation" represented by the second term in the right-hand member of (11).

In treating an eigenvalue E having a degeneracy of finite order g, we choose the subset (a) to contain g functions, since the inverse matrix $(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}$ will then exist. For a *nondegenerate level*, one has $g = 1$, and we will then choose $\mathbf{c}_a = c_1 = 1$. From (10), we obtain $E = \bar{H}_{11}$, or

$$E = H_{11} + \mathbf{H}_{1b}(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{b1}, \qquad (12)$$

which relation corresponds to (7). The right-hand member gives an explicit expression for the function $f(E)$, and we note that this function becomes infinite for the E values which are eigenvalues to $\mathbf{H}_{bb}$ but not simultaneously eigenvalues to $\mathbf{H}$. On the other hand, the function $f(E)$ is regular also for those E values which correspond to degenerate eigenvalues of $\mathbf{H}$. The matrix $(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}$ does not exist in this case, but the singularity is actually removed by the effect of the two factors $\mathbf{H}_{1b}$ and $\mathbf{H}_{b1}$.

The equation $E = f(E)$ may again be solved by iteration procedures of first and second order. The curve $y = E - f(E)$ is graphically illustrated in Fig. 1 together with the construction of the values $E^{(1)}$ and E^* from the starting value $E^{(0)}$. Even for the abstract operator formalism developed later, it is of value to keep the general shape of this curve in mind.

For a *degenerate level* of order g, we now return to the original approach in which we chose the subset (a) to consist of g functions. The vector $\mathbf{c}_a$ now contains g elements and, since there are g linearly independent vectors $\mathbf{c}$, these g elements are independent of each other. It is hence possible to make the following simple choice:

$$\mathbf{c}_a = \begin{pmatrix}1\\0\\0\\\vdots\end{pmatrix}, \begin{pmatrix}0\\1\\0\\\vdots\end{pmatrix}, \begin{pmatrix}0\\0\\1\\\vdots\end{pmatrix}, \cdots, \begin{pmatrix}0\\0\\0\\\vdots\\1\end{pmatrix}. \qquad (13)$$

Substitution into (10) gives immediately $\bar{H}_{11} = E$, $\bar{H}_{1l} = 0$, and more generally $\bar{H}_{kk} = E$, $\bar{H}_{kl} = 0$ for

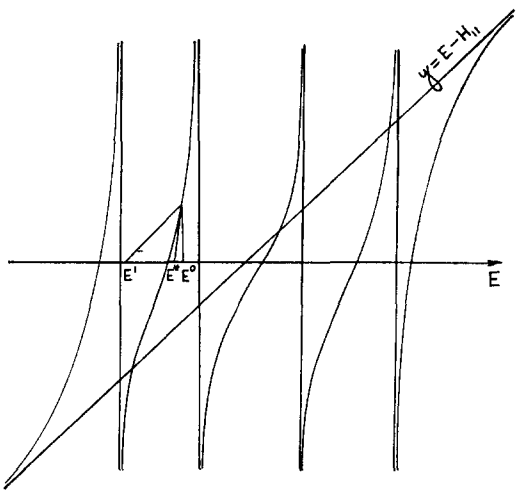


FIG. 1. The function $y = E - f(E)$ and the graphical construction of $E^{(1)}$ and E^* from $E^{(0)}$.

$k \neq l$, and $k, l = 1, 2, \cdots g$. The matrix $\bar{\mathbf{H}}_{aa}$ is hence diagonal of order g:

$$\bar{\mathbf{H}}_{aa} = \begin{pmatrix}\bar{H}_{11} & & & \\ & \bar{H}_{22} & & \\ & & \ddots & \\ & & & \bar{H}_{gg}\end{pmatrix} \qquad (14)$$

with all the diagonal elements equal: $\bar{H}_{11} = \bar{H}_{22} = \cdots = \bar{H}_{gg} = E$. For the energy we obtain in this way:

$$E = H_{11} + \mathbf{H}_{1b}(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{b1}, \qquad (15)$$

which relation is identical to (12) with the difference that the subspace (b) has now been reduced to such an extent—by removing g rows and columns from $\mathbf{H}$—that the inverse matrix $(E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}$ actually exists. It may be shown that (15) may be derived from (12) by a limiting procedure. Equation (12) is hence the general one but, in those cases when we would like to manipulate with the inverse matrix, the special form (15) is preferable. Our analysis shows that it is feasible to develop the partitioning technique for the nondegenerate levels and then extend the basic energy formula (12) to the degenerate levels, too, by a simple reinterpretation of the last term in the right-hand member.

Partitioning by Projection

In this section we carry through the partitioning of the vectors and matrices involved in the eigenvalue problem (2) by means of the two projection matrices:

$$\mathbf{0} = \begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}, \qquad \mathbf{P} = \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix}, \qquad (16)$$

which are self-adjoint and idempotent and fulfill the relations $\mathbf{\Theta} + \mathbf{P} = \mathbf{1}$, $\mathbf{\Theta P} = \mathbf{P\Theta} = \mathbf{0}$. One has immediately

$$\mathbf{c}'_a = \mathbf{\Theta c} = \begin{bmatrix} \mathbf{c}_a \\ \mathbf{0} \end{bmatrix}, \qquad \mathbf{c}'_b = \mathbf{Pc} = \begin{bmatrix} \mathbf{0} \\ \mathbf{c}_b \end{bmatrix}, \tag{17}$$

i.e., $\mathbf{c}'_a$ and $\mathbf{c}'_b$ are the complete vectors having $\mathbf{c}_a$ and $\mathbf{c}_b$, respectively, as the essential nonvanishing parts. For the energy matrix, we obtain similarly:

$$\begin{aligned} \mathbf{H}'_{aa} &= \mathbf{\Theta H\Theta} = \begin{bmatrix} \mathbf{H}_{aa} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ \mathbf{H}'_{ab} &= \mathbf{\Theta HP} = \begin{bmatrix} \mathbf{0} & \mathbf{H}_{ab} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \\ \mathbf{H}'_{ba} &= \mathbf{PH\Theta} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{H}_{ba} & \mathbf{0} \end{bmatrix}, \\ \mathbf{H}'_{bb} &= \mathbf{PHP} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{bb} \end{bmatrix}. \end{aligned} \tag{18}$$

The primed matrices are convenient to work with since they are defined in the total space, and they can hence be added and subtracted, so that, e.g., $\mathbf{c} = \mathbf{c}'_a + \mathbf{c}'_b$, $\mathbf{H} = \mathbf{H}'_{aa} + \mathbf{H}'_{ab} + \mathbf{H}'_{ba} + \mathbf{H}'_{bb}$, etc. Of essential importance in the theory is the "inverse of the corner" defined by the relation:

$$\mathbf{T}'_{bb} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1} \end{bmatrix}, \tag{19}$$

and the question is how it should be represented in terms of $\mathbf{\Theta}$ and $\mathbf{P}$. Starting out from the equation

$$\begin{bmatrix} \alpha\cdot\mathbf{1}_{aa} & \mathbf{0} \\ \mathbf{0} & (E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb}) \end{bmatrix}^{-1} = \begin{bmatrix} \alpha^{-1}\cdot\mathbf{1}_{aa} & \mathbf{0} \\ \mathbf{0} & (E\cdot\mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1} \end{bmatrix} \tag{20}$$

which is valid for all values of the constant $\alpha \neq 0$, and multiplying to the left and the right by $\mathbf{P}$, we obtain

$$\mathbf{T}'_{bb} = \mathbf{P}[\alpha\cdot\mathbf{\Theta} + \mathbf{P}(E\cdot\mathbf{1} - \mathbf{H})\mathbf{P}]^{-1}\mathbf{P}, \tag{21}$$

where the operator $\mathbf{T}'_{bb}$ is entirely independent of the value of $\alpha \neq 0$. For the sake of simplicity, it may be tempting to put $\alpha = 0$, but this leads actually to an improper notation, which will nevertheless often be used in the following, since it is now clear how this operator should be defined.

Using the projection matrices and (9), we will now drop the primes and write the eigenvector $\mathbf{c}$ in the form:

$$\mathbf{c} = (\mathbf{\Theta} + \mathbf{TH\Theta})\mathbf{C}, \tag{22}$$

where $\mathbf{C}$ is an arbitrary vector having $\mathbf{\Theta C} \neq \mathbf{0}$. Of particular interest is the matrix

$$\mathbf{\Omega} = \mathbf{\Theta} + \mathbf{TH\Theta}, \tag{23}$$

which applied to an arbitrary vector $\mathbf{C}$ gives an eigenvector $\mathbf{c}$, i.e., $\mathbf{c} = \mathbf{\Omega C}$. Since it satisfies the eigenvalue relation $\mathbf{H\Omega C} = E\mathbf{\Omega C}$ for arbitrary vectors $\mathbf{C}$, it is an eigenmatrix to $\mathbf{H}$, satisfying the relation $\mathbf{H\Omega} = E\mathbf{\Omega}$. It is further idempotent, so that $\mathbf{\Omega}^2 = \mathbf{\Omega}$, but it is neither self-adjoint nor normal. Using (10) and (11), we obtain finally for the energy:

$$E\mathbf{\Theta} = \mathbf{\Theta}(\mathbf{H} + \mathbf{HTH})\mathbf{\Theta}. \tag{24}$$

Formulas (23) and (24) are of such a general nature that they can be carried over into an abstract operator formalism which is independent of the choice of any specific representation based on a complete set $\{f_i\}$.

3. PROJECTION OPERATOR FORMALISM FOR SOLVING THE SCHRÖDINGER EQUATION

In this section, we will develop the partitioning technique in terms of a simple operator formalism, Let O be a self-adjoint projection operator which defines a certain subspace (a) of order g in the total Hilbert space, so that

$$O^2 = O; \qquad O^\dagger = O; \qquad \mathrm{Tr}\,(O) = g. \tag{25}$$

The operator $P = 1 - O$ satisfies the relations $P^2 = P$, $P^\dagger = P$, and $OP = PO = 0$ and defines a subspace (b), which we will call the orthogonal complement to the subspace (a).

Let us further introduce the operator T, which corresponds to the "inverse of the corner" (19) in matrix theory, by the definition:

$$T = P[\alpha\cdot O + P(E - H)P]^{-1}P, \tag{26}$$

where $\alpha \neq 0$ is an arbitrary number. The inverse operator in this expression exists for $\alpha \neq 0$ and, since $\partial T/\partial\alpha \equiv 0$, the operator T is entirely independent of the value of α. As before, it is hence tempting to put $\alpha = 0$ for the sake of simplicity, but this leads to a somewhat improper notation. In the following, we will often use the symbolic notation

$$T = P/(E - H), \tag{27}$$

instead of the complete expression (26). For the development of the theory, it is essential that T fulfills the following relations:

$$P(E - H)T = P, \tag{28}$$

$$OT = TO = 0. \tag{29}$$

The first one is obtained from the equation

$$[\alpha\cdot O + P(E - H)P][\alpha\cdot O + P(E - H)P]^{-1} = 1$$

by multiplying to the left and to the right by P, and the second one follows from the fact that $OP = PO = 0$.

Let us now consider the operator Ω defined by the relation

$$\Omega = O + THO \tag{30}$$

analogous to (23). We will prove that this operator is an eigenoperator to H, so that $H\Omega = E\Omega$, provided that E fulfills a certain condition. Using (28), we obtain the identity

$$P(H - E)\Omega = P(H - E)O + P(H - E)THO = PHO - PHO = 0,$$

for all values of E. Hence we have

$$\begin{aligned}(H - E)\Omega &= (O + P)(H - E)\Omega \\ &= O(H - E)\Omega = O(H - E)(O + THO) \\ &= O(H + HTH - E)O = 0,\end{aligned}$$

which gives the condition:

$$OEO = O(H + HTH)O. \tag{31}$$

This relation is completely analogous to (24). Formulas (30) and (31) form together the basis for the theory.

The operator Ω defined by (30) is an *idempotent eigenoperator* to H satisfying the relations

$$H\Omega = E\Omega, \qquad \Omega^2 = \Omega, \qquad \mathrm{Tr}\,(\Omega) = g. \tag{32}$$

The idempotency follows from the fact that, according to (29),

$$\Omega^2 = (O + THO)\cdot(O + THO) = O^2 + OTHO + THO^2 + THOTHO = O + THO = \Omega.$$

For the trace of Ω we have further

$$\mathrm{Tr}\,(\Omega) = \mathrm{Tr}\,(O + THO) = \mathrm{Tr}\,(O) + \mathrm{Tr}\,(OTH) = g.$$

The operator Ω is hence a *projection operator* but not of the conventional orthogonal type. It is neither self-adjoint nor normal, which is easily checked by considering the adjoint operator $\Omega^\dagger = O + OHT$. The operator Ω has actually a rather complicated character, and we note that it is a sum of an idempotent term O and a nilpotent term THO, satisfying the relation $(THO)^2 = 0$. In addition to Ω, every operator of the type $\Omega' = \Omega + \Omega A(1 - \Omega)$ is actually an idempotent eigenoperator to H. The particular importance of the form (30) comes from the connection with the perturbation theory. It fulfills further the simple relations:

$$\begin{aligned} O\Omega &= O; & \Omega O &= \Omega; \\ P\Omega &= \Omega - O; & \Omega P &= 0. \end{aligned} \tag{33}$$

We will now use the eigenoperator Ω to construct the eigenfunctions Ψ associated with the Hamiltonian H. Let us start by considering a *nondegenerate* level E, and let us choose $g = 1$ so that O is a one-dimensional projection operator which always (except for a constant factor) selects one and the same function. Let further Φ be an arbitrary trial function or "model" function which has a nonvanishing projection with respect to O. Let us put $O\Phi = \varphi$, and let us normalize this projection so that $\langle\varphi|\varphi\rangle = 1$, or $\langle\Phi\,|O|\,\Phi\rangle = 1$. The function

$$\Psi = \Omega\Phi = \varphi + TH\varphi \tag{34}$$

is now an exact eigenfunction to H since, according to (32), we have $H\Psi = H\Omega\Phi = E\Omega\Phi = E\Psi$. The eigenoperator Ω will hence from the trial function Φ project out the exact solution. This solution is characterized by a normalization condition of the type $\langle\varphi|\Psi\rangle = \langle O\Phi|\Omega\Phi\rangle = \langle\Phi\,|O|\,\Phi\rangle$, i.e.,

$$\langle\varphi\,|\,\Psi\rangle = 1, \tag{35}$$

whereas for the actual normalization integral we obtain $\langle\Psi|\Psi\rangle = \langle\Phi\,|\Omega^\dagger\Omega|\,\Phi\rangle = \langle\Phi\,|O + OHT^2HO|\,\Phi\rangle$, i.e.,

$$\langle\Psi\,|\,\Psi\rangle = \langle\varphi\,|1 + HT^2H|\,\varphi\rangle. \tag{36}$$

Multiplying the relation (31) to the left by Φ^* and to the right by Φ and integrating, we obtain

$$E = \left\langle\varphi\left|H + H\frac{P}{E - H}H\right|\varphi\right\rangle, \tag{37}$$

which equation is analogous to (12) and corresponds to the Schrödinger-Brillouin[10] formula in perturbation theory; the latter may be derived from (37) by expressing the inverse operator in T by means of a convenient power-series expansion. The corresponding wave function is given by (34). We note, however, that these power series expansions have to be convergent, whereas the condensed forms (34) and (37) are not subject to such a restriction and represent forceful alternatives to infinite-order perturbation theory.

In treating a *degenerate level* E of order g, we start out from a g-dimensional projection operator O and its orthogonal complement $P = 1 - O$. The

[10] L. Brillouin, J. phys. radium **33**, 373 (1932).

eigenoperator Ω defined by (30) will now project on the eigenspace of order g connected with the degeneracy. In order to follow the development previously sketched in the matrix formalism, we will now resolve O into g orthogonal components:

$$O = \sum_{k=1}^{g} O_k, \tag{38}$$

which are all one-dimensional projection operators fulfilling the relations:

$$O_k^2 = O_k, \qquad O_k^\dagger = O_k, \qquad \mathrm{Tr}\,(O_k) = 1, \tag{39}$$
$$O_k O_l = 0 \quad \text{for} \quad k \neq l.$$

For Ω we obtain the resolution $\Omega = \sum_k \Omega_k$, where the components $\Omega_k = \Omega O_k$ for $k = 1, 2, \cdots g$ form a set of eigenoperators to H satisfying the relations:

$$\Omega_k = O_k + THO_k, \tag{40}$$

$$H\Omega_k = E\Omega_k, \qquad \Omega_k^2 = \Omega_k, \qquad \mathrm{Tr}\,(\Omega_k) = 1. \tag{41}$$

Multiplying the energy relation (31) to the left and to the right by O_k and O_k, and by O_k and O_l respectively, we obtain

$$O_k E O_k = O_k(H + HTH)O_k, \tag{42}$$

$$O_k(H + HTH)O_l = 0, \tag{43}$$

which corresponds to the diagonalization achieved in (14). From (29), we get similarly $O_k T = TO_k = 0$ which relation is useful in the following.

In order to derive the eigenfunctions to H, we will let the projection operators $O_1, O_2, \cdots O_g$ select a set of orthogonal functions $\varphi_1, \varphi_2, \cdots \varphi_g$ which will be chosen normalized so that $\langle \varphi_k | \varphi_l \rangle = \delta_{kl}$. In the applications, the projection operators O_k are often defined essentially by means of these functions, which may here be chosen quite arbitrarily. However, once the set $\varphi_1, \varphi_2, \cdots \varphi_g$ is fixed, it defines also the orthogonal complement and the operators P and T. Let us now consider the g functions

$$\Psi_k = \Omega_k \varphi_k = \varphi_k + THO\varphi_k, \tag{44}$$

which are eigenfunctions to H according to (41). Since $\langle \varphi_k | \Psi_l \rangle = \delta_{kl}$, it is easily shown that the functions $\Psi_1, \Psi_2, \cdots \Psi_g$ are linearly independent. However, they are usually not orthonormal, as is shown by their metric integral:

$$\langle \Psi_k \mid \Psi_l \rangle = \langle \varphi_k \mid 1 + HT^2H \mid \varphi_l \rangle. \tag{45}$$

From (42) and (43), we obtain further

$$E = \left\langle \varphi_k \left| H + H \frac{P}{E - H} H \right| \varphi_k \right\rangle \tag{46}$$

and

$$\langle \varphi_k \mid H + HTH \mid \varphi_l \rangle = 0. \tag{47}$$

We note that formulas (44) and (46) are completely identical with (34) and (37) in the nondegenerate case with the difference that the projection operator P defining the "orthogonal complement" is here reduced to such an extent that the inverse operator in the definition (26) of T actually exists. Keeping this in mind, we will in the following not distinguish between degenerate and nondegenerate levels in discussing Eqs. (34) and (37) or the analogous Eqs. (44) and (46).

Iteration Procedures

In (37) and (46), the eigenvalue problem is given into an implicit form $E = f(E)$, where

$$f(E) \equiv \left\langle \varphi \left| H + H \frac{P}{E - H} H \right| \varphi \right\rangle. \tag{48}$$

It is natural to try to solve this problem by iteration procedures. These will be essentially the same as in the matrix representation, and they will be only briefly reviewed here in terms of the operator formalism. Of basic importance is the fact that the first derivative

$$f'(E) = -\left\langle \varphi \left| H \frac{P}{(E - H)^2} H \right| \varphi \right\rangle = -\langle THO\varphi \mid THO\varphi \rangle \leq 0, \tag{49}$$

is always negative; it equals the normalization integral of the second term in the wave function (34) with negative sign. The higher derivatives are also easily determined:

$$f^{(n)}(E) = (-1)^n\, n!\, \langle \varphi \mid HT^{n+1} H \mid \varphi \rangle; \tag{50}$$

they are useful in constructing iteration procedures of higher orders than the second.

The formula $E^{(k+1)} = f\{E^{(k)}\}$ leads to a *first-order* iteration procedure and a series of number $E^{(0)}, E^{(1)}, E^{(2)}, E^{(3)}, \cdots$, which may be convergent or divergent. Substitution of these numbers into the right-hand member of (48) leads to an expression for E in terms of a *continued fraction*. This expression is actually much more complicated than the classical continued fractions, since it involves operators and the formation of expectation values, but it preserves one very important feature of these fractions, since it approaches the limit E both from below and from above, if it is convergent. It is easy to study the series of numbers $E^{(0)}, E^{(1)}, E^{(2)}, \cdots$ directly, and we will put $E^{(k)} = E + \epsilon^{(k)}$, where $\epsilon^{(k)}$ is the error in the kth iteration. Using the mean-value theorem

$$f\{E + \epsilon^{(k)}\} = f(E) + \epsilon^{(k)} f'(E + \theta\epsilon^{(k)}),$$

where $0 \leq \theta \leq 1$, one obtains

$$\epsilon^{(k+1)} = \epsilon^{(k)} \cdot f'\{E + \theta\epsilon^{(k)}\}. \tag{51}$$

Since f' is always negative according to (49), the errors $\epsilon^{(k)}$ will alternate in sign, and this implies that the successive values of $E^{(k)}$ will alternately be upper and lower bounds to E. Between two consecutive values $E^{(k)}$ and $E^{(k+1)}$, there will hence always be at least one eigenvalue E, and we will refer to this as the *bracketing theorem*. The process will be convergent if $|f'| < 1$, and divergent if $|f'| > 1$.

Irrespective of whether the first-order process is convergent or divergent, one can easily go over to *second-order* iteration procedures having a much faster convergence, since the error in each step will be proportional to the *square* of the error in the preceding step. One can, for instance, solve the equation $y = E - f(E) = 0$ by the well-known Newton-Raphson procedure:

$$E^* = E^{(0)} - \frac{y^{(0)}}{y'^{(0)}} = E^{(0)} - \frac{E^{(0)} - E^{(1)}}{1 - f'\{E^{(0)}\}}, \tag{52}$$

which corresponds to the tangential construction in Fig. 1. The derivative $f'(E)$ is given by (49), and it is anyway evaluated in normalizing the eigenfunction. Introducing the approximate solution

$$\Psi^{(0)} = \varphi + \frac{P}{E^{(0)} - H} H\varphi, \tag{53}$$

one can transform (52) into the expectation value

$$E^* = \frac{\langle \Psi^{(0)} | H | \Psi^{(0)} \rangle}{\langle \Psi^{(0)} | \Psi^{(0)} \rangle}, \tag{54}$$

which shows the connection with the variation principle.

Extension of Wigner's Theorem.

In Brillouin's perturbation theory,[10] there is a famous theorem by Wigner[11] which gives a connection between the perturbation formulas and the variation principle of a somewhat different type. We will here derive an extension and generalization of this theorem in terms of the operator formalism.

The variation principle says that, if the wave function is affected by an error of the first order, the expectation value $\langle H \rangle$ is affected by an error of the second order. This implies that if the wave function in a perturbation calculation is correct up to order n with an error of order $(n + 1)$, the energy error is of order $2(n + 1)$ and $\langle H \rangle$ is hence correct up to order $2n + 1$. The explicit relations are easily shown by the operator formalism.

Let $\mathcal{E}$ be a trial value for the energy E, and let $\mathfrak{Z}$ be a certain approximation of the matrix $T = P/(\mathcal{E} - H)$ properly defined by (26). It is easily shown [see Appendix, (A10)] that, if $\mathfrak{Z}$ is affected by an error of the first order, then

$$\mathfrak{Z}^* = 2\mathfrak{Z} - \mathfrak{Z}(\mathcal{E} - H)\mathfrak{Z} \tag{55}$$

is a better approximation with an error of the second order. If, e.g., $\mathfrak{Z}$ is a power series expansion including terms up to order n, then $\mathfrak{Z}^*$ is represented by the same expansion including terms up to order $(2n + 1)$; see (A12). Instead of the operator Ω associated with the trial value $\mathcal{E}$, we have now the two approximate operators:

$$\tilde{\Omega} = O + \mathfrak{Z}HO, \qquad \tilde{\Omega}^* = O + \mathfrak{Z}^*HO. \tag{56}$$

They fulfill the following basic formulas

$$\begin{aligned} \tilde{\Omega}^\dagger\tilde{\Omega} &= O(1 + H\mathfrak{Z})(1 + \mathfrak{Z}H)O \\ &= O + OH\mathfrak{Z}^2HO, \end{aligned} \tag{57}$$

and

$$\begin{aligned} \tilde{\Omega}^\dagger(H - \mathcal{E})\tilde{\Omega} &= O(1 + H\mathfrak{Z})(H - \mathcal{E})(1 + \mathfrak{Z}H)O \\ &= O[(H - \mathcal{E}) + H\mathfrak{Z}(H - \mathcal{E}) \\ &\quad + (H - \mathcal{E})\mathfrak{Z}H + H\mathfrak{Z}(H - \mathcal{E})\mathfrak{Z}H]O \\ &= O[(H - \mathcal{E}) + H\{2\mathfrak{Z} - \mathfrak{Z}(\mathcal{E} - H)\mathfrak{Z}\}H]O \\ &= O[H - \mathcal{E} + H\mathfrak{Z}^*H]O = O[H\tilde{\Omega}^* - \mathcal{E}]O, \end{aligned} \tag{58}$$

which are used in calculating the expectation value of H with respect to the approximate wave function:

$$\tilde{\Psi} = \tilde{\Omega}\Phi = \varphi + \mathfrak{Z}H\varphi. \tag{59}$$

We obtain directly

$$\begin{aligned} \langle H - \mathcal{E} \rangle &= \frac{\langle \tilde{\Psi} | H - \mathcal{E} | \tilde{\Psi} \rangle}{\langle \tilde{\Psi} | \tilde{\Psi} \rangle} = \frac{\langle \Phi | \tilde{\Omega}^\dagger(H - \mathcal{E})\tilde{\Omega} | \Phi \rangle}{\langle \Phi | \tilde{\Omega}^\dagger\tilde{\Omega} | \Phi \rangle} \\ &= \frac{\langle \Phi | O(H + H\mathfrak{Z}^*H - \mathcal{E})O | \Phi \rangle}{\langle \Phi | O(1 + H\mathfrak{Z}^2H)O | \Phi \rangle} \\ &= \frac{\langle \varphi | H + H\mathfrak{Z}^*H | \varphi \rangle - \mathcal{E}}{\langle \varphi | 1 + H\mathfrak{Z}^2H | \varphi \rangle}; \end{aligned} \tag{60}$$

i.e.,

$$\langle H \rangle = \mathcal{E} + \frac{\langle \varphi | H + H\mathfrak{Z}^*H | \varphi \rangle - \mathcal{E}}{\langle \varphi | 1 + H\mathfrak{Z}^2H | \varphi \rangle}, \tag{61}$$

which relation holds for any trial value $\mathcal{E}$ for the energy. There is one specific $\mathcal{E}$ value, for which this relation takes a particularly elegant form, namely the value for which the numerator in the second term vanishes, so that

[11] E. Wigner, Math. naturw. Anz. ungar. Akad. Wiss. **53**, 477 (1935).

$$\varepsilon = \langle \varphi \,|H + H\mathfrak{J}^*H|\, \varphi\rangle, \tag{62}$$

since then $\langle H\rangle = \varepsilon$ for the wave function $\tilde{\Psi} = \varphi + \mathfrak{J}H\varphi$. In the special case, when $\mathfrak{J}$ is approximated by a power series up to order n, this gives Wigner's theorem. For practical purposes, however, the general form (61) is much more useful, since it is valid for *all* values of the parameter ε.

4. PERTURBATION THEORY

Let us now consider the case, when $H = H_0 + V$, where H_0 is the "unperturbed" Hamiltonian and V is an arbitrary, weak or strong "perturbation." It is now convenient to choose O as the eigenoperator to H_0 associated with the state under consideration, so that $H_0O = OH_0 = E_0O$, and O will hence project out the unperturbed eigenfunction φ_0. This means that, whenever H_0 is standing close to the operator O, it can be replaced by the number E_0. We note that we here introduce a *single* eigenfunction to H_0 and not the complete set of all its eigenfunctions, which is an essential simplification in both the theory and the applications. The "orthogonal complement" to φ_0 characterized by P may be introduced by orthogonalizing any complete set towards φ_0. According to (30) and (31), we now obtain

$$\Omega = O + THO = O + T(E_0 + V)O = (1 + TV)O; \tag{63}$$

$$OEO = O(H + HTH)O = O(E_0 + V + VTV)O. \tag{64}$$

Of particular interest is here the operator $W = 1 + TV$, which is called the *wave operator*, and the operator

$$t = V + VTV, \tag{65}$$

which is called the *reaction operator* associated with the perturbation V, the unperturbed Hamiltonian H_0, and the state under consideration. Using (37) and (64), we obtain

$$E = E_0 + \langle \varphi_0 \,|t|\, \varphi_0\rangle, \tag{66}$$

i.e., the energy shift is simply the expectation value $\langle t\rangle_0$ of the reaction operator with respect to the unperturbed state. We note further the relations $\Omega = WO$ and $t = VW$.

Brillouin-Type Formulas

These expansions are treated here mainly for historical reasons and to show the connection with other approaches. They are characterized by having denominators of the "mixed" type $(E - H_0)$ intermediate between the type $(E - H)$ treated previously and the type $(E_0 - H_0)$ used in the Schrödinger theory. Let us introduce the operator

$$T_0 = P[\alpha\cdot O + P(E - H_0)P]^{-1}P, \tag{67}$$

for which we will also use the symbolic notation

$$T_0 = P/(E - H_0). \tag{68}$$

Using the identity [see Appendix, (A1)]

$$(A - B)^{-1} = A^{-1} + A^{-1}B(A - B)^{-1}, \tag{69}$$

for $A = \alpha\cdot O + P(E - H_0)P$ and $B = PVP$, and multiplying to the left and to the right by P, we obtain

$$T = T_0 + T_0VT = T_0(1 + VT) \tag{70}$$

$$TV = T_0(V + VTV) = T_0t, \tag{71}$$

and

$$t = V + VT_0t. \tag{72}$$

This relation corresponds actually to the Lippmann-Schwinger[12] integral equation in scattering theory.[13] For t one finds the solution

$$t = (1 - VT_0)^{-1}V = V + VT_0V + VT_0VT_0V + \cdots, \tag{73}$$

and this solution is still not explicit, since each denominator in T_0 contains the energy $E = E_0 + \langle t\rangle_0$. The inverse operator in T_0 should further exist for the E value under consideration, and the power series should be convergent. Substitution of (73) into (63) and (64) gives the formal expansions:

$$\Omega = (1 + T_0t)O = (1 + T_0V + T_0VT_0V + T_0VT_0VT_0V + \cdots)O, \tag{74}$$

$$OEO = O(E_0 + V + VT_0V + VT_0VT_0V + \cdots)O. \tag{75}$$

Hence, we obtain for the eigenfunction Ψ and the associated eigenvalue E:

$$\Psi = (1 + T_0V + T_0VT_0V + \cdots)\varphi_0, \tag{76}$$

$$E = E_0 + \langle \varphi_0 \,|V + VT_0V + VT_0VT_0V + \cdots|\, \varphi_0\rangle, \tag{77}$$

[12] B. A. Lippmann and J. Schwinger, Phys. Rev. 79, 469 (1950).

[13] This derivation of the Lippmann-Schwinger equation using the partitioning technique was actually given by Dr. Kimio Ohno, Tokyo University, in a lecture in Uppsala in 1957 (unpublished).

which expansions are of Brillouin type. These formulas have the disadvantage that the total energy E occurs in the denominators $(E - H_0)$ to all the factors T_0, which gives a rather complicated implicit expression for the energy. A still more serious disadvantage comes from the fact that, in nuclear theory,[14] one has found that the terms in (77), do not stay proportional to the number N of particles when $N \to \infty$, but diverge. The use of the Brillouin type of formulas has for these reasons to large extent been abandoned.

Schrödinger-Type Formulas

Let us now study the possibilities of obtaining an explicit expression for the reaction operator t defined by (65). The denominator $(E - H)$ in T may be written in the form $(E_0 + \langle t\rangle_0 - H_0 - V)$, and hence we obtain

$$t = V + V \frac{P}{E_0 - H_0 - V + \langle t\rangle_0} V, \qquad (78)$$

which formula forms the basis for our discussion of the Schrödinger-type perturbation theory. For the energy shift, one gets directly:

$$\langle t\rangle_0 = \langle V\rangle_0 + \left\langle V \frac{P}{E_0 - H_0 - V + \langle t\rangle_0} V \right\rangle_0. \qquad (79)$$

The natural expression for the energy shift seems hence to be a *continued fraction* of the generalized type discussed previously in connection with the energy formula $E = f(E)$ and relation (48). We note, however, that the use of such a fraction is actually equivalent to the use of a *first-order* iteration procedure based on the formula $x^{(k+1)} = g\{x^{(k)}\}$, where

$$g(x) \equiv \langle V\rangle_0 + \left\langle V \frac{P}{E_0 - H_0 - V + x} V \right\rangle_0, \qquad (80)$$

and x is a real variable such that the solution of the equation $x = g(x)$ will give the true energy shift $\langle t\rangle_0$. Since $g(x) = f(E_0 + x) - E_0$, the discussion in connection with the Eqs. (48)–(54) is valid also here. Again there will be a "bracketing theorem" saying that any two consecutive numbers in the series $x^{(0)}, x^{(1)}, x^{(2)}, \cdots$ will bracket the true energy shift $\langle t\rangle_0$. This theorem will be used below in discussing upper and lower bounds to $\langle t\rangle_0$. According to (36), the normalization integral for the eigenfunction Ψ takes the form

$$\langle \Psi \mid \Psi\rangle = 1 + \langle \varphi_0 \mid VT^2V \mid \varphi_0\rangle, \qquad (81)$$

and, since the second term is identical to $-g'$, the first-order iteration procedure and the corresponding continued fraction will be convergent if $\langle\Psi|\Psi\rangle < 2$, whereas they will be divergent if $\langle\Psi|\Psi\rangle$ is larger than 2.

Irrespective of whether the first-order procedure is convergent or divergent, one can easily go over to a second order procedure analogous to (52):

$$x^* = x^{(0)} - \frac{x^{(0)} - x^{(1)}}{1 - g'\{x^{(0)}\}} \approx x^{(0)} - \frac{[x^{(0)} - x^{(1)}]^2}{x^{(0)} - 2x^{(1)} + x^{(2)}}, \qquad (82)$$

and we note that the expression in the middle is equivalent to the expectation value (54) used in the variation principle. In both the first- and second-order procedure, we have assumed that the inverse matrices occurring in T have been evaluated exactly, but later we will make us free from this restriction.

In order to get the connection with the conventional Schrödinger theory, we will now evaluate T and t by means of power-series expansions containing the denominator $(E_0 - H_0)$. For this purpose, we will introduce the operator

$$R_0 = P[\alpha \cdot O + P(E_0 - H_0)P]^{-1}P, \qquad (83)$$

which exists for $\alpha \neq 0$ and which is independent of the particular value of α. In the following, we will also use the symbolic expression

$$R_0 = P/(E_0 - H_0). \qquad (84)$$

For the sake of simplicity, we will assume that both the unperturbed and perturbed levels under consideration are nondegenerate, and the degenerate case will be discussed later. Using the identity (69) for $A = \alpha \cdot O + P(E_0 - H_0)P$ and $B = PV'P$, where

$$V' = V - \langle t\rangle_0, \qquad (85)$$

we obtain

$$T = R_0 + R_0V'T, \qquad (86)$$

which is the "integral equation" for the operator T. The solution

$$T = (1 - R_0V')^{-1}R_0 \qquad (87)$$

$$= R_0 + R_0V'R_0 + R_0V'R_0V'R_0 + \cdots \qquad (88)$$

may be expressed as a power series in R_0V', only if $|\langle R_0V'\rangle| < 1$ for all trial functions. Substitution of expansion (88) into (63) and (65) gives for the eigenoperator Ω and the reaction operator t the formal expressions

[14] See, e.g., K. A. Brueckner and C. A. Levinson, Phys. Rev. 97, 1344 (1955); J. Goldstone, Proc. Roy. Soc. (London) **A239**, 267 (1957); L. S. Rodberg, Ann. Phys. 2, 199 (1957).

$$\Omega = (1 + R_0V + R_0V'R_0V + R_0V'R_0V'R_0V + \cdots)O, \tag{89}$$

and

$$t = V + VR_0V + VR_0V'R_0V + VR_0V'R_0V'R_0V + \cdots = V + VR_0 \sum_{k=0}^{\infty} (V'R_0)^k V. \tag{90}$$

We note that, since V' contains $\langle t\rangle_0$, the formulas have an implicit character and have to be solved by iteration. Conventionally the results are arranged after powers of V and, in order to obtain t correct to at least the order n, one has to take n terms of the right-hand member of (90) and express $\langle t\rangle_0$ correct at least up to the order $(n-2)$. By using the partitioning of the integers, one can further easily collect the terms of various orders and obtain:

$$t = t_1 + t_2 + t_3 + t_4 + \cdots,$$

$$\begin{cases} t_1 = V, \\ t_2 = VR_0V, \\ t_3 = VR_0(V - \langle V\rangle_0)R_0V, \\ t_4 = VR_0(V - \langle V\rangle_0)R_0(V - \langle V\rangle_0)R_0V \\ \quad -\langle VR_0V\rangle VR_0^2V, \\ \cdots \end{cases} \tag{91}$$

For the wave operator $W = 1 + TV$, one gets correspondingly:

$$W = 1 + W_1 + W_2 + W_3 + \cdots;$$

$$\begin{cases} W_1 = R_0V, \\ W_2 = R_0(V - \langle V\rangle_0)R_0V, \\ W_3 = R_0(V - \langle V\rangle_0)R_0(V - \langle V\rangle_0)R_0V \\ \quad -\langle VR_0V\rangle R_0^2V, \\ \cdots \end{cases} \tag{92}$$

and we note that $t = VW$ and $t_k = VW_k$; the eigenoperator is given by the relation $\Omega = WO$. The energy shift $\langle t\rangle_0$ may now be written in the form

$$\langle t\rangle_0 = E_1 + E_2 + E_3 + E_4 + \cdots, \tag{93}$$

where $E_k = \langle\varphi_0\,|t_k|\,\varphi_0\rangle$ is the kth-order term.

We have in this way obtained condensed expressions which correspond to the basic formulas in Schrödinger's perturbation theory.[15] In order to get the conventional relations, one has further to choose the "orthogonal complement" to φ_0 characterized by the projection operator $P = 1 - O$ to consist of all the remaining eigenfunctions $\varphi_1, \varphi_2, \varphi_3, \cdots$ to the unperturbed Hamiltonian H_0, provided the entire set is complete. In this representation the operator R_0 is diagonal, and, denoting the eigenvalue of H_0 associated with φ_k by $E_k^{(0)}$, we obtain

$$P = \sum_{k\neq 0} |\varphi_k\rangle\langle\varphi_k|,$$
$$R_0 = \frac{P}{E_0^{(0)} - H_0} = \sum_{k\neq 0} \frac{|\varphi_k\rangle\langle\varphi_k|}{E_0^{(0)} - E_k^{(0)}}. \tag{94}$$

Substitution of this expression for R_0 into (91) and (92) leads to the conventional Schrödinger formulas, and we obtain, for instance,

$$E_2 = \langle\varphi_0\,|VR_0V|\,\varphi_0\rangle = \sum_{k\neq 0} \frac{V_{0k}V_{k0}}{E_0^{(0)} - E_k^{(0)}}, \tag{95}$$

where $V_{0k} = \langle\varphi_0\,|V|\,\varphi_k\rangle$. We note that the relations (91) and (92) have a more general character than the standard formulas, since the former are valid for *any* orthogonal complement to φ_0.[16]

Let us now discuss the normalization problem. According to (35), the eigenfunction $\Psi = \Omega\Phi$ determined by the eigenoperator $\Omega = WO$ satisfies the normalization condition $\langle\varphi_0|\Psi\rangle = 1$. Using (36), we obtain the actual normalization integral

$$\langle\Psi\,|\,\Psi\rangle = \langle\varphi_0\,|1 + VT^2V|\,\varphi_0\rangle, \tag{96}$$

which contains the operator

$$VT^2V = VR_0(1 - V'R_0)^{-1}R_0(1 - V'R_0)^{-1}V = VR_0^2V + VR_0^2V'R_0V + VR_0V'R_0^2V + VR_0^2V'R_0V'R_0V + VR_0V'R_0^2V'R_0V + \cdots. \tag{97}$$

This operator is easily systematized after powers of V, and we obtain

$$\langle\Psi\,|\,\Psi\rangle = 1 + \delta_2 + \delta_3 + \delta_4 + \cdots;$$

$$\begin{cases} \delta_2 = \langle VR_0^2V\rangle_0, \\ \delta_3 = \langle VR_0^2(V - \langle V\rangle_0)R_0V \\ \quad + VR_0(V - \langle V\rangle_0)R_0^2V\rangle_0, \\ \cdots \end{cases} \tag{98}$$

These relations will later be of importance in using the variation principle.

In concluding this section, we note that we have here considered a state which is nondegenerate both

[15] E. Schrödinger, Ann. Physik (4) **80**, 437 (1926); for the explicit form of the third- and fourth-order terms, see K. F. Niessen, Phys. Rev. **34**, 253 (1929); compare also K. A. Brueckner, *ibid.* **100**, 36 (1955).

[16] Compare, P. O. Löwdin, Technical Note 28, Uppsala Quantum Chemistry Group 1959 (unpublished).

in the unperturbed and perturbed system. When power-series expansions of T are used, the treatment of a *degeneracy* is complicated by the fact that the inverse matrices in both operators

$$T = P/(E - H), \qquad R_0 = P/(E_0 - H_0) \tag{99}$$

defined by (26) and (83), respectively, should exist. Starting out from an unperturbed level E_0 which has a degeneracy of order g, one has to choose the basic projection operator O to be of order g so that R_0 exists. It is feasible to introduce a resolution of O into g orthogonal components $O_1, O_2, \cdots O_g$ according to (38), and we will denote the corresponding functions by $\varphi_{01}, \varphi_{02}, \cdots \varphi_{0g}$. They are all eigenfunctions to H_0 associated with the level E_0. If also the perturbed level has a degeneracy of order g, there are no further complications, we can directly apply formulas (42) and (43), and the expansion relations (88)–(98) are valid in an unchanged form. However, if there is a *splitting* of the degeneracy which is usually the case, the matrix $\bar{\mathbf{H}}$ defined by (11) and having the elements

$$\bar{H}_{kl} = \left\langle \varphi_{0k} \left| E_0 + V + V \frac{P}{E - H} V \right| \varphi_{0l} \right\rangle \tag{100}$$

is no longer diagonal. This leads to a secular equation (10) of order g, which may again be conveniently treated by partitioning. This "repeated-partitioning" technique is actually of practical importance also in other connections.

Upper and Lower Bounds in Perturbation Theory

In connection with relation (51) in the general theory, we discussed a "bracketing" theorem which provided upper and lower bounds for the true eigenvalues E, and we will now study whether it is possible to apply this theorem in the case when $H = H_0 + V$. For the sake of simplicity, we will start by considering the *ground state*, for which one always has

$$E \leq \langle H_{\text{op}} \rangle_{\text{av}} \tag{101}$$

for all trial functions. For the zero-order function φ_0, one obtains in particular

$$E \leq E_0 + E_1, \tag{102}$$

which relation also tells us that the sum $E_2 + E_3 + E_4 + \cdots$ must be *negative*. Using (94), one can easily conclude that the term E_2 is negative, but the signs of the higher terms are usually harder to obtain.

Upper bounds with successively increased accuracy may further be derived by observing that formula (61) is valid for all values of ε. Let us start by considering the series expansion of the operator

$$\frac{P}{\varepsilon - H} = \frac{P}{(E_0 - H_0) - (V - \varepsilon + E_0)} = R_0[1 - (V - \varepsilon + E_0)R_0]^{-1} \tag{103}$$

in terms of powers of $(V - \varepsilon + E_0)R_0$, and let us introduce the partial sum

$$\mathfrak{I}_n = R_0 \sum_{k=0}^{n} [(V - \varepsilon + E_0)R_0]^k. \tag{104}$$

Application of (55) gives immediately $\mathfrak{I}_n^* = \mathfrak{I}_{2n+1}$, and formula (61) takes now the form

$$\langle H \rangle_{\text{av}} = \varepsilon + \frac{\langle \varphi_0 | E_0 + V + V\mathfrak{I}_{2n+1}V | \varphi_0 \rangle - \varepsilon}{\langle \varphi_0 | 1 + V\mathfrak{I}_n^2 V | \varphi_0 \rangle} \tag{105}$$

for arbitrary values of ε, which parameter, therefore, could be varied. Putting $n = 0$, we find that the right-hand member is independent of ε and the result

$$\langle H_{\text{op}} \rangle_{\text{av}} = E_0 + E_1 + (E_2 + E_3)/(1 + \delta_2) \tag{106}$$

as an improved upper bound. One sees often in the literature[17] the statement that all the odd-order energies $E_0 + E_1$, $E_0 + E_1 + E_2 + E_3$, $\cdots$ etc. would form upper bounds to the true energy. However, from (106) one could derive the inequality $\langle H \rangle < E_0 + E_1 + E_2 + E_3$, only if the sum $(E_2 + E_3)$ is positive which seldom seems to be the case.

Instead of minimizing the expression (105) with respect to ε, it is often simpler to choose ε equal to energy E correct to order n, so that $\varepsilon = E_0 + E_1 + E_2 + \cdots + E_n$. Application of (105) gives then $\langle H \rangle$ equal to the energy correct to order $(2n + 1)$ plus small correction terms of order $(2n + 2)$ and $(2n + 3)$; the latter may be omitted only if one definitely knows that the total correction is negative.

Let us now turn to the more difficult problem of evaluating a lower bound to the energy. Choosing $\varepsilon = \langle H \rangle$, substituting this value in the function $f(E)$, and applying the bracketing theorem, one obtains

$$E \geq \left\langle \varphi \left| H + H \frac{P}{\langle H \rangle - H} H \right| \varphi \right\rangle. \tag{107}$$

The right-hand member gives a better lower bound than the Temple value,[18] which is easily derived from the expression. By improving the value for

[17] See, e.g., P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. 2, p. 1120; O. Sinanoglu, Phys. Rev. **122**, 493 (1961).

[18] G. Temple, Proc. Roy. Soc. (London) **A119**, 276 (1928).

$\langle H \rangle$ according to (105), one can also improve this lower bound indefinitely. Putting $H = H_0 + V$, one obtains particularly

$$E \geq E_0 + E_1 + \left\langle \varphi_0 \left| V \frac{P}{\langle H \rangle - H_0 - V} V \right| \varphi_0 \right\rangle, \qquad (108)$$

where the last term is always negative. It seems as if formula (108) would form a good starting point for further research.

So far, our discussion concerning upper and lower bounds is valid only for the ground state. However, we can now make us free from this restriction and consider *any* state if we observe that, with a proper choice of the parameter ε, the two quantities (105) and (108) will bracket any true eigenvalue E. One of the quantities gives an upper bound and the other one a lower bound, and the order may vary from state to state.

In conclusion, it should be added that a great deal of research[19] has recently been carried out on the problem of the lower bounds in the special case when the perturbation V is *positive definite.*

Conclusions

This study of the eigenvalue problem shows that one can get a detailed insight into the structure of the various types of perturbation theories and their connection with the more condensed forms based on the use of inverse matrices of type (26) by means of the partitioning technique. The operator formalism is particularly transparent, but the analysis has, of course, only a formal character. We have made no attempts to analyze the nature of the eigenvalue spectrum itself or to prove any existence theorems. In certain connections, the formalism may seem to be too abstract, and it is then worthwhile to remember that the theory has been developed from a highly practical numerical method for solving secular equations. The approach is applicable both to one- and many-particle systems, but the special problems connected with the latter will further be discussed in a following paper.

APPENDIX. CALCULATION OF AN INVERSE MATRIX

It is clear that, if one is only interested in evaluating the corrections in Schrödinger's perturbation theory up to a certain finite order, i.e., the energy quantities

$$E_1 = \langle \varphi_0 | V | \varphi_0 \rangle, \qquad E_2 = \langle \varphi_0 | V R_0 V | \varphi_0 \rangle,$$
$$E_3 = \langle \varphi_0 | V R_0 V R_0 V | \varphi_0 \rangle - E_1 \langle \varphi_0 | V R_0^2 V | \varphi_0 \rangle,$$
$$\cdots, \text{ etc.},$$

and the corresponding corrections to the wave function, one does not have to invert any matrices or to solve any eigenvalue problems in addition to the unperturbed problem $H_0 \varphi_0 = E_0 \varphi_0$ for the specific state under consideration. If one introduces a complete discrete representation consisting of φ_0 and a conveniently chosen form of its orthogonal complement, the results desired may be obtained by *solving a series of equation systems* with fixed coefficients rendering a set of vectors representing the functions $R_0 V \varphi_0$, $R_0^2 V \varphi_0$, $R_0 V R_0 V \varphi_0$, $\cdots$ in the discrete basis.[20] The mathematical theory for the treatment of linear equation systems of infinite order is then valid.

However, if one is interested in obtaining a solution which would correspond to ∞-order perturbation theory, one has to evaluate the inverse matrix T defined by (26). This is a more difficult problem, and we will here briefly review some of the methods which are available for this purpose.

(a) Fundamental Identity

Let A and B be two arbitrary operators which usually do not commute. For the inverse of $(A - B)$, one has always the identity:

$$(A - B)^{-1} = A^{-1} + A^{-1} B (A - B)^{-1} \qquad (A1)$$
$$= A^{-1} + (A - B)^{-1} B A^{-1}, \qquad (A2)$$

provided that the inverse operators involved really exist. The identity is easily proven by multiplying it to the right (respectively, to the left) by $(A - B)$. Since the two right-hand members are equal, one has the relation

$$A^{-1} B (A - B)^{-1} = (A - B)^{-1} B A^{-1}, \qquad (A3)$$

which is useful in moving the factor $(A - B)^{-1}$ in long products. In perturbation theory, the relations (A1) and (A2) are often called "integral equations" depending on the fact that in a continuous representation, the operator products in the last term are evaluated by means of integrals.

(b) Power-Series Expansion

By repeated use of (A1) and (A2), we obtain

[19] N. W. Bazley, Proc. Natl. Acad. Sci. U. S. **45**, 850 (1959); Phys. Rev. **120**, 144 (1960); N. W. Bazley and D. W. Fox, Reports CF-2911 and CF-2928, Johns Hopkins Applied Physics Laboratory (1961) (unpublished).

[20] See, e.g., P. O. Löwdin, Technical Note No. 28, Part III, Uppsala Quantum Chemistry Group 1959 (unpublished).

$$(A-B)^{-1}$$
$$= A^{-1} + A^{-1}BA^{-1} + A^{-1}BA^{-1}B(A-B)^{-1}$$
$$= A^{-1} + A^{-1}BA^{-1} + A^{-1}B(A-B)^{-1}BA^{-1}$$
$$= A^{-1} + A^{-1}BA^{-1} + (A-B)^{-1}BA^{-1}BA^{-1}, \quad \text{(A4)}$$

where the three forms of the remainder are equal according to (A4). Repeating the procedure, one gets the well-known power-series expansion including a remainder term. The infinite series

$$(A-B)^{-1} = A^{-1} + A^{-1}BA^{-1} + A^{-1}BA^{-1}BA^{-1} + \cdots \quad \text{(A5)}$$

is convergent, if and only if, the absolute magnitude of all expectation values $\langle BA^{-1}\rangle$ are less than 1:

$$|\langle BA^{-1}\rangle| \leq q < 1. \quad \text{(A6)}$$

The power series has hence a limited range of applicability. In perturbation theory, it is used in deriving both the Brillouin-type and the Schrödinger-type expansions.

(c) Modified Power Series

Putting $A - B = A_1 - B_1 = A_2 - B_2 = \cdots$, one can modify the power series including remainder at any stage of the expansion:

$$(A-B)^{-1} = A^{-1} + A^{-1}BA_1^{-1} + A^{-1}BA_1^{-1}B_1A_2^{-1} + \cdots . \quad \text{(A7)}$$

This relation may be useful in identifying modifications of perturbation theory derived by other methods, since a great deal of variety is apparently possible.

(d) Eigenvalue Transformation

Let us now consider the special case when (A−B) is an Hermitian or normal operator, and let U denote the unitary transformation which brings $(A - B)$ to diagonal form λ, so that $U^\dagger(A - B)U = \lambda$. This gives immediately

$$(A-B)^{-1} = U\lambda^{-1}U^\dagger, \quad \text{(A8)}$$

which relation may be used for evaluating the inverse. In a discrete representation, one obtains

$$(A-B)^{-1}_{kl} = \sum_\alpha \frac{U_{k\alpha}U^*_{l\alpha}}{\lambda_\alpha}, \quad \text{(A9)}$$

This method for calculating the inverse is characteristic for a large part of the conventional perturbation theory, but it is often rather cumbersome, particularly if λ has a partly continuous spectrum. We note that, in deriving the Schrödinger form, one is using both the power series method for the expansion in V and the eigenvalue method for evaluating R_0.

(e) Second-Order Iteration Procedure

Let $\mathfrak{Z}$ be an approximation to the inverse $(A - B)^{-1}$, so that $\mathfrak{Z} = (A - B)^{-1} + \epsilon$, where ϵ is a first-order correction. For the new operator $\mathfrak{Z}^*$, defined by the relation

$$\mathfrak{Z}^* = 2\mathfrak{Z} - \mathfrak{Z}(A-B)\mathfrak{Z}, \quad \text{(A10)}$$

one obtains $\mathfrak{Z}^* = (A - B)^{-1} - \epsilon(A - B)\epsilon$, i.e., $\mathfrak{Z}^*$ is now correct to the second order. This leads to a convenient second-order iteration procedure for evaluating an inverse matrix. Formula (A10) is used in deriving the extension of Wigner's theorem given in (61).

If one chooses $\mathfrak{Z}$ to be the power series (A5) truncated at order n:

$$\mathfrak{Z}_n = A^{-1}\sum_{k=0}^{n}(BA^{-1})^k, \quad \text{(A11)}$$

application of (A10) gives immediately

$$\mathfrak{Z}_n^* = \mathfrak{Z}_{2n+1}, \quad \text{(A12)}$$

i.e., $\mathfrak{Z}^*$ equals the same power series up to order $(2n + 1)$.

(f) Infinite Product

Putting $\mathfrak{Z} = A^{-1}$ and using the iteration formula (A10), one obtains the infinite product

$$(A-B)^{-1} = A^{-1}(1 + BA^{-1})(1 + BA^{-1}BA^{-1}) \times (1 + BA^{-1}BA^{-1}BA^{-1}BA^{-1}) \cdots$$
$$= A^{-1}\prod_{k=0}^{\infty}\{1 + (BA^{-1})^{2^k}\}. \quad \text{(A13)}$$

The partial products equal the partial sums $\mathfrak{Z}_0$, $\mathfrak{Z}_1$, $\mathfrak{Z}_3$, $\mathfrak{Z}_7$, $\mathfrak{Z}_{15}$, $\cdots$ defined by (A11), and the infinite product is convergent and divergent at the same time as the infinite series (A5).

(g) Determinant Formula

If one introduces a discrete basis, the calculation of an inverse matrix is actually equivalent to the solution of a series of equation systems. In this connection, it is also worthwhile to remember the formula

$$(A-B)^{-1}_{kl} = ||A - B||_{lk}/||A - B||, \quad \text{(A14)}$$

where $||A - B||$ is the determinant of $(A - B)$ and the numerator is one of its minors. For an infinite

basis, one has to apply the theory for infinite determinants, and at least the quotient has to be convergent.

(h) Successive Partitioning

In quantum-mechanical applications to atoms, molecules, and solid-state, one has so far to a large extent used basic sets which have been truncated to finite orders. In this case, it is of importance to extend the basis, and this is conveniently done in the method of "successive partitioning" in which one utilizes the previously obtained results. For the inverse matrix, the method is based on the formula[21]:

$$\begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix}^{-1} = \begin{bmatrix} K_{11}^{-1} + \kappa_{12}\bar{K}_{22}^{-1}\kappa_{21}; & -\kappa_{12}\bar{K}_{22}^{-1} \\ -\bar{K}_{22}^{-1}\kappa_{21} & ; \quad \bar{K}_{22}^{-1} \end{bmatrix} \tag{A15}$$

where

$$\kappa_{12} = K_{11}^{-1}K_{12}, \qquad \kappa_{21} = K_{21}K_{11}^{-1},$$
$$\bar{K}_{22} = K_{22} - K_{21}K_{11}^{-1}K_{12}. \tag{A16}$$

One starts out in the upper left corner, adds one row and one column at a time, so that K_{22} stays one dimensional, and notes that K_{11}^{-1} has been evaluated in the preceding step of the calculation. The formula is easily derived by solving the equation $\mathbf{K}\cdot\mathbf{x} = \mathbf{1}$ by partitioning. The final result will actually be the same as if one applied the method of successive partitioning directly to the eigenvalue problem in the form (10).

(i) Chebyshev Expansions

Since this method has been treated in full elsewhere,[22] we will here make only a few comments. If α and x are two real numbers, one has the two expansions

$$(\alpha - x)^{-1} = [r/(1 - r^2)][1 + rC_1(x) + r^2C_2(x) + \cdots + r^nC_n(x) + \cdots]$$
$$= r[1 + rS_1(x) + r^2S_2(x) + \cdots + r^nS_n(x) + \cdots], \tag{A17}$$

where $1 + r^2 = \alpha r$,

$$r = \tfrac{1}{2}[\alpha - (\alpha^2 - 4)^{1/2}] = 2[\alpha + (\alpha^2 - 4)^{1/2}]^{-1}, \tag{A18}$$

and $C_n(x)$ and $S_n(x)$ are the standard Chebyshev polynomials

$$C_n(x) = 2\cos n\theta,$$
$$S_n(x) = \sin(n + 1)\theta/\sin\theta, \tag{A19}$$

with $x = 2\cos\theta$. The relations (A17) are easily derived from the expansion $(1 - re^{i\theta})^{-1} = \sum_k r^k e^{ik\theta}$ by separating it into its real and imaginary parts; they are convergent, if $r < 1$ and θ is real, i.e., $|\alpha| > 2$ and $|x| < 2$.

It is now possible to apply this technique to the operator $(A - B)^{-1}$ in several different ways. In the case when A^{-1} exists, we may write $(A - B)^{-1} = A^{-1}(1 - BA^{-1})^{-1}$ and consider the series expansion for the second factor. In the case when $|\langle BA^{-1}\rangle| \leq q < 1$, we will choose a parameter α situated in the interval $2 < \alpha \leq 2/q$ and use the formula:

$$(1 - BA^{-1})^{-1} = \alpha(\alpha - \alpha BA^{-1})^{-1} \tag{A20}$$
$$= \frac{\alpha r}{1 - r^2}\left[1 + \sum_{k=1}^{\infty} r^k C_k(\alpha BA^{-1})\right]$$
$$= \alpha r \sum_{k=0}^{\infty} r^k S_k(\alpha BA^{-1}), \tag{A21}$$

where r is given by (A18). We note that the right-hand members are independent of the value of α as long as the series are convergent. If the power-series (A5) have a quotient maximized by q, the Chebyshev expansions are characterized by a value of r which may be chosen as low as

$$r = q/[1 + (1 - q^2)^{1/2}];$$

the convergence is hence essentially improved.

Even combinations of the two formulas may be useful. The Chebyshev expansions are particularly convenient in treating operators and matrices having cyclic character, since they lead to nice closed expressions[22] that are then valid also outside the range of convergence of the original expansions.

In conclusion, we note that only a few of the methods described here have so far been used to investigate the various types of "perturbation theory" that could be obtained from the fundamental formulas of type (34) and (37) containing the matrix T by treating the inverse matrix in (26) in different ways. Particularly the Chebyshev expansions seem to offer an interesting starting point for further research.

---

[21] See, e.g., A. S. Householder, *Principles of Numerical Analysis* (McGraw-Hill Book Company, Inc., New York, 1953) p. 78; J. O. Hirschfelder and P. O. Löwdin, Molecular Phys. **2**, 229 (1959).

[22] P. O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. **1**, 461 (1960).

Variational Formulations of Equilibrium Statistical Mechanics*

CYRANO DE DOMINICIS†
Lyman Laboratory of Physics, Harvard University
(Received March 19, 1962)

Thermodynamical functions for classical and quantum systems are expressed in terms of the one-particle density n_1 and the two-particle correlation matrix C_{12} (or quantities in direct relation to them). Use is made of topological relations valid for the diagram representations of the grand partition function expansions. The result considered as a functional of n_1 and C_{12} is stationary under independent variations δn_1 and δC_{12}. In particular, the entropy functional of a classical system no longer contains any reference to the equilibrium parameters (or to the interactions) and the second functional derivative is a negative definite matrix. The entropy functional of a quantum system conserves traces of the equilibrium parameters in the Lee-Yang formulation; the Green's function formulation does not, but in this case the second functional derivative is no longer a negative definite matrix.

INTRODUCTION

IN the past three years, there have been several successful attempts[1-5] to express the thermodynamical functions of quantum-mechanical systems as explicit functionals of one-body functions bearing a more or less direct relation to the average occupation number. These formulations[1,5] remind us, in character, of the classical virial expansion[6] in terms of the one-particle density, and could be expected to present the same type of interest for the study of phase transitions, although some of these formulations[2-4] appear, in some respects, more like generalizations of usual zero-temperature (ground-state) expansions.[7] All of them, however share in common a variational property that was first established by Lee and Yang[1]; namely, that the grand partition function of the system is stationary under variations of the one-body function considered. This common feature has been traced back to a topological relation[5,8] satisfied by the various diagramatic representations and directly linked to the tree structure of these diagrams.

The next question arising naturally is whether analogous formulations exist in terms of one- and two-body functions (or matrices) directly related to the one- and two-particle density matrices.[9] Again, besides the advantages brought by their variational character, such formulations would be of particular interest in characterizing and studying second-order phase transitions.

The purpose of this work is to answer that question. For both the classical[10-12] and the quantum systems, the grand partition function is written as a stationary expression in terms of the one- and two-particle densities (or quantities in direct relation to them).

In Sec. I(i) a classical grand canonical ensemble is considered and we recall how its grand partition function $\ln Z(\alpha, \beta)$ can be expressed as a stationary functional of the one-particle density through the use of a topological relation valid for Yvon-Mayer diagrams. This section is essentially a repetition of Sec.[13] VI.1 of the last paper quoted in reference 5.

In Sec. I(ii) the interaction potential is expressed

* Supported in part by the Air Force Office of Scientific Research (ARDC), under contract number A.F. 49 (638)-589.

† On leave of absence from Service de Physique Mathematique, Centre d'Etudes Nucléaires de Saclay, Seine et Oise, France.

[1] T. D. Lee and C. N. Yang, Phys. Rev. **113**, 1165 (1959); **117**, 22 (1960). Hereafter we shall refer to the last reference as (L-Y). See also M. S. Green, Phys. Rev. Letters **1**, 409 (1958).

[2] P. Martin and J. Schwinger, Phys. Rev. **115**, 1342 (1959).

[3] R. Balian and C. De Dominicis, Nuclear Phys. **16**, 502 (1960); Compt. rend. **250**, 3285, 4111 (1960).

[4] J. Luttinger and J. Ward, Phys. Rev. **118**, 1417 (1960).

[5] R. Balian, C. Bloch, and C. De Dominicis, Compt. Rend. **250**, 2850 (1960); Nuclear Phys. **25**, 529 (1961); **27**, 294 (1961).

[6] H. Ursell, Proc. Cambridge Phil. Soc. **23**, 685 (1927); J. Yvon, Actualités sci. et ind. **203** (1935); J. Mayer, J. Chem. Phys. **5**, 67 (1937).

[7] J. Goldstone, Proc. Roy. Soc. (London) **A 239**, 267 (1957).

[8] C. Bloch, Physica **26**, 562 (1960).

[9] Notice that one could easily express the pressure in terms of the one- and two-particle density through the use of the virial theorem. Such an expression, however, has no stationarity property.

[10] T. Morita and K. Hiroike, Progr. Theoret. Phys. (Kyoto) **25**, 537 (1961).

[11] M. S. Green, J. Chem. Phys. **33**, 1403 (1960).

[12] The work of Morita and Hiroike (reference 10) was brought to the attention of the author by Professor Uhlenbeck, after this work had been completed. Morita and Hiroike study classical systems and establish Eq. (1.50). The method used here, however, covers both the classical and quantum cases by exhibiting the topological relation which underlies the stationarity properties; a feature particular to the entropy expression for classical systems which seems to have escaped Morita and Hiroike's attention is also discussed.

The functional expression derived for the grand partition function of classical systems in reference 11 is stationary but not maximal.

[13] Several results of that section are already contained in some early work of J. Yvon (1935).

as a functional of the two-particle density following, with slight variations, an analysis made by several authors.[10,14] Then a new topological relation is used to express ln $Z(\alpha, \beta)$ as a stationary functional of both the one and two particle densities. An interesting feature of the expression for the entropy is pointed out, with its hypothetical relation to a generalized expression for the Boltzmann H function. The effect of the introduction of many body forces is discussed.

In Sec. II(i), quantum systems are considered in the Lee-Yang formulation, which is described from the point of view introduced in Sec. I(i).

In Sec. II(ii) the results of Sec. I(ii) are extended to quantum systems in the Lee-Yang formulation. Similar results for the Green's function formulation[2,4] are quoted.

I. CLASSICAL SYSTEMS

i. Formulation in Terms of the One-Particle Density

a. Definitions, Yvon-Mayer Diagrams

A classical system of N identical particles is characterized by its Hamiltonian

$$H_N = \sum_i^N [\mathbf{p}_i^2 + u(\mathbf{r}_i)] + \sum_{i>j}^N v(\mathbf{r}_i - \mathbf{r}_j), \tag{1.1}$$

where $\mathbf{p}_i$ and $\mathbf{r}_i$ are the momentum and position of the ith particle; $u(\mathbf{r}_i)$ is a (one-body) external potential, $v(\mathbf{r}_i - \mathbf{r}_j)$ is a two-body interaction potential (for shortness we shall also use u_i and v_{ij}); and the mass is taken as $m = \frac{1}{2}$.

Grand partition function: It is defined by

$$Z(\alpha, \beta) = \sum_{N=0}^{\infty} \frac{e^{\alpha N}}{N!} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \, d\mathbf{p}_1 \cdots d\mathbf{p}_N \, e^{-\beta H_N} \tag{1.2}$$

$$= \sum_{N=0}^{\infty} \frac{1}{N!} \int d\mathbf{r}_1 \cdots d\mathbf{r}_N \prod_i^N n_i^0 \prod_{i>j}^N (1 + g_{ij}). \tag{1.3}$$

Here $e^{\alpha} = z$ is the (absolute) chemical activity, $\beta^{-1} = kT$ is the temperature (times the Boltzmann constant) and we have

$$g_{ij} = g(\mathbf{r}_i - \mathbf{r}_j) = \exp[-\beta v(\mathbf{r}_i - \mathbf{r}_j)] - 1 \tag{1.4}$$

$$n_i^0 = n^0(\mathbf{r}_i) = \lambda^{-3} \exp[\alpha - \beta u(\mathbf{r}_i)] \tag{1.5}$$

$$\lambda^{-3} = \int d\mathbf{p}_i \exp(-\beta \mathbf{p}_i^2). \tag{1.6}$$

Yvon-Mayer diagrams are then a convenient way of representing expansion (1.3). A labeled Yvon-Mayer diagram of order N is a set of N points, labeled 1, 2 $\cdots$ N, and of single lines joining some pairs of these points. To each diagram is associated an algebraic quantity calculated with the following rule (A):

(i) to each point i associate a factor n_i^0,
(ii) to each line linking points i and j associate a factor g_{ij},
(iii) integrate independently over N points with a weight $(N!)^{-1}$.

$Z(\alpha, \beta)$ is then given by the sum of the contributions associated with all distinct, labeled, Yvon-Mayer diagrams.

Densities: They are defined as the average of the density operators, over the grand canonical ensemble; the one- and two-particle densities are expressed as

$$n(\mathbf{r}) = Z^{-1}(\alpha, \beta)\Big[\sum_{N=0}^{\infty} \frac{1}{(N+1)!} \int d\mathbf{r}_1 \cdots d\mathbf{r}_{N+1} \times \prod_i^{N+1} n_i^0 \prod_{i>j}^{N+1} (1 + g_{ij}) \Big(\sum_i^{N+1} \delta(\mathbf{r} - \mathbf{r}_i)\Big)\Big] \tag{1.7}$$

$$n(\mathbf{r}, \mathbf{r}') = Z^{-1}(\alpha, \beta) \times \Big[\sum_{N=0}^{\infty} \frac{1}{(N+2)!} \int d\mathbf{r}_1 \cdots d\mathbf{r}_{N+2} \prod_i^{N+2} n_i^0 \times \prod_{i>j}^{N+2} (1 + g_{ij}) \Big(\sum_{i \neq j}^{N+2} \delta(\mathbf{r} - \mathbf{r}_i)\, \delta(\mathbf{r}' - \mathbf{r}_j)\Big)\Big]. \tag{1.8}$$

The expansion of $n(\mathbf{r}, \mathbf{r}')Z(\alpha, \beta)$ for example is represented by all distinct Yvon-Mayer diagrams with $N + 2$ points, of which N are labeled and two are distinguished and held fixed at $\mathbf{r}, \mathbf{r}'$. These points are the *roots* of the diagram and the diagram is a 2-rooted diagram (or a 2-diagram for short). The algebraic quantity associated with an N-labeled, 2-diagram, is calculated with the formally unchanged rule (A), as seen from inspection of (1.8).

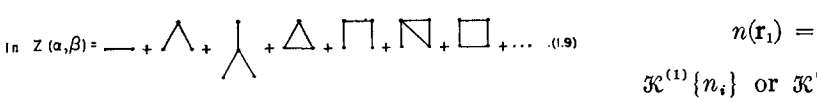
A diagram is *connected* if there exists at least one continuous line between any two points; a 2-diagram is root-connected if it is made of two disconnected parts, each one of them connected to a different root. Simple combinatorial arguments show[15] that: (i) in (1.7) and (1.8) the denominator $Z(\alpha, \beta)$ eliminates the contributions of all nonconnected

[14] J. Van Leuwen, J. Groeneveld, and J. de Boer, Physica **25**, 792 (1959). E. Meeron, Phys. Fluids **1**, 246 (1958); J. Math. Phys. **1**, 192 (1960). M. S. Green, Hugues Aircraft Company Report (1959) (unpublished). T. Morita and K. Hiroike, Progr. Theoret. Phys. (Kyoto) **23**, 1003 (1960). (To quote only the papers where a complete analysis of the two-particle density is performed). Early work in the same direction may be found in J. Yvon: Rev. sci. 662 (1939); Nuovo cimento Suppl. 9, 144, (1958).

[15] J. Mayer and M. Mayer, *Statistical Mechanics* (John Wiley & Sons, Inc., New York); see also, G. Ulhenbeck and G. Ford, Studies in Stat. Mech. 1 (1962).

(or non-root-connected) diagrams, (ii) $\ln Z(\alpha, \beta)$ is given by the sum of the contributions associated with the connected diagrams only.

$\ln Z(\alpha, \beta)$, $n(\mathbf{r})$, $n(\mathbf{r}, \mathbf{r}')$ are then represented by all the distinct, *connected*, (or root-connected) N-labeled, ζ-diagrams, with $\zeta = 0, 1, 2$, respectively. These quantities may also be represented in terms of the distinct, *free* (i.e., nonlabeled), (root-) connected ζ-diagrams. The weight $(N!)^{-1}$ is then replaced by $(s_\zeta)^{-1}$, where s_ζ is a symmetry factor for a ζ-diagram.[16] For example we represent the grand partition function expansion as



(1.9)

Notice that if we consider $Z(\alpha, \beta)$ as a functional of n_i^0 we have[17] ($\mathbf{r}_1 \neq \mathbf{r}_2$)

$$n(\mathbf{r}_1) = n^0(\mathbf{r}_1)[\delta/\delta n^0(\mathbf{r}_1)] \ln Z(\alpha, \beta), \tag{1.10}$$

$$n(\mathbf{r}_1; \mathbf{r}_2) = n^0(\mathbf{r}_1)n^0(\mathbf{r}_2)[\delta^2/\delta n^0(\mathbf{r}_1)\, \delta n^0(\mathbf{r}_2)] \times \ln Z(\alpha, \beta) + n(\mathbf{r}_1)n(\mathbf{r}_2). \tag{1.11}$$

From now on, unless specified, we deal with connected ζ-diagrams. We analyze ζ-diagrams using the following definitions:

An *articulation point* of multiplicity $m (m > 1)$ is such that by cutting a ζ-diagram at this point, we can separate at most m connected parts.[18]

A *1-irreducible ζ-diagram* is such that it has *no articulation point*. The 1-irreducible 0-diagrams are also known as stars, and a general 0-diagram is then a tree of stars.[15] We call $\mathcal{K}^{(1)}\{n_i^0\}$ the *contribution of all the 1-irreducible 0-diagrams*. The contribution of all the 1-irreducible 1-diagrams is then

$$n^0(\mathbf{r}_1)K^{(1)}\{\mathbf{r}_1; n_i^0\} = n^0(\mathbf{r}_1)[\delta/\delta n^0(\mathbf{r}_1)]\mathcal{K}^{(1)}\{n_i^0\}. \tag{1.12}$$

We now proceed to express $\ln Z(\alpha, \beta)$, $n(\mathbf{r}_1)$, $n(\mathbf{r}_1; \mathbf{r}_2)$ as functionals of the one-particle density.

b. *Densities*

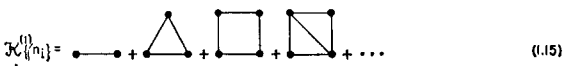
Consider the contribution to $n(\mathbf{r}_1)$ of the 1-diagrams where only $\mathbf{r}_1$ is allowed to be an articulation point, we get

$$n^0(\mathbf{r}_1)\left[1 + \sum_{m=1}^{\infty} \frac{1}{m!} (K^{(1)}\{\mathbf{r}_1; n_i^0\})^m\right] = n^0(\mathbf{r}_1) \exp (K^{(1)}\{\mathbf{r}_1; n_i^0\}), \tag{1.13}$$

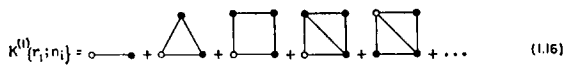
where m is the multiplicity of the articulation point $\mathbf{r}_1$. A general 1-diagram can be generated then by making the most general insertion at each point i other than the root $\mathbf{r}_1$ of these diagrams. This has the effect of replacing the factor $n^0(\mathbf{r}_i)$ by $n(\mathbf{r}_i)$ (cf. Appendix A) and yields the Yvon equation[13]

$$n(\mathbf{r}_1) = n^0(\mathbf{r}_1) \exp (K^{(1)}\{\mathbf{r}_1; n_i\}). \tag{1.14}$$

$\mathcal{K}^{(1)}\{n_i\}$ or $\mathcal{K}^{(1)}\{\mathbf{r}_1; n_i\}$ are now functionals of $n(\mathbf{r}_i)$ and are represented by all distinct 1-irreducible, 0- or 1-diagrams:



(1.15)



(1.16)

Associated contributions are now calculated with a rule (B) differing from rule (A) through point (i) which now reads:

B(i) to a root (represented by an empty dot) associate a factor of unity; to each point i (represented by a heavy dot) associate a factor n_i.

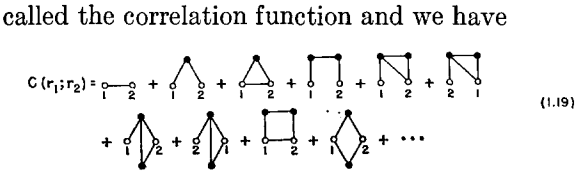
Naturally (1.16) is obtained by taking the functional derivative of (1.15)

$$K^{(1)}\{\mathbf{r}_1; n_i\} = [\delta/\delta n(\mathbf{r}_1)]\mathcal{K}^{(1)}\{n_i\}. \tag{1.17}$$

Similarly we may build the general 2-diagram by considering first the class of 1-irreducible 2-diagrams and by making the most general insertion at each point (including the root $\mathbf{r}_1$ and $\mathbf{r}_2$). If we write

$$n(\mathbf{r}_1, \mathbf{r}_2) = n(\mathbf{r}_1)n(\mathbf{r}_2)[1 + C(\mathbf{r}_1; \mathbf{r}_2)], \tag{1.18}$$

the first term represents the root-connected contributions, the second the connected one. $C(\mathbf{r}_1; \mathbf{r}_2)$ is called the correlation function and we have



(1.19)

$C(\mathbf{r}_1; \mathbf{r}_2)$ is represented by all the distinct, (connected), 1-irreducible, 2-diagrams, the contributions of which are calculated with rule (B).

---

[16] The number of N-labeled ζ-diagrams corresponding to one given free ζ-diagram is $N!/s_\zeta$. The symmetry number s_ζ is the order of the permutation group of the N free points of the diagram, which leaves the connections invariant. Labeled diagrams are useful for detailed proofs, free diagrams handier to describe representations of expansions and are used throughout the text.

[17] For convenience, in the following we use $n(\mathbf{r}_1)$, $n(\mathbf{r}_1, \mathbf{r}_2)$, or n_1, n_{12} instead of $n(\mathbf{r})$, $n(\mathbf{r}, \mathbf{r}')$.

[18] m-1 of these look like 1-diagrams, one like a $(\zeta + 1)$-diagram (or a ζ-diagram if the articulation point considered is a root).

c. Grand Partition Function

Of course, it could be obtained by direct integration of (1.14) since we have

$$\int d\mathbf{r}_1 n(\mathbf{r}_1) = (\partial/\partial\alpha) \ln Z(\alpha, \beta). \qquad (1.20)$$

We rather use a more detailed analysis which will eventually give more insight into the variational properties still to be established. $\ln Z(\alpha, \beta)$ is represented by the sum of distinct 0-diagrams (calculated with rule A) and we wish to express it in terms of $n(\mathbf{r}_i)$.

To that effect we consider the three following expressions:

$$\text{(a)} \qquad \int d\mathbf{r}_1 n(\mathbf{r}_1). \qquad (1.21)$$

This quantity can be thought of as represented by 0-diagrams (calculated with rule A) but with *one point distinguished* at $\mathbf{r}_1$, i.e., the contribution to (1.21) of the class of diagrams with $N(p)$ points is $N(p)$ times that of the same class to $\ln Z(\alpha, \beta)$. This is also obvious from relation (1.10).

$$\text{(b)} \qquad \mathcal{K}^{(1)}\{n_i\}. \qquad (1.22)$$

This quantity is represented [see (1.15)] by 1-irreducible 0-diagrams where one has made the most general insertion at each point. But it can also be represented by the most general 0-diagram (calculated with rule A) where one 1-*irreducible part* has been *distinguished*. Let $N(\mathcal{K}^{(1)})$ be the number of 1-irreducible parts of a diagram (i.e., the number of separate parts obtained by cutting all the articulation points of the diagram). The contribution of 0-diagrams with $N(\mathcal{K}^{(1)})$ 1-irreducible parts to $\mathcal{K}^{(1)}\{n_i\}$ is then $N(\mathcal{K}^{(1)})$ times the contribution to $\ln Z(\alpha, \beta)$.

(c) Consider finally a general 0-diagram where we *distinguish one point* at $\mathbf{r}_1$ and *one* 1-*irreducible part* rooted at $\mathbf{r}_1$). Such a diagram may be generated out of the distinguished 1-irreducible part, making the most general insertion at all points other than $\mathbf{r}_1$ (which provides a factor $K^{(1)}\{\mathbf{r}_1; n_i\}$) and at the root $\mathbf{r}_1$ yielding

$$\int d\mathbf{r}_1 n(\mathbf{r}_1) K^{(1)}\{\mathbf{r}_1; n_i\}. \qquad (1.23)$$

An ordinary point may be viewed as an articulation point with multiplicity $m = 1$. Let $N(p \cdot m)$ be, in a diagram, the number of articulation points weighted by their multiplicity, which is also the number of ways one can choose a 1-irreducible part rooted at a point $\mathbf{r}_1$ on the diagram. Again, a 0-diagram contributing to (1.23) carries an extra weight $N(p \cdot m)$ as compared with its contribution to $\ln Z(\alpha, \beta)$.

We use now a general relation between the introduced weights,

$$N(p) - N(p \cdot m) + N(\mathcal{K}^{(1)}) = +1. \qquad (1.24)$$

This relation valid for a tree of 1-irreducible parts is easily shown by induction[5,8,19]; it allows us to construct the proper weight for 0-diagrams contributing to $\ln Z(\alpha, \beta)$ out of (1,21), (1.22), and (1.23), thereby yielding $\ln Z(\alpha, \beta)$ in terms of $n(\mathbf{r}_i)$;

$$\ln Z(\alpha, \beta) = \int d\mathbf{r}_1 n(\mathbf{r}_1) - \int d\mathbf{r}_1 n(\mathbf{r}_1) K^{(1)}\{\mathbf{r}_1; n_i\} + \mathcal{K}^{(1)}\{n_i\}. \qquad (1.25)$$

Using (1.5), (1.6), and (1.14) we get

$$\ln Z(\alpha, \beta) = \int d\mathbf{r}_1 n(\mathbf{r}_1)[1 - \ln n(\mathbf{r}_1) + \alpha - \ln \lambda^3 - \beta u(\mathbf{r}_1)] + \mathcal{K}^{(1)}\{n_i\}. \qquad (1.26)$$

A more suggestive form is obtained if one uses the *one-particle density in phase space* $\mu(\mathbf{r}_1, \mathbf{p}_1) = \mu_1$ with the result

$$\ln Z(\alpha, \beta) = \int d1 \mu_1 (1 - \ln \mu_1) + \mathcal{K}^{(1)}\{\mu_i\} + \int d1 \mu_1 (\alpha - \beta \mathbf{p}_1^2 - \beta u_1), \qquad (1.27)$$

where $d1$ stands for $d\mathbf{r}_1\, d\mathbf{p}_1$.

d. Variational Properties

Consider the right-hand side of (1.25) as a functional $\Phi\{\bar{K}_i^{(1)}\}$ of an arbitrary function $\bar{K}^{(1)}(\mathbf{r}_i)$ through the relation

$$\bar{n}(\mathbf{r}_1) = n^0(\mathbf{r}_1) \exp[\bar{K}^{(1)}(\mathbf{r}_1)], \qquad (1.28)$$

it is verified that $\Phi\{\bar{K}_i^{(1)}\}$ has an extremum when (1.17) is satisfied, i.e., $\bar{K}^{(1)}(\mathbf{r}_1) = K^{(1)}\{\mathbf{r}_1; n_i\}$. This stationary property is traced to relation (1.24), i.e., to the tree structure of the diagrams.[5,8] Indeed the three terms of the variation $\delta\Phi\{\bar{K}_i^{(1)}\}$ generated by $\delta\bar{K}^{(1)}(\mathbf{r}_1)$ may again be represented diagramatically in three different ways as in points (a, b, c) above, but these diagrams now have an "external" 1-irreducible part $\delta\bar{K}^{(1)}(\mathbf{r}_1)$ rooted at $\mathbf{r}_1$. Such an "external" 1-irreducible part cannot play the role of a distinguished 1-irreducible part in generating diagrams

[19] G. Ford and G. Uhlenbeck, Proc. Natl. Acad. Sci. U. S. **42**, 122 (1956). See also M. S. Green, J. Math. Phys. **1**, 391, (1960).

(point b) and hence for the diagrams representing $\delta\Phi\{\bar{K}_i^{(1)}\}$ we have the relation

$$\bar{N}(p) - \bar{N}(p\cdot m) + \bar{N}(\mathcal{K}^{(1)}) = 0 \qquad (1.29)$$

equivalent to the stationary condition

$$[\delta/\delta\bar{K}^{(1)}(\mathbf{r}_1)]\Phi\{\bar{K}_i^{(1)}\} = 0, \qquad (1.30)$$

which yields

$$\bar{K}^{(1)}(\mathbf{r}_1) = K^{(1)}\{\mathbf{r}_1; n_i\} = [\delta/\delta n(\mathbf{r}_1)]\mathcal{K}^{(1)}\{n_i\} \qquad (1.31)$$

$$\Phi\{K_i^{(1)}\} = \ln Z(\alpha, \beta). \qquad (1.32)$$

The variational property is more conveniently expressed in terms of $n(\mathbf{r}_i)$ [or $\mu(\mathbf{r}_i; \mathbf{p}_i)$]. Let the right-hand side of (1.26), regarded as a functional of $\bar{n}(\mathbf{r}_i)$, be $\Psi\{\bar{n}_i\}$. It is then stationary with respect to variations $\delta\bar{n}(\mathbf{r}_1)$ when $\bar{n}(\mathbf{r}_1) = n(\mathbf{r}_1)$, and hence.

$$\Psi\{n_i\} = \ln Z(\alpha, \beta). \qquad (1.33)$$

In the next section we shall see that this extremum is a maximum for $\ln Z(\alpha, \beta)$.

(ii) Formulation in Terms of the One- and Two-Particle Density

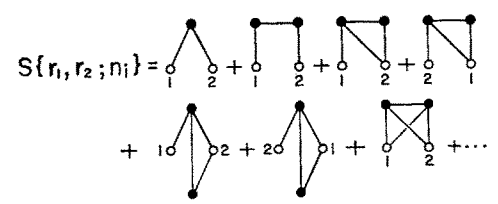
a. Definitions

$\ln Z(\alpha, \beta)$, $n(\mathbf{r}_1)$, $n(\mathbf{r}_1; \mathbf{r}_2)$ have been expressed in terms of $n(\mathbf{r}_i)$ and represented by 1-irreducible ζ-diagrams. In the following, unless mentioned explicitly we only deal with such 1-irreducible ζ-diagrams. We introduce now the following definitions:

An *articulation pair* of multiplicity $\mu(\mu > 1)$ is such that by cutting the diagram at the points (i, j) of the pair, we can separate at most μ connected parts.[20] A single line directly linking (i, j) is not counted as a separate part.

A *2-irreducible ζ-diagram* is such that it has *no articulation pair.*

A *simple 2-diagram* is such that the *roots* $\mathbf{r}_1$, $\mathbf{r}_2$ are *not directly linked and do not constitute an articulation pair.* The contribution of simple 2-diagrams, is called $S\{\mathbf{r}_1; \mathbf{r}_2; n_i\}$:



(1.34)

Like C_{12}, S_{12} is a symmetrical function of $\mathbf{r}_1$ and $\mathbf{r}_2$.

[20] $\mu - 1$ of these parts look like 2-diagrams, one like a $(\zeta + 2)$-diagram [a $(\zeta + 1)$- or a ζ-diagram if the articulation pair contains one or two roots].

b. Two-Particle Density

We discard for the moment 2-diagrams containing a direct link between the roots $\mathbf{r}_1$, $\mathbf{r}_2$ [i.e., a factor $g(\mathbf{r}_1 - \mathbf{r}_2)$]; we write then the contribution of such 2-diagrams where the pair (1, 2) is allowed to be an articulation pair of any multiplicity μ

$$\sum_{\mu=1}^{\infty} \frac{1}{\mu!}[S(\mathbf{r}_1; \mathbf{r}_2)]^{\mu}.$$

Adding the contribution of 2-diagrams with a direct (1, 2) link, we get for the correlation function

$$C(\mathbf{r}_1; \mathbf{r}_2) = [1 + g(\mathbf{r}_1 - \mathbf{r}_2)] \times \{\exp[S(\mathbf{r}_1; \mathbf{r}_2)] - 1\} + g(\mathbf{r}_1 - \mathbf{r}_2). \qquad (1.35)$$

To further analyze S_{12} we introduce the following definitions: A *nodal point* is such that if we cut a 2-diagram at this point, we can separate two parts; one containing the root $\mathbf{r}_1$, the other the root $\mathbf{r}_2$. *A nodal diagram* contains at least one nodal point. Simple diagrams fall into two classes: nodal and elementary (i.e., simple, non-nodal) diagrams.

Nodal 2-diagrams: Let $N(\mathbf{r}_1; \mathbf{r}_2)$ be their contribution. Between two successive nodal points i and j there is, by definition, a non-nodal part described by $X(\mathbf{r}_i; \mathbf{r}_j)$

$$X(\mathbf{r}_i; \mathbf{r}_j) = C(\mathbf{r}_i; \mathbf{r}_j) - N(\mathbf{r}_i; \mathbf{r}_j). \qquad (1.36)$$

The contribution of nodal diagrams can then be written in short, as the solution of the integral equation

$$N(\mathbf{r}_1; \mathbf{r}_2) = \int d\mathbf{r}_3 X(\mathbf{r}_1; \mathbf{r}_3)n(\mathbf{r}_3)X(\mathbf{r}_3; \mathbf{r}_2) + \int d\mathbf{r}_3 X(\mathbf{r}_1; \mathbf{r}_3)n(\mathbf{r}_3)N(\mathbf{r}_3; \mathbf{r}_2). \qquad (1.37)$$

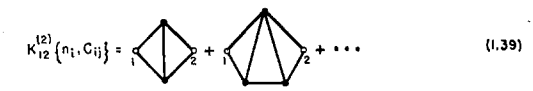
Substitution of (1.36) into (1.37) yields

$$N(\mathbf{r}_1; \mathbf{r}_2) = \int d\mathbf{r}_3 C(\mathbf{r}_1; \mathbf{r}_3)n(\mathbf{r}_3)C(\mathbf{r}_3; \mathbf{r}_2) - \int d\mathbf{r}_3 N(\mathbf{r}_1; \mathbf{r}_3)n(\mathbf{r}_3)C(\mathbf{r}_3; \mathbf{r}_2), \qquad (1.38)$$

an equation expressing N_{12} in terms of n_i and C_{ij}. Notice that for a homogeneous system, n_i is a constant, $C_{ij} = C(\mathbf{r}_i - \mathbf{r}_j)$, and the Fourier transforms of (1.37, 38) reduce to algebraic equations linear in $\tilde{N}(\mathbf{k}) \propto \int d\mathbf{r} \exp(i\mathbf{k}\cdot\mathbf{r})N(\mathbf{r})$.

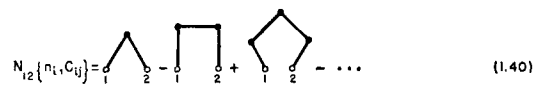
Elementary 2-diagrams: They can be generated from the 2-irreducible 2-diagrams $K^{(2)}(\mathbf{r}_1; \mathbf{r}_2)$ (containing no articulation pairs) by the following procedure: Replace each line directly linking any pair (i, j) by the most general insertion connecting (i, j);

algebraically it amounts to replace the factor g_{ij} by C_{ij} (cf. Appendix A). We thus have, diagramatically,



(1.39)

and by iteration of (1,38) we obtain the alternate series



(1.40)

Algebraic quantities associated with these 2-diagrams are calculated now with rule (C) differing from rule (B) through point (ii) which now reads:

C(ii) to each heavy line linking the pair (i, j) associate a factor C_{ij}.

Notice that the series of diagrams giving N_{12} carries an extra weight $(-)^{p+1}$ where p is the order of the diagram.

We thus have expressed S_{12} in terms of n_i and C_{ij},

$$S_{12}\{n_i, C_{ij}\} = N_{12}\{n_i, C_{ij}\} + K_{12}^{(2)}\{n_i, C_{ij}\}, \quad (1.41)$$

and with (1.18, 35) the two-particle density itself in terms of n_i and C_{ij}.

c. *Grand Partition Function*

Direct integrations of $n(\mathbf{r}_1; \mathbf{r}_2)$ over α would be much less straightforward here. We use the same type of analysis as in Sec. I(i), i.e., we list quantities easily expressed in terms of n_i and C_{ij} and represented by the same diagrams as $\ln Z(\alpha, \beta)$ except for varying weights:

$$\text{(a)} \qquad \int d\mathbf{r}_1\, d\mathbf{r}_2\, n(\mathbf{r}_1)n(\mathbf{r}_2)C(\mathbf{r}_1; \mathbf{r}_2); \qquad (1.42)$$

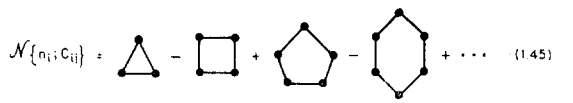
This quantity can be thought of as represented by all (reducible or not) 0-diagrams calculated with rule (A), but with *one pair of points distinguished* at $\mathbf{r}_1$, $\mathbf{r}_2$; contribution to (1.42) of diagrams with $N(\pi)$ pairs is $N(\pi)$ times that of the same diagrams contribution to $\ln Z(\alpha, \beta)$ which is also obvious from (1.11).

(b) In Sec. I(i) we considered $K_1^{(1)}$ as a functional derivative of $\mathcal{K}^{(1)}$, here we shall consider N_{12} and $K_{12}^{(2)}$ as related to functionals $\mathfrak{N}\{n_i; C_{ij}\}$ and $\mathcal{K}^{(2)}\{n_i; C_{ij}\}$ in the following fashion (Appendix B)

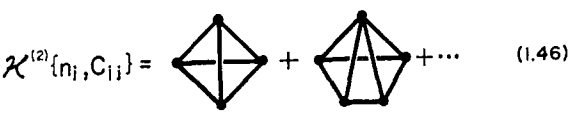
$$n(\mathbf{r}_1)n(\mathbf{r}_2)N\{\mathbf{r}_1; \mathbf{r}_2; n_i; C_{ij}\} = 2[\delta/\delta C(\mathbf{r}_1; \mathbf{r}_2)]\mathfrak{N}\{n_i; C_{ij}\} \quad (1.43)$$

$$n(\mathbf{r}_1)n(\mathbf{r}_2)K^{(2)}\{\mathbf{r}_1; \mathbf{r}_2; n_i; C_{ij}\} = 2[\delta/\delta C(\mathbf{r}_1; \mathbf{r}_2)]\mathcal{K}^{(2)}\{n_i; C_{ij}\}. \quad (1.44)$$

Diagramatically we have



(1.45)



(1.46)

$\mathfrak{N}$ is represented by polygonal diagrams (with alternate signs) and $\mathcal{K}^{(2)}$ by 2-irreducible diagrams, built with heavy dots and lines (which represent the most general insertions n_i or C_{ij}), and calculated with rule (C). Thus, $\mathcal{S} = \mathfrak{N} + \mathcal{K}^{(2)}$ can also be represented by a general (reducible or not) 0-diagram, calculated with rule (A), with *a polygonal or a 2-irreducible skeleton distinguished.* Let $N(\mathfrak{N}) + N(\mathcal{K}^{(2)})$ be the number of ways this choice can be done, it is also the extra weight these diagrams carry as compared to the contribution of the same diagrams to $\ln Z(\alpha, \beta)$.

(c) Consider a general (reducible or not) 0-diagram where we now *distinguish one pair of points at* $\mathbf{r}_1$, $\mathbf{r}_2$ and *one simple part rooted at* $\mathbf{r}_1$, $\mathbf{r}_2$. These diagrams may be generated from a skeleton nodal or 2-irreducible 2-diagram, rooted at $\mathbf{r}_1$, $\mathbf{r}_2$, by making the most general insertion at all points and pairs except $\mathbf{r}_1$, $\mathbf{r}_2$ (which provides a factor $N_{12}\{n_i, C_{ij}\} + K_{12}^{(2)}\{n_i, C_{ij}\} = S_{12}\{n_i, C_{ij}\}$) and at the roots yielding

$$\int d\mathbf{r}_1\, d\mathbf{r}_2\, n(\mathbf{r}_1)n(\mathbf{r}_2)[1 + C(\mathbf{r}_1; \mathbf{r}_2)] \times S\{\mathbf{r}_1; \mathbf{r}_2; n_i C_{ij}\}. \quad (1.47)$$

Let $N(\pi \cdot \mu)$ be, in a diagram, the number of articulation pairs weighted by their multiplicity (an ordinary pair being viewed as an articulation pair with $\mu = 1$). $N(\pi \cdot \mu)$ is also the number of ways one can choose a simple part rooted at $\mathbf{r}_1$, $\mathbf{r}_2$ in the diagram, and a 0-diagram contributing to (1.47) thus carries an extra weight $N(\pi \cdot \mu)$ as compared to with its contribution to $\ln Z(\alpha, \beta)$.

The introduced weights can be shown (Appendix C) to satisfy a relation valid for any 0-diagrams

$$N(\mathcal{K}^{(1)}) = \tfrac{1}{2}[N(\pi) - N(\pi \cdot \mu)] + N(\mathfrak{N}) + N(\mathcal{K}^{(2)}). \quad (1.48)$$

With (1.13) we get

$$N(p) - N(p \cdot m) + \tfrac{1}{2}[N(\pi) - N(\pi \cdot \mu)] + N(\mathfrak{N}) + N(\mathcal{K}^{(2)}) = +1, \quad (1.49)$$

which allows us to build the proper weight for

0-diagrams contributing to $\ln Z(\alpha, \beta)$ as

$$\ln Z(\alpha, \beta) = \int d\mathbf{r}_1\, n(\mathbf{r}_1) - \int d\mathbf{r}_1\, n(\mathbf{r}_1) K^{(1)}\{\mathbf{r}_1; n_i\} + \tfrac{1}{2}\int d\mathbf{r}_1\, d\mathbf{r}_2\, n(\mathbf{r}_1) n(\mathbf{r}_2)[C(\mathbf{r}_1; \mathbf{r}_2) - [1 + C(\mathbf{r}_1; \mathbf{r}_2)] S\{\mathbf{r}_1; \mathbf{r}_2; n_i; C_{ij}\}] + \mathbb{S}\{n_i, C_{ij}\}. \quad (1.50)$$

Expressed in another form, we see that relation (1.48) is equivalent to

$$\mathcal{K}^{(1)}\{n_i\} = \tfrac{1}{2}\int d\mathbf{r}_1\, d\mathbf{r}_2\, n(\mathbf{r}_1) n(\mathbf{r}_2)[C(\mathbf{r}_1; \mathbf{r}_2) - [1 + C(\mathbf{r}_1; \mathbf{r}_2)] S\{\mathbf{r}_1, \mathbf{r}_2; n_i C_{ij}\}] + \mathbb{S}\{n_i C_{ij}\}, \quad (1.51)$$

where $C(\mathbf{r}_1; \mathbf{r}_2)$ can be considered as a functional of n_i through repeated iterations of (1.35), which we rewrite as

$$1 + C(\mathbf{r}_1; \mathbf{r}_2) = [1 + g(\mathbf{r}_1 - \mathbf{r}_2)] \times \exp\left(S\{\mathbf{r}_1, \mathbf{r}_2; n_i, C_{ij}\}\right). \quad (1.52)$$

Using this expression and (1.14) to eliminate S_{12} and $K_1^{(1)}$, respectively, from (1.50), we get

$$\ln Z(\alpha, \beta) = \int d\mathbf{r}_1\, n(\mathbf{r}_1)[1 - \ln n(\mathbf{r}_1) + \alpha - \ln \lambda^3 - \beta u(\mathbf{r}_1)] + \tfrac{1}{2}\int d\mathbf{r}_1\, d\mathbf{r}_2\, n(\mathbf{r}_1) n(\mathbf{r}_2)\Big[C(\mathbf{r}_1; \mathbf{r}_2) - [1 + C(\mathbf{r}_1; \mathbf{r}_2)] \ln \frac{1 + C(\mathbf{r}_1; \mathbf{r}_2)}{1 + g(\mathbf{r}_1 - \mathbf{r}_2)}\Big] + \mathbb{S}\{n_i; C_{ij}\}. \quad (1.53)$$

If we introduce the one- and two-particle density in phase space $\mu(\mathbf{r}_1, \mathbf{p}_1)$ and

$$\mu_{12} = \mu(\mathbf{r}_1, \mathbf{p}_1; \mathbf{r}_2, \mathbf{p}_2) = \mu(\mathbf{r}_1, \mathbf{p}_1)\mu(\mathbf{r}_2, \mathbf{p}_2)[1 + C(\mathbf{r}_1; \mathbf{r}_2)], \quad (1.54)$$

we get

$$\ln Z(\alpha, \beta) = \int d1\, \mu_1(1 - \ln \mu_1) + \tfrac{1}{2}\int d1\, d2\, \mu_1\mu_2[C_{12} - (1 + C_{12}) \ln (1 + C_{12})] + \mathbb{S}\{n_i, C_{ij}\} + \int d1\, \mu_1(\alpha - \beta \mathbf{p}_1^2 - \beta\mu_1) - (\beta/2)\int d1\, d2\, u_{12} v_{12}. \quad (1.55)$$

d. Variational Properties

Discussion of the stationarity properties of the right-hand side Υ of (1.50) could be carried out in very much the same way as it was done for (1.25) in Sec. I(i). Here it would be necessary to consider variations of the quantity $\bar{S}_{ij}$, $\Upsilon\{\bar{n}_i, \bar{S}_{ij}\}$ being expressed in terms of $\bar{n}_i$ and $\bar{S}_{ij}$ through relation (1.52). Again the stationarity property

$$[\delta/\delta\bar{S}(\mathbf{r}_1; \mathbf{r}_2)]\Upsilon\{\bar{n}_i, \bar{S}_{ij}\} = 0 \quad (1.56)$$

could be traced to the structure of the diagrams which entails (1.49).

It is more convenient to consider the right-hand side of (1.53) as a functional $\Omega\{\bar{n}_i, \bar{C}_{ij}\}$ of $\bar{n}_i$ and $\bar{C}_{ij}$. If we write the particular values n_i and $\bar{C}_{ij}$ of these functions which render Ω stationary with respect to *independent* variations δn_i and δC_{ij} we obtain

$$n(\mathbf{r}_1) = n^0(\mathbf{r}_1) \exp\left\{\int d\mathbf{r}_2\, n(\mathbf{r}_2)\Big[C(\mathbf{r}_1, \mathbf{r}_2) - [1 + C(\mathbf{r}_1, \mathbf{r}_2)] \ln \frac{1 + C(\mathbf{r}_1, \mathbf{r}_2)}{1 + g(\mathbf{r}_1 - \mathbf{r}_2)}\Big] + [\delta/\delta n(\mathbf{r}_1)]\, \mathbb{S}\{n_i, C_{ij}\}\right\} \quad (1.57)$$

$$\frac{1 + C(\mathbf{r}_1, \mathbf{r}_2)}{1 + g(\mathbf{r}_1 - \mathbf{r}_2)} = \exp\left(2\,\frac{\delta}{\delta C(\mathbf{r}_1; \mathbf{r}_2)}\, \mathbb{S}\{n_i, C_{ij}\}\right); \quad (1.58)$$

(1.57) is identical to the expression of $n(\mathbf{r}_1)$ in terms of n_i and C_{ij} obtained with (1.14), (1.17), and (1.50); (1.58) is identical to (1.52). The stationary value of the functional itself

$$\Omega\{n_i, C_{ij}\} = \ln Z(\alpha, \beta). \quad (1.59)$$

We determine now the sign of the second functional derivatives. From expression (1.26) we get

$$\frac{\delta^2}{\delta n(\mathbf{r}_1)\, \delta n(\mathbf{r}_2)}\Psi\{n_i\} = -\frac{1}{n(\mathbf{r}_1)}\,\delta(\mathbf{r}_1 - \mathbf{r}_2) + X(\mathbf{r}_1, \mathbf{r}_2). \quad (1.60)$$

Here $X(\mathbf{r}_1, \mathbf{r}_2)$ is the second functional derivative of $\mathcal{K}^{(1)}\{n_i\}$, i.e., it is represented by the 1-irreducible, non-nodal, 2-diagrams, and hence identical with (1.36). From (1.36) and (1.38) we get an integral equation satisfied by $X(\mathbf{r}_1, \mathbf{r}_2)$ (Yvon[14]),

$$X(\mathbf{r}_1, \mathbf{r}_2) = C(\mathbf{r}_1, \mathbf{r}_2) - \int d\mathbf{r}_3\, X(\mathbf{r}_1, \mathbf{r}_3) n(\mathbf{r}_3) C(\mathbf{r}_3, \mathbf{r}_2). \quad (1.61)$$

On the other hand, the number of particles fluctuations, an essentially positive definite matrix is given by

$$\varphi(\mathbf{r}_1, \mathbf{r}_2) = \left\langle \left(\sum_i^N \delta(\mathbf{r}_1 - \mathbf{r}_i) \right) \left(\sum_j^N \delta(\mathbf{r}_2 - \mathbf{r}_j) \right) \right\rangle - \left\langle \sum_i^N \delta(\mathbf{r}_1 - \mathbf{r}_i) \right\rangle \left\langle \sum_j^N \delta(\mathbf{r}_2 - \mathbf{r}_j) \right\rangle , \quad (1.62)$$

where the average value is over the grand canonical ensemble, i.e.,

$$\varphi(\mathbf{r}_1, \mathbf{r}_2) = n(\mathbf{r}_1)\ \delta(\mathbf{r}_1 - \mathbf{r}_2) + n(\mathbf{r}_1) n(\mathbf{r}_2) C(\mathbf{r}_1, \mathbf{r}_2). \quad (1.63)$$

It is a straightforward matter to verify then, using (1.61), that the second functional derivative (1.60) is the inverse matrix of $-\varphi(\mathbf{r}_1, \mathbf{r}_2)$, and hence a negative definite matrix. The extremum of $\Psi\{\bar{n}_i\}$, where it takes the value $\ln Z(\alpha, \beta)$, is thus a *maximum.*

The second-order variation of the functional $\Omega\{\bar{n}_i, \bar{C}_{ij}\}$, around the extremum $\bar{n}_i = n_i, \bar{C}_{ij} = C_{ij}$, is also shown (Appendix D) to be a negative definite quadratic form in $\delta\bar{n}_i$ and $\delta\bar{C}_{ij}$.

e. Entropy. Generalization to Many-Body Forces

The entropy of the grand canonical ensemble has a particularly interesting form. From (1.55) we have

$$S = \ln Z(\alpha, \beta) - \alpha\langle N\rangle + \beta\langle H\rangle = \int d\mathbf{1}\ \mu_1(1 - \ln \mu_1) + \tfrac{1}{2} \int d\mathbf{1}\, d\mathbf{2}\ \mu_1\mu_2[C_{12} - (1 + C_{12}) \ln (1 + C_{12})] + \mathfrak{N} + \mathfrak{K}^{(2)}. \quad (1.64)$$

Here $\mathfrak{N}$ and $\mathfrak{K}^{(2)}$ are calculated with the series (1.45, 46) and rule (C), where $n(\mathbf{r}_i)$ is replaced by $\mu(\mathbf{r}_i, \mathbf{p}_i)$. Again we may consider the right-hand side of (1.64) as a functional $\Xi\{\bar{\mu}_i, \bar{C}_{ij}\}$. If we variate $\bar{\mu}_i$ and $\bar{C}_{ij}$ with the constraints

$$\langle N\rangle = \int d\mathbf{1}\ \mu_1 \quad (1.65)$$

$$\langle H\rangle = \int d\mathbf{1}\ \mu_1(p_1^2 + u_1) + \tfrac{1}{2} \int d\mathbf{1}\, d\mathbf{2}\ \mu_1\mu_2(1 + C_{12})v_{12}, \quad (1.66)$$

Ξ has a maximum (equal to the entropy) for the values of $\mu(\mathbf{r}_1, \mathbf{p}_1)$ and $C(\mathbf{r}_1, \mathbf{r}_2)$ obtained from (1.57) and (1.58).

The potentials u_i and v_{ij} no longer appear explicitly in $\Xi\{\bar{\mu}_i, \bar{C}_{ij}\}$; perhaps more significant is the fact that *the equilibrium parameters α and β have completely disappeared.* Only the maximum value of Ξ, through the constraints (1.65, 66) and the Lagrange multipliers, will introduce explicitly those quantities.

This raises the interesting question: Is it possible to give a physical interpretation to $\Xi\{\bar{\mu}_i, \bar{C}_{ij}\}$ outside its maximum? More precisely, *is it possible to identify $-(\Xi/k)$ with a generalized Boltzmann H function,* for a system outside equilibrium, in its kinetic stage? We do not know, as yet, the answer to this question. Assuming that $\bar{\mu}_i$ and $\bar{C}_{ij}$ depend upon the time t, the time derivative of Ξ is written as

$$-\frac{d\Xi}{dt} = \int d\mathbf{1} \ln \left[\bar{\mu}_1 \exp \left(-\frac{\delta}{\delta\bar{\mu}_1} S\{\bar{\mu}_i, \bar{C}_{ij}\} \right) \right] \frac{d}{dt} \bar{\mu}_1 + \tfrac{1}{2} \int d\mathbf{1}\, d\mathbf{2} \ln [(1 + \bar{C}_{12}) \times \exp (-S_{12}\{\bar{\mu}_i, \bar{C}_{ij}\})] \frac{d}{dt} [\bar{\mu}_1\bar{\mu}_2(1 + \bar{C}_{12})] \quad (1.67)$$

and may suggest evolution equations for $\bar{\mu}(\mathbf{r}_1, \mathbf{p}_1, t)$ and $\bar{C}(\mathbf{r}_1, \mathbf{r}_2, t)$ that imply $(-d\Xi/dt) \geq 0$.

The introduction of 3-body forces in the Hamiltonian would render the entropy expression more complicated and β dependent. However it is possible to further refine the analysis of the diagrams by considering triplets of points, articulation triplets, 3-irreducible diagrams and by the same method to recast Eq. (1.67) into the following form, now β independent:

$$S = \int d\mathbf{1}\ \mu_1(1 - \ln \mu_1) + \frac{1}{2!} \int d\mathbf{1}\, d\mathbf{2}\ \mu_1\mu_2 \times [C_{12} - (1 + C_{12}) \ln (1 + C_{12})] + \frac{1}{3!} \int d\mathbf{1}\, d\mathbf{2}\, d\mathbf{3}\ \mu_1\mu_2\mu_3 \times (1 + C_{12})(1 + C_{23})(1 + C_{31}) \times [\Gamma_{123} - (1 + \Gamma_{123}) \ln (1 + \Gamma_{123})] + \mathfrak{J}\{n_i, C_{ij}, \Gamma_{ijk}\}, \quad (1.68)$$

here we have defined the 3-particle density in phase space as

$$\mu_{123} = \mu_1\mu_2\mu_3(1 + C_{12}) \times (1 + C_{23})(1 + C_{31})(1 + \Gamma_{123}), \quad (1.69)$$

$\mathfrak{J}$ contains two classes of diagrams, one having some similarity with the polygonal diagrams $\mathfrak{N}$, the other being mainly made of 3-irreducible diagrams.

II. QUANTUM SYSTEMS

(i) Formulation in Terms of the One-Particle Density

We now want to extend to quantum systems the results established so far. Among the various possible formulations of quantum statistical mechanics we shall consider a formulation of Lee and Yang,[1]

which, besides its simplicity, remains closer to the classical formulation. At the end we shall also quote the corresponding results for the Green's function formulation used by Luttinger and Ward.[4]

In this section we use the same procedure as Sec. I(i) to rederive Lee and Yang's results, expressing the grand partition function as a stationary functional of a quantity in direct relation to the one-particle density. In the next section, the method is extended to produce a stationary expression of quantities directly related to the one and two particle densities. We have tried to keep the same symbols as in the classical case, for equivalent quantities.

a. Definitions. Lee-Yang Diagrams

We start from the Lee-Yang description of statistical mechanics of translationally invariant systems in terms of the (contracted) ζ-diagrams (L-Y Sec. IV), which, for short, we call ζ-diagrams. Such diagrams are built up with α-vertices connected by lines bearing an arrow. An α-vertex ($\alpha \geq 2$) connects α incoming lines and α outgoing lines. A ζ-diagram has ζ incoming and ζ outgoing external lines. We shall only be concerned with *connected* diagrams. We recall the rule giving the algebraic quantity associated with a labeled diagram (rule A). Each of the N internal lines bears an integer i and a corresponding momentum $\mathbf{k}_i$, each external line (if any) bears a pregiven momentum:

(i) associate to each internal line i a factor

$$m(\mathbf{k}_i) = z[1 - \epsilon z \exp(-\beta \mathbf{k}_i^2)]^{-1}, \tag{2.1}$$

where $\epsilon = +1$ or -1 for Bose or Fermi systems' respectively.

(ii) associate to each α-vertex, a factor

$$\langle \mathbf{k}_{B_1} \cdots \mathbf{k}_{B_\alpha} | \Upsilon_\alpha | \mathbf{k}_{A_1} \cdots \mathbf{k}_{A_\alpha} \rangle$$

where $\mathbf{k}_A$ and $\mathbf{k}_B$ are momenta associated with the incoming and outgoing lines of the vertex. This factor can be said to describe the evolution of α particles between the values 0 and β of the inverse of the temperature, as if the rest of the medium were absent. It is defined in more detail for Bose or Fermi systems in L-Y.

(iii) sum over all internal momenta with a weight $(N!)^{-1}$ (for *labeled* diagrams) or a weight $(s_\zeta)^{-1}$ (for *free* diagrams[21]).

(iv) assign a factor ϵ^P where P is the order of the permutation

$$\mathbf{k}_{A_1} \to \mathbf{k}_{B_1}, \cdots \mathbf{k}_{A_\alpha} \to \mathbf{k}_{B_\alpha}, \cdots$$

of all the initial coordinates into the final coordinates of all the vertex functions Υ_α taken together.

The grand partition function defined as

$$Z(\alpha, \beta) = \mathrm{Tr} \exp(\alpha \mathbf{N} - \beta \mathbf{H}), \tag{2.2}$$

where $\mathbf{N}$ and $\mathbf{H}$ are the number and the Hamiltonian operators, is then expressed as

$$\ln Z(\alpha, \beta) = \sum_{\mathbf{k}_1} \epsilon \ln [z^{-1} m(\mathbf{k}_1)] + \sum (\text{all distinct 0-diagrams}). \tag{2.3}$$

The *one-particle density (diagonal) matrix* is defined as

$$\langle \mathbf{k}_1' | n_1 | \mathbf{k}_1 \rangle = \delta_{\mathbf{k}_1', \mathbf{k}_1} n(\mathbf{k}_1) = Z^{-1}(\alpha, \beta) \mathrm{Tr} \exp(\alpha \mathbf{N} - \beta \mathbf{H}) a_{\mathbf{k}_1'}^+ a_{\mathbf{k}_1}, \tag{2.4}$$

where $a_{\mathbf{k}_1}^+$, $a_{\mathbf{k}_1}$ are the creation and annihilation operators for a particle in a state defined by its momentum $\mathbf{k}_1$ ($\hbar = 1$). It is conveniently expressed in terms of

$$M(\mathbf{k}_1) = z[1 + \epsilon n(\mathbf{k}_1)] \tag{2.5}$$

$$M(\mathbf{k}_1) = m(\mathbf{k}_1) + m^2(\mathbf{k}_1) \sum (\text{all distinct, 1-diagrams}). \tag{2.6}$$

Equations (2.1) to (2.6) are introduced or derived in L-Y and we shall consider them our starting point. We also introduce the *two-particle density matrix* as

$$\langle \mathbf{k}_1' \mathbf{k}_2' | n_{12} | \mathbf{k}_1 \mathbf{k}_2 \rangle = Z^{-1}(\alpha, \beta) \times \mathrm{Tr} \exp(\alpha \mathbf{N} - \beta \mathbf{H}) a_{\mathbf{k}_1'}^+ a_{\mathbf{k}_2'}^+ a_{\mathbf{k}_2} a_{\mathbf{k}_1} \tag{2.6}$$

and express it in terms of Lee-Yang 2-diagrams

$$\langle \mathbf{k}_1' \mathbf{k}_2' | n_{12} | \mathbf{k}_1 \mathbf{k}_2 \rangle = (\delta_{\mathbf{k}_1', \mathbf{k}_1} \delta_{\mathbf{k}_2', \mathbf{k}_2} + \epsilon \delta_{\mathbf{k}_2', \mathbf{k}_1} \delta_{\mathbf{k}_1', \mathbf{k}_2}) n(\mathbf{k}_1) n(\mathbf{k}_2) + z^{-2} m(\mathbf{k}_1') m(\mathbf{k}_2') m(\mathbf{k}_1) m(\mathbf{k}_2) \times \sum (\text{all distinct, 2-diagrams}). \tag{2.7}$$

To analyze the diagrams, the following definitions are introduced:

An *articulation lace* $\mathbf{k}_1$ of multiplicity $m (m > 1$; also called necklace in reference 5) is a sequence of lines bearing the same momentum $\mathbf{k}_1$ (by virtue of momentum conservation) in a ζ-diagram. Removal of one articulation lace splits the diagram into m disconnected parts. A *lace* $\mathbf{k}_1$ is meant to be either an articulation lace $\mathbf{k}_1 (m > 1)$ or a nonrepeated line $\mathbf{k}_1$.

A *1-irreducible ζ-diagram* contains no articulation lace. $\mathfrak{K}^{(1)}\{m_i\}$ is the class of 1-irreducible 0-dia-

[21] Compare with footnote 16.

grams; the class of 1-irreducible 1-diagrams is then represented by

$$K^{(1)}\{\mathbf{k}_1; m_i\} = [\delta/\delta m(\mathbf{k}_1)]\mathcal{K}^{(1)}\{m_i\}. \quad (2.8)$$

b. Densities

One-particle density: The contribution to $M(\mathbf{k}_1)$ of the 1-diagrams where only $\mathbf{k}_1$ is allowed to be an articulation lace, is given by

$$m(\mathbf{k}_1)\left[1 + \sum_{m=1}^{\infty} [\epsilon K^{(1)}\{\mathbf{k}_1; m_i\} m(\mathbf{k}_1)]^m\right], \quad (2.9)$$

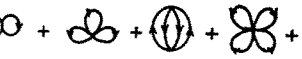
where m is the mulitplicity of the lace. The general 1-diagram is generated by making the most general insertion at all lines other than $\mathbf{k}_1$ with the effect of replacing $m(\mathbf{k}_1)$ by $M(\mathbf{k}_1)$.

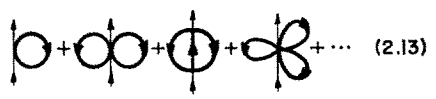
$$M(\mathbf{k}_1) = m(\mathbf{k}_1) + \epsilon m(\mathbf{k}_1) K^{(1)}\{\mathbf{k}_1; M_i\} M(\mathbf{k}_1) \quad (2.10)$$

or

$$M(\mathbf{k}_1) = m(\mathbf{k}_1)(1 - \epsilon m(\mathbf{k}_1) K^{(1)}\{\mathbf{k}_1; M_i\})^{-1}. \quad (2.11)$$

Relation (2.11) between $M(\mathbf{k}_1)$ and $K^{(1)}(\mathbf{k}_1)$ is homographic instead of exponential as in (1.14) because there is an ordering imposed at an articulation lace which does not exist at an articulation point. $\mathcal{K}^{(1)}\{M_i\}$ or $K^{(1)}\{\mathbf{k}_1; M_i\}$ is given by the sum of 1-irreducible 0- or 1-diagrams

$\mathcal{K}^{(1)}_{\{M_i\}}$ =  (2.12)

$K^{(1)}\{k_1; M_i\}$ =  (2.13)

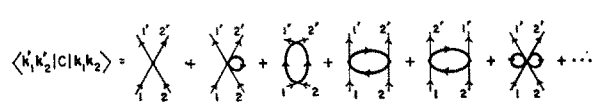
calculated with rule (B) which differs from rule (A) through point (i):

B(i) to each internal line i (drawn with a heavy line) associate a factor M_i.

Two-particle density: We have

$$z^2[\langle \mathbf{k}_1'\mathbf{k}_2' | n_{12} | \mathbf{k}_1\mathbf{k}_2\rangle - (\delta_{\mathbf{k}'_1\mathbf{k}_1}\delta_{\mathbf{k}'_2\mathbf{k}_2} + \epsilon\,\delta_{\mathbf{k}'_2\mathbf{k}_1}\delta_{\mathbf{k}'_1\mathbf{k}_2}) n(\mathbf{k}_1) n(\mathbf{k}_2)]$$
$$= M(\mathbf{k}_1')M(\mathbf{k}_2')\langle \mathbf{k}_1'\mathbf{k}_2' | C | \mathbf{k}_1\mathbf{k}_2\rangle M(\mathbf{k}_1)M(\mathbf{k}_2), \quad (2.14)$$

where the matrix $C_{12}\{M_i\}$ is represented by the sum of the 1-irreducible 2-diagrams

$\langle k_1'k_2'|C|k_1k_2\rangle$ =  (2.15)

calculated with rule (B).

c. Grand Partition Function

Instead of computing $\ln Z(\alpha, \beta)$ by direct integration of $\sum_{\mathbf{k}_1} n(\mathbf{k}_1)$ over α, we consider three expressions which are also represented by 0-diagrams (with some element distinguished) and expressed in terms of $M(\mathbf{k}_1)$.

(a) A general 0-diagram with *one lace* $\mathbf{k}_1$ *distinguished*, carries an extra weight $N(l)$ equal to its number of laces. The contribution of such diagrams is[22]

$$\sum_{m=1}^{\infty} \frac{\epsilon}{m} [\epsilon K^{(1)}\{\mathbf{k}_1; M_i\} m(\mathbf{k}_1)]^m, \quad (2.16)$$

where m is the multiplicity of the lace.

(b) A general 0-diagram with a *1-irreducible part distinguished* has an extra weight $N(\mathcal{K}^{(1)})$. This number can be obtained in the following way: Suppress an articulation lace $\mathbf{k}_1$, close each one of these m parts obtained by a $\mathbf{k}_1$ line, and repeat the operation inside each one of these m parts; the total number of 1-irreducible parts obtained is $N(\mathcal{K}^{(1)})$. Such diagrams are generated from $\mathcal{K}^{(1)}\{m_i\}$ by making the most general insertion in all lines yielding

$$\mathcal{K}^{(1)}\{M_i\}. \quad (2.17)$$

(c) Consider a general 0-diagram with one single line $\mathbf{k}_1$ distinguished, or equivalently with *one lace* $\mathbf{k}_1$ *and one 1-irreducible part* $K^{(1)}(\mathbf{k}_1)$ distinguished in the lace. Such a 0-diagram carries an extra weight $N(l \cdot m)$, which is the number of laces weighted by their multiplicity (and also the number of lines). Since such diagrams are generated by starting from a 1-irreducible part, making the most general insertion in all laces but $\mathbf{k}_1$ (i.e., $K^{(1)}\{\mathbf{k}_1; M_i\}$) and then in the lace $\mathbf{k}_1$ (i.e., $M(\mathbf{k}_1)$), their contribution is

$$\sum_{\mathbf{k}_1} K^{(1)}\{\mathbf{k}_1; M_i\} M(\mathbf{k}_1). \quad (2.18)$$

These weights satisfy a simple relation which is immediately proven by induction and which follows from the tree structure of the diagrams

$$N(l) - N(l \cdot m) + N(\mathcal{K}^{(1)}) = +1. \quad (2.19)$$

Such a relation, together with Eqs. (2.16) to (2.18), implies

$$\ln Z(\alpha, \beta) - \sum_{\mathbf{k}_1} \epsilon \ln (z^{-1} m(\mathbf{k}_1))$$
$$= -\sum_{\mathbf{k}_1} \epsilon \ln [1 - \epsilon K^{(1)}\{\mathbf{k}_1; M_i\} m(\mathbf{k}_1)]$$
$$- \sum_{\mathbf{k}_1} K^{(1)}\{\mathbf{k}_1; M_i\} M(\mathbf{k}_1) + \mathcal{K}^{(1)}\{M_i\}. \quad (2.20)$$

[22] A diagram made of m 1-irreducible parts $K^{(1)}(\mathbf{k}_1)$ connected by an articulation lace $\mathbf{k}_1$ is invariant under circular permutation of these m parts giving rise to a symmetry number $s = m$.

Using relations (2.3) and (2.11) one gets [L-Y. Eq. (IV.32)]

$$\ln Z(\alpha, \beta) = \sum_{\mathbf{k}_1} \epsilon \ln (z^{-1} M(\mathbf{k}_1)) - \sum_{\mathbf{k}_1} \epsilon [M(\mathbf{k}_1) - m(\mathbf{k}_1)][m(\mathbf{k}_1)]^{-1} + \mathcal{K}^{(1)}\{M_i\}. \quad (2.21)$$

The right-hand side Φ of (2.20), considered through (2.11) as a functional of $\bar{K}_i^{(1)}$ [or the right-hand side Ψ of (2.21) as a functional of $\bar{M}_i$], is stationary under variations $\delta\bar{K}_i^{(1)}$ (or $\delta\bar{M}_i$), a property which could be traced to relation (2.19).

Using exactly the same method and argument as in the classical case [Sec. I(ii), the relevant matrix X_{12} is defined in the following section] it is shown that the functional $\Psi\{\bar{M}_i\}$ has a negative second derivative[1] at its extremum where $\bar{M}_i = M_i$. It is to be remembered however that the proof of Appendix D does not apply as such to quantum systems; indeed, second functional derivatives of $\ln Z(\alpha, \beta)$ with respect to potentials do not necessarily behave like fluctuations, i.e., have no obvious negative definite character.

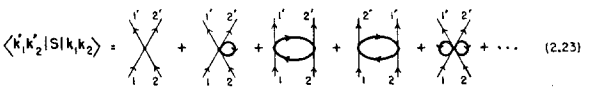
(ii) Formulation in Terms of the One- and Two-Particle Densities

a. Definitions

The grand partition function and the one- and two-particle densities are now expressed as functionals of M_i and represented by 1-irreducible ζ-diagrams calculated with rule (B) (in the following, unless mentioned, we only deal with 1-irreducible diagrams). We now introduce further definitions along the same lines as in Sec. I(ii).

Consider in a 0-diagram a pair of lines (i, j) which we may think of as being cut open: $\mathbf{k}_i$, $\mathbf{k}_j$ for the incoming pieces, $\mathbf{k}_i'$, $\mathbf{k}_j'$ for the outgoing pieces. Consider now all the possible pairs (i_1, j_1), (i_2, j_2) $\cdots$, which after being further cut open, lead to a partition of the original diagram into two parts; *one containing the incoming lines* $\mathbf{k}_i$, $\mathbf{k}_j$, *the other the outgoing lines* $\mathbf{k}_i'$, $\mathbf{k}_j'$. The set of μ pairs of lines $(i, j; i_1, j_1; \cdots i_{\mu-1}, j_{\mu-1})$ plays the same role as an articulation pair for the classical systems and is called an articulation sequence of pairs of laces or *articulation sequence*, of multiplicity μ. The same definition applies to a 2-diagram after the incoming external lines have been closed with the outgoing external lines yielding a 0-diagram.

A *simple 2-diagram* is such that the external lines considered as an opened pair are not part of an articulation sequence and the contribution of all the distinct, simple diagrams we represent by the matrix $\langle \mathbf{k}_1'\mathbf{k}_2' \,|S|\, \mathbf{k}_1\mathbf{k}_2\rangle$



(2.23)

Like C_{12}, the matrix S_{12} is a symmetrical function of $\mathbf{k}_1$, $\mathbf{k}_2$ (and of $\mathbf{k}_1'$, $\mathbf{k}_2'$).

b. Two-Particle Density Matrix

The contribution to $\langle \mathbf{k}_1'\mathbf{k}_2' \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle$ of diagrams where the opened pair $(\mathbf{k}_1'\mathbf{k}_2'; \mathbf{k}_1\mathbf{k}_2)$ may be part of an articulation sequence of increasing multiplicity, is written in short as (Appendix A)

$$S_{12}\left[1 + \sum_{\mu=1}^{\infty} (\tfrac{1}{2} S_{12} M_1 M_2)^{\mu}\right],$$

or equivalently, the matrix C_{12} is expressed as the solution of the integral equation

$$\langle \mathbf{k}_1'\mathbf{k}_2' \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle = \langle \mathbf{k}_1'\mathbf{k}_2' \,|S|\, \mathbf{k}_1\mathbf{k}_2\rangle + \tfrac{1}{2} \sum_{1_1 1_2} \langle \mathbf{k}_1'\mathbf{k}_2' \,|S|\, 1_1 1_2\rangle M(1_1) M(1_2) \langle 1_1 1_2 \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle. \quad (2.24)$$

Further definitions are needed to analyze S_{12}: A *nodal pair* in a 2-diagram, is a pair of internal lines (i, j) such that by cutting them open, the 2-diagram separates into two 2-diagrams *each one containing one incoming external line* ($\mathbf{k}_1$ or $\mathbf{k}_2$) *and outgoing external line* ($\mathbf{k}_1'$ or $\mathbf{k}_2'$). A *nodal diagram* contains at least one nodal pair. Simple 2-diagrams are classified into nodal and elementary (i.e., simple, non-nodal) diagrams.

Nodal diagrams: Let $\langle \mathbf{k}_1'\mathbf{k}_2' \,|n|\, \mathbf{k}_1\mathbf{k}_2\rangle$ be the contribution of those nodal 2-diagrams where by cutting a nodal pair one disconnects $(\mathbf{k}_1\mathbf{k}_1')$ from $(\mathbf{k}_2\mathbf{k}_2')$; the contribution of all nodal 2-diagrams is

$$\langle \mathbf{k}_1'\mathbf{k}_2' \,|N|\, \mathbf{k}_1\mathbf{k}_2\rangle = \langle \mathbf{k}_1'\mathbf{k}_2' \,|n|\, \mathbf{k}_1\mathbf{k}_2\rangle + \epsilon\langle \mathbf{k}_2'\mathbf{k}_1' \,|n|\, \mathbf{k}_1\mathbf{k}_2\rangle. \quad (2.25)$$

The matrix $\langle \mathbf{k}_1'\mathbf{k}_2' \,|n|\, \mathbf{k}_1\mathbf{k}_2\rangle$ is generated by iterating non-nodal parts described by the matrix

$$\langle \mathbf{k}_1'\mathbf{k}_2' \,|X|\, \mathbf{k}_1\mathbf{k}_2\rangle = \langle \mathbf{k}_1'\mathbf{k}_2' \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle - \langle \mathbf{k}_1'\mathbf{k}_2' \,|N|\, \mathbf{k}_1\mathbf{k}_2\rangle. \quad (2.26)$$

In the iteration these non-nodal parts are connected by nodal pairs of lines. N_{12} may be expressed as a solution of the integral equation

$$\langle \mathbf{k}_1\mathbf{k}_1' \,|N|\, \mathbf{k}_2\mathbf{k}_2'\rangle = \sum_{\mathbf{k}_3\mathbf{k}_3'} \langle \mathbf{k}_1\mathbf{k}_1' \,|X|\, \mathbf{k}_3\mathbf{k}_3'\rangle M(\mathbf{k}_3) M(\mathbf{k}_3') \langle \mathbf{k}_3\mathbf{k}_3' \,|X|\, \mathbf{k}_2\mathbf{k}_2'\rangle + \sum_{\mathbf{k}_3\mathbf{k}_3'} \langle \mathbf{k}_1\mathbf{k}_1' \,|X|\, \mathbf{k}_3\mathbf{k}_3'\rangle M(\mathbf{k}_3) M(\mathbf{k}_3') \langle \mathbf{k}_3\mathbf{k}_3' \,|N|\, \mathbf{k}_2\mathbf{k}_2'\rangle, \quad (2.27)$$

where we have

$$(\mathbf{k}_i\mathbf{k}_i' \mid \quad \mid \mathbf{k}_j\mathbf{k}_j') = \langle \mathbf{k}_i'\mathbf{k}_i' \mid \quad \mid \mathbf{k}_j\mathbf{k}_j\rangle; \qquad (2.28)$$

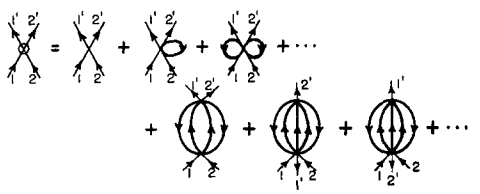
with (2.26) and (2.28) we can transform (2.27) into

$$(\mathbf{k}_1\mathbf{k}_1' \mid N \mid \mathbf{k}_2\mathbf{k}_2') = \sum_{\mathbf{k}_3\mathbf{k}_3'} (\mathbf{k}_1\mathbf{k}_1' \mid C \mid \mathbf{k}_3\mathbf{k}_3')M(\mathbf{k}_3)M(\mathbf{k}_3')(\mathbf{k}_3\mathbf{k}_3' \mid C \mid \mathbf{k}_2\mathbf{k}_2') - \sum_{\mathbf{k}_3\mathbf{k}_3'} (\mathbf{k}_1\mathbf{k}_1' \mid N \mid \mathbf{k}_3\mathbf{k}_3')M(\mathbf{k}_3)M(\mathbf{k}_3')(\mathbf{k}_3\mathbf{k}_3' \mid C \mid \mathbf{k}_2\mathbf{k}_2'), \qquad (2.29)$$

thereby expressing N_{12} in terms of C_{12}.

Elementary diagrams: Here the analysis differs from the classical case because 3, 4 $\cdots$ n-vertices may play the role of many-body interactions. We find convenient to distinguish two subclasses among the (simple, non-nodal) elementary 2-diagrams:

(*i*) *"2-vertex"-like 2-diagrams*: They contain no articulation sequence. Their classical equivalent reduces to the direct link between points (1, 2) a diagram which was excluded from the class of simple 2-diagrams. Their contribution is written as $\langle \mathbf{k}_1'\mathbf{k}_2' \mid V \mid \mathbf{k}_1\mathbf{k}_2\rangle$ and symbolized by a general 2-vertex

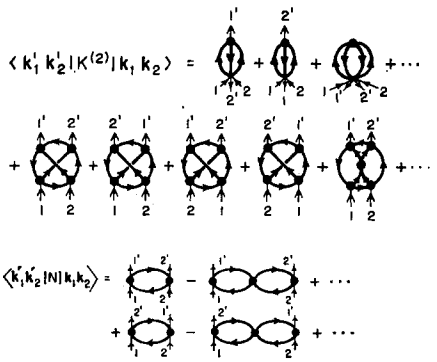


(2.30)

(*ii*) *2-irreducible 2-diagrams*: They contain no articulation sequence of multiplicity $\mu > 2$. Notice that, as simple 2-diagrams, their external lines (considered as an opened pair) cannot be part of an articulation sequence, even with $\mu = 2$. Classically this subtlety in the classification entered through the special treatment allowed to the direct (i, j) links.

These 2-diagrams contribute a matrix $\langle \mathbf{k}_1'\mathbf{k}_2' \mid K^{(2)} \mid \mathbf{k}_1\mathbf{k}_2\rangle$. They are built of parts playing the role of "effective" two-body forces and represented [Eq. (2.30)] as generalized 2-vertices and of parts corresponding to "effective" many-body forces.

Elementary 2-diagrams which do not belong to the "2-vertex" subclass are obtained from the 2-irreducible 2-diagrams by the following procedure: Each generalized 2-vertex part linking the four internal lines $\mathbf{k}_i'$, $\mathbf{k}_j'$ and $\mathbf{k}_i$, $\mathbf{k}_j$ is replaced by the most general insertion connecting these lines; algebraically it amounts to replace each factor $\langle \mathbf{k}_i'\mathbf{k}_j' \mid V \mid \mathbf{k}_i\mathbf{k}_j\rangle$ by $\langle \mathbf{k}_i'\mathbf{k}_j' \mid C \mid \mathbf{k}_i\mathbf{k}_j\rangle$ (Appendix A) which is represented as a heavy dot with two incoming $\mathbf{k}_i$, $\mathbf{k}_j$ and two outgoing $\mathbf{k}_i'$, $\mathbf{k}_j'$ lines. In terms of diagrams we have



(2.31)

(2.32)

Contributions associated with these 2-diagrams are calculated with rule (C) differing from rule (B) through point (ii):

C(ii) To each heavy dot 2-vertex, associate a matrix $\langle \mathbf{k}_i'\mathbf{k}_j' \mid C \mid \mathbf{k}_i\mathbf{k}_j\rangle$, where $\mathbf{k}_i'$, $\mathbf{k}_j'$ and $\mathbf{k}_i$, $\mathbf{k}_j$ are the momenta of the outgoing and incoming lines; to each α-vertex ($\alpha > 2$) associate a matrix Υ_α.

The result of this section is summarized by the symbolic matrix equations

$$C_{12} = (1 - \tfrac{1}{2}S_{12}M_1M_2)^{-1}S_{12}, \qquad (2.33)$$

$$S_{12}\{M_i, C_{ij}\} = N_{12}\{M_i, C_{ij}\} + V_{12}\{M_i\} + K_{12}^{(2)}\{M_i, C_{ij}\}. \qquad (2.34)$$

c. Grand Partition Function

Following the pattern used for the classical systems, we consider three expressions which are easily expressed as functionals of M_i and C_{ij} and represented by general 0-diagrams with one element distinguished.

(a) Consider a general 0-diagram with *one articulation sequence* (multiplicity μ) *distinguished*. As an iteration of simple parts its contribution, in terms of the matrix S_{12}, is written,[23] with obvious notations,

$$\mathrm{Tr}_{12}\left[S_{12}M_1M_2 + 2\sum_{\mu=2}^{\infty}\frac{1}{\mu}(\tfrac{1}{2}S_{12}M_1M_2)^\mu\right]; \qquad (2.35)$$

Compared with its contribution to $\ln Z(\alpha, \beta)$, here a 0-diagram carries an extra weight $N(\sigma)$ equal to its number of articulation sequences.

(b) Just as we introduced the functional $\mathfrak{K}^{(1)}\{M_i\}$, here we define functionals of the matrix C_{12}, the derivative of which restores the matrices N_{12}, V_{12}, and $K_{12}^{(2)}$, namely (Appendix B),

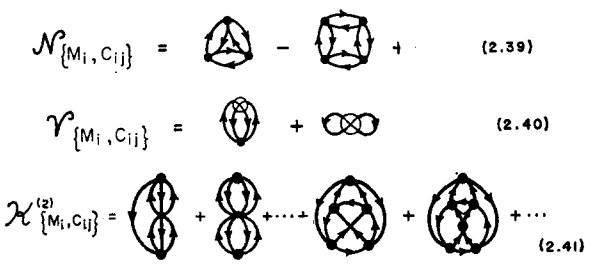
$$M(\mathbf{k}_1')M(\mathbf{k}_2')\langle \mathbf{k}_1'\mathbf{k}_2' \mid N \mid \mathbf{k}_1\mathbf{k}_2\rangle M(\mathbf{k}_1)M(\mathbf{k}_2) = 4[\delta/\delta\langle \mathbf{k}_1'\mathbf{k}_2' \mid C \mid \mathbf{k}_1\mathbf{k}_2\rangle]\mathfrak{K}\{M_i, C_{ij}\} \qquad (2.36)$$

[23] Compare with footnote 22 for the factor $(\mu)^{-1}$; see also Appendix A for factors 1/2.

$$M(\mathbf{k}_1')M(\mathbf{k}_2')\langle \mathbf{k}_1'\mathbf{k}_2' \,|V|\, \mathbf{k}_1\mathbf{k}_2\rangle M(\mathbf{k}_1)M(\mathbf{k}_2)$$
$$= 4[\delta/\delta\langle \mathbf{k}_1'\mathbf{k}_2' \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle]\mathcal{U}\{M_i, C_{ij}\} \quad (2.37)$$

$$M(\mathbf{k}_1')M(\mathbf{k}_2')\langle \mathbf{k}_1'\mathbf{k}_2' \,|K^{(2)}|\, \mathbf{k}_1\mathbf{k}_2\rangle M(\mathbf{k}_1)M(\mathbf{k}_2)$$
$$= 4[\delta/\delta\langle \mathbf{k}_1'\mathbf{k}_2' \,|C|\, \mathbf{k}_1\mathbf{k}_2\rangle]\mathcal{K}^{(2)}\{M_i, C_{ij}\}. \quad (2.38)$$

Diagrammatically we have



$\mathfrak{N}$ is represented by polygonal diagrams (with alternate signs), $\mathcal{K}^{(2)}$ by 2-irreducible diagrams (no articulation pair of multiplicity $\mu > 2$). $\mathcal{U}$ is defined as containing a term independent of the matrix C_{12}. Contributions of these diagrams are calculated with rule (C). The quantity $\mathcal{S} = \mathfrak{N} + \mathcal{U} + \mathcal{K}^{(2)}$ is again represented by the most general 0-diagram, since the heavy dots and lines are equivalent to the most general insertions, but now a *nodal, 2-vertex or 2-irreducible skeleton is distinguished.* A general 0-diagram thus carries a weight

$$N(\mathcal{S}) = N(\mathfrak{N}) + N(\mathcal{U}) + N(\mathcal{K}^{(2)}) \quad (2.42)$$

counting the number of ways it can be generated out of nodal, 2-vertex or 2-irreducible skeletons, respectively.

(c) Consider a general 0-diagram where we *distinguish one pair of laces,* or equivalently, one articulation sequence and one simple part (e.g., the simple part S_{12} which follows the pair of laces distinguished). Such a 0-diagram carries an extra weight $N(\sigma\cdot\mu)$ which is the number of articulation sequences weighted by their multiplicity, or also, the number of pair of laces (a pair of laces which is not part of an articulation sequence is counted with multiplicity $\mu = 1$). These diagrams may be generated by making insertions in the simple part ($S_{12}\{M_i, C_{ij}\}$) and in the distinguished pair of laces yielding

$$\mathrm{Tr}_{12}\, S_{12}(M_1M_2 + \tfrac{1}{2}M_1M_2C_{12}M_1M_2). \quad (2.43)$$

We can then immediately construct the grand partition function as a functional of M_i and C_{ij}. To that effect we use a second topological relationship valid for Lee-Yang diagrams (Appendix E) which relates the weights introduced in (a, b, c) to $N(\mathcal{K}^{(1)})$ (number of ways a 0-diagram may be generated out of 1-irreducible skeletons).

$$N(\mathcal{K}^{(1)}) = \tfrac{1}{2}[N(\sigma) - N(\sigma\cdot\mu)] + N(\mathfrak{N}) + N(\mathcal{U}) + N(\mathcal{K}^{(2)}). \quad (2.44)$$

Or equivalently, considering (2.19) we have

$$N(l) - N(l\cdot m) + \tfrac{1}{2}[N(\sigma) - N(\sigma\cdot\mu)] + N(\mathcal{S}) = +1. \quad (2.45)$$

Together with (2.20), (2.35), (2.39) to (2.41), and (2.43), this relation implies

$$\ln Z(\alpha, \beta) - \mathrm{Tr}_1[\epsilon \ln (z^{-1}m_1)] = \mathrm{Tr}_1[-\epsilon \ln (1 - \epsilon K_1^{(1)}m_1) - K_1^{(1)}M_1] + \mathrm{Tr}_{12}[-\epsilon \ln (1 - \tfrac{1}{2}S_{12}M_1M_2) - \tfrac{1}{2}S_{12}M_1M_2(1 + \tfrac{1}{2}C_{12}M_1M_2)] + \mathcal{S}\{M_i, C_{ij}\}, \quad (2.46)$$

$$\mathcal{S}\{M_i, C_{ij}\} = \mathfrak{N}\{M_i, C_{ij}\} + \mathcal{U}\{M_i, C_{ij}\} + \mathcal{K}^{(2)}\{M_i, C_{ij}\}. \quad (2.47)$$

Notice that (2.44) is equivalent to the relation

$$\mathcal{K}^{(1)}\{M_i\} = -\mathrm{Tr}_{12}[\ln (1 - \tfrac{1}{2}S_{12}\{M_i, C_{ij}\}M_1M_2) + \tfrac{1}{2}S_{12}\{M_i, C_{ij}\}M_1M_2\,(1 + \tfrac{1}{2}C_{12}M_1M_2)] + \mathcal{S}\{M_i, C_{ij}\}, \quad (2.48)$$

where in the right-hand side C_{ij} may be considered as a functional of M_i through repeated iterations of (2.26). Using (2.26) and (2.11) to eliminate $K_1^{(1)}$ and S_{12} from (2.47), we get the desired result:

$$\ln Z(\alpha, \beta) = \mathrm{Tr}_1[\epsilon \ln (z^{-1}M_1) - \epsilon(M_1/m_1 - 1)] + \mathrm{Tr}_{12}[\ln (1 + \tfrac{1}{2}C_{12}M_1M_2) - \tfrac{1}{2}C_{12}M_1M_2] + \mathcal{S}\{M_i, C_{ij}\}. \quad (2.49)$$

d. Variational Properties

The right-hand side of (2.47) can be verified to be stationary under changes of the functions $K_i^{(1)}$ and S_{ij} in a way traceable to the diagram structure as in Sec. I. We rather consider the right-hand side of (2.49) as a functional $\Omega\{\bar{M}_i, \bar{C}_{ij}\}$ of the matrices $\bar{M}_i$ and $\bar{C}_{ij}$. The particular values M_i, C_{ij} of those matrices which render stationary Ω under variations $\delta\bar{M}_i$, $\delta\bar{C}_{ij}$ satisfy then the following equations:

$$\epsilon(1/M_1 - 1/m_1) + \mathrm{Tr}_2[C_{12}M_2(1 + \tfrac{1}{2}C_{12}M_1M_2)^{-1} - C_{12}M_2] + (\delta/\delta M_1)\mathcal{S}\{M_i, C_{ij}\} = 0 \quad (2.50)$$

$$\tfrac{1}{4}M_1M_2C_{12}M_1M_2(1 + \tfrac{1}{2}C_{12}M_1M_2)^{-1} - (\delta/\delta C_{12})\mathcal{S}\{M_i, C_{ij}\} = 0. \quad (2.51)$$

Equation (2.50) together with (2.11) expresses $K_1^{(1)}$ as a functional of M_i and C_{ij}, an expression which could be also obtained from (2.48) and (2.26). Equation (2.51) is an immediate consequence of (2.16) and (2.39) to (2.41). For those particular values of $\bar{M}_i$, $\bar{C}_{ij}$ equal to the actual matrices M_i, C_{ij} immediately related to the one- and two-particle densities through (2.5) and (2.14), the functional Ω has an extremum equal to the grand partition function. The study of the quadratic form in $\delta\bar{M}_i$ and $\delta\bar{C}_{ij}$ representing the second-order variation of Ω is deferred to another paper.

(iii) Other Formulations. Conclusions

A stationary expression of $\ln Z(\alpha, \beta)$ in terms of the one-particle Green's function has been derived by Luttinger and Ward.[4] Analysis and proofs similar to the ones carried out above could be extended to this formalism resulting in a doubly stationary expression in terms of quantities directly related to the one- and two-particles densities. We shall just quote the results.

We define the one- and two-particles Green's functions in the Fourier series representation[2,4,24] and try to keep the same notations as before for equivalent quantities:

$$G_1^0 = (-\omega_1 + \mathbf{k}_1^2 - \alpha/\beta)^{-1} \quad (2.52)$$

$$G_1 = G_1^0 + G_1^0K_1^{(1)}G_1 \quad (2.53)$$

$$G_{12} = G_1G_2(\delta_{11'}\,\delta_{22'} + \epsilon\,\delta_{12'}\,\delta_{21'}) + \tfrac{1}{2}G_1G_2S_{12}G_{12}. \quad (2.54)$$

We have used shorthand notation G_1 for $G(\omega_1, \mathbf{k}_1)$, G_{12} for $\langle\omega_1'\mathbf{k}_1', \omega_2'\mathbf{k}_2'\,|G|\,\omega_1\mathbf{k}_1, \omega_2\mathbf{k}_2\rangle$ etc $\cdots$. The energy variable ω takes only discrete values

$$\omega = i\pi\beta^{-1}[2n + (1 - \epsilon)/2], \quad (2.55)$$

n is a positive or negative integer, $K_1^{(1)}$ is the mass operator, $\frac{1}{2}S_{12}$ the interaction operator. We introduce the correlation operator C_{12} by the equation

$$\tfrac{1}{2}S_{12}G_{12} = C_{12}G_1G_2, \quad (2.56)$$

which after taking into account Eq. (2.54) and the ϵ-symmetry of $\langle\omega_1\mathbf{k}_1,\ \omega_2\mathbf{k}_2^1\,|S|\,\omega_1\mathbf{k}_1,\ \omega_2\mathbf{k}_2\rangle$ under exchange of 1, 2 of 1′, 2′, may be written as

$$C_{12} = (1 - \tfrac{1}{2}S_{12}G_1G_2)^{-1}S_{12}. \quad (2.57)$$

Using a topological relation between various weights as in the previous sections leads to

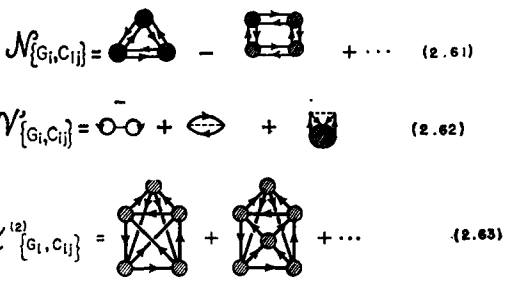
$$\ln Z(\alpha, \beta) = \mathrm{Tr}_1\,[\,\epsilon \ln (G_1) - G_1K_1^{(1)}] - \mathrm{Tr}_{12}\,[\ln (1 - \tfrac{1}{2}S_{12}G_1G_2) + \tfrac{1}{2}S_{12}(G_1G_2 + \tfrac{1}{2}G_1G_2C_{12}G_1G_2)] + \mathcal{S}\{G_i, C_{ij}\}, \quad (2.58)$$

with

$$G_1G_2S_{12}G_1G_2 = 4[\delta/\delta C_{12})\mathcal{S}\{G_i, C_{ij}\} \quad (2.59)$$

and

$$\mathcal{S}\{G_i, C_{ij}\} = \mathfrak{N}\{G_i, C_{ij}\} + \mathfrak{V}\{G_i, C_{ij}\} + \mathfrak{K}^{(2)}\{G_i, C_{ij}\} \quad (2.60)$$



Here the polygonal ($\mathfrak{N}$) and the 2-irreducible ($\mathfrak{K}^{(2)}$) diagrams are *Feynman* diagrams constructed with heavy lines (associated with G_i) and bubbles with two incoming and two outgoing lines (associated with matrices C_{ij}). The vertex-like diagrams ($\mathfrak{V}$) also contain a single dotted horizontal line (associated with the original two-body potential matrix $-(\beta/2)\langle\mathbf{k}_i'\mathbf{k}_j'\,|v|\,\mathbf{k}_i\mathbf{k}_j\rangle$). Algebraic contributions are calculated then by summing over momenta and energies (with conservation of these quantities through each vertex or bubble) as indicated elsewhere.[2,4] Equation (2.58) may also be rewritten as

$$\ln Z(\alpha, \beta) = \mathrm{Tr}_1\,\epsilon\,[\ln (G_1^{-1}) + (\omega_1 - \mathbf{k}_1^2 + \alpha/\beta)G_1] + \mathrm{Tr}_{12}\,[\ln (1 + \tfrac{1}{2}C_{12}G_1G_2) - \tfrac{1}{2}C_{12}G_1G_2] + \mathcal{S}\{G_i, C_{ij}\}, \quad (2.64)$$

an expression stationary with respect to changes in G_i and C_{ij}.

It is possible to eliminate the dependence upon the equilibrium parameters in the entropy functional through a scale transformation. If we write

$$\beta\omega_1 = \tilde{\omega}_1 = i\pi[2n + (1 - \epsilon)/2]$$
$$\beta^{-1}G_1 = \tilde{G}_1 \quad (2.65)$$
$$\beta^{-2}G_{12} = \tilde{G}_{12},$$

and also

[24] E. Montroll and J. Ward, Phys. Fluids **1**, 55 (1958).

$$\begin{aligned} \beta K_1^{(1)} &= \tilde{K}_1^{(1)} \\ \beta^2 S_{12} &= \tilde{S}_{12} \\ \beta^2 C_{12} &= \tilde{C}_{12}, \end{aligned} \tag{2.66}$$

we get for the entropy

$$\begin{aligned} S &= \mathrm{Tr}_1\, \epsilon\, [\ln(\tilde{G}_1) + \tilde{\omega}_1 \tilde{G}_1] \\ &+ \mathrm{Tr}_{12}\, [\ln(1 + \tfrac{1}{2}\tilde{C}_{12}\tilde{G}_1\tilde{G}_2) - \tfrac{1}{2}\tilde{C}_{12}\tilde{G}_1\tilde{G}_2] \\ &+ \mathfrak{N}\{\tilde{G}_i, \tilde{C}_{ij}\} + \mathfrak{K}^{(2)}\{\tilde{G}_i, \tilde{C}_{ij}\}. \end{aligned} \tag{2.67}$$

This expression is free of any reference to the equilibrium parameters (or the interactions) and is stationary with respect to variations of $\tilde{G}_1$, $\tilde{C}_{12}$ under the constraints of constant particle number and energy

$$\langle N \rangle = \mathrm{Tr}_1\, \tilde{G}_1 \tag{2.68}$$

$$\langle H \rangle = \mathrm{Tr}_1\, k_1^2 \tilde{G}_1 + \tfrac{1}{2}\,\mathrm{Tr}_{12}\, v_{12}\tilde{G}_{12}; \tag{2.69}$$

the one- and two-particle Green's functions then satisfy their equilibrium equations [$\beta^2\tilde{G}_{12}$ for example satisfies (2.54), (2.57), and (2.59) to (2.63). However, the second functional derivative of (2.67) with respect to $\tilde{G}_1$, $\tilde{C}_{12}$, is not a negative definite matrix[25] as contrasted with the classical case.

Doubly stationary expressions involving the average occupation number as in reference 5 and an average two-particle density matrix, are more involved to establish. They require extending the disentangling theorem[5] to operators involving two creation or two annihilation operators. Such a formulation would presumably furnish the closest analog to the classical result towards which it would tend term by term as $\hbar$ tends to zero. An extension of the disentangling theorem would also furnish a doubly stationary formulation in terms of densities of "quasi-particles".[3] Notice that the entropy as expressed in terms of "quasi-particle" occupation numbers, was formally identical to the entropy of "free" "quasi-particles," and hence independent of the equilibrium parameters (in contrast, the energy constraint contained a complicated functional). Moreover, "quasi-particle" occupation numbers are not directly related to physical quantities (except at zero temperature and at the Fermi surface) and one would not expect any positive definiteness character for the second functional derivatives.

In this work we have thus given an answer to the question raised in the introduction by explicitly writing the thermodynamical functions as doubly stationary functionals of the one- and two-particle densities (or quantities in direct relation to them). It is not clear whether the suggestive forms obtained for the entropy (or yet to be derived in the formulations mentioned just above) will lead to a generalization of Boltzmann's H function.

From a more practical point of view it is hoped that the doubly stationary expressions obtained will prove useful for the study of phase transitions of classical[10,11] and quantum systems. For example within the framework of Lee-Yang theory of superfluid transitions in bose systems, it would be necessary first to rewrite the above results in the x-ensemble formulation.[26] Superfluid transitions in a Fermi system would require, to take care of pair condensation, a generalization of the x-ensemble, which the doubly stationary formulations seem particularly suited to carry out.[27]

ACKNOWLEDGMENTS

The author is deeply indebted to Professor C. N. Yang for having brought the problem to his attention and for several illuminating discussions. He would like also to thank Professor Uhlenbeck for discussing the results of this work and Professor P. C. Martin for helpful criticism and an important remark in connection with Sec. II(iii). Finally, he wishes to express his gratitude to Professor J. R. Oppenheimer for the generous hospitality shown him during his stay at the Institute for Advanced Studies where this work was performed. Support from Harvard University and a leave of absence from the French Commissariat à l'Energie Atomique are gratefully acknowledged.

APPENDIX A. INSERTIONS

Classical systems: A general insertion made at points (i) or (i and j) on a skeleton diagram may be factorized and leads to the substitutions $n^0(\mathbf{r}_i) \to n(\mathbf{r}_i)$ or[28] $g(\mathbf{r}_i - \mathbf{r}_i) \to C(\mathbf{r}_i, \mathbf{r}_i)$. We briefly mention the proof.

A skeleton ζ-diagram with $N + \zeta$ points has N labeled points; the insertion, a diagram of $N' + 1$ or $N' + 2$ points, is made at the labeled point(s) $\mathbf{r}_i$ or $\mathbf{r}_i$ and $\mathbf{r}_j$. There are

$$\frac{(N + N')!}{N!\,N'!} \tag{A1}$$

[25] A expression equivalent to (2.21) in this formulation is stationary under variations of G_i but its second functional derivative is not necessarily a positive definite matrix (contrary to an assertion made in reference 4). The reason is that, here, the matrix C_{12} is *not* a fluctuation matrix. The stationarity equation has to be supplemented by a stability condition.

[26] T. Lee and C. Yang, Phys. Rev. **117**, 897 (1960).

[27] C. N. Yang (private communication).

[28] Or multiplication by a factor C_{ij} if in the skeleton the points (i, j) are not directly linked.

distinct labeled ζ-diagrams giving the same contribution and obtained by permuting in every possible way the labels of the skeleton and those of the insertion. The definition of a ζ-diagram contribution and the form of (A1) then lead to the factorization of the skeleton and the insertion contributions (when $\mathbf{r}_i$ or $\mathbf{r}_i$ and $\mathbf{r}_j$ are held fixed) and, therefore to the substitution mentioned above. The same sort of proof may be used when the roots of the skeleton play the role of $\mathbf{r}_i$, $\mathbf{r}_j$.

Quantum systems: Substitutions used in Secs. II(i) and (ii) are treated similarly. In particular the proof is the same for the insertions made on a line (i) leading to the substitution $m(\mathbf{k}_i) \rightarrow M(\mathbf{k}_i)$, and for insertions relative to the internal incoming $(\mathbf{k}_i, \mathbf{k}_j)$ and outgoing $(\mathbf{k}_i', \mathbf{k}_j')$ lines of a (generalized) 2-vertex in a skeleton leading to the substitution

$$\langle \mathbf{k}_i'\mathbf{k}_j' \mid V \mid \mathbf{k}_i\mathbf{k}_j\rangle \rightarrow \langle \mathbf{k}_i'\mathbf{k}_j' \mid C \mid \mathbf{k}_i\mathbf{k}_j\rangle.$$

A slight modification occurs, in a 2-diagram, when the pairs of lines involved belong to an articulation sequence comprising the external lines $(\mathbf{k}_1\mathbf{k}_2; \mathbf{k}_1'\mathbf{k}_2')$ like it occurs in relation (2.24).

Consider the class of 2-diagrams with $N' + N + 2$ internal lines containing two *simple* parts S_{12}' and S_{12} with N' and N internal lines (which we symbolize in Fig. 1). For each diagram of this class there are

$$\frac{(N' + N + 2)!}{(N' + 2)!\, N!} \tag{A2}$$

distinct labeled 2-diagrams obtained by permuting the label of what we arbitrarily call the skeleton (e.g., simple part S_{12}' plus lines $\mathbf{l}_1$, $\mathbf{l}_2$) and of the insertion (simple part S_{12}). If, in the skeleton we do not distinguish diagrams where lines $\mathbf{l}_1$, $\mathbf{l}_2$ bear different labels, we get a counting factor

$$\frac{(N' + 2)(N' + 1)}{2} \frac{(N' + N + 2)!}{(N' + 2)!\, N!}. \tag{A3}$$

The factor $\frac{1}{2}$ arises because, due to the symmetry of the simple parts, exchange of labels between $\mathbf{l}_1$, and $\mathbf{l}_2$ does not lead to distinct diagrams. Expression (A3) leads to the factorization

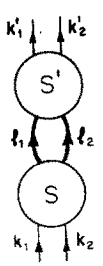


FIG. 1. Class of diagrams with two *simple* parts S_{12}' and S_{12}.

$$\tfrac{1}{2} \sum_{\mathbf{l}_1\mathbf{l}_2} \langle \mathbf{k}_1'\mathbf{k}_2' \mid S' \mid \mathbf{l}_1\mathbf{l}_2\rangle M(\mathbf{l}_1) M(\mathbf{l}_2) \langle \mathbf{l}_1\mathbf{l}_2 \mid S \mid \mathbf{k}_1\mathbf{k}_2\rangle$$

and consequently to (2.24), (2.36), and (2.43).

APPENDIX B: FUNCTIONAL DERIVATIVES

Classical systems: Given a functional $\mathcal{K}^{(2)}\{C_{ij}\}$ of the symmetrical function $C(\mathbf{r}_i, \mathbf{r}_j)$, its functional derivative may be defined as

$$\lim_{\Delta \rightarrow 0} \left[\frac{\mathcal{K}^{(2)}\{\bar{C}_{ij}\} - \mathcal{K}^{(2)}\{C_{ij}\}}{\Delta(\mathbf{r}_1, \mathbf{r}_2)} \right] = \frac{\delta}{\delta C(\mathbf{r}_1; \mathbf{r}_2)} \mathcal{K}^{(2)}\{C_{ij}\}, \tag{B1}$$

where

$$\bar{C}_{ij} = C(\mathbf{r}_i, \mathbf{r}_j) + \tfrac{1}{2} \Delta(\mathbf{r}_1, \mathbf{r}_2) \times [\delta(\mathbf{r}_1 - \mathbf{r}_i)\, \delta(\mathbf{r}_2 - \mathbf{r}_j) + \delta(\mathbf{r}_1 - \mathbf{r}_j)\, \delta(\mathbf{r}_2 - \mathbf{r}_i)]. \tag{B2}$$

If $\mathcal{K}^{(2)}$ is represented by 0-diagrams [calculated with rule (C), for example], we show that *twice* the right-hand side of (B1) is represented by the sum of *all distinct* 2-diagrams which can be generated out of the 0-diagrams [calculated with rule (C)] by suppressing a link C_{ij} and letting the pair (i, j) be fixed at $(\mathbf{r}_1, \mathbf{r}_2)$.

Consider the family of $(N + 2)$-labeled, distinct, 0-diagrams associated with a given free diagram.[16] Choose a pair of points (i, j) in the free diagram. In the family of $(N + 2)$-labeled, distinct, 0-diagrams, that pair is labeled in $(N + 2)(N + 1)/2$ ways if i and j play a symmetrical role, or in $(N + 2)(N + 1)$ ways if i and j do not play a symmetrical role. But there is only one distinct way of identifying (i, j) with $(\mathbf{r}_1, \mathbf{r}_2)$ in the first case and two ways in the second. The original weight $[(N + 2)!]^{-1}$ assigned to $(N + 2)$-labeled 0-diagrams thus becomes $2(N!)^{-1}$ for the N-labeled, 2-diagrams generated in the fashion described above, which proves the assertion.

Quantum systems: The extension to functionals of matrices $\langle \mathbf{k}_i'\mathbf{k}_j' \mid C \mid \mathbf{k}_i\mathbf{k}_j\rangle$ is immediate. The matrix C_{ij} is now invariant under exchange of $(\mathbf{k}_i, \mathbf{k}_j)$ *or* $(\mathbf{k}_i', \mathbf{k}_j')$. The argument applied above to the pair $(\mathbf{r}_i, \mathbf{r}_j)$, applies here both to the incoming pair $(\mathbf{k}_i, \mathbf{k}_j)$ and to the outgoing pair $(\mathbf{k}_i', \mathbf{k}_j')$. Consequently, if a functional $\mathcal{K}^{(2)}$, of the matrix C_{ij}, is represented by 0-diagrams the sum of all *distinct* 2-diagrams which can be generated out of the 0-diagrams by suppressing a heavy dot C_{ij} and letting the incoming pair $(\mathbf{k}_i, \mathbf{k}_j)$ and the outgoing pair $(\mathbf{k}_i', \mathbf{k}_j')$ fixed, respectively, at $(\mathbf{k}_1, \mathbf{k}_2)$ and $(\mathbf{k}_1', \mathbf{k}_2')$, represents *four times* the functional derivative.

APPENDIX C: PROOF OF RELATION (1.48)

It suffices to prove that the relation

$$\tfrac{1}{2}[N(\pi) - N(\pi\cdot\mu)] + N(\mathfrak{N}) + N(\mathfrak{K}^{(2)}) = +1 \quad \text{(C1)}$$

is true for any 1-irreducible 0-diagram. Indeed consider a 0-diagram where we choose a pair of points (i, j) with i and j belonging to different 1-irreducible parts. The pair (i, j) has then a multiplicity $\mu = 1$ and only pairs inside the same 1-irreducible part will contribute a nonvanishing $[N(\pi) - N(\pi\cdot\mu)]$ to (1.48). The quantity $N(\mathfrak{N}) + N(\mathfrak{K}^{(2)})$ counting the number of ways a diagram can be generated out of a polygonal skeleton (1.45) or a 2-irreducible skeleton (1.46) can also be broken up into contributions referring to each separate 1-irreducible parts, since the generating diagrams (1.45) and (1.46) have no articulation points. As a consequence, the relation (1.48) itself can be broken into a sum of relations (C1), one for each 1-irreducible part of the diagram.

In the following we only consider 1-irreducible 0-diagrams. The proof is inductive. We assume that (C1) is verified for all 1-irreducible 0-diagrams of order up to p, we show that (C1) is still verified for diagrams of order $p + 1$.

1. Consider in a 0-diagram of order $p' < p$, a pair of points (i, j) connected by a direct link. We generate a 0-diagram of order $p + 1$ by (Fig. 2)

(a) inserting at (i, j) a 2-irreducible 2-diagram, or

(b) replacing the direct link (i, j) by a 2-irreducible, 2-diagram.

Under transformations (a) or (b), the weights involved in relation (C1) undergo changes which we now examine. In either cases the multiplicity of the pair (i, j) is increased by one unit. Each new pair has multiplicity one. Thus,

$$\delta[N(\pi) - N(\pi\cdot\mu)] = -1, \quad \text{(C2)}$$

and obviously

$$\delta N(\mathfrak{N}) = 0 \quad \text{(C3)}$$

$$\delta N(\mathfrak{K}^{(2)}) = 0. \quad \text{(C4)}$$

Relation (C1) remains verified for the new diagram.

2. Consider in a 0-diagram of order p, a pair of points (i, j) connected by a direct link. We generate a 0-diagram of order $p + 1$ by (Fig. 3)

(a) inserting at (i, j) the nodal 2-diagram of lowest order, or

(b) replacing the direct link (i, j) by the nodal 2-diagram of lowest order.

These transformations generate changes of the weights

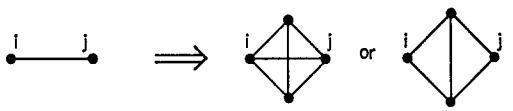


FIG. 2. Illustration by a simple example of substitution 1 considered in Appendix C.

$$\delta N(\pi) = p. \quad \text{(C5)}$$

(i) Let m be that $(p + 1)$th point of the new diagram. The pair (i, j) has a multiplicity increased by one; any pair including point m has multiplicity one except when transformation (b) is applied to a pth-order diagram of the following character: after removal of the direct link (i, j), the remaining 2-diagrams, rooted at points (i, j) is a *nodal 2-diagram*. Excepting such a case for the moment, we have

$$\delta N(\pi\cdot\mu) = p + 1, \quad \text{(C6)}$$

and obviously

$$\delta N(\mathfrak{N}) = 1 \quad \text{(C7)}$$

$$\delta N(\mathfrak{K}^{(2)}) = 0, \quad \text{(C8)}$$

which leaves relation (C1) invariant.

(ii) We return to the exceptional case. The pth-order diagram is obtained by adding a direct link (i, j) to the nodal 2-diagram rooted at (i, j). By virtue of the proofs given in operations 1 and 2(i) if suffices to consider the case of a pth-order polygonal diagram. It is then just as easy to verify directly relation (C1) for any polygonal diagram. We have

$$\tfrac{1}{2}N(\pi) = \binom{2}{p} \quad \text{(C10)}$$

$$\tfrac{1}{2}N(\pi\cdot\mu) = 2\binom{2}{p} - \binom{1}{p}; \quad \text{(C11)}$$

(C11) states that any pair of points has multiplicity 2 except if the two points are adjacent summits (in which case it has multiplicity one).

$$N(\mathfrak{K}^{(2)}) = 0 \quad \text{(C12)}$$

$$N(\mathfrak{N}) = -\left[\sum_{q=0}^{p}(-)^q\binom{q}{p}\right] + \left[\binom{2}{p} - \binom{1}{p} + \binom{0}{p}\right] = \binom{2}{p} - \binom{1}{p} + \binom{0}{p}; \quad \text{(C13)}$$

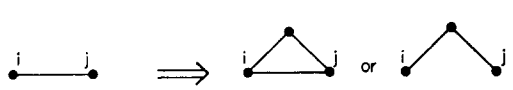


FIG. 3. Illustration of substitution 2 considered in Appendix C.

(C13) gives the number of ways one can generate such a diagram out of the polygonal diagrams (of order $\leq$ p) of (1.45). Here it suffices to choose in in the p th-order diagram the summits we want to keep in the generating one. Successive terms in the first bracket corresponds to the choice of q summits with the factor $(-)^{q+1}$ attached to polygonal diagrams in (1.45). The second bracket takes care of the fact that the lowest order polygonal diagram is of order 3. Expressions (C10) to (C13) again imply (C1).

Clearly now we can generate a general 0-diagram, of order $p + 1$, out of diagrams of lower order by applications of operations 1 or 2; relation (C1) being checked for the simplest diagram ($p = 2$) is valid for any diagram.

APPENDIX D: SECOND VARIATIONS

We show that, for classical systems, the quadratic form of the second variations obtained from the functional $\Xi\{\bar{n}_i, \bar{C}_{ij}\}$ (or Ω) is positive definite at the extremum $\bar{n}_i = n_i$, $\bar{C}_{ij} = C_{ij}$. The proof follows some of the argument used in reference 5 for a single functional and a quantum system.

From the definition of Ξ (1.64) we have

$$\delta\Xi/\delta n_i = w_i \tag{D1}$$

$$\delta\Xi/\delta n_{ij} = w_{ij}, \tag{D2}$$

where we have considered Ξ as a functional of n_i, n_{ij} and we let

$$\begin{aligned} w_i &= \beta u_i \\ w_{ij} &= \tfrac{1}{2}\beta v_{ij}. \end{aligned} \tag{D3}$$

The second functional derivatives are given by the matrix

$$\delta w_A/\delta n_{A'} \tag{D4}$$

where $A = (i)$ or (ij). The quadratic form representing the second variation of Ξ is thus

$$\sum_{A,A'} \int \delta n_A \frac{\delta w_A}{\delta n_{A'}} \delta n_{A'}. \tag{D5}$$

From the definition of ln Z (1.3) we also have

$$\delta \ln Z/\delta w_A = -n_A. \tag{D6}$$

The second functional derivatives are given by the matrix

$$\delta n_A/\delta w_{A'} \tag{D7}$$

and give rise to a quadratic form

$$\sum_{A,A'} \int \delta w_A \frac{\delta n_A}{\delta w_{A'}} \delta w_{A'} \geq 0. \tag{D8}$$

This form is *positive definite* because successive functional derivatives of ln Z with respect to w_A give rise to successive fluctuations. It suffices to recognize that the matrix (D4) is the inverse and opposite of (D7), to prove that the quantity (D5) is negative definite.

APPENDIX E: PROOF OF RELATION (2.44)

It suffices to prove that relation

$$\tfrac{1}{2}[N(\sigma) - N(\sigma\cdot\mu)] + N(\mathfrak{N}) + N(\mathfrak{V}) + N(\mathfrak{K}^{(2)}) = +1 \tag{E1}$$

holds for any 1-irreducible 0-diagram. Consider two laces (i, j), each one belonging to a different 1-irreducible part of the diagram. Laces i and j cannot belong to an articulation sequence of multiplicity $\mu > 1$: indeed after having cut open the lines i and j ($\mathbf{k}_i$, $\mathbf{k}_j$) incoming lines; $\mathbf{k}'_i$, $\mathbf{k}'_j$ outgoing lines), we may further separate the diagram into 2 pieces by separating away, for example, the 1-irreducible part to which ($\mathbf{k}_i$, $\mathbf{k}'_i$) belong. We then have two pieces containing ($\mathbf{k}_i$, $\mathbf{k}'_i$)($\mathbf{k}_j$, $\mathbf{k}'_j$), respectively. The 2-diagram obtained after opening the laces i and j thus had the structure of a nodal (simple) diagram. Hence, only pairs of laces belonging to the same 1-irreducible part will contribute a nonvanishing $[N(\sigma) - N(\sigma\cdot\mu)]$ to (2.37). Counting the number of ways a 0-diagram can be generated out of the polygonal diagrams (2.39), the vertex-like diagrams (2.40) or the 2-irreducible diagrams (2.41) may also be broken up into contributions referring to each separate 1-irreducible part, since the generating diagrams have no articulation lace. Consequently, (2.44) itself may be broken up into a sum of relations like (E1), one for each of the 1-irreducible parts constituting the diagram.

In the following we only consider 1-irreducible 0-diagrams the inductive proof given here follows closely the proof of Appendix C relative to classical systems. We assume that (E1) is valid for any 1-irreducible diagram of order up to p (the order is the sum of the orders α_i of each α-vertex). We show that (E1) is still verified for diagrams of order $p + 1$.

We consider the following possibilities of generating a 0-diagram of order p out of a diagram of lower order.

1. The diagram is obtained by replacing an ordinary 2-vertex by a *generalized 2-vertex* (Fig. 4). We label by m or n internal lines of the insertion, by s internal lines of the rest of the diagram. Insertion lines are labeled $\mathbf{k}_i\mathbf{k}_j$, $\mathbf{k}'_i\mathbf{k}'_j$; considering only 1-irreducible diagrams these momenta must all be dif-

ferent. We recall that a pair (m, n) is part of an articulation sequence ($\mu > 1$) if, after cutting open (m, n) into incoming $(\mathbf{k}_m, \mathbf{k}_n)$ and outgoing $(\mathbf{k}'_m, \mathbf{k}'_n)$ pieces, we can separate the diagram into two disconnected parts containing $(\mathbf{k}_m, \mathbf{k}_n)$ and $(\mathbf{k}'_m, \mathbf{k}'_n)$, respectively, by cutting open one more pair only. From the definition of the generalized 2-vertex part no pair $(m, n)(m, s)$, (m, i), or (m, j) may be part of an articulation sequence (of multiplicity $\mu > 1$) since it takes cutting more than four internal lines to split apart a generalized 2-vertex.

We then have

$$\delta[N(\sigma) - N(\sigma\cdot\mu)] = 0, \tag{E2}$$

and similarly

$$\delta N(\mathfrak{N}) = \delta N(\mathfrak{V}) = \delta N(\mathfrak{K}^{(2)}) = 0. \tag{E3}$$

Relation (E1) is thus unchanged under the considered operation. Naturally, substitution of a generalized 2-vertex by another one leads to the same conclusion.

2. The diagram is obtained by replacing an ordinary 2-vertex by a 2-irreducible 2-diagram (Fig. 5). The 2-irreducible 2-diagram inserted may be thought as built with ν generalized 2-vertices besides "effective" α-vertices ($\alpha > 2$).

By definition, no pair (m, n) may be part of the same articulation sequence as $(\mathbf{k}_i, \mathbf{k}_j; \mathbf{k}'_i, \mathbf{k}'_j)$; but 2ν pairs (m, n), (m, i), or (m, j) form ν articulation sequences of multiplicity 2. Further a pair (m, s) cannot be part of an articulation sequence: Indeed, suppose we cut open the pair (m, s); to separate the diagram into two parts containing $(\mathbf{k}_m, \mathbf{k}_s)$ and $(\mathbf{k}'_m, \mathbf{k}'_s)$ we have to separate the inserted 2-irreducible diagram which requires cutting at least three more lines. Hence, we have

$$\delta[N(\sigma) - N(\sigma\cdot\mu)] = -\nu. \tag{E4}$$

The new diagram may be generated now in $\nu - 1$ extra ways out of the diagrams of $\mathfrak{V}$, and one extra way out of the diagrams of $\mathfrak{K}^{(2)}$

$$\delta N(\mathfrak{N}) = 0, \tag{E5}$$

$$\delta N(\mathfrak{V}) = \nu - 1, \tag{E6}$$

$$\delta N(\mathfrak{K}^{(2)}) = 1, \tag{E7}$$

leaving thus (E1) unchanged.

3. The diagram is obtained by replacing an ordinary 2-vertex by two of them connected ladderwise (Fig. 6). The pair (m, n) is part of the articulation sequence containing $(\mathbf{k}_i, \mathbf{k}_j; \mathbf{k}'_i, \mathbf{k}'_j)$ and its order is increased by one unit. A pair (m, i), (m, j), or (m, s) cannot be part of an articulation sequence:

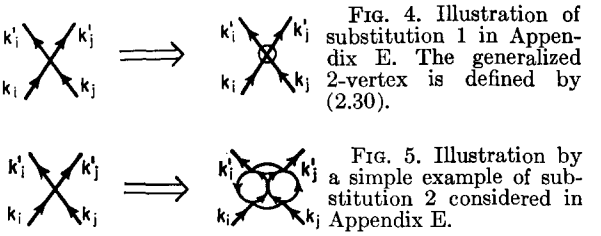


FIG. 4. Illustration of substitution 1 in Appendix E. The generalized 2-vertex is defined by (2.30).

FIG. 5. Illustration by a simple example of substitution 2 considered in Appendix E.

Indeed, after having cut the line m into $\mathbf{k}_m$ and $\mathbf{k}'_m$, to separate into different parts $\mathbf{k}_m$ from $\mathbf{k}'_m$, we need to cut also the line n into $\mathbf{k}_n$, $\mathbf{k}'_n$; thus, the only way the line m can be made part of an articulation sequence is by being paired with the line n. Hence the change in $N(\sigma) - N(\sigma\cdot\mu)$ reduces to

$$\delta[N(\sigma) - N(\sigma\cdot\mu)] = -1. \tag{E8}$$

Also we have one more possibility of generating the diagram out of $\mathfrak{V}$

$$\delta N(\mathfrak{V}) = 1, \tag{E9}$$

and

$$\delta N(\mathfrak{N}) = \delta N(\mathfrak{K}^{(2)}) = 0, \tag{E10}$$

thus leaving preserved relation (E1).

Notice that the substitution considered has no classical equivalent.

4. The diagram is obtained by replacing an ordinary 2-vertex by two of them connected "nodewise" (Fig. 7). Notice that we shall not need to consider apart the substitution corresponding to case 2(a) of Appendix C since it can be obtained by competition of operations 3 and 4.

We consider two cases:

(i) $(\mathbf{k}_i, \mathbf{k}_j; \mathbf{k}'_i, \mathbf{k}'_j)$ are part of an articulation sequence of multiplicity $\mu > 2$; the resulting diagram has the structure shown in Fig. 8 for $\mu = 3$, for example. Pairs (m, n) or (m, s) cannot be part of an articulation sequence. This is obvious for (m, n). For (m, s), let us cut (m, s) open into $(\mathbf{k}_m, \mathbf{k}_s)$ and $\mathbf{k}'_m, \mathbf{k}'_s)$; to put $\mathbf{k}_m$ and $\mathbf{k}'_m$ into different parts we certainly need to cut n open, then to perform the separation of the diagram into two parts, we shall need to cut more than one line. On the other

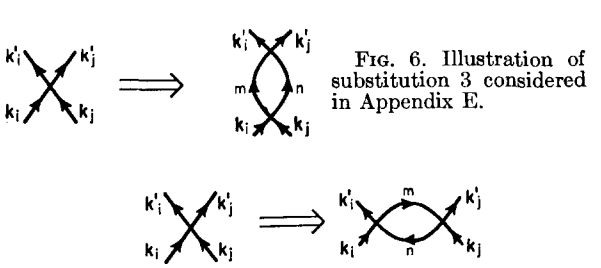


FIG. 6. Illustration of substitution 3 considered in Appendix E.

FIG. 7. Illustration of substitution 4 considered in Appendix E.

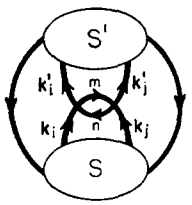


FIG. 8. Pairs of lines ($\mathbf{k}_i$ $\mathbf{k}_j$) or ($\mathbf{k}_i'$ $\mathbf{k}_j'$) are part of an articulation sequence with $\mu = 3$. The bubbles figure *simple* parts.

hand the transformed diagram contains two more articulation sequences ($\mathbf{k}_m$, $\mathbf{k}_j$; $\mathbf{k}_n$, $\mathbf{k}_i'$) and ($\mathbf{k}_n$, $\mathbf{k}_i$; $\mathbf{k}_m$, $\mathbf{k}_i'$) so that we have

$$\delta[N(\sigma) - N(\sigma\cdot\mu)] = -2. \tag{E11}$$

But we also have one more possibility to generate the diagram out of $\mathcal{U}$ and out of $\mathfrak{N}$

$$\delta N(\mathfrak{N}) = \delta N(\mathcal{U}) = 1, \tag{E12}$$

$$\delta N(\mathcal{K}^{(2)}) = 0, \tag{E13}$$

which leaves (E1) unchanged.

(ii) ($\mathbf{k}_i$, $\mathbf{k}_j$; $\mathbf{k}_i'$, $\mathbf{k}_j'$) is an articulation sequence of multiplicity $\mu = 2$. By virtue of points 1 to 3 it suffices to consider the particular case of polygonal diagrams. It is then just as easy to show directly that relation (E1) is satisfied.

Consider a polygonal diagram (2.39) with p summits ($2p$ lines), it is immediately found that

$$\tfrac{1}{2}N(\sigma) = \binom{2}{2p} \tag{E14}$$

$$\tfrac{1}{2}N(\sigma\cdot\mu) = \left[\binom{2}{2p} - 2\binom{2}{p}\right] + \binom{2}{p}. \tag{E15}$$

Further, we have

$$N(\mathcal{U}) = \binom{1}{p}, \tag{E16}$$

$$N(\mathcal{K}^{(2)}) = 0. \tag{E17}$$

Thus

$$\tfrac{1}{2}(N(\sigma) - N(\sigma\cdot\mu)) + N(\mathcal{U}) + N(\mathcal{K}^{(2)}) = -\binom{2}{p} + \binom{1}{p}. \tag{E18}$$

The calculation for $N(\mathfrak{N})$ being exactly the same as for the classical case, relation (E1) is thus proved for the polygonal diagrams.

By using operations 1 to 4 we can build any diagram of order $p + 1$ out of diagrams of lower order and relation (E1) being verified for the simplest diagram (single 2-vertex) is verified for any diagram.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Analytic Properties of the Quantum Corrections to the Second Virial Coefficient*

HUGH E. DEWITT
Lawrence Radiation Laboratory, University of California, Livermore, California
(Received October 2, 1961)

The usefulness of the perturbation expansion and the Wigner-Kirkwood expansion of the quantum-mechanical partition function is discussed for various interaction potentials. It is shown that, contrary to what is expected from the Wigner-Kirkwood expansion, quantum-mechanical diffraction corrections at high temperature to the classical partition function may involve nonanalytic forms of $\hbar^2$. This occurs when the second-order perturbation term is finite in the classical limit, and the interaction potential has a cusp or singularity in any derivative. The second-order perturbation term is evaluated exactly for the exponential, screened Coulomb, and square barrier potentials, and the nonanalytic form $(\hbar^2)^{1/2}$ is found. For potentials more singular than $1/r$ at the origin, the diffraction corrections are analytic in $\hbar^2$.

A new method of deriving the Wigner-Kirkwood expansion from the perturbation expansion is given. The method allows one to subtract off any order of the perturbation expansion which may be evaluated separately, and is particularly useful for the screened Coulomb potential.

The classical second virial coefficient and the $O(\hbar^2)$ and $O(\hbar^4)$ diffraction corrections are evaluated for the singular potential, $u(r) = (g_p/r^p)e^{-r/r_0}$, by using the Mellin transform of $e^{-\beta u}$.

I. INTRODUCTION

THE problem of calculating small quantum corrections at high temperature to classical thermodynamic quantities has been discussed extensively during the three decades since the classic papers of Wigner and Kirkwood.[1] In this paper the same problem is considered again, but with the purpose of establishing the analytic properties of the partition function with respect to Planck's constant for various interaction potentials. The quantum corrections to the classical partition function to be considered are those due to the operation of the uncertainty principle. Effects due to quantum statistics will not be treated here. Thus, we consider a gas of distinguishable particles interacting according to the laws of wave mechanics. Such quantum corrections will be referred to as diffraction effects.

The fundamental problem of quantum statistical mechanics is the evaluation of the partition function, $Z = \mathrm{Tr}\exp(-\beta H)$, where $H = \sum_i (-\hbar^2/2m_i)\nabla_i^2 + \sum_{i<j} u(r_{ij})$, and $\beta = 1/kT$ is the reciprocal temperature. Since the partition function may be evaluated directly for only a very limited set of interaction potentials, it is necessary in general to resort to some expansion procedure. One method is to expand in powers of the interaction potential. Such a perturbation expansion is appropriate when $u(r)$ is small in some sense compared with the kinetic energy. When the terms of the perturbation expansion are evaluated, the diffraction corrections to the classical limit of the nth order appear as some function of $\hbar$ multiplying the nth power of the coupling constant of the interaction. A second method in common use involves expanding in powers of $\hbar^2\nabla^2$; thus the kinetic energy is considered small compared with the potential energy. This second method is appropriate when $u(r)$ is very singular at $r = 0$. The expansion in powers of the kinetic energy is the well-known Wigner-Kirkwood expansion[1] (hereafter to be referred to as the WK expansion). Using the WK method, the partition function (written for one particle) is:

$$\mathrm{Tr}\, e^{-\beta H} = \left(\frac{2\pi mkT}{h^2}\right)^{3/2} \int d^3r\, e^{-u} \times \left\{1 - \frac{\lambda\!\!\!^{-2}}{12}(\nabla U)^2 + \frac{\lambda\!\!\!^{-4}}{1440}[((\nabla U)^2)^2 - 8(\nabla U)^2\nabla^2 U + 12(\nabla^2 U)^2] - \cdots\right\}, \quad (1)$$

where $U = \beta u(r)$, and $\lambda\!\!\!^{-} = \hbar/(2mkT)^{1/2}$ is the thermal de Broglie wavelength.[2] The $\lambda\!\!\!^{-4}$ term in Eq. (1) is the form obtained by Yaglom.[3] The evaluation of the terms of the WK expansion is

* Work performed under the auspices of the U. S. Atomic Energy Commission.

[1] E. P. Wigner, Phys. Rev. **40**, 747 (1932); J. G. Kirkwood, *ibid.* **44**, 31 (1933).

[2] In statistical-mechanics textbooks, the thermal wavelength is often defined as $h/(2\pi mkT)^{1/2}$ and denoted by the symbol λ, so that the ideal gas partition function reads $(V/\lambda^3)^N$.

[3] A. M. Yaglom, Teoriya Veroyatnostei i ee Primeniya **1**, 161 (1956). For an English language summary of Yaglom's method see S. G. Brush, Revs. Modern Phys. **33**, 79 (1961).

quite lengthy and no terms beyond the λ^4 term are known to exist in the literature. Equation (1) has had considerable practical application in the calculation of quantum corrections to the equation of state of nonideal gases.[4,5]

At first glance it would appear from the structure of Eq. (1) that the partition function is an analytic function of λ^2, and in consequence there seems to be a common and erroneous belief among physicists that diffraction corrections necessarily involve only even powers of Planck's constant. The argument for the nonideal gas calculations is that any reasonable form of the intermolecular interaction potential is strongly repulsive near $r = 0$; usually r^{-12} is assumed. Hence e^{-U} goes to zero much faster as $r \to 0$ than the terms of the expansion go to ∞, so that the configuration space integrals are finite. Although little is known about the convergence of the resulting series, it seems reasonable that at high temperature the first few terms give the diffraction corrections accurately.

It is not true, however, that for all potential interactions the partition function is analytic in λ^2. A simple counter example is the exponential potential, $u(r) = g_0 e^{-r/r_0}$. Since this potential form is finite at $r = 0$, one cannot depend on the e^{-U} factor for the existence of the coefficients of powers of λ^2. The mth term of the WK expansion includes $(\nabla^2 U)^m$, and since $\nabla^2 e^{-r} = (1 - 2/r)e^{-r}$, one sees that the coefficient of λ^{2m} includes at least one term of order $r^{-m}e^{-mr}$. Thus, after the integration over $\mathbf{r}$ the coefficients of λ^2 and λ^4 are finite, but that of λ^6 is logarithmically divergent and all higher coefficients are more strongly divergent. The exponential potential is an example, albeit not very interesting for physical problems, for which the WK expansion may not be used. Instead, one must evaluate the terms of the perturbation expansion, and it will be found that the coefficient of each power of the coupling constant is a nonanalytic function of $\hbar^2$. It will be shown that the nonanalyticity takes the form of terms of order $(\hbar^2)^{m+1/2}$ in addition to the expected terms of order $\hbar^{2m}$. A more interesting, though less obvious example, is a potential with an r^{-1} singularity at the origin. Evaluation of the second-order perturbation term for the screened Coulomb potential yields again a function with both even and odd powers of $\hbar$ in its expansion. In view of these examples it seems worthwhile to examine the question of when the partition function is analytic in $\hbar^2$ and when it is not.

Before taking up the analyticity of Z as a function of $\hbar^2$, its analyticity with respect to two other quantities should be considered, namely, the particle number density ρ and the coupling constant of the interaction g. In this paper we will consider only potentials such that the cluster integrals of the Mayer cluster expansion exist.[6] With this restriction the pressure is an analytic function of ρ; i.e., it is given by a power series expansion in ρ, the usual virial expansion.[7] One is next interested in the analyticity of the virial coefficients as functions of g and $\hbar^2$. In this paper, only the second virial coefficient will be studied since the methods used may be easily extended to the higher virial coefficients. The second virial coefficient is the sum of all two-body interactions, and is defined as:

$$B_2 = -\frac{(4\pi\lambda)^{3/2}}{2!}\left[\mathrm{Tr}\, e^{-\beta(H_0+u)} - \mathrm{Tr}\, e^{-\beta H_0}\right] \underset{\hbar\to 0}{\longrightarrow} -\frac{1}{2!} \times \int d^3r(e^{-\beta u} - 1), \qquad (2)$$

where $H_0 = -(\hbar^2/2\mu)\nabla^2$ and μ is the reduced mass of the two interacting particles.

We will be primarily interested in the evaluation of B_2 for repulsive singular potentials of the form:

$$u(r) = (g_p/r^p)e^{-r/r_0}, \qquad (3)$$

where the coupling constant g_p has dimensions EL^p. The exponential screening function is chosen for mathematical convenience. Other screening functions such as the Gaussian form e^{-r^2/r_0^2} may also be used. The analyticity of the classical form of B_2 as a function of g is obvious for nonsingular potentials, say, of the form $r^m e^{-r/r_0}$. The first-order singularity, $p = 1$ in Eq. (3), is the very interesting case of the screened Coulomb potential. For this potential the first two terms of the perturbation expansion are finite because of the three-dimensional volume element $4\pi r^2\, dr$. The third order is logarithmically divergent, and the higher orders more strongly divergent. The exact evaluation of B_2 for the screened Coulomb potential yields:

$$B_2 = -2\pi r_0^3\{-\beta g_1/r_0 + (1/4)(\beta g_1/r_0)^2 + (1/6)(\beta g_1/r_0)^3[\ln(\beta g_1/r_0) + \text{const}] + \cdots\}. \qquad (4)$$

[4] J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), Chap. 6, pp. 419–424.

[5] Oppenheim and A. S. Friedman, J. Chem. Phys. **35**, 35 (1961).

[6] For the unscreened Coulomb potential, the cluster integrals are all divergent because of the infinite range of e^2/r. The correct pressure expression includes the nonanalytic forms $\rho^{3/2}$ (the Debye-Hückel term) and $\rho^2 \ln \rho$.

[7] T. E. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), pp. 141–144.

Thus, the divergence of the third and higher orders of the perturbation expansion gives rise to the nonanalytic form $g_1^3 \ln g_1$. For a second-order singularity, $p = 2$, the first term of the perturbation expansion is finite, and all higher orders are divergent. The exact result for B_2 contains the nonanalytic form $g_2^2 \ln g_2$. Similarly, for $p = 3$ all orders of the perturbation expansion are infinite, and the exact result for B_2 begins with $g_3 \ln g_3$. For $p > 3$, B_2 begins with $(\beta g_p)^{3/p}$.

A simple dimensional analysis gives quickly some information about the analyticity of B_2 as a function of $\hbar^2$. For the singular potentials defined by Eq. (3), the fundamental lengths which completely determine B_2 are: the classical interaction length $l = (\beta g_p)^{1/p}$, the thermal wavelength $\lambda = \hbar(\beta/2m)^{1/2}$, and the screening length r_0. From these lengths we may form two independent dimensionless parameters which will be taken to be any two of the ratios:

$$\Lambda = l/r_0, \qquad \eta = \lambda/l, \qquad \gamma = \lambda/r_0.$$

By using Eqs. (1) and (2) the WK expansion of the second virial coefficient in terms of the parameters Λ and η is:

$$B_2 = -2\pi l^3[C_0(\Lambda) - \eta^2 C_1(\Lambda) + \eta^4 C_2(\Lambda) - \cdots], \quad (5)$$

where the expansion coefficients are:

$$C_0(\Lambda) = \int_0^\infty x^2\, dx\, (e^{-U} - 1),$$
$$C_1(\Lambda) = \frac{1}{12}\int_0^\infty x^2\, dx\, e^{-U}(\nabla_x U)^2, \quad (6)$$

with $x = r/l$, and $U = x^{-p}e^{-\Lambda x}$. $C_0(\Lambda)$ gives the classical second virial coefficient. The coefficients of the diffraction corrections, $C_1(\Lambda) \cdots C_m(\Lambda)$, are finite for all $p \geq 1$ in the limit of no screening, $r_0 = \infty$ or $\Lambda = 0$. From Eq. (5) we see that the parameter of smallness for the WK expansion is η^2. Since its dependence on the coupling constant is $g_p^{-2/p}$, it is clear that the WK expansion is a *strong coupling* expansion in contrast to the perturbation expansion. The temperature dependence of η^2 is $\beta^{1-2/p}$, and hence the diffraction corrections vanish at high temperature when $p \geq 3$. The radius of convergence of the power series in η^2 of Eq. (5) is not known, but it seems clear that a third-order singularity in the potential is sufficient to guarantee that diffraction corrections at high temperature involve only powers of η^2, and hence only even powers of $\hbar^2$.

The less singular cases, $p = 2$ and $p = 1$, must be considered separately from $p \geq 3$. For $p = 2$, η^2 has no temperature dependence, and hence the WK expansion would indicate that diffraction corrections do not depend on temperature. Finally, for $p = 1$, the screened Coulomb potential, the temperature dependence of η^2 is β^{-1}. Note that for $p \geq 3$ the thermal wavelength is small compared with the classical interaction length at high temperature, whereas for $p = 1$ the order is reversed, $\lambda \gg l$. Thus the terms of the WK expansion diverge in the high-temperature limit for $p = 1$. This behavior indicates that B_2 cannot be an analytic function of $\hbar^2$ for $p = 1$. The r^{-1} singularity at first appears to be too weak to allow an expansion in which the kinetic energy is treated as small compared with $u(r)$. It will be shown in Secs. III and V, however, that the nonanalyticity in $\hbar^2$ for $p = 1$ appears only in the second-order perturbation term. The diffraction corrections involve $\gamma = \eta\Lambda$, and in the second-order theory both odd and even powers of γ appear. The diffraction corrections to the sum of all higher orders of the perturbation expansion involve only even powers of γ, and the coefficients may be calculated by a modification of the WK expansion.

In Sec. II the perturbation expansion is developed in some detail, and a method of deriving the WK expansion from the perturbation expansion is given. In Sec. III the second-order perturbation term is evaluated explicitly for a number of different potentials in order to illustrate the condition for which it is or is not analytic in γ^2. In Sec. IV some of the coefficients of the WK expansion are evaluated by a very convenient technique, the use of the Mellin transform. In Sec. V the special case of the screened Coulomb potential is considered in some detail.

II. THE PERTURBATION EXPANSION AND ITS USE FOR DERIVING THE WK EXPANSION

The perturbation expansion of B_2 is most easily developed with the help of the resolvent operator. One uses

$$e^{-\beta H} = \frac{1}{2\pi i}\int_C \frac{dz\, e^{-\beta H}}{z - H},$$

where the contour C goes from right to left in the upper half-plane and left to right in the lower half. Thus, it encloses the simple poles on the real axis at the eigenvalues of H when the trace is taken. This method was used by Glassgold, Heckrotte, and Watson for the linked cluster expansion of the complete partition function.[8] If we put $H =$

[8] A. E. Glassgold, W. Heckrotte, and K. M. Watson, Phys. Rev. 115, 1374 (1959).

$(H - E_0) + E_0$, where $E_0 = p^2/2\mu$ is the unperturbed kinetic energy of relative motion, and use the resolvent of $e^{-\beta(H-E_0)}$, then the second virial coefficient is

$$B_2 = -\frac{(4\pi\lambda^2)^{3/2}}{2!}\int \frac{d^3p e^{-\beta E_0}}{(2\pi\lambda)^3}\frac{1}{2\pi i}\int_C dz\, e^{-\beta z} \times\left\langle p\left|\frac{1}{z-(H_0+u-E_0)}-\frac{1}{z-(H_0-E_0)}\right|p\right\rangle. \tag{7}$$

Expanding in powers of u gives

$$B_2 = -\frac{(4\pi\lambda^2)^{3/2}}{2!}\sum_{n=1}\int\frac{d^3p e^{-\beta E_0}}{(2\pi\hbar)^3}\frac{1}{2\pi i} \times\int_C\frac{dz\, e^{-\beta z}}{z}\left\langle p\left|\left(\frac{1}{z-(H_0-E_0)}u\right)^n\right|p\right\rangle = \sum_{n=1} B_{2n}. \tag{8}$$

Since Boltzmann statistics have been assumed, the individual particle momenta may be transformed to center-of-mass and relative momenta, and the center-of-mass momentum integrated out. Thus, H_0 is the free-particle Hamiltonian for relative motion. $H_0 = -(\hbar^2/2\mu)\nabla^2$, and $|p\rangle$ indicates its eigenfunctions.

The operator product in Eq. (8) is written out in momentum space to give the nth order of B_2 as

$$B_{2n} = -\frac{(4\pi\lambda^2)^{3/2}}{2!}\int\frac{d^3p e^{-p^2/2\mu\kappa T}}{(2\pi\hbar)^3} \times\int\frac{V^n\, d^3k_1\cdots d^3k_n}{(2\pi)^{3(n-1)}}\,\delta(\mathbf{k}_n-\mathbf{k}_1-\cdots-\mathbf{k}_{n-1}) \times u(k_1)\cdots u(k_n)\frac{1}{2\pi_i}\int_C\frac{dz\, e^{-\beta z}}{z} \times\frac{1}{z-(p_{n-1}^2-p^2)/2\mu}\cdots\frac{1}{z-(p_1^2-p^2)/2\mu}, \tag{9}$$

where

$$\mathbf{p}_1 = \mathbf{p} + \hbar\mathbf{k}_1, \cdots, \mathbf{p}_{n-1} = \mathbf{p} + \hbar(\mathbf{k}_1 + \cdots + \mathbf{k}_{n-1}),$$

and

$$u(k) = V^{-1}\int d^3r e^{i\mathbf{k}\cdot\mathbf{r}}u(r)$$

is the Fourier transform of the potential. The quantities $\hbar\mathbf{k}_1, \cdots, \hbar\mathbf{k}_n$ are the momentum transfers at the respective n interactions. The δ function ensures momentum conservation in the final interaction. The contour integration in Eq. (9) may be performed and the result represented as:

$$\frac{1}{2\pi i}\int_C\frac{dz\, e^{-\beta z}}{z}\cdots = \int_0^\beta d\beta_n\int_0^{\beta_n}d\beta_{n-1}\cdots\int_0^{\beta_2}d\beta_1 \times\exp-[(\beta_n-\beta_{n-1})(p_{n-1}^2-p^2)/2\mu + \cdots(\beta_2-\beta_1)(p_1^2-p^2)/2\mu]. \tag{10}$$

The multiple integral form [Eq. (10)] is the nth term of the familiar ordered exponential expansion commonly used in field theory calculations and in recent years also in quantum statistical mechanics.[9] Equation (10) shows the equivalence of the linked cluster expansion of the partition function in the resolvent operator formalism[8] and the work of Bloch and de Dominicis.[10]

The relative momentum integration may be readily performed when Eq. (10) is used in Eq. (9). The result is

$$B_{2n} = -\frac{(-\beta)^n}{2!}\int\frac{V^n\, d^3k_1\cdots d^3k_n}{(2\pi)^{3(n-1)}} \times\delta(\mathbf{k}_n-\mathbf{k}_1-\cdots-\mathbf{k}_{n-1}) \times u(k_1)\cdots u(k_n)F_n(\lambda\mathbf{k}_1,\cdots,\lambda\mathbf{k}_{n-1}), \tag{11}$$

where

$$F_n(\lambda\mathbf{k}_1,\cdots,\lambda\mathbf{k}_{n-1}) = \int_0^1 dv_n\cdots\int_0^{v_2}dv_1 \times\exp-\lambda^2\{[(v_2-v_1)k_1^2+\cdots + (v_n-v_{n-1})(\mathbf{k}_1+\cdots+\mathbf{k}_{n-1})^2] - [(v_2-v_1)\mathbf{k}_1+\cdots + (v_n-v_{n-1})(\mathbf{k}_1+\cdots+\mathbf{k}_{n-1})]^2\}. \tag{12}$$

All quantum-mechanical diffraction effects are contained in the functions F_n. The F_n are entire functions of λ^2. The first term of F_n when expanded in powers of λ^2 is $1/n!$, and consequently in the classical limit, $\lambda \to 0$, Eq. (11) reduces to

---

[9] The first use of the ordered exponential expansion in statistical mechanics was by M. L. Goldberger and E. N. Adams, J. Chem. Phys. 20, 240 (1952). They pointed out that if a and b are any two operators, the expansion of $e^{-(a+b)}$ may be written as:

$$e^{-a}\sum_{n=0}(-1)^n\int_0^1 dv_n\cdots\int_0^{v_2}dv_1\, e^{av_n} \times be^{-a(v_n-v_{n-1})}b\cdots be^{-a(v_2-v_1)}be^{-av_1}.$$

When b is chosen to be the kinetic energy operator and a the potential energy, the WK expansion is obtained. Conversely, when b is the potential energy, the perturbation expansion is obtained.

[10] C. Bloch and C. de Dominicis, Nuclear Phys. 7, 459 (1959). Their work is much more general than this paper in that quantum statistics are included. Also, they develop the perturbation expansion of the grand partition function, whereas in this paper only the expansion of the relative-motion one-particle canonical partition function is needed for the second virial coefficient.

$$B_{2n\,\text{classical}} = -\frac{(-\beta)^n}{2n!}\int d^3r\, u(r)^n. \tag{13}$$

Also since $F_1 \equiv 1$, the first-order perturbation term has no diffraction effects.

The actual evaluation of Eq. (11) for any given potential is tractable only in second order since for $n = 2$ there is only one integration variable. This integration for a few examples is discussed in the next section. For higher order terms the evaluation of Eq. (11) is very difficult because it requires integration over the $n - 1$ wave vectors, $\mathbf{k}_1 \cdots \mathbf{k}_{n-1}$. Because of this complexity it may be wondered whether or not the integration is simpler in configuration space rather than in wave-number space. Two different forms may be obtained directly from Eq. (11). One of these is

$$\begin{aligned} B_{2n} = {} & -\tfrac{1}{2}(4\pi\lambda\!\!\!^{-2})^{3/2}(-\beta)^n \int \cdots \int d^3r_1 \cdots d^3r_n \\ & \times u(r_1) \cdots u(r_n) \int_0^1 dv_n \cdots \int_0^{v_2} dv_1 \\ & \times \frac{\exp\left[-(\mathbf{r}_1 - \mathbf{r}_n)^2/4\lambda\!\!\!^{-2}(1 - v_n + v_1)\right]}{[4\pi\lambda\!\!\!^{-2}(1 - v_n + v_1)]^{3/2}} \\ & \times \frac{\exp\left[-(\mathbf{r}_2 - \mathbf{r}_1)^2/4\lambda\!\!\!^{-2}(v_2 - v_1)\right]}{[4\pi\lambda\!\!\!^{-2}(v_2 - v_1)]^{3/2}} \cdots \\ & \times \frac{\exp\left[-(\mathbf{r}_n - \mathbf{r}_{n-1})^2/4\lambda\!\!\!^{-2}(v_n - v_{n-1})\right]}{[4\pi\lambda\!\!\!^{-2}(v_n - v_{n-1})]^{3/2}}. \end{aligned} \tag{14}$$

The details required to turn Eq. (11) into Eq. (14) are not given here since this form and its derivation are adequately discussed by Goldberger and Adams[9] and also by Green.[11] Unfortunately, the evaluation of B_{2n} in the form of Eq. (14) appears to be even harder than in the form of Eq. (11) since one must still integrate over the n vectors $\mathbf{r}_1 \cdots \mathbf{r}_n$ which represent the separation of the two particles at the "times" $v_1 \cdots v_n$.

The second form of B_{2n} as a configuration space integral is obtained by using the familiar representation of the three-dimensional delta function:

$$\delta(\mathbf{k}) = \frac{1}{(2\pi)^3}\int d^3r \exp(i\mathbf{k}\cdot\mathbf{r})$$

and noting that in the power series expansion of F_n the wave numbers $\mathbf{k}_1 \cdots \mathbf{k}_{n-1}$ become the differential operators, $i\nabla_1 \cdots i\nabla_{n-1}$. Thus, one obtains

$$B_{2n} = -\tfrac{1}{2}(-\beta)^n \int d^3r\, u_n(r) F_n(i\lambda\!\!\!^{-}\nabla_1, \cdots, i\lambda\!\!\!^{-}\nabla_{n-1}) \times u_1(r) \cdots u_{n-1}(r). \tag{15}$$

The subscripts 1 through n in Eq. (15) are for bookkeeping purposes. They may be erased after the differential operators $\nabla_1 \cdots \nabla_{n-1}$ in the expansion of F_n are applied, respectively, on $u_1 \cdots u_{n-1}$.

B_{2n} in the form (15) is not particularly useful for explicit evaluation, but it is useful as a means of obtaining the WK expansion from the perturbation expansion. For this derivation the expansion of F_n as defined by Eq. (12) in powers of $\lambda\!\!\!^{-2}$ is needed. The exponent of the integrand of Eq. (12) may be written as:

$$\begin{aligned} &\{[(v_2 - v_1)k_1^2 + \cdots + (v_n - v_{n-1})(\mathbf{k}_1 + \cdots + \mathbf{k}_{n-1})^2] \\ &\qquad - [(v_2 - v_1)\mathbf{k}_1 + \cdots + (v_n - v_{n-1}) \\ &\qquad \times (\mathbf{k}_1 + \cdots + \mathbf{k}_{n-1})]^2\} \\ &\qquad = \sum_{r=1}^{n-1} a_{nrr}k_r^2 + 2\sum_{r_2>r_1}^{n-1} a_{nr_2r_1}\mathbf{k}_{r_2}\cdot\mathbf{k}_{r_1}, \end{aligned}$$

where

$$a_{nr_2r_1} = (v_n - v_{r_2})[1 - (v_n - v_{r_1})].$$

The differences, $v_n - v_r$, are a measure of the "times" from the interactions with momentum transfer $\hbar\,\mathbf{k}_r$ to the final interaction. The multiple v integrations in Eq. (12) represent an average over the duration of these excitations. It may be shown that

$$\langle a_{nr_2r_1}\rangle_{\text{av}} \equiv \int_0^1 dv_n \cdots \int_0^{v_2} dv_1\, a_{nr_2r_1} = \frac{(r_1 + 1)(n - r_2)}{(n + 2)!}. \tag{16}$$

The expansion of $F_n(\lambda\!\!\!^{-}\mathbf{k}_1, \cdots, \lambda\!\!\!^{-}\mathbf{k}_{n-1})$ is

$$\begin{aligned} F_n &= \int_0^1 dv_n \cdots \int_0^{v_2} dv_1 \Big\{1 - \lambda\!\!\!^{-2}\Big[\sum_{r=1}^{n-1} a_{nrr}k_r^2 \\ &\qquad + 2\sum_{r_2>r_1}^{n-1} a_{nr_2r_1}\mathbf{k}_{r_2}\cdot\mathbf{k}_{r_1}\Big] + O(\lambda\!\!\!^{-4}) \cdots\Big\} \\ &= \frac{1}{n!} - \lambda\!\!\!^{-2}\Big[\sum_{r=1}^{n-1}\langle a_{nrr}\rangle_{\text{av}}k_r^2 \\ &\qquad + 2\sum_{r_2>r_1}^{n-1}\langle a_{nr_2r_1}\rangle_{\text{av}}\mathbf{k}_{r_2}\cdot\mathbf{k}_{r_1}\Big] + \cdots. \end{aligned} \tag{17}$$

When Eq. (17) is put into B_{2n} in the form of Eq. (15) and the summations over r_1 and r_2 carried out, one obtains

[11] H. S. Green, J. Chem. Phys. 20, 1274 (1952). In his work Green goes one step further and obtains a result for the intermediate temperature integrals. The result, however, does not appear useful for the remaining configuration space integrations required in the evaluation of Eq. (14).

$$B_{2n} = \sum_{m=0}^{\infty} B_{2n}^{(m)}$$

$$= -\tfrac{1}{2}\int d^3r\Big\{\frac{(-U)^n}{n!} - \frac{\lambda^2}{(n+2)!} \times \Big[\frac{(n-1)n(n+4)}{6}\nabla^2 U(-U)^{n-1} - \frac{(n-2)(n-1)n(n+5)}{12} \times (\nabla U)^2(-U)^{n-2}\Big] + \cdots\Big\}. \quad (18)$$

The $O(\lambda^{2m})$ term of the WK expansion is obtained by summing $B_{2n}^{(m)}$ for all orders of perturbation theory. Clearly $B_{2n}^{(0)}$ summed from $n = 1$ to ∞ gives the classical second virial coefficient. The method will be carried out explicitly only for the $O(\lambda^2)$ term. In order to obtain the usual form of the $O(\lambda^2)$ term it is necessary to integrate the $\nabla^2 U(-U)^{n-1}$ portion of Eq. (19) by parts. Since we are interested primarily in potentials which are singular at the origin, the integration by parts is done by excluding a sphere of radius δ about the origin. The result for $B_{2n}^{(1)}$ in Eq. (19) including a surface term from the integration by parts is:

$$B_{2n}^{(1)} = -\frac{\lambda^2}{2}\lim_{\delta\to 0}\Big\{\int \frac{d^3r(\nabla U)^2(-U)^{n-2}}{12(n-2)!} - \frac{(n-1)n(n+4)}{6(n+2)!}\,4\pi\,\delta^2 U'(\delta)(-U(\delta))^{n-1}\Big\}. \quad (19)$$

When Eq. (19) is summed over n, the surface term gives no contribution, and one obtains:

$$B_2 = -\tfrac{1}{2}\int d^3r\Big\{(e^{-U} - 1) - \frac{\lambda^2}{12}(\nabla U)^2 e^{-U} + \cdots\Big\} \quad (20)$$

in agreement with Eq. (1). In order to obtain higher-order terms in the WK expansion, one needs formulae analogous to Eq. (16) for powers and products of the $a_{nr_1r_2}$. These may be calculated readily but laboriously.

The usual method for the derivation of the WK expansion described in textbooks[12] follows the procedure used by Kirkwood[1] in which the Bloch equation, $\partial f/\partial\beta = -Hf$, is solved by iteration in powers of $\hbar\nabla$ subject to the condition that $f = \exp -\beta[p^2/2m + u(r)]$ at $\hbar = 0$. This method is very analogous to the WKB method for solving the Schrödinger equation. It is straightforward but very tedious. Other methods have been given by Goldberger and Adams,[9] Oppenheim and Ross,[13] Chester,[14] Siegert,[15] and Yaglom.[3] All these methods require considerable effort even to obtain the $O(\lambda^4)$ term. The very elegant method of Yaglom is probably the most useful as judged by the ease in which the $O(\lambda^4)$ term is obtained. In this method the solution of the Bloch equation is expressed as a Wiener functional integral which is expanded in a Taylor series in powers of λ.

The derivation given in the preceding paragraphs in which a given power of λ^2 in the perturbation expansion is summed, has been described in some detail, not as an addition to the list of methods of obtaining the WK expansion, but because of an important advantage that it has. With this method one may subtract out any number of lower orders of the perturbation expansion which have a finite classical limit and make a WK expansion on the remainder. This procedure must be used for potentials with r^{-2} and r^{-1} singularities at the origin.

III. ANALYTICITY OF SECOND-ORDER PERTURBATION THEORY AS A FUNCTION OF $\hbar^2$

In the previous section, three different forms were given for the nth-order perturbation term B_{2n}. The explicit evaluation of B_{2n} when $n > 2$ for any potential is in general a formidable task. The second-order term, however, is sufficiently simple that the evaluation may often be accomplished. In this section, B_{22} will be evaluated for a few potentials in order to exhibit the diffraction corrections, and to indicate the analyticity as a function of λ^2, hence also of $\hbar^2$. For the evaluation of B_{22} the momentum space form (11) is the easiest to use.

The unpleasant function $F_n(\lambda\mathbf{k}_1, \cdots, \lambda\mathbf{k}_{n-1})$ defined by Eq. (12) may be expressed in terms of known functions for $n = 2$. It is:

$$F_2(\lambda\mathbf{k}) = \tfrac{1}{2}\int_0^1 dv\, e^{-\kappa^2 v(1-v)},$$

$$= \tfrac{1}{2}(2/\kappa)e^{-\kappa^2/4}\,\mathrm{Erfi}\,(\kappa/2), \quad (21)$$

$$= \tfrac{1}{2}\sum_{s=1}\frac{(-1)^s\kappa^{2s}}{2^s(2s+1)!!},$$

where $\kappa = \lambda k$, and $\mathrm{Erfi}\,(a) = \int_0^a dt e^{t^2}$ is the imaginary

[12] I. D. Landau and E. M. Lifshitz, *Statistical Physics* (Pergamon Press, New York, 1958), pp. 96–100.

[13] I. Oppenheim and J. Ross, Phys. Rev. **107**, 28 (1957).
[14] G. V. Chester, Phys. Rev. **93**, 606 (1954).
[15] A. J. F. Siegert, J. Chem., Phys. **20**, 572 (1952).

error function. F_2 has the following asymptotic expansion for large real κ:

$$F_2 = \frac{1}{2}\left[\frac{2}{\kappa^2} + \frac{2^2}{\kappa^4} + \cdots + \frac{2^n(2n-3)!!}{\kappa^{2n}}\right]. \quad (22)$$

The expression (11) for the second-order term may also be found in the work of Montroll and Ward.[16] They develop their results for certain terms in the perturbation expansion of the partition function of a many-body system by using the pair interaction propagator. For Boltzmann statistics the pair propagator is

$$G(\lambda\mathbf{k}, \beta' - \beta'') = N\beta \exp[-\lambda^2 k^2 v(1-v)],$$

where N is the number of particles of the system and $v = (\beta' - \beta'')/\beta$. Because of the symmetry property of the propagator $G(\boldsymbol{\kappa}, \beta - (\beta' - \beta'')) = G(\boldsymbol{\kappa}, \beta' - \beta'')$, it may be expanded in Fourier series, $G = N\beta \sum_l L_l(\kappa^2) \exp(2\pi i l v)$. The Fourier components are:

$$L_l(\kappa^2) = \int_0^1 dv \exp[2\pi i l v - \kappa^2 v(1-v)]. \quad (23)$$

Thus, $F_2(\boldsymbol{\kappa})$ is the 0th component of the pair propagator $F_2 \equiv \frac{1}{2}L_0(\kappa^2)$.

The second-order perturbation term to be evaluated is:

$$B_{22} = -\frac{\beta^2}{4}\int_0^\infty \frac{4\pi k^2\, dk}{(2\pi)^3}(Vu(k))^2 L_0(\lambda^2 k^2). \quad (24)$$

Note from the series expansion (21) of L_0 that the integrand of Eq. (24) is an analytic function of λ^2. The resulting function of λ after the integration is not necessarily analytic in λ^2.

As the first example we consider B_{22} for the exponential potential, $u(r) = g_0 e^{-r/r_0}$. Its Fourier transform is $Vu(k) = 8\pi r_0^3 g_0/(1 + k^2 r_0^2)^2$. After changing the integration variable to $x = kr_0$, B_{22} becomes:

$$B_{22} = -\frac{r_0^3(8\pi\beta g_0)^2}{8\pi^2}\int_0^\infty \frac{x^2\, dx L_0(\gamma^2 x^2)}{(x^2+1)^4} \quad (25)$$

with $\gamma = \lambda/r_0$. The diffraction effects may not be obtained by integrating term by term the expansion of $L_0(\gamma^2 x^2)$ since all terms beyond $O(\gamma^4)$ are divergent. Integrals of this type may be evaluated in the following manner:

$$\begin{aligned} H_r(\gamma) &= \int_0^\infty \frac{x^2\, dx L_0(\gamma^2 x^2)}{(x^2+1)^r}, \\ &= \int_0^\infty x^2\, dx L_0(\gamma^2 x^2) \\ &\qquad \times \frac{(-1)^{r+1}}{(r-1)!}\frac{d^{r-1}}{da^{r-1}}\frac{1}{(x^2+a)}\bigg|_{a=1}, \\ &= \frac{(-1)^{r+1}}{(r-1)!}\frac{d^{r-1}}{da^{r-1}} aH(\gamma, a)\bigg|_{a=1}, \end{aligned} \quad (26)$$

with

$$\begin{aligned} H(\gamma, a) &= \int_0^\infty \frac{dx\, L_0(\gamma^2 x^2)}{x^2 + a} \\ &= \int_0^1 dv \int_0^\infty \frac{dx \exp[-\gamma^2 x^2 v(1-v)]}{x^2 + a}, \\ &= \int_0^1 dv\, \theta\left\{\frac{\pi}{2}\frac{e^{a\theta^2}}{a^{1/2}\theta} - \sqrt{\pi}\, e^{a\theta^2}\frac{\mathrm{Erf}(a^{1/2}\theta)}{a^{1/2}\theta}\right\}, \end{aligned} \quad (27)$$

where $\theta^2 = \gamma^2 v(1-v)$.[17] After carrying out the differentiations indicated in Eq. (26), one obtains for Eq. (25):

$$\begin{aligned} B_{22} &= -\frac{2\pi}{3} r_0^3(\beta g_0)^2 \\ &\times \left\{\sum_{m=0} \frac{(m+1/2)(m-1/2)(m-3/2)\gamma^{2m}}{2^m(2m+1)!!} \right. \\ &\left. - \frac{(\pi)\gamma^5 e^{\gamma^2/4}}{64}\right\}. \end{aligned} \quad (28)$$

This result (28) consists of an expected analytic function of γ^2 plus another analytic function multiplied by the nonanalytic form $(\gamma^2)^{1/2}$. The summation in Eq. (28) is the expansion of

$$3/8 - \gamma^2/16 + \gamma^4/32 + (\gamma^6/64)L_0(-\gamma^2),$$

where $L_0(-\gamma^2)$ as defined by Eq. (23) for imaginary argument is $(2/\gamma)e^{\gamma^2/4}\,\mathrm{Erf}\,(\gamma/2)$. Consequently, the odd and even powers of γ in the braces of Eq. (28) may be combined into one function:

$$\begin{aligned} \{\ \} &= 3/8 - \gamma^2/16 + \gamma^4/32 \\ &\quad - (\gamma^5/32)e^{\gamma^2/4}\,\mathrm{Erfc}\,(\gamma/2). \end{aligned} \quad (29)$$

As a contrast to the exponential potential

[16] E. W. Montroll and J. C. Ward, Phys. Fluids 1, 55 (1958). For a more complete discussion of the properties of propagators and their Fourier components, see the article by Montroll in *La theorie des gaz neutres et ionises* (Hermann et Cie, Paris, 1960). In place of $F_2(\lambda\mathbf{k})$ they obtain $1/2\sum_l L_l^2(\beta\hbar^2k^2/2m)$ where m is the mass of one particle. An easily proved identity valid for classical statistics, $\sum_l L_l^2(x^2) = L_0(2x^2)$, establishes the correspondence between their form and Eq. (21) of this paper. The factor of 2 in $L_0(2x^2)$ represents the change to the reduced mass of the two interacting particles, i.e., $2\beta\hbar^2k^2/2m = \beta\hbar^2k^2/2\mu = \lambda^2k^2$.

[17] The x integral in Eq. (27) after the change of variable $y = x^2$., is the Laplace transform of $y^{-1/2}(y+a)^{-1}\exp(-\theta^2 y)$ and is $(\pi/2)e^{1/2}e^{a\theta^2}\,\mathrm{Erfc}(a^{1/2}\theta)$. See *Bateman Project Tables* (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I, p. 136, No. 24. In Eq. (27) we have used $\mathrm{Erfc}\,(y) = \pi^{1/2}/2 - \mathrm{Erf}\,(y)$.

we next consider the Gaussian potential $u(r) = g_0 e^{-r^2/r_0^2}$, which has the Fourier transform $Vu(k) = \pi^{3/2} r_0^3 g_0 \exp -(kr_0)^2/4$. Since $u(k)$ is also Gaussian, B_{22} may be evaluated by expanding $L_0(\gamma^2 x^2)$ and integrating term by term. One obtains:

$$\begin{aligned} B_{22} &= -\frac{(\beta g_0)^2 r_0^3}{8\pi^2} \int_0^\infty dx\, x^2 (\pi^{3/2} e^{-x^2/4})^2 L_0(\gamma^2 x^2), \\ &= -\frac{\pi}{8} (\beta g_0)^2 r_0^3 \sum_{m=0} \frac{(-1)^m \gamma^{2m}}{2^m (2m+1)!!} \\ &\qquad \times \int_0^\infty dx\, x^{2m+2} e^{-x^2/2}, \\ &= -\frac{(2\pi)^{3/2}}{32} \frac{(\beta g_0)^2 r_0^3}{1+\gamma^2/2}, \end{aligned} \tag{30}$$

which is an analytic function of γ^2.

The reason for the different analyticity properties of the two forms (28) and (30) for the two potentials, lies in the behavior as $r \to 0$. The Gaussian form and all of its derivatives are smooth as $r \to 0$, whereas the exponential form has a cusp at $r = 0$. As functions of a complex variable z, one notes that e^{-z^2} is analytic at $z = 0$, while $e^{-|z|}$ is nonanalytic since its first derivative is discontinuous at $z = 0$. In the Fourier transform the cusp of e^{-r/r_0} is manifested by the second-order pole of $u(k) \propto (k^2 + 1/r_0^2)^{-2}$ at i/r_0.

In general, the second-order perturbation term will be analytic function $\hbar^2$ for any potential that is smooth at $r = 0$, for example, $r^m e^{-r^2/r_0^2}$, while some nonanalytic form of $\hbar^2$ will appear for any potential that has a cusp in any derivative. For example, $r^m e^{-r/r_0}$ has a cusp in the mth derivative at $r = 0$. (Its Fourier transform has a pole of order $m + 2$ at i/r_0.) B_{22} can be evaluated for $u(r) \propto r^m e^{-r/r_0}$ for any integer m using Eq. (26); the first nonanalytic form to appear is of order γ^{4m+5}.

An interesting example of a potential with a cusp not at $r = 0$ is the square barrier:

$$\begin{aligned} u(r) &= g_0 \qquad r < r_0 \\ &= 0 \qquad r > r_0 \end{aligned}$$

which has the Fourier transform

$$Vu(k) = 4\pi g_0 r_0^3 (kr_0)^{-1} j_1(kr_0),$$

where $j_1(x)$ is a spherical Bessel function. The second-order perturbation term for this potential is:

$$\begin{aligned} B_{22} &= -\frac{r_0^3 (4\pi\beta g_0)^2}{8\pi^2} \int_0^\infty dx\, j_1^2(x) L_0(\gamma^2 x^2) \\ &= -\pi r_0^3 (\beta g_0)^2 [1/3 - (\gamma/4)\,\mathrm{Erf}\,(2/\gamma) \\ &\quad + (\gamma^3/32)\,\mathrm{Erf}\,(2/\gamma) - (\gamma^2/16) e^{-4/\gamma^2}]. \end{aligned} \tag{31}$$

Some details of the integration required to obtain Eq. (31) are given in the Appendix. For small γ the expression in brackets in Eq. (31) becomes $[\,] = \frac{1}{3} - \pi^{1/2}\gamma/8 + \pi^{1/2}\gamma^3/64$. Again the diffraction corrections to the classical result are nonanalytic in γ^2.

Next we consider the screened Coulomb potential $u(r) = (g_1/r) e^{-r/r_0}$, which has the Fourier transform $Vu(k) = 4\pi g_1 r_0^2/(k^2 + 1/r_0^2)$. This potential has not just a cusp, but an infinite spike at $r = 0$. According to the rule discussed in a previous paragraph a result that is nonanalytic in γ^2 is to be expected; the first nonanalytic diffraction correction should be of $O(\gamma)$. Also, as was discussed in the Introduction, nonanalytic behavior for the $1/r$ singularity is indicated even by the WK expansion parameter η^2 which increases linearly with temperature. The integration of B_{22} for this potential is easily worked out by using Eq. (26). The result is

$$\begin{aligned} B_{22} &= -\frac{r_0^3 (4\pi\beta g_1/r_0)^2}{8\pi^2} \int_0^\infty \frac{x^2\, dx\, L_0(\gamma^2 x^2)}{(x^2+1)^2} \\ &= -\tfrac{1}{2}\pi r_0^3 (\beta g_1/r_0)^2 \{[1 + (\gamma^2/2) L_0(-\gamma^2)] \\ &\qquad - \pi^{1/2} (\gamma/2) e^{\gamma^2/4}\}. \end{aligned} \tag{32}$$

Equation (32) has the expected form similar to the result for the exponential potential, i.e., the power-series expansion of the function in braces contains both even and odd powers of γ. By using the definition of $L_0(-\gamma^2)$ in terms of the error function, the expression in braces of Eq. (32) may be written as

$$\{\ \} = 1 - \gamma e^{\gamma^2/4}\,\mathrm{Erfc}\,(\gamma/2). \tag{33}$$

The second form (33) is convenient for obtaining an asymptotic expansion for large γ; it begins with $2/\gamma^2$. This limit means $\lambda\!\!\!^{-} \gg r_0$ and is not physically interesting, since quantum statistics have not been considered.

Some remarks about the electron gas at finite temperature are in order at this point. In the electron gas the interaction potential is the unscreened Coulomb potential e^2/r. Electrical neutrality is maintained by the assumption of a continuous background of positive charge equal to the charge of N electrons in a volume V. Since every term of the perturbation expansion of the partition function in powers of e^2/r is divergent (because of the infinite range of the interaction), finite results for the free energy are obtained by selective summation of terms in perturbation expansion. It is well known that the sum of the most divergent part of each cluster integral, the sum of the ring diagrams, gives the Debye-Hückel free energy. The fundamental lengths of the electron gas are $l = \beta e^2$, the Debye screening length $\lambda_D = (4\pi\beta e^2 \rho)^{-1/2}$ which replaces r_0, and the thermal wavelength. The free energy of the classical gas is a function of the ratio of the two classical lengths $\Lambda = \beta e^2/\lambda_D = 2\pi^{1/2}\beta^{3/2}\rho^{1/2}e^3$.

The Debye-Hückel contribution to $\beta F = -\ln Z$ is $-N\Lambda/3$. It is the leading interaction term when $\Lambda \gg 1$; note that it is nonanalytic in ρ and e^2, i.e., Λ involves $\rho^{1/2}$ and $(e^2)^{3/2}$. Diffraction corrections will be in a function of the ratio $\gamma = \lambda\!\!\!^{-}/\lambda_D$ multiplying the classical Debye term. The WK expansion cannot be used to find the diffraction corrections since the WK expansion parameter, $\eta^2 = (\lambda\!\!\!^{-}/\beta e^2)^2$, diverges as β^{-1} at high temperature. Instead, the diffraction corrections must be found by an evaluation of the quantum mechanical ring sum.[16,18] There is an analogy between the quantum ring sum for the electron gas and the second-order perturbation term for the static screened Coulomb potential. However, the mathematical expression for the ring sum is far more complicated than B_{22}, and important additional physical effects due to plasma oscillations are described by it. Since the Coulomb potential has a spike at $r = 0$ [and its Fourier transform, $Vu(k) = 4\pi e^2/k^2$, has a double pole at $k = 0$], it is to be expected that the function of γ^2 multiplying the classical Debye term will be nonanalytic in γ^2 in exactly the same manner as Eq. (32) is nonanalytic. This nonanalyticity, the appearance of both even and odd powers of γ in the diffraction corrections to the Debye term, has already been reported.[19] The explicit evaluation will be given in a forthcoming publication. Because of the complexity of the mathematical expressions in the quantum ring sum, it is not possible to obtain the diffraction corrections in closed form as in Eqs. (32) or (33), but instead only as two convergent series, one involving γ^{2m} and the other $(\gamma^2)^{m+1/2}$.

IV. EVALUATION OF TERMS IN THE WK EXPANSION

In this section the evaluation of a few terms of the WK expansion will be described for the singular potential $\beta u(r) = (\beta g_p/r^p)e^{-r/r_0} = x^{-p}e^{-\Lambda x}$ with $x = r/l$. Specifically, we need the coefficients $C_m(\Lambda)$ of η^{2m} in Eq. (5). The coefficients C_0, C_1, and C_2 have been evaluated for the Lennard-Jones potential in the form of infinite series of gamma functions, and for other more complicated potentials used in the theory of nonideal gases they have been evaluated numerically.[4] The usefulness of these results is somewhat limited by the fact that little is known about the convergence of the WK expansion. The simple singular potential to be discussed here does not correspond well to any physical problem, but the results do show the dependence of the coefficients $C_m(\Lambda)$ on the order of the singularity and thus give a little more information about the convergence of the expansion.

The coefficients in Eq. (5) may be evaluated readily by the use of the Mellin transform, an elegant and useful method in statistical mechanics recently pointed out by Iwata.[20] The Mellin transformation of the exponential series is

$$\sum_{n=r} \frac{(-U)^n}{n!} = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} ds\ \Gamma(s)U^{-s} \qquad -r < \sigma < -(r-1). \tag{34}$$

In our use of this transform, the contour of integration is deformed to enclose the entire negative real axis to the left of $-(r-1)$. For the evaluation of the $C_m(\Lambda)$, the exponential e^{-U} in the integrand is expanded with Eq. (34) and the order of x and s integration inverted.

The classical second virial coefficient, $C_0(\Lambda)$ from Eq. (6) is

$$C_0(\Lambda) = \frac{1}{2\pi i}\int_C ds\ \Gamma(s)\int_0^\infty x^2\,dx\,(x^{-p}e^{-\Lambda x})^{-s} \tag{35}$$

subject to $-1 < \mathrm{Re}(s) < 0$ since $e^{-U} - 1$ is being expanded. With the change of variable $y = -\Lambda s x$, the x integral in Eq. (35) becomes a gamma function, and the result is:

$$C_0(\Lambda) = \frac{1}{2\pi i}\int_C ds\ \Gamma(s)\ \Gamma(ps+3)(-s\Lambda)^{-(ps+3)}. \tag{36}$$

The result is obtained by summing the residues of all poles to the left of $s = 0$. The integrand has first-order poles when $ps + 3 = 0, -1, -2, \cdots, -t$ but s is not an integer; it has second-order poles when $s = -1, -2, \cdots$. The residues of the first-order poles are $O(\Lambda^t)$ while the residues of the second-order poles are nonanalytic in Λ. The complete result is:

$$\begin{aligned} C_0(\Lambda) = {} & \frac{\pi}{p}\sum_{\gamma_p=0}^{p-1}{}' \frac{(-1)^{\gamma_p+1}}{\sin \pi(\gamma_p+3)/p} \\ & \times \sum_{t=0}^{\infty} \frac{(-1)^{t(p+1)}\{[t+(\gamma_p+3)/p]\Lambda\}^{pt+\gamma_p}}{\Gamma[t+1+(\gamma_p+3)/p]\Gamma(1+pt+\gamma_p)} \\ & + \sum_{t=0} \frac{(-1)^{(p+1)t+p-1}[(t+1)\Lambda]^{pt+p-3}}{\Gamma(t+2)\Gamma(pt+1+p-3)} \\ & \times \Big\{\ln(t+1)\Lambda + 1 - 3/p(t+1) \\ & - \frac{\Gamma'(t+2)}{p\Gamma(t+2)} - \frac{\Gamma'(pt+p-2)}{\Gamma(pt+p-2)}\Big\}, \end{aligned} \tag{37}$$

[18] H. E. DeWitt, J. Nuclear Energy, Part C; Plasma Physics 2, 27 (1961).
[19] H. E. DeWitt, Bull. Am. Phys. Soc., 5, 7 (1960).
[20] G. Iwata, Progr. Theoret. Phys. (Kyoto) 24, 1118 (1960).

where $\gamma_p = 0, 1, 2, \cdots, p-4, p-2, p-1$. The prime on the summation indicates that the value $\gamma_p = p-3$ is to be excluded; for this value the integrand has double poles, and the second summation in Eq. (37) gives these residues. This expression (37) is general for $p \geq 3$. For $p = 2$, however, in addition to Eq. (37) there is a residue from the simple pole at $s = -1$ which is $-\Lambda^{-1}$. This additional term is the first-order perturbation term, i.e., $\int_0^\infty x^2\, dx(-x^{-2}e^{-\Lambda x}) = -\Lambda^{-1}$. Similarly, for the screened Coulomb potential, $p = 1$, the first two orders of perturbation theory are finite and are given by the residues of the simple poles at $s = -1$ and -2.

The coefficient of the first diffraction correction, $C_1(\Lambda)$, is evaluated by the same method. It is:

$$C_1(\Lambda) = \frac{1}{12}\int_0^\infty x^2\, dx \exp\left(-x^{-p}e^{-\Lambda x}\right)\left(\frac{d}{dx}\, x^{-p}e^{-\Lambda x}\right)^2,$$

$$= \frac{1}{12}\int_C \frac{ds}{2\pi i}\,\Gamma(s)\int_0^\infty x^2\, ds(x^{-p}e^{-\Lambda x})^{-s}$$

$$\times [p^2x^{-(2p+2)} + 2p\Lambda x^{-(2p+1)} + \Lambda^2 x^{-2p}]. \quad (38)$$

The contour for Eq. (38) crosses the real s axis to the right of $s = 0$. With the change of variable $y = -(s-2)\Lambda x$, Eq. (38) becomes

$$C_1(\Lambda) = \frac{1}{12}\int_C \frac{ds}{2\pi i}\,\Gamma(s)$$

$$\times \{p^2[-(s-2)\Lambda]^{-(ps-2p+1)}\Gamma(ps-2p+1)$$

$$+ 2p\Lambda[-(s-2)\Lambda]^{-(ps-2p+2)}\Gamma(ps-2p+2)$$

$$+ \Lambda^2[-(s-2)\Lambda]^{-(ps-2p+3)}\Gamma(ps-2p+3)\}. \quad (39)$$

The first few terms in Λ are

$$C_1(\Lambda) = \frac{1}{12}\left\{[p\Gamma(2-1/p) - (\Lambda^2/2p)\right.$$

$$\times \Gamma(2-3/p) + \cdots] + \frac{(-1)^{3(p+1)}(3\Lambda)^{3p-1}}{\Gamma(3p)\Gamma(2)}$$

$$\left.\times\left[\ln 3\Lambda + 1 - 1/3p - \frac{\Gamma'(2)}{p\Gamma(2)} - \frac{\Gamma'(3p)}{\Gamma(3p)}\right] + \cdots\right\}. \quad (40)$$

The complete result for Eq. (39) is easily obtained by summing all residues from the first- and second-order poles of the integrand. It is a lengthy result and is not written down since it is not needed.

The coefficient of η^4 as defined by Eq. (1) is:

$$C_2(\Lambda) = \frac{1}{1440}\int_0^\infty x^2\, dx\, e^{-U}$$

$$\times [((\nabla_x U)^2)^2 - 8(\nabla_x U)^2\nabla_x^2 U + 12(\nabla_x^2 U)^2]. \quad (41)$$

A more convenient form for computational purposes is obtained by using $\nabla_x^2 U = U'' + (2/x)U'$, and then integrating by parts the terms U'^2U'' and $U'U''/x$. The result is

$$C_2(\Lambda) = \frac{1}{120}\int_0^\infty x^2\, dx\, e^{-U}$$

$$\times\left[-\frac{5}{36}U'^4 + \frac{10}{3}\frac{U'^3}{x} + \frac{2U'^2}{x^2} + U''^2\right]. \quad (42)$$

With $U = x^{-p}e^{-\Lambda x}$, the same procedure used for $C_1(\Lambda)$ gives, after some algebra,

$$C_2(\Lambda) = \frac{1}{1440}\int_C \frac{ds}{2\pi i}\,\Gamma(s)\{-\tfrac{5}{3}[p^4\Gamma(ps-4p-1)[-(s-4)\Lambda]^{-(ps-4p-1)}$$

$$+ 4p^3\Lambda\Gamma(ps-4p)[-(s-4)\Lambda]^{-(ps-4p)} + 6p^2\Lambda^2\Gamma(ps-4p+1)[-(s-4)\Lambda]^{-(ps-4p+3)}$$

$$+ 4p\Lambda^2\Gamma(ps-4p+2)[-(s-4)\Lambda]^{-(ps-4p+2)} + \Lambda^4\Gamma(ps-4p+3)[-(s-4)\Lambda]^{-(ps-4p+3)}]$$

$$- (40/3)[p^3\Gamma(ps-3p-1)[-(s-3)\Lambda]^{-(ps-3p-1)} + 3p^2\Lambda\Gamma(ps-3p)[-(s-3)\Lambda]^{-(ps-3p)}$$

$$+ 3p\Lambda^2\Gamma(ps-3p+1)[-(s-3)\Lambda]^{-(ps-3p+1)} + \Lambda^3\Gamma(ps-3p+2)[-(s-3)\Lambda]^{-(ps-3p+2)}]$$

$$+ 12[p^2((p+1)^2+2)\Gamma(ps-2p-1)[-(s-2)\Lambda]^{-(ps-2p-1)} + (4p^2(p+1)+4p)$$

$$\times \Lambda\Gamma(ps-2p)[-(s-2)\Lambda]^{-(ps-2p)} + (2p(p+1)+4p^2+2)\Lambda^2\Gamma(ps-2p+1)$$

$$\times [-(s-2)\Lambda]^{-(ps-2p+1)} + 4p^2\Lambda^2\Gamma(ps-2p+2)[-(s-2)\Lambda]^{-(ps-2p+2)}$$

$$+ \Lambda^4\Gamma(ps-2p+3)[-(s-2)\Lambda]^{-(ps-2p+3)}]\}. \quad (43)$$

Each term in this lengthy expression may be evaluated by summing the residues of first- and second-order poles. We give only the value of the leading term:

$$C_2(\Lambda) = \frac{p(2p^2-11p+21)}{1440}\,\Gamma(2+1/p) + \cdots. \quad (44)$$

Collecting the previous results gives the second virial coefficient for $U = x^{-p}$ valid for $p > 3$ as

$$B_2 = -2\pi l^3\Big\{-\frac{\pi}{p\sin 3\pi/p}\frac{1}{\Gamma(1+3/p)} - \frac{p\eta^2}{12}\Gamma(2-1/p) + \frac{(2p^3-11p^2+21p)\eta^4}{1440} \times \Gamma(2+1/p) + \cdots\Big\}. \quad (45)$$

Although it has not been possible to obtain a general term for this expansion, it is clear that the form of the general term for large p is $p^{2m-1}\eta^{2m}$. Thus, the convergence of the WK expansion for any given value of η^2 depends strongly on the order of the singularity of the repulsive core of the potential.

The limit of large p is interesting because the potential $g_p r^{-p}$ becomes equivalent to a hard sphere with radius $r_0 = \lim (\beta g_p)^{1/p}$ as $p \to \infty$. The first term in Eq. (45) reduces to $(2/3)\pi r_0^3$, the classical hard-sphere second virial coefficient. Thus, WK expansion when fully evaluated could give the diffraction corrections to the hard sphere second virial coefficient at high temperature. For large p Eq. (45) becomes:

$$B_2 = -2\pi r_0^3\{-1/3 - \eta[p\eta/12 - (p\eta)^3/720 + \cdots]\}, \quad (46)$$

where $\eta = \lambda\!\!\!^{-}/r_0$. It appears from the numerical values of the first two diffraction terms in Eqs. (45) and (46) that the WK expansion is a convergent series in powers of $p\eta$, although nothing can be said about the radius of convergence. It is possible that the limit of the square bracket in Eq. (46) as $p \to \infty$ is finite and nonzero, in which case the diffraction corrections to the hard sphere virial coefficient are nonanalytic in η^2. This result seems very probable in view of the nonanalytic result (31) for B_{22} with a barrier potential of finite height. It should be noted that most recent work on the quantum-mechanical hard sphere gas has been at low temperature so that $\lambda\!\!\!^{-}$ is much greater than the hard sphere radius. Thus, in the work of Yang and Lee[21] the expansion parameter is $r_0/\lambda\!\!\!^{-}$, rather than $\lambda\!\!\!^{-}/r_0$. We hope to study the hard sphere gas at high temperature, $\lambda\!\!\!^{-} \ll r_0$, in more detail in a later publication.

V. THE SCREENED COULOMB POTENTIAL

The screened Coulomb potential must be considered separately from the more singular potentials treated in the previous section since the WK expansion parameter η^2 is large at high temperature. The diffraction corrections to the classical limit of the second virial coefficient for this potential must be expressed in powers of $\gamma = \eta\Lambda = \lambda\!\!\!^{-}/r_0$, which goes to zero as $\beta^{1/2}$ at high temperature. The classical value of B_2 is obtained from Eq. (37) with $p = 1$ and with the additional residues of the two simple poles at $s = -1$ and -2. It is:

$$B_2 \text{ classical} = -2\pi(\beta g_1)^3 C_0(\Lambda) = -2\pi r_0^3\Big\{-\Lambda + \tfrac{1}{4}\Lambda^2 + \sum_{r=1}\frac{\Lambda^{r+2}(r+2)^{r-1}}{\Gamma(r+3)\Gamma(r)} \times \Big[\ln(r+2)\Lambda + 2C - 2h_r + \frac{r^3+2}{r(r+1)(r+2)} - \frac{1}{r+2}\Big]\Big\}. \quad (47)$$

In obtaining Eq. (47) from Eq. (37) the relation $\Gamma'(r+1)/\Gamma(r+1) = -C + h_r$ with $h_r = 1 + \frac{1}{2} + \cdots + 1/r$ has been used. The first two terms of Eq. (47) are the first and second orders of the perturbation expansion (from the residues of the simple poles at $s = -1$ and -2). The higher orders of the perturbation expansion are individually infinite, but their sum gives the nonanalytic form $\Lambda^{r+2}\ln\Lambda$. The summation in Eq. (47) is identical with Iwata's[20] result, the S_2 integral of Abe's[22] modified cluster expansion for the classical electron gas.

This section is devoted to obtaining diffraction corrections to Eq. (47). The first-order perturbation term is always classical. The diffraction corrections to the second-order term were obtained in closed form in Sec. III, Eqs. (32) and (33) and found to be nonanalytic in γ^2, i.e., they involve both even and odd powers of γ. Our problem then is to find diffraction corrections to the $\Lambda^{r+2}\ln\Lambda$ terms in Eq. (47). One conceivable approach is to evaluate every order of the quantum perturbation expansion and sum them. (The third order begins with $\Lambda^3\ln\gamma$ and the higher orders with Λ^n/γ^{n-3}.) Such an approach is approximately as difficult as solving a quantum mechanical scattering problem by calculating the nth order Born approximation and summing the Born series. Instead, it will be shown how the WK expansion may be used, even though at first glance the $1/r$ singularity appears to be too weak to allow the WK expansion.

[21] T. D. Lee and C. N. Yang, Phys. Rev. **105**, 1119 (1957).

[22] R. Abe, Progr. Theoret. Phys. (Kyoto) **22**, 213 (1959). A similar expansion for ionic solutions was developed by H. F. Friedman, Molecular Phys. **2**, 23 (1959). Abe's giant cluster expansion for the electron gas is a special case of the very general Meeron nodal expansion, E. Meeron, Phys. Fluids **1**, 139 (1958); E. Meeron and E. R. Rodemich, *ibid.* **1**, 246 (1958).

Let us first consider what happens when the screened Coulomb potential is used blindly in the evaluation of the WK expansion coefficients $C_m(\Lambda)$. The general form may be shown to be:

$$C_m(\Lambda) = \sum_{s=0}^{2m-1} a_{ms}\Lambda^s + \Lambda^{2m-1} \times \sum_{r=1}^{\infty} \Lambda^r(b_{mr} \ln \Lambda + c_{mr}). \quad (48)$$

The calculation of the coefficients a_{ms}, b_{mr}, and c_{mr} in Eq. (48) is feasible in any order with such expressions as Eq. (39) and (43), but is very tedious even for $m = 2$. For example, the complete result for $C_1(\Lambda)$ obtained from Eq. (39) with $p = 1$ is:

$$C_1(\Lambda) = \frac{1}{12}(1 - \Lambda/2) - \frac{1}{12}\sum_{r=1}^{\infty} \frac{\Lambda^{r+1}(r+2)^{r-1}}{\Gamma(r+2)\Gamma(r)} \left[\ln (r+2)\Lambda + 2C - 2h_r - \frac{r^3+2}{r(r+1)(r+2)}\right]. \quad (49)$$

The series in Eq. (49) closely resembles the series in the classical expression (47). Both series are rapidly convergent. The limit of $C_1(\Lambda)$ as $\Lambda \to 0$ is 1/12.

The infinite sum in Eq. (48) comes from the residues of second-order poles, and the finite sum comes from the $2m$ simple poles all lying to the right of the second-order poles on the real s axis. The diffraction corrections from the finite sum are of order $\eta^{2m}\Lambda^s$ and have temperature dependence of β^{-m+s}; thus they become large at high temperature for $s < m$. The diffraction corrections from the infinite sum, however, are of order $\eta^{2m}\Lambda^{2m-1+r} = \gamma^{2m}\Lambda^{r-1}$ and are small at high temperature. In fact, the finite sum in Eq. (48) contributes only to the second-order perturbation term B_{22}, and the divergence at high temperature of $\eta^{2m}\Lambda^s$ is just what is needed to give the nonanalytic form $(\gamma^2)^{1/2}$ which appears in B_{22}. The infinite sum gives the quantum corrections to the third and higher orders of the perturbation expansion, i.e., the desired diffraction corrections to the $\Lambda^{r+2} \ln \Lambda$ terms in Eq. (47). These diffraction corrections are analytic in γ^2.

There is no point in giving a direct proof that $\sum_m \eta^{2m}C_m(\Lambda)$ with only the finite sum part of Eq. (48) does indeed reproduce B_{22}. Instead, the proper procedure is to subtract out of the second virial coefficient the first and second orders of the perturbation expansion, and make a WK expansion of the remainder. Thus we define:

$$B_2' = \sum_{n=3} B_{2n} = -2\pi(\beta g_1)^3 \sum_{m=0} \eta^{2m}C_m'(\Lambda), \quad (50)$$

and use Eq. (18) for B_{2n}. The coefficients $C_m'(\Lambda)$ where the prime indicates the removal of first- and second-order perturbation theory are defined as:

$$-2\pi(\beta g_1)^3\eta^{2m}C_m'(\Lambda) = \sum_{n=3} B_{2n}^{(m)}.$$

In Sec. II, only $B_{2n}^{(0)}$ and $B_{2n}^{(1)}$ were explicitly evaluated. By summing Eq. (18) and (19) from $n = 3$ one obtains for the modified second virial coefficient

$$B_2' = -\frac{1}{2}\left\{\int d^3r\, [(e^{-U} - 1 + U - U^2/2) - (\chi^2/12)(\nabla U)^2(e^{-U} - 1) + O(\chi^4)] + (\chi^2/12)\lim_{\delta\to 0} 4\pi\, \delta^2 U'(\delta)U(\delta)\right\}, \quad (51)$$

instead of Eq. (20). The surface term in Eq. (19) must be retained in order that the $O(\chi^2)$ term in Eq. (51) be finite. The singularity in the integral coefficient of χ^2 is canceled by the surface term for the $1/r$ potential. The terms of the WK expansion of B_2' are calculated as described in the previous section with the help of the Mellin transform of the exponential series. Thus $C_0'(\Lambda)$ is defined as in Eq. (35), but the condition on the contour of the s integration is $-3 < \mathrm{Re}(s) < -2$ where the contour crosses the real axis. Thus the simple poles at $s = -1$ and -2 are not included and the result for $C_0'(\Lambda)$ is the infinite sum in Eq. (47). Similarly, $C_1'(\Lambda)$ is given by Eq. (39), but with the restriction that $-1 < \mathrm{Re}(s) < 0$ where the contour crosses the real axis. Again this restriction eliminates the simple poles and leaves only the second-order poles within the contour. Thus $C_1'(\Lambda)$ is equal to the infinite sum in Eq. (49). Similarly, for arbitrary m the subtraction of second-order perturbation theory leaves only the second-order poles within the contour, and $C_m'(\Lambda)$ is equal to the infinite sum indicated in Eq. (48).

In this paper only the $O(\chi^2)$ corrections to B_2' have been evaluated explicitly. The $O(\chi^4)$ corrections may be obtained as the residues of the second-order poles of Eq. (43) with $p = 1$. Higher-order corrections must await the evaluation of more terms of the WK expansion. Our complete result for B_2 for the screened Coulomb potential including the second-order term is

$$
\begin{aligned}
B_2 &= B_{21} + B_{22} + B_2' \\
&= -2\pi r_0^3 \Big\{ -\Lambda + \tfrac{1}{4}\Lambda^2 [1 + \tfrac{1}{2}\gamma^2 L_0(-\gamma^2) \\
&\quad - \tfrac{1}{2}\pi^{1/2}\gamma e^{\gamma^2/4}] + \sum_{r=1} \frac{\Lambda^{r+2}(r+2)^{r-1}}{(r+2)!\,(r-1)!} \\
&\quad \times \Big[\ln (r+2)\Lambda + 2C - 2h_r \\
&\quad + \frac{r^2+2}{r(r+1)(r+2)} - \frac{1}{r+2} \Big] \\
&\quad + \gamma^2 \sum_{r=1} \frac{\Lambda^{r+2}(r+2)^r}{(r+2)!\,(r-1)!} \\
&\quad \times \Big[\ln (r+2)\Lambda + 2C - 2h_r \\
&\quad + \frac{r^2+2}{r(r+1)(r+2)} \Big] + O(\gamma^4) \Big\}. \qquad (52)
\end{aligned}
$$

In Sec. III it was pointed out that the quantum mechanical ring sum for the electron gas and B_{22} for the screened Coulomb potential were rather similar. Diffraction corrections to both are nonanalytic in γ^2. In the same way there is a considerable similarity between B_2' for the screened Coulomb potential and the quantum mechanical generalization of the Abe $S_2(\Lambda)$ contribution to the electron gas free energy.[22] The S_2 term is the next step in the rearrangement of the perturbation expansion of the partition function after the ring terms have been grouped together. It is the sum of three and more effective interactions between two electrons in the plasma. Each effective interaction is the sum of all possible chains of Coulomb interactions; the result is a screened Coulomb interaction with $r_0 = \lambda_D$ in the classical limit. Thus the classical form of S_2, Abe's result, is identical in form to B_2' for the screened Coulomb potential.

The quantum theory of S_2 has not been completely developed yet, although it is implicit in the article by Montroll.[16] It seems clear, however, that diffraction corrections to S_2 must be calculated in the same manner that B_2' in Eq. (52) was obtained, that is, by a WK expansion of S_2. This calculation is being carried out now.

VI. CONCLUDING REMARKS

The main point of this article has been to show with specific examples that nonanalytic forms of $\hbar^2$ may appear in the diffraction corrections to the classical partition function of an interacting gas for some potentials. The analysis here has been limited for simplicity to the second virial coefficient, although some of our conclusions will apply also to the higher virial coefficients. No attempt has been made to give an exhaustive specification of what nonanalytic forms of $\hbar^2$ may appear. The following statements seem to be valid conclusions from the examples worked out. If the second-order perturbation term B_{22} has a finite classical limit for a given potential, then the diffraction corrections to that classical limit include nonanalytic forms of $\hbar^2$ when the potential has a cusp or singularity in any derivative. For the examples of the square wall potential and the form $r^m e^{-r/r_0}$ this nonanalytic form is $(\hbar^2)^{1/2}$. This statement applies also to the Coulomb potential for a gas in three dimensions, since in three dimensions the spatial volume element $4\pi r^2\,dr$ assures the finiteness of B_{22} for the screened Coulomb potential and of the ring sum for the electron gas.

For potentials more singular at the origin than $1/r$, B_{22} is infinite, and one must evaluate the entire second virial coefficient. The virial coefficient is nonanalytic in the coupling constant of the interaction, but the diffraction corrections are analytic in $\hbar^2$ and may be obtained as the first few terms of the WK expansion. Hence, the WK expansion is quite justified for calculating diffraction corrections to the virial coefficients of nonideal gases. The convergence of the expansion, however, depends strongly on the order of the singularity assumed in the intermolecular potential.

Any sharp corners in the potential will result in diffraction corrections that are nonanalytic in $\hbar^2$. The reason is that the WK expansion fails since its coefficients are integrals over derivatives of the potential and thus are delta functions and derivatives of delta functions. Thus the second virial coefficient for the square barrier potential has nonanalytic diffraction corrections, and so also does the hard sphere potential (a special case of the square barrier with the height of the barrier put to ∞), give rise to nonanalytic form $(\hbar^2)^{1/2}$.

The screened Coulomb interaction in three dimensions is particularly interesting since its second virial coefficient has two parts with different types of diffraction corrections. B_{22} is finite classically, but because of the $1/r$ singularity its diffraction corrections involve both $\hbar^2$ and $(\hbar^2)^{1/2}$. The remainder of B_2, i.e., all higher orders of the perturbation expansion, is nonanalytic in the coupling constant $(g_1^3 \ln g_1)$, but the diffraction corrections are analytic in $\hbar^2$ since they may be calculated with the WK expansion. In one and two dimensions, however, all diffraction corrections to B_2 are analytic in $\hbar^2$ since B_{22} is infinite. With this mathematical

structure in mind, it is easy to make the extension to the electron gas for which $u(r) = e^2/r$. The ring sum which is analogous to B_{22} must have non-analytic diffraction corrections, while the diffraction corrections to the remaining orders of the perturbation expansion when appropriately grouped together (the Abe expansion) involve only powers of $\hbar^2$. It is believed that the method described in this article for using the WK expansion will have considerable utility in evaluating the theory of the quantum mechanical electron gas.

ACKNOWLEDGMENTS

Much of this work was inspired by a conversation with Irwin Oppenheim. Useful conversations with B. W. Knight are also gratefully acknowledged.

APPENDIX

The integral required for the second-order perturbation term (31) for the square wall potential is:

$$I = \int_0^\infty dx\, j_1(x)^2 L_0(\gamma^2 x^2). \tag{A1}$$

In order to evaluate it, the square of the spherical Bessel function is written in terms of trigonometric functions,

$$j_1(x)^2 = x^{-4}[\tfrac{1}{2}(1 - \cos 2x) - x \sin 2x + \tfrac{1}{2}x^2(1 + \cos 2x)],$$

and expanded in powers of x. Also the integral representation (23) of $L_0(\gamma^2 x^2)$ is used. Equation (A1) becomes

$$I = \int_0^\infty dx\, 2^3 \sum_{n=3} \frac{(-1)^n (2x)^{2n-4}}{(2n-4)!}\left(\frac{1}{2n} - \frac{1}{4}\right) \times \int_0^1 dv\, e^{-\gamma^2 x^2 v(1-v)},$$

$$= \pi^{1/2} \sum_{n=3} \frac{(-1)^n 2^{2n-3}}{(2n-3)}\left(\frac{1}{n!} - \frac{1}{2(n-1)!}\right) \times \int_0^{\pi/2} d(\sin\theta)(\gamma\cos\theta)^{-2n+3}. \tag{A2}$$

The second line of Eq. (A2) is obtained by doing the x integration over the Gaussian function first, and with the change of variable $v = (1 + \sin\theta)/2$.

For each value of n, the θ integral in Eq. (A2) is divergent. This trouble is avoided by using the Mellin transform of the series in Eq. (A2); it becomes

$$I = -\pi^{1/2} \int_C \frac{ds}{2\pi i} \frac{2^{-2s+3}}{2s+3} [\Gamma(s) + \tfrac{1}{2}\Gamma(s+1)] \times \int_0^{\pi/2} d(\sin\theta)(\gamma\cos\theta)^{2s+3}, \tag{A3}$$

where the contour C encloses the entire real axis to the left of the point -3. The θ integration for arbitrary s is

$$\int_0^{\pi/2} d(\sin\theta)\cos\theta^{2s+3} = \frac{\pi^{1/2}}{2}\frac{\Gamma(s+5/2)}{\Gamma(s+3)},$$

so that Eq. (A3) becomes

$$I = -\frac{\pi}{2}\int_C \frac{ds}{2\pi i}\frac{\gamma^{2s+3}}{2^{2s+3}} \times \frac{\Gamma(s+5/2)}{(s+2)(s+1)(2s+3)}\left(\frac{1}{s} + \frac{1}{2}\right). \tag{A4}$$

The integrand of Eq. (A4) has only simple poles. After calculating the residues, one obtains

$$I = \frac{\pi}{2}\sum_{r=0} \frac{(-1)^r (2/\gamma)^{2r+2}}{(2r+3)(2r+5)(r+1)!}. \tag{A5}$$

Equation (A5) is the expansion of

$$\frac{\pi}{2}\left[\frac{1}{3} - \frac{1}{2a}\int_0^a dt\, e^{-t^2} + \frac{1}{2a^3}\int_0^a dt\, t^2 e^{-t^2}\right], \tag{A6}$$

with $a = 2/\gamma$. Integrating the last term of Eq. (A6) by parts gives the form recorded in Eq. (31).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Derivation of Generalized Master Equations*†

ROBERT J. SWENSON
Lehigh University, Bethlehem, Pennsylvania
(Received February 9, 1962)

We present a simplified derivation of a generalized master equation for the diagonal part of the occupation probability density. This derivation is valid for systems of arbitrary volume. It does not require the use of perturbation expansions nor the use of a diagonal singularity condition. In addition, a similar derivation is presented of a generalized master equation for the nondiagonal part of the occupation probability density. These equations become identical to the generalized master equations of Van Hove and Janner, respectively, if a perturbation expansion is made, if a diagonal singularity condition is assumed, and if the limit of infinite volume is taken.

RECENT advances in the understanding of the nonequilibrium behavior of many-particle quantum systems have been made by Van Hove,[1] by Prigogine and his co-workers,[2] and by others. They have used dynamical arguments to obtain kinetic equations which describe the irreversible evolution of systems from particular initial states. In general, these kinetic equations are non-Markovian, but in some limiting cases they can be approximated by the Markovian Pauli equations.[3] Central in the work of these authors is a many-body perturbation theory applied to an infinite system. In particular, Van Hove has obtained a generalized master equation for the diagonal part of the occupation probability density (g.m.e.d.) by making a perturbation expansion and assuming a diagonal singularity condition. Also, the limit of an infinite system is essential to the analysis so that certain terms can be ignored.

It is clearly desirable to derive a g.m.e.d. without recourse to perturbation theory.[4] We wish to present a simplified derivation of a g.m.e.d. by making use of reasoning similar to that of Heitler and others in the treatment of damping theory.[5] This derivation does not make use of perturbation expansions or of Van Hove's diagonal singularity condition and it is valid for systems of finite volume. The difference between our result and Van Hove's will be discussed below.

Recently, Janner[6] has used the reasoning and assumptions of Van Hove to derive a generalized master equation for the nondiagonal part of the occupation probability density (g.m.e.n.). A simplified derivation of a g.m.e.n. is given below and this g.m.e.n. is compared to Janner's.

We will first present a derivation of the following g.m.e.d.:

$$\frac{dP_E(t/\alpha\alpha_0)}{dt} = f_E(t/\alpha)\ \delta(\alpha - \alpha_0) + 2\pi \sum_{\alpha'} \int_0^t dt' [w_E(t - t'/\alpha\alpha')P_E(t'/\alpha'\alpha_0) - w_E(t - t'/\alpha'\alpha)P_E(t'/\alpha\alpha_0)], \quad (1)$$

where $\delta(\alpha - \alpha_0)$ is a Kronecker delta, and the partial transition probability $P_E(t/\alpha\alpha_0)$ is related to the transition probability $P(t/\alpha\alpha_0)$ for $t > 0$ by

$$\int_{-\infty}^{\infty} dE P_E(t/\alpha\alpha_0) = P(t/\alpha\alpha_0) \equiv \langle\alpha_0] \exp(iHt/\hbar)\ [\alpha\rangle\ \langle\alpha] \exp(-iHt/\hbar)\ [\alpha_0\rangle, \quad (2)$$

with $P(0/\alpha\alpha_0) = \delta(\alpha - \alpha_0)$ and the quantities w_E and f_E are defined below. (Our notation will be the same as that used by Van Hove and Janner.) The Hamiltonian, which is assumed not to depend explicitly on the time, is written as a sum of an unperturbed part H_0 and a perturbation H_1,

$$H = H_0 + H_1, \quad (3)$$

with

$$H_0\ [\alpha\rangle = \epsilon_\alpha\ [\alpha\rangle, \quad (4)$$

* This work was supported by the National Science Foundation.

† Based on part of the author's dissertation submitted to Lehigh University in partial fulfillment of the requirements for the degree of Doctor of Philosophy, October, 1961.

[1] L. Van Hove, Physica **23**, 441 (1957).

[2] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).

[3] I. Prigogine and P. Resibois, Physica **27**, 629 (1961).

[4] Similar motivations appear in R. W. Zwanzig, in *Lectures in Theoretical Physics*, edited by W. E. Brittin, B. W. Downs, and J. Downs (Interscience Publishers, Inc., New York, 1961), vol. III; E. W. Montroll, in *Fundamental Problems in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1962).

[5] W. Heitler, *The Quantum Theory of Radiation* (Oxford University Press, New York, 1953), 3rd ed., p. 164.

[6] A. Janner, Helv. Phys. Acta (to be published).

and we use the eigenstates $[\alpha\rangle$ to compute matrix elements.

Our analysis will be based on the properties of the resolvent operator R_l which has been discussed by Van Hove[1,7] and by Hugenholtz,[8] and is defined by

$$R_l = (H - l)^{-1}, \tag{5}$$

where l is a complex number. The resolvent is related to the time unitary transformation operator according to

$$\exp(-iHt/\hbar) = -(2\pi i)^{-1} \int_C dl\, R_l \exp(-ilt/\hbar), \tag{6}$$

where the contour C is taken counterclockwise around a sufficiently large portion of the real axis.

Introducing Eq. (6) into Eq. (2), we obtain

$$P(t/\alpha\alpha_0) = -(2\pi)^{-2} \int_C dl\, dl' \times \exp[i(l - l')t/\hbar]\, \langle\alpha_0]\, R_l\, [\alpha\rangle\, \langle\alpha]\, R_{l'}\, [\alpha_0\rangle,$$

and defining a quantity X,

$$X_{l'l}(\alpha_0\alpha) = X_{ll'}(\alpha\alpha_0) \equiv \langle\alpha_0]\, R_l\, [\alpha\rangle\, \langle\alpha]\, R_{l'}\, [\alpha_0\rangle, \tag{7}$$

we have

$$P(t/\alpha\alpha_0) = -(2\pi)^{-2} \int_C dl\, dl' \times \exp[i(l - l')t/\hbar]\, X_{ll'}(\alpha\alpha_0). \tag{8}$$

The g.m.e.d. is a mathematical consequence of the following equation for X:

$$(l - l')X_{ll'}(\alpha\alpha_0) = F_{ll'}(\alpha)\, \delta(\alpha - \alpha_0) - i \sum_{\alpha'} [\tilde{W}_{ll'}(\alpha\alpha')X_{ll'}(\alpha'\alpha_0) - \tilde{W}_{ll'}(\alpha'\alpha)X_{ll'}(\alpha\alpha_0)], \tag{9}$$

where we have introduced

$$\tilde{W}_{ll'}(\alpha\alpha') \equiv iF_{ll'}(\alpha)W_{ll'}(\alpha\alpha') \tag{10}$$

and F and W are defined below.

We first derive this equation for X and then show how the g.m.e.d. follows from it. In the representation furnished by H_0 we write the resolvent as the sum of its diagonal and nondiagonal parts; e.g. (the complex arguments l and l' will not be explicitly indicated except when needed for clarity)

$$R = R_d + R_{nd}, \tag{11}$$

d and nd denoting the diagonal and nondiagonal parts, respectively. Defining D and U by $R_d = D$ and $R_{nd} = DUD$ (note that U is nondiagonal), we can write

$$R = (I + DU)\, D, \tag{12}$$

where I is the unit operator. Equation (12) can be rearranged to yield

$$I = (H_0 + H_1 - l)(I + DU)\, D, \tag{13}$$

and if the diagonal part of this equation is taken, we obtain an expression for D,

$$D_l = (H_0 + G_l - l)^{-1}, \tag{14}$$

with

$$G_l = [H_1 + H_1\, D_l U_l]_d. \tag{15}$$

If the nondiagonal part of Eq. (13) is taken, we obtain

$$0 = [H_1 + (H_0 - l)\, DU + H_1\, DU]_{nd},$$

and using Eq. (14) we can write

$$U = -[H_1 + H_1\, DU - G\, DU]_{nd}.$$

Substitution of Eq. (15) into this expression yields

$$U = -[H_1 + [H_1]_{nd}\, DU - [H_1\, DU]_d\, DU]_{nd}. \tag{16}$$

We find after substitution of Eq. (12) into Eq. (7) and recalling that U is nondiagonal

$$X_{ll'}(\alpha\alpha_0) = D_l(\alpha)\, D_{l'}(\alpha)[\delta(\alpha - \alpha_0) + D_l(\alpha_0)\, D_{l'}(\alpha_0)\, \langle\alpha_0]\, U_l\, [\alpha\rangle\, \langle\alpha]\, U_{l'}\, [\alpha_0\rangle], \tag{17}$$

with $D_l(\alpha) \equiv \langle\alpha]\, D_l\, [\alpha\rangle$. An identity for $D_l(\alpha)\, D_{l'}(\alpha)$ is obtained from Eq. (14):

$$\begin{aligned} F_{ll'}(\alpha) &\equiv D_l(\alpha) - D_{l'}(\alpha) \\ &= [\epsilon_\alpha + G_l(\alpha) - l]^{-1} - [\epsilon_\alpha + G_{l'}(\alpha) - l']^{-1} \\ &= D_l(\alpha)\, D_{l'}(\alpha)[l - l' - G_l(\alpha) + G_{l'}(\alpha)], \end{aligned}$$

which can be solved to give

$$D_l(\alpha)\, D_{l'}(\alpha) = F_{ll'}(\alpha)[l - l' - G_l(\alpha) + G_{l'}(\alpha)]^{-1}. \tag{18}$$

Substitution of Eq. (18) into Eq. (17) yields, after some simplification,

$$(l - l')X_{ll'}(\alpha\alpha_0) = F_{ll'}(\alpha)\, \delta(\alpha - \alpha_0) + F_{ll'}(\alpha)\, D_l(\alpha_0)\, D_{l'}(\alpha_0)U_l(\alpha_0\alpha)U_{l'}(\alpha\alpha_0) + [G_l(\alpha) - G_{l'}(\alpha)]X_{ll'}(\alpha\alpha_0), \tag{19}$$

where we introduced $\langle\alpha_0]\, U_l\, [\alpha\rangle \equiv U_l(\alpha_0\alpha)$, etc. Let us define an operator W by the equation

---

[7] L. Van Hove, in *La théorie des gas neutres et ionisés*, edited by C. DeWitt and J. G. Detoeuf (John Wiley & Sons, Inc., New York, 1959).

[8] N. M. Hugenholtz, in *Lectures in Theoretical Physics*, edited by W. E. Brittin and B. W. Downs (Interscience Publishers, Inc., New York, 1960), Vol. II.

$$W_{ll'}(\alpha\alpha_0) = U_l(\alpha_0\alpha)U_{l'}(\alpha\alpha_0) - \sum_{\alpha'} W_{ll'}(\alpha\alpha') \times D_l(\alpha')\, D_{l'}(\alpha')U_l(\alpha_0\alpha')U_{l'}(\alpha'\alpha_0). \quad (20)$$

This equation can be iterated to yield

$$W_{ll'}(\alpha\alpha_0) = U_l(\alpha_0\alpha)U_{l'}(\alpha\alpha_0) - \sum_{\alpha'} U_l(\alpha'\alpha)U_{l'}(\alpha\alpha') \times D_l(\alpha')\, D_{l'}(\alpha')U_l(\alpha_0\alpha')U_{l'}(\alpha'\alpha_0) + \cdots, \quad (21)$$

from which it is evident that W has the symmetry

$$W_{ll'}(\alpha\alpha_0) = W_{l'l}(\alpha_0\alpha). \quad (22)$$

We can obtain an expression for the second quantity on the right-hand side of Eq. (19) in terms of X and W as follows: Multiply Eq. (20) by $D_l(\alpha_0)\, D_{l'}(\alpha_0)$ and rearrange terms to obtain

$$\sum_{\alpha'} W_{ll'}(\alpha\alpha')\, D_l(\alpha')\, D_{l'}(\alpha')[\delta(\alpha' - \alpha_0) + D_l(\alpha_0)\, D_{l'}(\alpha_0)U_l(\alpha_0\alpha')U_{l'}(\alpha'\alpha_0)] = D_l(\alpha_0)\, D_{l'}(\alpha_0)U_l(\alpha_0\alpha)U_{l'}(\alpha\alpha_0),$$

which with the aid of Eq. (17) gives the desired relation

$$\sum_{\alpha'} W_{ll'}(\alpha\alpha')X_{ll'}(\alpha'\alpha_0) = D_l(\alpha_0)\, D_{l'}(\alpha_0)U_l(\alpha_0\alpha)U_{l'}(\alpha\alpha_0). \quad (23)$$

This enables us to write Eq. (19) as

$$(l - l')X_{ll'}(\alpha\alpha_0) = F_{ll'}(\alpha)\,\delta(\alpha - \alpha_0) + F_{ll'}(\alpha)\sum_{\alpha'} W_{ll'}(\alpha\alpha')X_{ll'}(\alpha'\alpha_0) + [G_l(\alpha) - G_{l'}(\alpha)]X_{ll'}(\alpha\alpha_0). \quad (24)$$

Now we obtain an expression for $[G_l(\alpha) - G_{l'}(\alpha)]$ in terms of $W_{ll'}$. From the definition of R_l, it follows that

$$R_l - R_{l'} = (H - l)^{-1} - (H - l')^{-1} = R_{l'}R_l(l - l'),$$

and by taking a diagonal element we obtain

$$F_{ll'}(\alpha) = (l - l')\sum_{\alpha_0} X_{ll'}(\alpha\alpha_0). \quad (25)$$

If we sum Eq. (24) over α_0 and use Eq. (25), we obtain

$$\sum_{\alpha'} F_{ll'}(\alpha')W_{ll'}(\alpha\alpha') = -[G_l(\alpha) - G_{l'}(\alpha)]. \quad (26)$$

Instead of using this to eliminate $[G_l(\alpha) - G_{l'}(\alpha)]$ in Eq. (24), we will use

$$\sum_{\alpha'} F_{ll'}(\alpha')W_{ll'}(\alpha'\alpha) = -[G_l(\alpha) - G_{l'}(\alpha)]. \quad (26')$$

Equation (26′) is obtained as follows: We interchange the complex arguments l and l' in Eq. (17) and follow the same analysis that led to Eq. (26). This will yield

$$\sum_{\alpha'} F_{l'l}(\alpha')W_{l'l}(\alpha\alpha') = -[G_{l'}(\alpha) - G_l(\alpha)], \quad (26'')$$

and using the symmetry property of W expressed by Eq. (22) and multiplying by minus one we obtain Eq. (26′).

Using Eq. (26′) to eliminate $[G_l(\alpha) - G_{l'}(\alpha)]$ in Eq. (24), we write

$$(l - l')X_{ll'}(\alpha\alpha_0) = F_{ll'}(\alpha)\,\delta(\alpha - \alpha_0) - i\sum_{\alpha'}[\tilde{W}_{ll'}(\alpha\alpha')X_{ll'}(\alpha'\alpha_0) - \tilde{W}_{ll'}(\alpha'\alpha)X_{ll'}(\alpha\alpha_0)], \quad (27)$$

which is the desired equation, Eq. (9), for X. As indicated by Van Hove,[1] a g.m.e.d. is a mathematical consequence of this equation and for the sake of completeness we demonstrate this in detail in the appendix. In the appendix we also discuss the role played by the inhomogeneous term on the right-hand side of Eq. (27).

Let us now derive a g.m.e.n., which is

$$\frac{dI_E(t/\alpha\alpha'\alpha_0)}{dt} = g_E(t/\alpha\alpha'\alpha_0) + 2\pi\int_0^t dt' \sum_{\alpha''}[w_E(t - t'/\alpha\alpha'')I_E(t'/\alpha''\alpha'\alpha_0) - w_E(t - t'/\alpha''\alpha)I_E(t'/\alpha\alpha'\alpha_0)], \quad (28)$$

$I_E(t/\alpha\alpha'\alpha_0)$ is related to $I(t/\alpha\alpha'\alpha_0)$ for $t > 0$ by

$$\int_{-\infty}^{\infty} dE\, I_E(t/\alpha\alpha'\alpha_0) = I(t/\alpha\alpha'\alpha_0) \equiv \langle\alpha'] \exp(iHt/\hbar)\,[\alpha\rangle\,\langle\alpha] \exp(-iHt/\hbar)\,[\alpha_0\rangle \quad (29)$$

with $I(0/\alpha\alpha'\alpha_0) = 0$, w_E is the same quantity that appears in Eq. (1), and the inhomogeneous term $g_E(t/\alpha\alpha'\alpha_0)$ will be defined below. Introduce Eq. (6) into Eq. (29) to obtain

$$I(t/\alpha\alpha'\alpha_0) = -(2\pi)^{-2}\int_C dl\, dl' \times \exp[i(l - l')t/\hbar]\; Y_{ll'}(\alpha\alpha'\alpha_0), \quad (30)$$

with

$$Y_{ll'}(\alpha\alpha'\alpha_0) \equiv \langle\alpha']\, R_l\,[\alpha\rangle\,\langle\alpha]\, R_{l'}\,[\alpha_0\rangle. \quad (31)$$

The g.m.e.n. (28) is a mathematical consequence of the following equation for Y:

$$(l - l')Y_{ll'}(\alpha\alpha'\alpha_0) = F_{ll'}(\alpha)V_{ll'}(\alpha\alpha'\alpha_0) - i\sum_{\alpha''}[\tilde{W}_{ll'}(\alpha\alpha'')Y_{ll'}(\alpha''\alpha'\alpha_0) - \tilde{W}_{ll'}(\alpha''\alpha)Y_{ll'}(\alpha\alpha'\alpha_0)], \quad (32)$$

where F and $\tilde{W}$ are the same quantities which appear in Eq. (9) and V will be defined below.

We now derive Eq. (32) in essentially the same way we obtained Eq. (27). Substitute Eq. (12) into Eq. (31) to obtain

$$\begin{aligned} Y_{ll'}(\alpha\alpha'\alpha_0) &= D_l(\alpha)\, D_{l'}(\alpha)[D_l(\alpha')U_l(\alpha'\alpha)\ \delta(\alpha-\alpha_0) \\ &+ D_{l'}(\alpha_0)U_{l'}(\alpha\alpha_0)\ \delta(\alpha-\alpha') \\ &+ D_l(\alpha')U_l(\alpha'\alpha)U_{l'}(\alpha\alpha_0)\ D_{l'}(\alpha_0)], \end{aligned} \tag{33}$$

and use Eq. (18) to write

$$\begin{aligned} (l-l')Y_{ll'}(\alpha\alpha'\alpha_0) &= F_{ll'}(\alpha)[D_l(\alpha')U_l(\alpha'\alpha)\ \delta(\alpha-\alpha_0) \\ &+ D_{l'}(\alpha_0)U_{l'}(\alpha\alpha_0)\ \delta(\alpha-\alpha') \\ &+ D_l(\alpha')U_l(\alpha'\alpha)U_{l'}(\alpha\alpha_0)\ D_{l'}(\alpha_0)] \\ &+ [G_l(\alpha)-G_{l'}(\alpha)]Y_{ll'}(\alpha\alpha'\alpha_0). \end{aligned} \tag{34}$$

Let us define an operator V by the equation

$$\begin{aligned} V_{ll'}(\alpha\alpha'\alpha_0) &= D_l(\alpha')U_l(\alpha'\alpha)\ \delta(\alpha-\alpha_0) \\ &+ D_{l'}(\alpha_0)U_{l'}(\alpha\alpha_0)\ \delta(\alpha-\alpha') \\ &+ D_l(\alpha')U_l(\alpha'\alpha)U_{l'}(\alpha\alpha_0)\ D_{l'}(\alpha_0) \\ &- \sum_{\alpha''} D_l(\alpha'')\ D_{l'}(\alpha'') \\ &\times U_l(\alpha''\alpha)U_{l'}(\alpha\alpha'')V_{ll'}(\alpha''\alpha'\alpha_0). \end{aligned} \tag{35}$$

Multiply by $D_l(\alpha)\ D_{l'}(\alpha)$ and rearrange terms to obtain

$$\begin{aligned} &\sum_{\alpha''} D_l(\alpha)\ D_{l'}(\alpha)[\delta(\alpha-\alpha'')+D_l(\alpha'')\ D_{l'}(\alpha'') \\ &\quad\times U_l(\alpha''\alpha)U_{l'}(\alpha\alpha'')]V_{ll'}(\alpha''\alpha'\alpha_0) \\ &= D_l(\alpha)\ D_{l'}(\alpha)[D_l(\alpha')U_l(\alpha'\alpha)\ \delta(\alpha-\alpha_0) \\ &\quad+ D_{l'}(\alpha_0)U_{l'}(\alpha\alpha_0)\ \delta(\alpha-\alpha') \\ &\quad+ D_l(\alpha')U_l(\alpha'\alpha)U_{l'}(\alpha\alpha_0)\ D_{l'}(\alpha_0)]. \end{aligned} \tag{36}$$

The factor multiplying $V_{ll'}(\alpha''\alpha'\alpha_0)$ is recognized from Eq. (17) to be $X_{ll'}(\alpha\alpha'')$ and the right-hand side is recognized from Eq. (33) to be $Y_{ll'}(\alpha\alpha'\alpha_0)$, so we obtain the relation

$$\sum_{\alpha''} X_{ll'}(\alpha\alpha'')V_{ll'}(\alpha''\alpha'\alpha_0) = Y_{ll'}(\alpha\alpha'\alpha_0). \tag{37}$$

Using Eqs. (23) and (37), we write Eq. (35) as

$$\begin{aligned} &D_l(\alpha')U_l(\alpha'\alpha)\ \delta(\alpha-\alpha_0)+D_{l'}(\alpha_0)U_{l'}(\alpha\alpha_0)\ \delta(\alpha-\alpha') \\ &\quad+ D_l(\alpha')U_l(\alpha'\alpha)U_{l'}(\alpha\alpha_0)\ D_{l'}(\alpha_0) \\ &= V_{ll'}(\alpha\alpha'\alpha_0)+\sum_{\alpha''} W_{ll'}(\alpha\alpha'')Y_{ll'}(\alpha''\alpha'\alpha_0). \end{aligned} \tag{38}$$

Substituting this into the right-hand side of Eq. (34) and using Eq. (26′) we obtain the desired equation

$$\begin{aligned} (l-l')Y_{ll'}(\alpha\alpha'\alpha_0) &= F_{ll'}(\alpha)V_{ll'}(\alpha\alpha'\alpha_0) \\ &- i\sum_{\alpha''}[\tilde{W}_{ll'}(\alpha\alpha'')Y_{ll'}(\alpha''\alpha'\alpha_0) \\ &- \tilde{W}_{ll'}(\alpha''\alpha)Y_{ll'}(\alpha\alpha'\alpha_0)]. \end{aligned} \tag{39}$$

The g.m.e.n. (28) is a mathematical consequence of this equation, as discussed in the Appendix.

It is thus possible to obtain generalized master equations for the diagonal and nondiagonal parts of the occupation probability density which are valid for finite systems without using perturbation theory and without requiring that the potential satisfy a diagonal singularity condition. Of course in any practical calculation it is convenient and for some things necessary to take the limit of an infinite system and then sums can be replaced by integrals. Furthermore, in this limit the analytic behavior of the resolvent and related operators simplifies, since a set of dense poles along the real axis condense into a branch cut.[1,8]

In order to compare the above results with the work of Van Hove[1,7] and Janner[6] we express the resolvent operator as an expansion in powers of the "interaction" H_1 (see reference 1). If this is done, we obtain the following expression for W:

$$\begin{aligned} W_{ll'}(\alpha\alpha_0) &= \{\langle\alpha_0](H_1 - H_1\ D_l H_1 \\ &+ H_1\ D_l H_1\ D_l H_1 - \cdots)\ [\alpha\rangle\ \langle\alpha] \\ &\times (H_1 - H_1\ D_{l'} H_1 + \cdots)\ [\alpha_0\rangle\}_{id}, \end{aligned} \tag{40}$$

where *id* stands for "irreducible diagonal" and means that all intermediate states are unequal to the initial state and no two intermediate states are equal. Van Hove's expression for W [see Eq. (9.22) in reference 7] is the same as Eq. (40) except for the meaning of *id*. In his work the definition of id is that all intermediate states are "nonidentical" to the initial state and no two intermediate states are "identical"; however, intermediate states may become "occasionally equal." This leads to some terms being counted twice, but Van Hove demonstrates[7] that if a diagonal singularity condition is assumed the error introduced is proportional to one over the volume and thus vanishes in the limit of an infinite system. If we make the same assumptions as Van Hove, then the *id* which appears in Eq. (40) can be replaced by Van Hove's *id*.

We conclude that our results become identical to Van Hove's and Janner's if we adopt their assumptions. However, in some cases it is more convenient to expand the resolvent in terms of some quantity other than the interaction[9]; e.g., if the

[9] R. J. Swenson, J. Math. Phys. (to be published).

interaction contains a hard core it is useful to expand in terms of a scattering matrix.

ACKNOWLEDGMENTS

The author expresses his thanks to Professor James A. McLennan, Jr., for his encouragement, many discussions, and critical reading of the manuscript. He also thanks Dr. Janner for sending a preprint of his recent work.

APPENDIX

We wish to demonstrate how a g.m.e.d. can be obtained from Eq. (27). The same arguments apply for obtaining Eq. (28) from Eq. (39). From Eq. (8) we have

$$P(t/\alpha\alpha_0) = -(2\pi)^{-2}\int_\gamma dl\int_\gamma dl' \times \exp[i(l-l')t/\hbar]X_{ll'}(\alpha\alpha_0), \quad \text{(A1)}$$

and for $t > 0$ we can deform the contours to obtain (see reference 7)

$$P(t/\alpha\alpha_0) = (2\pi)^{-2}\int_{-\infty}^{\infty} dE^1\int_{-\infty}^{\infty} dE^2 \times \exp[i(E^1-E^2-2i\eta)t/\hbar]X_{E^1-i\eta,E^2+i\eta}(\alpha\alpha_0). \quad \text{(A2)}$$

Let us make the change of variables

$$E^1 - E^2 = 2E' \quad \text{and} \quad E^1 + E^2 = 2E,$$

which yields, for $P(t/\alpha\alpha_0)$,

$$P(t/\alpha\alpha_0) = (2\pi^2)^{-1}\int_{-\infty}^{\infty} dE\int_{-\infty}^{\infty} dE' \times \exp[2i(E'-i\eta)t/\hbar]X_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha_0), \quad \text{(A3)}$$

or

$$P(t/\alpha\alpha_0) = \int_{-\infty}^{\infty} dE\, P_E(t/\alpha\alpha_0), \quad \text{(A4)}$$

with $P_E(t)$ defined as

$$P_E(t/\alpha\alpha_0) = (2\pi^2)^{-1}\int_{-\infty}^{\infty} dE' \times \exp[2i(E'-i\eta)t/\hbar]X_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha_0). \quad \text{(A5)}$$

Taking the time derivative of Eq. (A5) yields

$$\frac{dP_E(t/\alpha\alpha_0)}{dt} = i(2\pi^2\hbar)^{-1}\int_{-\infty}^{\infty} dE' \exp[2i(E'-i\eta)t/\hbar] \times 2(E'-i\eta)X_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha_0), \quad \text{(A6)}$$

and from Eq. (27) we obtain an expression for $2(E'-i\eta)\,X_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha_0)$; i.e.,

$$\begin{aligned}2(E'-i\eta)X_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha_0) &= F_{E+E'-i\eta,E-E'+i\eta}(\alpha)\,\delta(\alpha-\alpha_0) \\ &\quad - i\sum_{\alpha'}[\tilde{W}_{(\ldots)}(\alpha\alpha')X_{(\ldots)}(\alpha'\alpha_0) \\ &\quad - \tilde{W}_{(\ldots)}(\alpha'\alpha)X_{(\ldots)}(\alpha\alpha_0)], \quad \text{(A7)}\end{aligned}$$

where the arguments indicated by $(\cdots)$ are the same as the arguments of F. Equation (A7) enables us to write the time derivative of $P_E(t)$ as

$$\begin{aligned}\frac{dP_E(t/\alpha\alpha_0)}{dt} &= i(2\pi^2\hbar)^{-1}\int_{-\infty}^{\infty} dE' \exp[2i(E'-i\eta)t/\hbar] \\ &\quad \times \{F_{E+E'-i\eta,E-E'+i\eta}(\alpha)\,\delta(\alpha-\alpha_0) \\ &\quad - i\sum_{\alpha'}[\tilde{W}_{(\ldots)}(\alpha\alpha')X_{(\ldots)}(\alpha'\alpha_0) \\ &\quad - \tilde{W}_{(\ldots)}(\alpha'\alpha)X_{(\ldots)}(\alpha\alpha_0)]\}. \quad \text{(A8)}\end{aligned}$$

Now we define w_E and f_E by

$$f_E(t/\alpha) = i(2\pi^2\hbar)^{-1}\int_{-\infty}^{\infty} dE' \times \exp[2i(E'-i\eta)t/\hbar]F_{E+E'-i\eta,E-E'+i\eta}(\alpha) \quad \text{(A9)}$$

and

$$w_E(t/\alpha\alpha') = (2\pi^2\hbar^2)^{-1}\int_{-\infty}^{\infty} dE' \times \exp[2i(E'-i\eta)t/\hbar]\tilde{W}_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha') \quad \text{(A10)}$$

and we use these definitions in Eq. (A8) to write a g.m.e.d.

$$\begin{aligned}\frac{dP_E(t/\alpha\alpha_0)}{dt} &= f_E(t/\alpha)\,\delta(\alpha-\alpha_0) \\ &\quad + 2\pi\sum_{\alpha'}\int_0^t dt'[w_E(t-t'/\alpha\alpha')P_E(t'/\alpha'\alpha_0) \\ &\quad - w_E(t-t'/\alpha'\alpha)P_E(t'/\alpha\alpha_0)]. \quad \text{(A11)}\end{aligned}$$

We have made use of the following convolution theorem in obtaining Eq. (A11):

$$\begin{aligned}&\int_0^t dt'\, w_E(t-t'/\alpha\alpha')P_E(t'/\alpha'\alpha_0) \\ &\quad = (4\pi^3\hbar)^{-1}\int_{-\infty}^{\infty} dE' \exp[2i(E'-i\eta)t/\hbar] \\ &\qquad \times \tilde{W}_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha')X_{(\ldots)}(\alpha'\alpha_0). \quad \text{(A12)}\end{aligned}$$

Let us prove the above convolution theorem. Substitute the definitions of w_E and P_E into the left-hand side and perform the time integration,

$$\begin{aligned}&\int_0^t dt'\, w_E(t-t'/\alpha\alpha')P_E(t'/\alpha'\alpha_0) \\ &\quad = i(8\pi^4\hbar)^{-1}\mathcal{P}\int_{-\infty}^{\infty} dE'\,dE''\,\tilde{W}_{E+E'-i\eta,E-E'+i\eta}(\alpha\alpha') \\ &\qquad \times X_{E+E''-i\eta,E-E''+i\eta}(\alpha'\alpha_0)(E''-E')^{-1} \\ &\qquad \times \{\exp[2i(E'-i\eta)t/\hbar] - \exp[2i(E''-i\eta)t/\hbar]\}.\end{aligned}$$

Since the integrand is analytic on the real axis nothing is changed by taking the principal parts $\mathcal{P}$ of the integrals. We do the E'' integration in the first term and the E' integration in the second term to obtain the desired relation (A12). For example, the E'' integration in the first term is

$$B \equiv \mathcal{P} \int dE'' \, X_{E+E''-i\eta, E-E''+i\eta}(\alpha'\alpha_0)(E'' - E')^{-1}$$

and closing the contour in the lower half plane (since for large E'', $X \sim E''^{-2}$) we obtain

$$B = \mathcal{P} \int_C dz \, X_{E+z-i\eta, E-z+i\eta}(\alpha'\alpha_0)(z - E')^{-1}.$$

The singularities of $X_{E+z-i\eta, E-z+i\eta}(\alpha'\alpha_0)$ lie above the real E'' axis (as follows from the definition of X), so that the only singularity of the integrand is a simple pole at $z = E'$. The integral is easily evaluated to give (for E' real, we use the definition of the principal part)

$$B = -i\pi X_{E+E'-i\eta, E-E'+i\eta}(\alpha'\alpha_0).$$

The E' integration in the second term is done in exactly the same way and Eq. (A12) is the result.

The derivation of the g.m.e.n. follows in the same manner with g_E defined by

$$g_E(t/\alpha\alpha'\alpha_0) = i(2\pi^2\hbar)^{-1} \int_{-\infty}^{\infty} dE' \times \exp[2i(E' - i\eta)t/\hbar] F_{E+E'-i\eta, E-E'+i\eta}(\alpha) \times V_{E+E'-i\eta, E-E'+i\eta}(\alpha\alpha'\alpha_0), \quad \text{(A13)}$$

and a convolution theorem identical to Eq. (A12) with I_E and $Y(\alpha\alpha'\alpha_0)$ replacing P_E and $X(\alpha\alpha_0)$.

Van Hove[1] has shown in the limit of weak interaction that the inhomogeneous term in Eq. (A11) serves to specify the initial value of $P(t)$. We wish to show that this is an exact result. To see this, we integrate Eq. (1) over all E and obtain

$$\frac{dP(t/\alpha\alpha_0)}{dt} = \int_{-\infty}^{\infty} dE \, f_E(t/\alpha) \, \delta(\alpha - \alpha_0) + \int_{-\infty}^{\infty} dE \int_0^t dt'[\cdots]. \quad \text{(A14)}$$

Designating the inhomogeneous term by Δ and substituting Eq. (A9), we obtain

$$\Delta \equiv \delta(\alpha - \alpha_0) \int_{-\infty}^{\infty} dE \, i(2\pi^2\hbar)^{-1} \int_{-\infty}^{\infty} dE' \times \exp[2i(E' - i\eta)t/\hbar] F_{E+E'-i\eta, E-E'+i\eta}(\alpha).$$

From the definition of the resolvent, we obtain the identities

$$R_{E+E'-i\eta} - R_{E-E'+i\eta} = 2(E' - i\eta) R_{E+E'-i\eta} R_{E-E'+i\eta}$$

and

$$\int_{-\infty}^{\infty} dE \, R_{E+E'-i\eta} R_{E-E'+i\eta} = -i\pi(E' - i\eta)^{-1},$$

from which it follows that

$$\int_{-\infty}^{\infty} dE \, [R_{E+E'-i\eta} - R_{E-E'+i\eta}] = -2\pi i.$$

Taking a diagonal part and recalling the definition of $F(\alpha)$,

$$\int_{-\infty}^{\infty} dE \, F_{E+E'-i\eta, E-E'+i\eta}(\alpha) = -2\pi i,$$

allows us to write Δ as

$$\Delta = \delta(\alpha - \alpha_0)(\pi\hbar)^{-1} \int_{-\infty}^{\infty} dE' \exp[2i(E' - i\eta)t/\hbar],$$

which reduces to

$$\Delta = \delta(\alpha - \alpha_0) \, \delta(t).$$

Thus, Eq. (A14) becomes

$$\frac{dP(t/\alpha\alpha_0)}{dt} = \delta(\alpha - \alpha_0) \, \delta(t) + \int_{-\infty}^{\infty} dE \int_0^t dt' \, [\cdots], \quad \text{(A15)}$$

and it is clear that the inhomogeneous term does not contribute to the time development of $P(t)$.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

On the Number of Electron Levels in a One-Dimensional Random Lattice

J. A. MORRISON
Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey
(Received March 7, 1962)

Let the potential of a one-dimensional scalar particle be $V(x) = V_0 \sum_{j=-\infty}^{\infty} \delta(x - x_j)$, $-\infty < x < \infty$, where $V_0 < 0$, and where the sequence (x_j) is random, with a Poisson distribution. This paper investigates analytically the number N of electron levels per atom below energy $E = -\hbar^2\kappa^2/2m$, when $0 < n/\kappa_0 \ll 1$ and $0 < \kappa/\kappa_0 < 2$, where n is the expected density of atoms and $\kappa_0 = -mV_0/\hbar^2$. The region $\kappa/\kappa_0 \approx 1$, with n/κ_0 small, is of considerable interest, and some previous numerical computations have been inaccurate in this region. Explicit bounds on N^{-1} may be written down which give the asymptotic behavior of N, as $\kappa_0/n \to \infty$, for $0 < \kappa/\kappa_0 < 1$ and $1 < \kappa/\kappa_0 < (2 - \delta)$, $\delta > 0$.

1. INTRODUCTION AND SUMMARY

We are concerned with the distribution of energy levels of a one-dimensional electron (scalar particle) moving in a one-dimensional random array of atoms. The atoms, all of one kind, have (randomly) fixed positions, and the electron atom potentials are assumed to be δ functions. The Schrödinger equation for an electron of mass m and energy E is then

$$-\frac{\hbar^2}{2m}\frac{d\psi}{dx} + V_0 \sum_{i=-\infty}^{\infty} \delta(x - x_i)\psi(x) = E\psi(x), \quad (1.1)$$

where $V_0 < 0$ is the strength of each electron-atom interaction and the sequence (x_i) of positions of the atoms is random, with a Poisson distribution.

It was shown by Frisch and Lloyd[1] that the random variables $z_i = \psi'(x_i - 0)/\psi(x_i)$, $-\infty < j < \infty$, constitute an ergodic stationary Markov process. The stationary density $T(z)$ of the (z_i) was shown to satisfy the functional equation

$$(z^2 - \kappa^2)T(z) = n\left[N - \int_z^{z+2\kappa_0} T(\zeta)\, d\zeta\right], \quad (1.2)$$

with

$$T(z) \geq 0; \qquad \int_{-\infty}^{\infty} T(z)\, dz = 1, \quad (1.3)$$

where, using the notation of Lax and Phillips,[2] n is the expected density of atoms, N is the number of electron levels per atom below energy E and

$$\kappa_0 = -mV_0/\hbar^2; \qquad E = -\hbar^2\kappa^2/2m. \quad (1.4)$$

The quantity of interest is N. Numerical calculations, based on the Monte Carlo method, were made by Lax and Phillips.[2] On the other hand, Frisch and Lloyd[1] obtained their numerical results by integrating the (complex) second-order linear differential equation satisfied by the Fourier transform of $T(z)$. Their results were least accurate when $\epsilon \ll 1$ and $\nu \approx 1$, where

$$0 < \epsilon = n/\kappa_0; \qquad \nu = \kappa/\kappa_0. \quad (1.5)$$

This is a region of considerable interest and is one we consider here.

We study directly the first-order differential difference equation satisfied by

$$f(x) = N^{-1} \int_{-\infty}^{\kappa_0 x} T(\zeta)\, d\zeta, \quad (1.6)$$

when $0 < \epsilon \ll 1$ and $0 < \nu < 2$, $\nu \neq 1$. We show that

$$N^{-1} = \tfrac{1}{4}[3 - (1 - \nu)^{\epsilon/\nu}]^2\,[1 + O(\epsilon)]; \qquad 0 < \nu < 1, \quad (1.7)$$

and

$$N^{-1} = \frac{[3 - (\nu - 1)^{\epsilon/\nu}]^2}{4[1 - (\nu - 1)^{\epsilon/\nu}]}\,[1 + O(\epsilon)]; \qquad 1 < \nu < (2 - \delta), \quad \delta > 0, \quad (1.8)$$

where explicit bounds may be written down for the terms of $O(\epsilon)$. For Eq. (1.7), these bounds may be obtained from Eqs. (2.21) to (2.23), (2.25), (2.26), and (2.28), and for Eq. (1.8) from Eqs. (3.12), (3.14) to (3.16), (3.18), (3.22) to (3.25), (3.27), (3.28), and (3.30). The leading terms in Eqs. (1.7) and (1.8) are precisely those given by Lax and Phillips[2] as a modification of the formulas given by Schmidt,[3] who considered $\nu \approx 1$ and

[1] H. L. Frisch and S. P. Lloyd, Phys. Rev. **120**, 1175 (1960).
[2] M. Lax and J. C. Phillips, Phys. Rev. **110**, 41 (1958).
[3] H. Schmidt, Phys. Rev. **105**, 425 (1957).

instead of ϵ/ν had just ϵ (allowing for a misprint). However, Schmidt's derivation was not rigorous, in that no bounds on the error terms were given.

Actually, our results are somewhat stronger. Thus, for $0 < \nu < 1$, we show that

$$0 \leq \left[N^{-1} - 2 - \Psi(-2) + A\left(\frac{2+\nu}{2-\nu}\right)^{\epsilon/2\nu}\right] \leq \frac{\epsilon}{2\nu} \ln\left[\frac{(1+\nu)(2+\nu)}{(2-\nu)}\right] \leq \epsilon, \quad (1.9)$$

where $\Psi(-2)$ and A are given explicitly, Eqs. (2.13) and (2.15). Moreover, we indicate how to obtain an explicit expression for N^{-1}, correct to $O(\epsilon^2)$, when $0 < \nu < 1$.

2. THE CASE $0 < \nu < 1$

From Eqs. (1.2), (1.3), (1.5), and (1.6),

$$(x^2 - \nu^2)f'(x) = \epsilon[1 - f(x+2) + f(x)], \quad (2.1)$$

$$f(-\infty) = 0; \qquad f(+\infty) = N^{-1}; \qquad f'(x) \geq 0. \quad (2.2)$$

We first consider $x > \nu$. Then,

$$0 \leq f'(x) \leq \epsilon/(x^2 - \nu^2); \qquad x > \nu. \quad (2.3)$$

Integration of Eq. (2.3) gives

$$0 \leq [N^{-1} - f(x)] \leq \frac{\epsilon}{2\nu} \ln\left(\frac{x+\nu}{x-\nu}\right); \quad x > \nu. \quad (2.4)$$

Also, From Eq. (2.3) we may obtain

$$[f(x+2) - f(x)] \leq \frac{\epsilon}{2\nu} \ln\left[\frac{(x+2-\nu)(x+\nu)}{(x+2+\nu)(x-\nu)}\right]; \quad x > \nu. \quad (2.5)$$

Substitution of this inequality into Eq. (2.1) leads, upon integration, to

$$[N^{-1} - f(x)] \geq \left\{\frac{\epsilon}{2\nu} \ln\left(\frac{x+\nu}{x-\nu}\right) - \frac{\epsilon^2}{2\nu} \times \int_x^\infty \ln\left[\frac{(y+2-\nu)(y+\nu)}{(y+2+\nu)(y-\nu)}\right] \frac{dy}{(y^2-\nu^2)}\right\}; \quad x > \nu. \quad (2.6)$$

Iteration of the above procedures leads to successive upper and lower bounds on $[N^{-1} - f(x)]$, for $x > \nu$.

From Eqs. (2.1) and (2.2) we may obtain

$$f(x) = \left\{\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\left[1 - \epsilon\int_x^\infty \frac{[N^{-1} - f(y+2)]}{(y^2-\nu^2)} \times \left(\frac{y+\nu}{y-\nu}\right)^{\epsilon/2\nu} dy\right] + (N^{-1} - 1)\right\}; \quad x > \nu. \quad (2.7)$$

Also

$$f(x) = \left[\epsilon\left(\frac{\nu-x}{\nu+x}\right)^{\epsilon/2\nu} \int_{-\nu}^x \frac{f(y+2)}{(\nu^2-y^2)} \times \left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy - 1\right]; \qquad |x| < \nu, \quad (2.8)$$

and

$$f(x) = \left[\epsilon\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu} \int_x^{-\nu} \frac{f(y+2)}{(y^2-\nu^2)} \times \left(\frac{y+\nu}{y-\nu}\right)^{\epsilon/2\nu} dy - 1\right]; \qquad x < -\nu. \quad (2.9)$$

Now, Eq. (2.3) also holds for $x < -\nu$. Thus,

$$0 \leq f(x) \leq \frac{\epsilon}{2\nu} \ln\left(\frac{x-\nu}{x+\nu}\right); \qquad x < -\nu. \quad (2.10)$$

Moreover, Eq. (2.5) holds for $x < -(\nu+2)$, and substitution of this inequality into Eq. (2.1) leads, upon integration, to

$$f(x) \geq \left\{\frac{\epsilon}{2\nu} \ln\left(\frac{x-\nu}{x+\nu}\right) - \frac{\epsilon^2}{2\nu} \times \int_{-\infty}^x \ln\left[\frac{(y+2-\nu)(y+\nu)}{(y+2+\nu)(y-\nu)}\right] \frac{dy}{(y^2-\nu^2)}\right\}; \quad x < -(\nu+2). \quad (2.11)$$

Let us now consider the case $0 < \nu < 1$. Combination of Eqs. (2.7) to (2.9) gives

$$N^{-1} = \left[2 + f(x) + \Psi(x) + F(x) + (B - A)\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\right]; \quad 0 < \nu < 1; \quad -(\nu+2) < x < (\nu-2), \quad (2.12)$$

where

$$\Psi(x) = \epsilon^2\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu} \int_{(2+x)}^{\nu} \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu} \times \left\{\int_{-\nu}^{w}\left[1 - \left(\frac{y+2-\nu}{y+2+\nu}\right)^{\epsilon/2\nu}\right]\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} \frac{dy}{(\nu^2-y^2)}\right\} \frac{dw}{[(2-w)^2-\nu^2]}, \quad (2.13)$$

$$F(x) = \epsilon^3\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu} \int_{(2+x)}^{\nu} \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu}\left\{\int_{-\nu}^{w}\left(\frac{y+2-\nu}{y+2+\nu}\right)^{\epsilon/2\nu}\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} \times \left[\int_{(y+2)}^{\infty} \frac{[N^{-1} - f(z+2)]}{(z^2-\nu^2)}\left(\frac{z+\nu}{z-\nu}\right)^{\epsilon/2\nu} dz\right] \frac{dy}{(\nu^2-y^2)}\right\} \frac{dw}{[(2-w)^2-\nu^2]}, \quad (2.14)$$

and

$$A = \epsilon \int_{\nu}^{(2-\nu)} \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{w-\nu}{w+\nu}\right)^{\epsilon/2\nu} \times \frac{dw}{[(2-w)^2-\nu^2]}, \quad (2.15)$$

$$B = \epsilon^2 \int_{\nu}^{(2-\nu)} \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{w-\nu}{w+\nu}\right)^{\epsilon/2\nu} \times \left[\int_{w}^{\infty} \frac{[N^{-1}-f(y+2)]}{(y^2-\nu^2)}\left(\frac{y+\nu}{y-\nu}\right)^{\epsilon/2\nu} dy\right] \times \frac{dw}{[(2-w)^2-\nu^2]}. \quad (2.16)$$

But, from Eq. (2.4),

$$0 \leq [N^{-1} - f(x)] \leq (\epsilon/2\nu)\ln(1+\nu); \qquad x \geq (2+\nu). \quad (2.17)$$

Hence, from Eqs. (2.13) through (2.17),

$$0 \leq B \leq (\epsilon/2\nu)\ln(1+\nu)[(1-\nu)^{\epsilon/2\nu} - A], \quad (2.18)$$

and

$$0 \leq F(x) \leq (\epsilon/2\nu)\ln(1+\nu)\Psi(x). \quad (2.19)$$

Also, it is seen from Eq. (2.13) that

$$0 \leq \Psi(x) \leq \left[1 - (1-\nu)^{\epsilon/2\nu}\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\right]. \quad (2.20)$$

Hence, from Eqs. (2.10), (2.12), and (2.18) to (2.20),

$$0 \leq \left[N^{-1} - 2 - \Psi(x) + A\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\right] \leq \frac{\epsilon}{2\nu}\ln\left[\frac{(1+\nu)(x-\nu)}{(x+\nu)}\right]; \qquad 0<\nu<1; \quad -(\nu+2) < x < (\nu-2). \quad (2.21)$$

We remark that if we let $x \to -(\nu+2)$ in Eqs. (2.11) and (2.12), from opposite directions of course, then from Eqs. (2.4), (2.6), (2.10), and (2.13) to (2.16) we can obtain an explicit expression for N^{-1} which is correct to $O(\epsilon^2)$, rather than to $O(\epsilon)$ as in Eq. (2.21).

We now set $x = -2$ (for convenience) in Eq. (2.21) and examine $\Psi(-2)$ and A. From Eq. (2.15) it may be shown that

$$0 \leq \left[A\left(\frac{1+\nu}{1-\nu}\right)^{\epsilon/2\nu} - 1\right] \leq 2[(1+\nu)^{\epsilon/2\nu} - 1]. \quad (2.22)$$

Also, from Eq. (2.13),

$$0 \leq [\Psi(-2) - IJ] \leq \left[\left(\frac{2+\nu}{2-\nu}\right)^{\epsilon/2\nu} - 1\right] \times \left[\left(\frac{2-\nu}{2+\nu}\right)^{\epsilon/2\nu} - (1-\nu)^{\epsilon/2\nu}\right], \quad (2.23)$$

where

$$I = \epsilon\left(\frac{2+\nu}{2-\nu}\right)^{\epsilon/2\nu} \int_0^{\nu} \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu} \times \left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu} \frac{dw}{[(2-w)^2-\nu^2]}, \quad (2.24)$$

and

$$J = \epsilon \int_{-\nu}^{0} \left[1 - \left(\frac{y+2-\nu}{y+2+\nu}\right)^{\epsilon/2\nu}\right]\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} \frac{dy}{(\nu^2-y^2)} = \left\{\left[1 - \left(\frac{2-\nu}{2+\nu}\right)^{\epsilon/2\nu}\right] + \left(\frac{2-\nu}{2+\nu}\right)^{\epsilon/2\nu} I\right\}. \quad (2.25)$$

It remains to estimate I. It may be shown that

$$(\tfrac{1}{2})^{\epsilon/2\nu}(1-L) \leq I \leq (1-L); \qquad M \leq L \leq [\tfrac{1}{2}(2+\nu)]^{\epsilon/2\nu} M, \quad (2.26)$$

where

$$M = \frac{\epsilon}{2\nu}\int_0^1 \left[1 - \frac{\nu x}{(2-\nu)}\right]^{\epsilon/2\nu}(1-x)^{(\epsilon/2\nu-1)}\,dx = F\left[-\frac{\epsilon}{2\nu}, 1; 1+\frac{\epsilon}{2\nu}; \frac{\nu}{(2-\nu)}\right]. \quad (2.27)$$

But it may be shown that

$$\tfrac{1}{2}\left[1 + \left(\frac{2-2\nu}{2-\nu}\right)^{\epsilon/\nu}\right] \leq M \leq \tfrac{1}{2}\left(\frac{2}{2-\nu}\right)^{\epsilon/2\nu}\left[1 + \left(\frac{2-2\nu}{2-\nu}\right)^{\epsilon/\nu}\right]. \quad (2.28)$$

Hence, combining the various results,

$$N^{-1} \sim [2 + \tfrac{1}{4}[1-(1-\nu)^{\epsilon/\nu}]^2 - (1-\nu)^{\epsilon/\nu}] = \tfrac{1}{4}[3-(1-\nu)^{\epsilon/\nu}]^2, \qquad (0<\nu<1), \quad (2.29)$$

where the error is $O(\epsilon)$ and explicit bounds on it may be written down.

3. THE CASE $1 < \nu < 2$

We here consider the case $1 < \nu \leq (2-\delta)$, $\delta > 0$. We also assume that $\epsilon/2\nu < 1$. For $|x| < \nu$, we may write Eq. (2.8) in the form

$$[1 + f(x) - f(\nu+2)] = \epsilon\left(\frac{\nu-x}{\nu+x}\right)^{\epsilon/2\nu} \times \int_{-\nu}^{x} \frac{[f(y+2)-f(\nu+2)]}{(\nu^2-y^2)}\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy. \quad (3.1)$$

But, from Eq. (2.3),

$$0 \leq [f(\nu+2) - f(y+2)] \leq \frac{\epsilon}{2\nu}\ln\left[\frac{(y+2+\nu)}{(1+\nu)(y+2-\nu)}\right]; \qquad (\nu-2) < y \leq \nu. \quad (3.2)$$

Hence, $[f(\nu+2)-f(y+2)]/(\nu-y)$ remains finite as $y \to \nu - 0$. Thus, for $\epsilon/2\nu < 1$,

$$f(x) = [f(\nu+2)-1] - \alpha\left(\frac{\nu-x}{\nu+x}\right)^{\epsilon/2\nu} + \epsilon\left(\frac{\nu-x}{\nu+x}\right)^{\epsilon/2\nu}\int_x^\nu \frac{[f(\nu+2)-f(y+2)]}{(\nu^2-y^2)} \times \left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy; \qquad |x| < \nu, \tag{3.3}$$

where

$$\alpha = \epsilon\int_{-\nu}^{\nu} \frac{[f(\nu+2)-f(y+2)]}{(\nu^2-y^2)}\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy. \tag{3.4}$$

From Eqs. (2.7), (2.9), and (3.3),

$$N^{-1} = \{2 + f(x) + [N^{-1} - f(\nu+2)] + \alpha P(x) - Q(x)\}; \qquad -(4-\nu) < x < -\nu, \tag{3.5}$$

where

$$P(x) = \epsilon\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\int_{(2+x)}^{(2-\nu)}\left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu} \times \left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\frac{dw}{[(2-w)^2-\nu^2]}, \tag{3.6}$$

and

$$Q(x) = \epsilon^2\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu} \times \int_{(2+x)}^{(2-\nu)}\left(\frac{2-\nu-w}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu} \times \left\{\int_w^\nu \frac{[f(\nu+2)-f(y+2)]}{(\nu^2-y^2)} \times \left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy\right\}\frac{dw}{[(2-w)^2-\nu^2]}. \tag{3.7}$$

Also, from Eqs. (3.3) and (3.4),

$$\alpha(1-\sigma) = [(\nu-1)^{\epsilon/2\nu} + (\beta-\gamma)], \tag{3.8}$$

where

$$\sigma = \epsilon\int_{(2-\nu)}^{\nu}\left(\frac{y-2+\nu}{2+\nu-y}\right)^{\epsilon/2\nu} \times \left(\frac{\nu-y}{\nu+y}\right)^{\epsilon/2\nu}\frac{dy}{(\nu^2-y^2)}, \tag{3.9}$$

$$\beta = \epsilon\int_{(\nu-2)}^{\nu}\frac{[f(\nu+2)-f(y+2)]}{(\nu^2-y^2)} \times \left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy, \tag{3.10}$$

and

$$\gamma = \epsilon^2\int_{(2-\nu)}^{\nu}\left(\frac{w-2+\nu}{2+\nu-w}\right)^{\epsilon/2\nu}\left(\frac{\nu-w}{\nu+w}\right)^{\epsilon/2\nu} \times \left\{\int_w^\nu \frac{[f(\nu+2)-f(y+2)]}{(\nu^2-y^2)} \times \left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy\right\}\frac{dw}{[\nu^2-(2-w)^2]}. \tag{3.11}$$

We now set $x = -2$ in Eq. (3.5) and obtain, using Eqs. (2.4) and (2.10),

$$0 \le [N^{-1} - 2 - \alpha P(-2) + Q(-2)] \le \frac{\epsilon}{2\nu}\ln\left[\frac{(1+\nu)(2+\nu)}{(2-\nu)}\right]. \tag{3.12}$$

We begin by estimating $Q(-2)$. From Eq. (3.2)

$$0 \le \frac{[f(\nu+2)-f(y+2)]}{(\nu-y)} \le \frac{\epsilon}{2\nu^2}\ln\left[\frac{(2+\nu)}{(1+\nu)(2-\nu)}\right] \quad 0 \le y < \nu. \tag{3.13}$$

Hence, for $\epsilon/2\nu < 1$, it may be shown from Eq. (3.7) that

$$0 \le Q(-2) \le \frac{\epsilon^2}{2\nu^2(1-\epsilon/2\nu)}\ln\left[\frac{(2+\nu)}{(1+\nu)(2-\nu)}\right]. \tag{3.14}$$

Also, from Eq. (3.11),

$$0 \le \gamma \le \frac{\epsilon^2(\nu-1)^{\epsilon/2\nu}}{2\nu^2(1-\epsilon/2\nu)}\ln\left[\frac{(2+\nu)}{(1+\nu)(2-\nu)}\right]. \tag{3.15}$$

Further, it may be shown from Eq. (3.6) that

$$(\nu/2)^{\epsilon/2\nu}(1-V) \le P(-2) \le (1-V); \qquad (1-W) \le V \le [(2+\nu)/2\nu]^{\epsilon/2\nu}(1-W), \tag{3.16}$$

where

$$W = F\left[-\frac{\epsilon}{2\nu}, 1; 1+\frac{\epsilon}{2\nu}; \frac{(2-\nu)}{\nu}\right], \tag{3.17}$$

and hence,

$$\tfrac{1}{2}\left[1+\left(\frac{2\nu-2}{\nu}\right)^{\epsilon/\nu}\right] \le W \le \tfrac{1}{2}\left(\frac{2}{\nu}\right)^{\epsilon/2\nu}\left[1+\left(\frac{2\nu-2}{\nu}\right)^{\epsilon/\nu}\right]. \tag{3.18}$$

It remains to estimate α from Eq. (3.8). Bounds on γ are given by Eq. (3.15). Now, for $\epsilon/2\nu < 1$,

$$[(\nu-1)^{\epsilon/2\nu} + \beta] = \int_{(\nu-2)}^{\nu} f'(y+2)\left(\frac{\nu+y}{\nu-y}\right)^{\epsilon/2\nu} dy, \tag{3.19}$$

as may be shown from Eq. (3.10) by integrating by parts, and using the relation $[f(\nu+2)-f(\nu)]=1$, which follows from Eq. (2.1). But, from Eq. (2.1) and Eq. (2.7), again by integration by parts, it follows that

$$f'(x)=\frac{\epsilon}{(x^2-\nu^2)}\left(\frac{x-\nu}{x+\nu}\right)^{\epsilon/2\nu}\times\left[1+\int_x^\infty f'(y+2)\left(\frac{y+\nu}{y-\nu}\right)^{\epsilon/2\nu}dy\right];\qquad x>\nu. \tag{3.20}$$

From Eq. (2.3) it may be shown that, for $x>\nu$ and $\epsilon/2\nu<1$,

$$0\le\int_x^\infty f'(y+2)\left(\frac{y+\nu}{y-\nu}\right)^{\epsilon/2\nu}dy\le\frac{\epsilon}{2\nu}(\nu+1)^{\epsilon/2\nu}\left[\left(1-\frac{\epsilon}{2\nu}\right)^{-1}+\ln\left(\frac{2+\nu}{2}\right)\right]=\chi(\epsilon). \tag{3.21}$$

Thus, from Eqs. (3.19) to (3.21),

$$\xi\le[(\nu-1)^{\epsilon/2\nu}+\beta]\le\xi[1+\chi(\epsilon)], \tag{3.22}$$

where

$$\xi=(\eta+\zeta)=\epsilon\left[\int_\nu^{(\nu+1)}+\int_{(\nu+1)}^{(\nu+2)}\right]\left(\frac{y-2+\nu}{2+\nu-y}\right)^{\epsilon/2\nu}\times\left(\frac{y-\nu}{y+\nu}\right)^{\epsilon/2\nu}\frac{dy}{(y^2-\nu^2)}. \tag{3.23}$$

It may be shown that

$$0\le\zeta\le\frac{\epsilon(2\nu)^{\epsilon/2\nu}}{(2\nu+1)(1-\epsilon/2\nu)}. \tag{3.24}$$

Also,

$$(2)^{-\epsilon/2\nu}(1-y)\le\left(\frac{2\nu+1}{2\nu-1}\right)^{\epsilon/2\nu}\eta\le(1-y);$$

$$(1-z)\le y\le[(2\nu+1)/2\nu]^{\epsilon/2\nu}(1-z), \tag{3.25}$$

where

$$z=F\left[-\frac{\epsilon}{2\nu},1;1+\frac{\epsilon}{2\nu};\frac{1}{(2\nu-1)}\right], \tag{3.26}$$

and hence

$$\tfrac{1}{2}\left[1+\left(\frac{2\nu-2}{2\nu-1}\right)^{\epsilon/\nu}\right]\le z\le\tfrac{1}{2}\left(\frac{2\nu}{2\nu-1}\right)^{\epsilon/2\nu}\times\left[1+\left(\frac{2\nu-2}{2\nu-1}\right)^{\epsilon/\nu}\right]. \tag{3.27}$$

The final quantity to be estimated is σ, as given by Eq. (3.9). Now,

$$\sigma=(\nu-1)^{\epsilon/\nu}(1-R), \tag{3.28}$$

where

$$R=\frac{\epsilon}{2\nu}\int_0^1\{1-(1-x)^{\epsilon/2\nu}\times[1+(\nu^2-1)x]^{-\epsilon/2\nu}\}x^{(\epsilon/2\nu-1)}\,dx. \tag{3.29}$$

It may be shown that

$$0\le R\le\{(\nu^2+1)(\epsilon/2\nu)^2+[1-(\tfrac{1}{2})^{\epsilon/2\nu}]^2\}. \tag{3.30}$$

Thus,

$$(1-\sigma)=\{[1-(\nu-1)^{\epsilon/\nu}]+O(\epsilon^2)\}. \tag{3.31}$$

Combining the results of this section, we obtain

$$N^{-1}=\left\{2+\frac{[1+(\nu-1)^{\epsilon/\nu}]^2}{4[1-(\nu-1)^{\epsilon/\nu}]}\right\}[1+O(\epsilon)]=\frac{[3-(\nu-1)^{\epsilon/\nu}]^2}{4[1-(\nu-1)^{\epsilon/\nu}]}[1+O(\epsilon)];\qquad 1<\nu<(2-\delta),\quad \delta>0, \tag{3.32}$$

and explicit bounds on the error term may be written down.

Exact Solution for the Vibrations of a Nonlinear Continuous Model String

NORMAN J. ZABUSKY
*Bell Telephone Laboratories, Inc., Whippany, New Jersey**
(Received April 30, 1962)

An *exact* solution is given for the partial differential equation

$$y_{tt} = [1 + \epsilon y_x]^{\alpha} y_{xx},$$

which describes the standing vibrations of a finite, continuous, and nonlinear string. The nonlinearity studied, $[1 + \epsilon y_x]^{\alpha}$, was motivated by the work of Fermi, Pasta, and Ulam (1955), where they reported on numerical studies of the "equipartition of energy" in nonlinear systems. To obtain the solution, the above equation is transformed into a linear equation by inverting the roles of the dependent ($u = y_x$ and $v = y_t$) and independent (x and t) variables. Riemann's method of integration is applied to the problem and the solutions for t and x are written as integrals. The nature of the "inverse Riemann plane," how it is related to the initial conditions, and how one unfolds it, are discussed in detail. A general procedure is described for *reinverting* the solution, so that y can be written as a function of x and t. It is illustrated to order ϵ for the above problem. *It is demonstrated that y_{xx} becomes singular, that is, y_x develops a discontinuity after an elapsed time of order* $(1/\epsilon)$. The methods described are applicable to any nonlinear string where the coefficient of y_{xx} is a function of y_x only. The effect of higher spatial derivatives on the formation of the singularity is discussed.

1. INTRODUCTION AND SUMMARY

THE phenomena associated with wave propagation and standing oscillations in nonlinear media (or fields) have been of interest to physicists for a long time.[1] The equations of hydrodynamics[2] and nonlinear electrodynamics (for example, the Born-Infeld theory[3]) are pertinent illustrations.

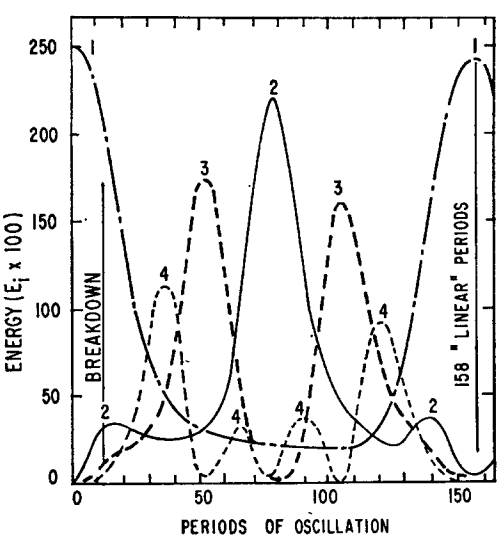


FIG. 1. The energy (arbitrary units) in the low modes of the nonlinear string:
$\partial^2 y_i/\partial\tau^2 = (y_{i+1} - 2y_i + y_{i-1})\{1 + \alpha(y_{i+1} - y_{i-1})\}$.
Reproduced from Fermi, Pasta, and Ulam,[4] Fig. 1 ($N = 32$, $\alpha = 1/4$, and $a = 1$).

Fermi, Pasta, and Ulam[4] (FPU) published a report in 1955, wherein they studied, numerically, the phenomena associated with the vibration of a *nonlinear, loaded* (or beaded), and finite string. They used the nonlinear string as a model of a dynamical system which they expected would exhibit certain ergodic properties. In particular, they hoped to establish computationally the rate of approach to "equipartition of energy" among the various modes—that is, the normal modes or degrees of freedom of the equivalent linear system. They employed three different types of nonlinear perturbations and started from a variety of initial conditions. However, the same "unexpected" phenomenon occurred in all calculations, namely, that the system of particles showed "... very little, if any, tendency toward equipartition of energy among all degrees of freedom ...."

This is exhibited in Fig. 1, which is taken from FPU (their Fig. 1), where the energy in a given mode is plotted as a function of time. If initially the energy resided in a few of the low modes (in Fig. 1 only the first mode is initially excited), then during the subsequent vibrations the energy "flowed"

* This work was done at the Princeton University Plasma Physics Laboratory under the auspices of the U. S. Atomic Energy Commission. The results were reported at the New York meeting of the American Physical Society, January, 1962. See Bull. Am. Phys. Soc. 7, 31 (1962).

[1] T. Taniuti, Progr. Theoret. Phys., Suppl. No. 9, 69 (1958).

[2] R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves* (Interscience Publishers, Inc., New York, 1948).

[3] M. Born and L. Infeld, Proc. Roy. Soc. (London) **A 144**, 425 (1934); **A 147**, 522 (1934); **A 150**, 141 (1935).

[4] E. Fermi, J. R. Pasta, and S. Ulam, "Studies of Nonlinear Problems I," Los Alamos Report No. 1940, May, 1955 (unpublished). The problem studied in this report is described briefly in *A Collection of Mathematical Problems* by S. Ulam (Interscience Publishers, Inc., New York, 1960), Chap. 7, paragraph 8.

to the neighboring higher modes due to the nonlinear coupling. After a large number of oscillation periods of the equivalent linear string (the abscissa of Fig. 1), the energy-flow direction was reversed such that the original initial state was almost reached again.

In formulating their problem, they wished to consider a "one-dimensional continuum."[5] However, for the purposes of numerical computation they employed the model of a loaded string, which is equivalent to replacing the spatial partial derivatives by spatial differences. Two nonlinear differential-difference equations which FPU studied are given in Eq. (1.1), the "quadratic" nonlinearity, and Eq. (1.2), the "cubic" nonlinearity:

$$\partial^2 y_i/\partial\tau^2 = (y_{i+1} - 2y_i + y_{i-1}) + \alpha[(y_{i+1} - y_i)^2 - (y_i - y_{i-1})^2] \tag{1.1}$$

or

$$\partial^2 y_i/\partial\tau^2 = (y_{i+1} - 2y_i + y_{i-1}) \times [1 + \alpha(y_{i+1} - y_{i-1})];$$

and

$$\partial^2 y_i/\partial\tau^2 = (y_{i+1} - 2y_i + y_{i-1}) + \beta[(y_{i+1} - y_i)^3 - (y_i - y_{i-1})^3] \tag{1.2}$$

or

$$\partial^2 y_i/\partial\tau^2 = (y_{i+1} - 2y_i + y_{i-1})\{1 + (\tfrac{1}{2}\beta) \times [(y_{i+1} - y_i)^2 + (y_{i+1} - y_{i-1})^2 + (y_i - y_{i-1})^2]\}.$$

y_i represents the displacement of the ith mass point. In some of their calculations they considered as many as $N = 64$ equimass particles. These equations are normalized in the sense that the ratio of the tension to the particle mass has been set equal to unity. The spacing between particles has been absorbed into the normalized time τ.

Kruskal first drew the author's attention to this problem at a lecture (January, 1961), where he presented a procedure which readily yields an approximate solution to the problem of the vibrations of the *continuous* (see Sec. 2) nonlinear string.[6] After thirteen equivalent linear oscillations, his solution of (1.1) exhibited a *discontinuity* in the first derivative of the displacement. The energy in the first three modes was obtained by spatially Fourier analyzing the amplitude of vibration. These energies, when graphed, are in good agreement with the FPU plots up until the breakdown (discontinuity) time.

2. PROBLEM STATEMENT

In this paper we will determine the *exact* solution for the vibrations of a continuous nonlinear string. The continuum equations which are equivalent to (1.1) and (1.2) are obtained by taking the limit $h \to 0$ (h = particle spacing) and $N \to \infty$ such that: $L = Nh$ remains finite and (α/N) remains finite and approaches $\frac{1}{2}\epsilon$.[7] Thus,

$$(y_{i+1} - y_i) \to (y_i - y_{i-1}) \to \tfrac{1}{2}(y_{i+1} - y_{i-1}) \to h(\partial y/\partial\xi)$$

and

$$(y_{i+1} - 2y_i + y_{i-1}) \to h^2(\partial^2 y/\partial\xi^2).$$

We normalize by introducing

$$t = \tau/N \tag{2.1}$$

$$x = \xi/L \tag{2.2}$$

and obtain the partial differential equations:

$$\partial^2 y/\partial t^2 = (\partial^2 y/\partial x^2)[1 + \epsilon\,\partial y/\partial x], \tag{2.3}$$

$$\partial^2 y/\partial t^2 = (\partial^2 y/\partial x^2)[1 + \epsilon'(\partial y/\partial x)^2], \tag{2.4}$$

where

$$\epsilon = 2\alpha/N \quad \text{and} \quad \epsilon' = 3\beta/N^2. \tag{2.5}$$

ϵ and ϵ' are taken as finite in the above limit, $N \to \infty$. The initial conditions are prescribed over the range $0 < x < 1$ as

$$y(x, 0) = y_0(x), \qquad \partial y/\partial t|_{t=0} = 0. \tag{2.6}$$

The procedure which will be described is applicable to the more general initial condition, where the temporal derivative at $t = 0$ is finite.

The boundary conditions in the FPU report and in the problem considered below are

$$y(0, t) = y(1, t) = 0. \tag{2.7}$$

[5] E. Fermi, J. R. Pasta, and S. Ulam, reference 4, p. 3, "We imagine a one-dimensional continuum . . . ."

[6] This procedure will be described in a forthcoming paper. The properties of the solution which are derived will be studied, and the variation of the energy in the "normal" modes will be presented.

[7] The equation which describes the large-amplitude, *pure* longitudinal vibrations of a continuous string is given as $\rho y_{tt} = T_x$. In general, the tension T is a function of y_x, and its specific form depends upon the stress-strain relationship of the string material. For a linear stress-strain relation, $T = T_0[1 + y_x]$. The model string, (2.3), employed by FPU and analyzed here, implies that $T = T_0[1 + y_x + \frac{1}{2}\epsilon\, y_x^2]$.

It is the aim of this paper to derive the exact solution of (2.3) subject to the initial conditions (2.6) and boundary conditions (2.7). It will be shown that at values of t of order $\approx 1/(\epsilon a \pi^2)$ (a is the maximum amplitude of the initial condition) the continuous nonlinear string develops a discontinuity in the first derivative, $\partial y/\partial x$.

The procedure for deriving the exact solution depends only on the fact that the governing partial differential equation can be written in the form

$$\partial^2 y/\partial t^2 = [F(\partial y/\partial x)]^2 (\partial^2 y/\partial x^2). \tag{2.8}$$

Equation (2.8) is also the Lagrangian representation for describing longitudinal waves in a nonlinear, hysteresis-free, solid continuum.[8] Here, $y = y(x, t)$ is the displacement of an element of the material from an initial reference state and $F^2(\partial y/\partial x)$ is proportional to the derivative of the stress with respect to the strain [strain $= (\partial y/\partial x) - 1$]. For ordinary metals, F is an even function of the strain and vanishes as the strain gets large.

3. TRANSFORMATION TO AN EQUIVALENT LINEAR REPRESENTATION

A. The Equivalent Linear Partial Differential Equation

If we define

$$u = y_x \qquad v = y_t,^9 \tag{3.1}$$

then (2.8) can be written as two coupled partial differential equations of first order:

$$u_t - v_x = 0 \tag{3.2}$$

$$v_t - F^2(u)u_x = 0. \tag{3.3}$$

Equation (3.2) is a consistency condition, whereas (3.3) describes the nonlinear behavior as given in (2.8). If we multiply (3.2) by $F(u)$ and add and subtract the result to (3.3), we obtain the pair of equations:

$$v_t \pm Fu_t \mp Fv_x - F^2u_x = 0 \tag{3.4}$$

or

$$r_t - F(u)r_x = 0 \tag{3.5}$$

$$s_t + F(u)s_x = 0, \tag{3.6}$$

where

$$\begin{matrix} r \\ s \end{matrix} = \pm v + \int^u F(u')\,du' = \pm v + B(u). \tag{3.7}$$

The variables r and s are commonly known as the *Riemann invariants*, in that r is invariant along the characteristic $dx/dt = -F$ and s is invariant along the characteristic $dx/dt = +F$.

We can express u in terms of $(r + s)$ by adding the equations of (3.7) and taking the inverse function of the result,

$$u = B^{-1}(\tfrac{1}{2}[r + s]). \tag{3.8}$$

We now recognize that an inversion transformation will linearize exactly the resulting equation. In this so-called "speedgraph"[10] transformation, the dependent variables are x and t and are expressed as functions of the independent variables r and s. Thus, we write

$$\begin{aligned} r_x &= jt_s \qquad & r_t &= -jx_s \\ s_x &= -jt_r \qquad & s_t &= jx_r, \end{aligned} \tag{3.9}$$

where the Jacobian of the transformation from the (x, t) to the (r, s) plane is given by

$$j = r_x s_t - s_x r_t. \tag{3.10}$$

Note that

$$j = J^{-1} = [x_r t_s - x_s t_r]^{-1}. \tag{3.11}$$

The resulting equations in the speedgraph plane are

$$\begin{aligned} x_s + F(u)t_s &= 0 \\ x_r - F(u)t_r &= 0. \end{aligned} \tag{3.12}$$

For example, if

$$F^2 = (1 + \epsilon u)^\alpha, \tag{3.13}$$

then

$$\begin{matrix} r \\ s \end{matrix} = \pm v + [\tfrac{1}{2}\epsilon(\alpha + 2)]^{-1}(1 + \epsilon u)^{\{1 + (1/2)\alpha\}}. \tag{3.14}$$

Thus, adding the equations of (3.14) and taking the inverse yields

$$u = -\epsilon^{-1} + \epsilon^{-1}[\tfrac{1}{4}\epsilon(\alpha + 2)(r + s)]^{\{2/(2+\alpha)\}}. \tag{3.15}$$

Note that v is obtained by differencing the equations of (3.14)

$$v = \tfrac{1}{2}(r - s). \tag{3.16}$$

If we substitute (3.15) into (3.13) we can write

$$F = [\beta(r + s)]^{[\alpha/(2+\alpha)]}, \tag{3.17}$$

where

$$\beta = \tfrac{1}{4}\epsilon(\alpha + 2). \tag{3.18}$$

[8] R. Courant and K. O. Friedrichs, reference 2. See paragraphs 97 and 98.

[9] We now employ the subscript notation to designate partial derivatives. For example, $r_t = \partial r/\partial t$.

[10] R. von Mises, *Mathematical Theory of Compressible Fluid Flow* (Academic Press Inc., New York, 1958).

In our case, $\alpha = 1$ and the resulting expressions become

$$\begin{matrix} r \\ s \end{matrix} = \pm v + (2/3\epsilon)(1 + \epsilon u)^{3/2}; \tag{3.19}$$

$$F(u) = [\beta(r + s)]^{1/3}, \quad \text{where} \quad [\beta = 3\epsilon/4]; \tag{3.20}$$

$$j = -2r_x s_x[\beta(r + s)]^{1/3}. \tag{3.21}$$

The symmetrical equations become

$$x_s + [\beta(r + s)]^{1/3} t_s = 0; \tag{3.22}$$

$$x_r - [\beta(r + s)]^{1/3} t_r = 0. \tag{3.23}$$

If we eliminate x from (3.22) and (3.23), we obtain the canonical partial differential equation for t, namely,

$$t_{rs} + [n/(r + s)](t_r + t_s) = 0, \tag{3.24}$$

where

$$n = \tfrac{1}{2}\alpha[2 + \alpha]^{-1}$$
$$= 1/6 \quad \text{for} \quad \alpha = 1. \tag{3.25}$$

The equation for x (obtained by eliminating t from the pair) is identical with (3.24) except that n is replaced by $-n$.

Equation (3.24) is identically the equation obtained in the one-dimensional nonlinear hydrodynamics of an isentropic medium and was first derived by Riemann[11] in 1860. It is commonly referred to in the literature as the Euler-Poisson-Darboux equation.

Table I summarizes some of the values of α which yield integral n. A great simplification is obtained when n is integral, for the solution of (3.24) and its adjoint differential equation (defined below) can be expressed as the ratio of polynomials.

In what follows we deal with the x and t equations separately, since it is easier to visualize the application of the boundary conditions.

4. AN INTEGRAL REPRESENTATION FOR THE EXACT SOLUTION

A. General Considerations

As observed above, the method for obtaining the characteristic solutions of the x and t equations has been given by Riemann. The general concepts involved in Riemann's method of integration are adequately described in several classic textbooks[12] on mathematical physics and, therefore, only a brief outline of it is given below.

If one is given the linear partial differential equation

$$L(t) = t_{rs} + at_r + bt_s + ct = 0, \tag{4.1}$$

then the operator M is defined by

$$M(G) = G_{rs} - (aG)_r - (bG)_s + cG = 0, \tag{4.2}$$

and is known as the *adjoint* of L. $G = G(r, s; \xi, \eta)$ is commonly called the Riemann function. As described below, $r = \xi$ and $s = \eta$ are two lines along which G satisfies specific conditions.

One verifies that

$$GL(t) - tM(G) = R_r + S_s, \tag{4.3}$$

where

$$R = \tfrac{1}{2}(Gt_s - tG_s) + aGt \tag{4.4}$$

$$S = \tfrac{1}{2}(Gt_r - tG_r) + bGt. \tag{4.5}$$

If the functions t and G are such that $L(t) = 0$ and $M(G) = 0$ in a region bounded by a sufficiently smooth curve Γ, and if we apply Green's theorem to (4.3), we obtain

$$0 = \int dr\, ds[R_r + S_s] = \int_\Gamma [n_r R + n_s S]\, dl, \tag{4.6}$$

where dl represents differential arc length and n_r and n_s are components in the r and s directions of the outward pointing normal, as shown in Fig. 2.

Equation (4.6) can be written as

$$\int_{P_A}^{P_B} [n_r R + n_s S]\, dl + \int_{P_B}^{P} R\, ds - \int_{P}^{P_A} S\, dr = 0. \tag{4.7}$$

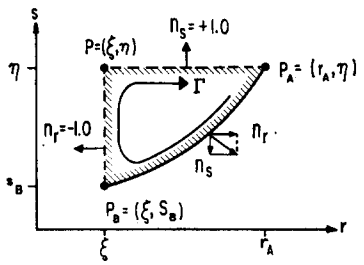


FIG. 2. Cauchy problem of the first kind along an arbitrary line.

TABLE I. Values of α corresponding to integral n.

n	−3	−2	−1	0	1	2	3
α	−12/7	−8/5	−4/3	0	−4	−8/3	−12/5

[11] Riemann's *Gesammelte Mathematische Werke*, 1876, I. See Chap. VIII, paragraph 2. First published in Abhandl. Ges. Wiss. Göttingen, Math.-physik. Kl. **8**, 43 (1860).

[12] R. Courant and D. Hilbert, *Methoden der Mathematischen Physik, II* (Verlag Julius Springer, Berlin, 1937). Chap. 5, paragraph 4; A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press Inc., New York, 1949). See Chap. 3, paragraph 11.

If we require G, the solution of (4.2), to satisfy the conditions:

$$G_s - aG = 0 \quad \text{on the line } r = \xi; \tag{4.8}$$

$$G_r - bG = 0 \quad \text{on the line } s = \eta; \tag{4.9}$$

$$G = 1 \quad \text{at the point } P(r = \xi, s = \eta); \tag{4.10}$$

then the right side of (4.7) simplifies and we obtain the classic result

$$t(\xi, \eta) = t(P) = \tfrac{1}{2}[G(P_B)\cdot t(P_B) + G(P_A)\cdot t(P_A)] + \int_{P_A}^{P_B} (n_r R + n_s S)\, dl. \tag{4.11}$$

Thus, the value of t at any point P in the region bounded by P_AP_B, by the horizontal characteristic through P_A, and by the vertical characteristic through P_B can be expressed as an integral. This integration can be performed provided that t and the partial derivative t_r or t_s are given along P_AP_B.

B. Application to a Specific Problem

Consider the differential equation (3.24) with r and s defined by (3.19). We take (2.6) as our initial conditions and (2.7) as our boundary conditions in the *physical* (x, t) plane. The problem is simplified by writing it purely as an initial condition problem. This is accomplished if we require the solution to be odd about the origin and periodic over the spatial interval, $\Delta x' = 2$. Thus, the new initial conditions are written as

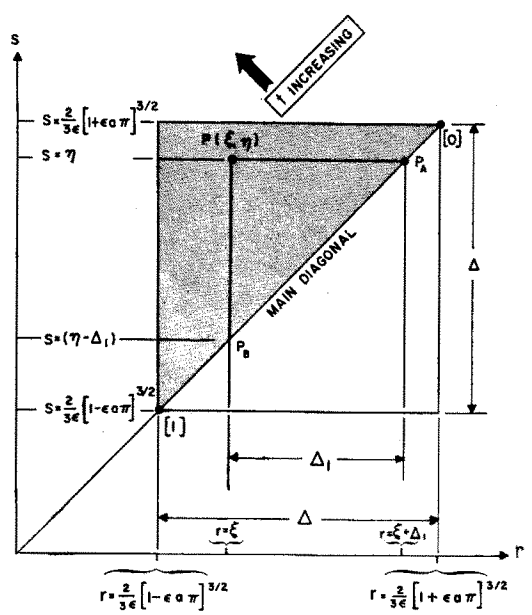


FIG. 3. The (r, s) plane for a periodic initial condition.

$$y(x', 0) = y_0(x') \quad \text{and} \quad y_t(x', 0) = 0, \quad \text{for} \quad -\infty < x' < \infty. \tag{4.12}$$

Let us examine the transformed line in the *inverse* (r, s) plane. Since $y_t(x', 0) = v = 0$ at $t = 0$, the initial condition line is the straight line $r = s$, as shown in Fig. 3. We call this line the "main diagonal." If $y_0(x') = a \sin \pi x'$, as it was in the FPU calculations, then along the main diagonal the mapping is given by

$$r|_{t=0} = \rho(x') = s|_{t=0} = \sigma(x') = (2/3\epsilon)f^{3/2}(x'), \tag{4.13}$$

where

$$f(x) = [1 + \epsilon a\pi \cos \pi x]. \tag{4.14}$$

We have used ρ and σ to designate the values of r and s at $t = 0$, that is, along the main diagonal of the inverse plane. Distance along the string at $t = 0$ is indicated by x'.

Let us now emphasize and clarify the distinction between the coordinate system and the points on the plane. We have used r and s as a *generic* designation for the points along the horizontal and vertical coordinate axes, respectively. The point $r = \xi$, $s = \eta$ is an arbitrary point in the (r, s) plane at which the solution is evaluated. As r and s are invariant along vertical and horizontal lines, respectively, then

$$\xi = (2/3\epsilon)f^{3/2}(x_B); \qquad \eta = (2/3\epsilon)f^{3/2}(x_A). \tag{4.15}$$

Hence, the t and x solutions obtained by Riemann integration can be expressed in terms of ξ and η or, equivalently, x_A and x_B.

Each point in the region $0 < x' < 1$ of the physical plane is mapped uniquely into a point along the main diagonal of the (r, s) plane. The solution at $P(\xi, \eta)$ is "influenced" by the initial conditions along P_AP_B.

If, at $t = 0$, one proceeds outside the main interval, $0 < x' < 1$, of the physical plane, then r and s will oscillate *to* and *fro* over the region shown in Fig. 3, because $u|_{t=0}$ was taken as a periodic function. This suggests a *many-sheeted* (r, s) plane. interval, $0 < x' < 1$, of the physical plane, r and s will oscillate *to* and *fro* over the region shown in Fig. 3, because u was taken as a periodic function, (4.14). This suggests a *many-sheeted* (r, s) plane. Procedures for continuing the solutions from one sheet to another by "unfolding" the (r, s) plane have been developed by Ludford[13,14] and will be applied to this problem in Sec. 5.

[13] G. S. S. Ludford, Proc. Cambridge Phil. Soc. **48**, 499 (1952).
[14] G. S. S. Ludford, J. Rat. Mech. **3**, 77 (1954).

For any point in the shaded region of Fig. 3 we can immediately write the solution for x and t by applying Riemann's method to this Cauchy problem of the first kind. Along P_AP_B, $t = 0$ and $dt = 0$. Thus,

$$r_x|_{t=0} = \rho_{x'} = (x'_\rho)^{-1} = s_x|_{t=0} = \sigma_{x'} = (x'_\sigma)^{-1}. \quad (4.16)$$

Applying (4.16) to (3.9) and (3.21), we obtain the result that along P_AP_B

$$t_\rho = -t_\sigma = \{\rho_{x'}[\beta(\rho + \sigma)]^{1/3}\}^{-1}. \quad (4.17)$$

Substituting (4.16) and (4.17) into (4.11) gives the solution for t as

$$t(\xi, \eta) = -(\tfrac{1}{2})^{1/2} \int_{P_A}^{P_B} Gt_\rho \, dl, \quad (4.18)$$

where we have taken $n_r = -n_s = (\frac{1}{2})^{1/2}$. G is the Riemann function associated with the t equation, (3.24). This integral can be expressed as the integral over the dummy variable x' by substituting into (4.18) the results

$$dl = -\sqrt{2}\, d\rho = -\sqrt{2}\,(\rho_{x'}\, dx') = -\sqrt{2}\,(\sigma_{x'}\, dx')$$

or

$$dl = \sqrt{2}\, a\pi^2 f^{1/2}(x') \sin \pi x' \, dx', \quad (4.19)$$

and

$$t_r \left.\frac{dr}{dx}\right|_0 = t_\rho \rho_{x'} = \tfrac{1}{2} f^{-1/2}(x'). \quad (4.20)$$

We obtain

$$t(\xi, \eta) = \tfrac{1}{2} \int_{x_A}^{x_B} Gf^{-1/2}(x') \, dx', \quad (4.21)$$

where $x_A = x(P_A)$ and $x_B = x(P_B)$.

Since $x_r = x_s$ along P_AP_B, the x solution is expressed by the general equation (4.11) as

$$x(\xi, \eta) = \tfrac{1}{2}[x(P_A)\Gamma(P_A) + x(P_B)\Gamma(P_B)] + \tfrac{1}{2} \int_{x_A}^{x_B} (\Gamma_\sigma - \Gamma_\rho) x' \rho_{x'} \, dx', \quad (4.22)$$

where Γ is the Riemann function associated with the x equation. For the problem being considered we can write

$$G = G^{(+)} \quad \text{and} \quad \Gamma = G^{(-)}, \quad (4.23)$$

since the t and x equations differ only by the sign of n. [See Eqs. (3.24) and (3.25).] Courant and Friedrichs[15] give $G^{(\pm)}$ as

$$G^{(\pm)} = [(\rho + \sigma)/(\xi + \eta)]^{\pm n} P_{(\pm n - 1)}(q), \quad (4.24)$$

where P_ν is the Legendre function of order ν [which has the property $P_\nu(q) = P_{-\nu-1}(q)$] and

$$q = 1 + [2(\rho - \xi)(\sigma - \eta)]/[(\xi + \eta)(\rho + \sigma)]. \quad (4.25)$$

The terms $(\rho + \sigma)$ and $(\xi + \eta)$ are both of order $(1/\epsilon)$, whereas $(\rho - \xi)$ and $(\sigma - \eta)$ are both of order (ϵ^0). At $t = 0$, $\rho = \sigma$ and

$$\tfrac{1}{2}(q - 1) = [(\rho - \xi)(\rho - \eta)]/[2\rho(\xi + \eta)] = \mathcal{O}(\epsilon^2). \quad (4.26)$$

From (4.25) one concludes that q and any function of q are even periodic functions of x with period 2. This follows more directly from (4.13), since both ρ and σ were assumed to be even periodic functions of x'.

The solutions for t and x are written in more explicit form by using the above definitions:

$$t(\xi, \eta) = \tfrac{1}{2}(C_1)^{-1} \int_{x_A}^{x_B} P_{-5/6}(q) f^{-1/4}(x') \, dx'; \quad (4.27)$$

$$x(\xi, \eta) = \tfrac{1}{2}C_1\{x_A f^{-1/4}(x_A) + x_B f^{-1/4}(x_B)\} - C_2 \int_{x_A}^{x_B} x' \Pi(x') \, dx'; \quad (4.28)$$

where C_1 and C_2 are functions of ξ and η and are given by

$$C_1 = +[(3\epsilon/4)(\xi + \eta)]^{1/6} = \mathcal{O}(\epsilon^0) \quad (4.29)$$

$$C_2 = +a(3\pi\epsilon/4)^2(\xi - \eta) \times [(3\epsilon/4)(\xi + \eta)]^{-5/6} = \mathcal{O}(\epsilon^2). \quad (4.30)$$

$\Pi(x')$ is an even periodic function of x' with period 2,

$$\Pi(x') = (dP_{1/6}/dq)(\sin \pi x') f^{-5/4}(x'). \quad (4.31)$$

To illustrate the essence of these equations, we recover the solution of the linear equation by setting $\epsilon = 0$. Therefore, $F(y_x) = 1$, $q = 1$, and $P_\nu(1) = +1$. Equations (4.27) and (4.28) become

$$t = \tfrac{1}{2}(x_B - x_A) \quad \text{and} \quad x = \tfrac{1}{2}(x_B + x_A), \quad (4.32)$$

because the integral contribution in the x equation, (4.28), is of $\mathcal{O}(\epsilon^2)$ and $C_1 \to 1$. We rewrite this set as

$$x_A = (x - t) \quad \text{and} \quad x_B = (x + t). \quad (4.33)$$

We form the difference $2v = (\xi - \eta)$, and obtain

$$2v = 2y_t = (2/3\epsilon)[f^{3/2}(x_B) - f^{3/2}(x_A)]. \quad (4.34)$$

Neglecting terms of $\mathcal{O}(\epsilon)$ yields

$$y_t = \tfrac{1}{2}a\pi[\cos \pi x_B - \cos \pi x_A]$$

[15] R. Courant and K. O. Friedrichs, reference 2, paragraph 82, Eq. (82.18).

or

$$y_t = -a\pi \sin \pi x \sin \pi t, \tag{4.35}$$

which is the solution for the linear wave equation $y_{tt} - y_{xx} = 0$, with the initial conditions: $y(x', 0) = a \sin \pi x'$ and $y_t(x', 0) = 0$; and with the boundary conditions: $y(0, t) = y(1, t) = 0$. An analogous procedure is used for obtaining the solution of the nonlinear problem. This is given below.

5. THE UNFOLDED INVERSE RIEMANN PLANE

A. General Considerations

In the previous section we have shown that the periodic initial conditions cause the entire line $-\infty < x' < \infty$ at $t = 0$ to be projected onto the main diagonal of Fig. 3 between the points [0] and [1]. The line [0] $\rightarrow$ [1] is many valued, and the region formed by the intersecting characteristics through these points is *many sheeted*. Each sheet corresponds to an interval of length 1 in the physical plane. As we proceed from $x' = 0$ to $x' = 1$ in the physical plane, we move in the negative r and s directions from [0] to [1]. If we continue from $x' = 1$ to $x' = 2$ in the physical plane, the transformed point moves from [1] to [0] in the (r, s) plane, etc. The direction of movement in the (r, s) plane is evident from (4.19).

A unidirectional movement in the physical plane is transformed into a to-and-fro movement in the (r, s) plane. The latter is rendered unidirectional by "unfolding" the main diagonal. Thus, proceeding from $x' = 0$ to positive values of x' is equivalent to proceeding from [0] down to the left. Passing through $x' = 1$ toward $x' = 2$ corresponds to passing through [1] and into the neighboring square of the unfolded (r, s) plane, where *the positive directions of r and s are reversed.* (See Fig. 4.) The reversal in direction of the coordinate system is

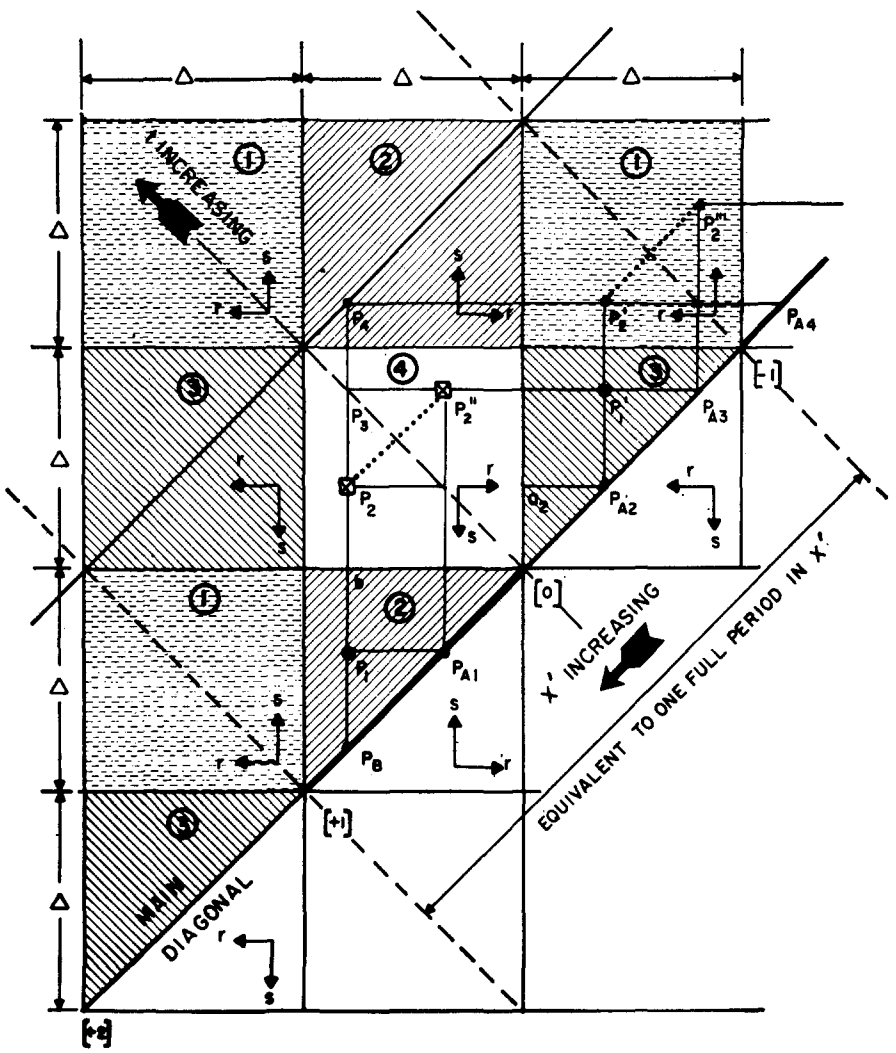


FIG. 4. The unfolded (r, s) plane. (Circled numbers correspond to barred numbers in the text.)

represented by the right angles in the lower right-hand corner of each square. Thus, the single square with sides of length Δ is unfolded into a sequence of squares along a line of slope $+1$, these squares having, *alternately*, the same properties.

The region above and to the left[16] of $[-1] \rightarrow [0]$ is designated as a $\bar{3}$ region, whereas that above and to the left of $[0] \rightarrow [1]$ is a $\bar{2}$ region. In the figures the barred quantities are designated by circled quantities. The Jacobian j given by (3.21) reduces to

$$j = -2[a\pi^2 \sin \pi x' f^{3/4}(x')]^2 \qquad (5.1)$$

along the main diagonal, and thus vanishes at the points $\ldots [-1], [0], [1], \ldots$. This vanishing is manifest in the linear problem as well, and doesn't introduce any difficulty.

In the unfolded plane we require the continuity of r along a vertical characteristic and the continuity of s along a horizontal characteristic. Thus, the remainder of the unfolded (r, s) plane is subdivided into squares as shown in Fig. 4. This introduces two additional regions: In the direction of "increasing x'" a $\bar{1}$ region lies between a $\bar{2}$ and a $\bar{3}$ region, whereas a $\bar{4}$ region lies between a $\bar{3}$ and a $\bar{2}$ region.

The solution at P_1 (in region $\bar{2}$) or P_1' (in region $\bar{3}$) is readily obtained from (4.27) and (4.28) by giving the appropriate initial conditions along $P_B P_{A1}$ or $P_{A2} P_{A3}$, respectively. The points P_2 (in region $\bar{4}$) or P_2' (in region $\bar{1}$) are outside the sequence of

[16] This is the direction of increasing t. This follows from (4.19), where one sees that t_ρ is of the same sign as $\rho_{x'}$, and, therefore, negative in the region $0 < x' < 1$. If $\Delta t = t_r\, dr$ is to be > 0, then dr must be < 0.

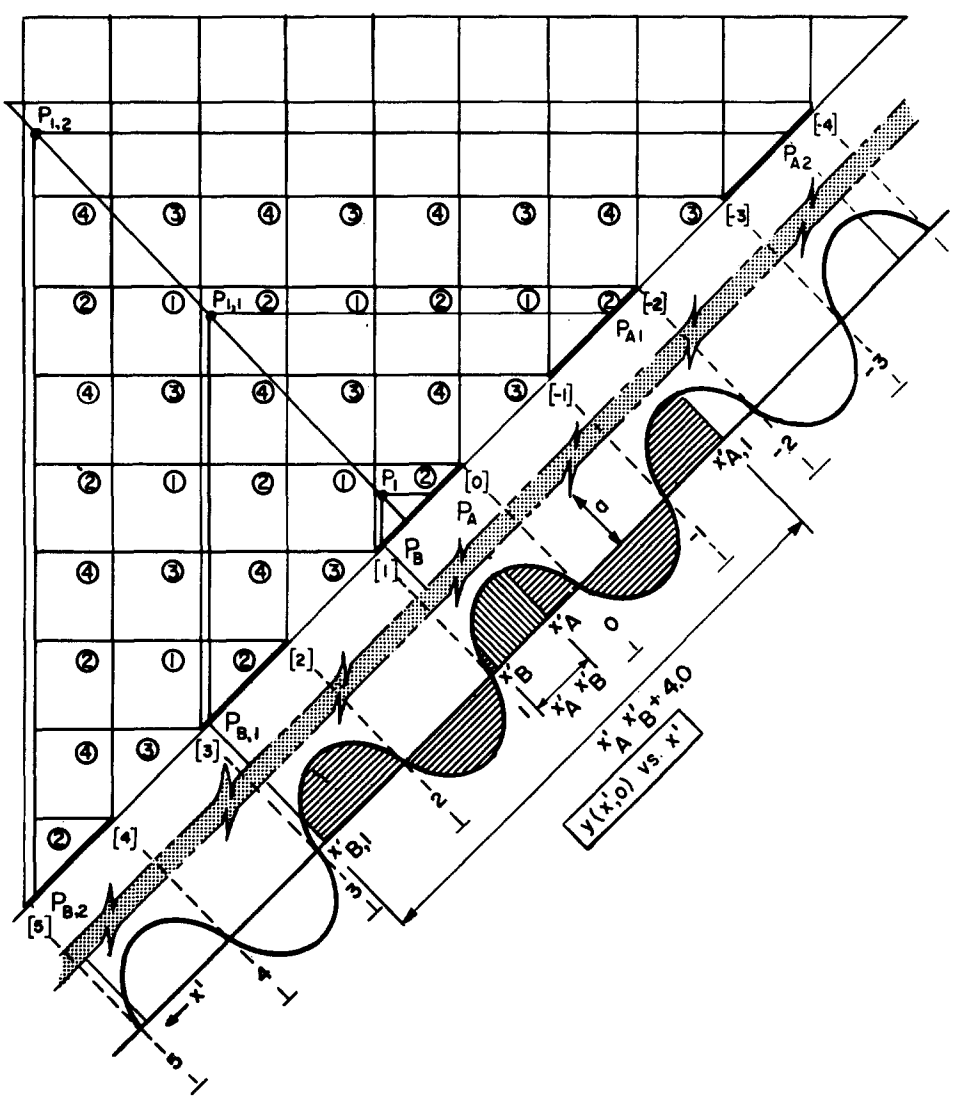


FIG. 5. Evolution of the solution in the unfolded (r, s) plane. (Circled numbers correspond to barred numbers in the text.)

squares along the main diagonal. The solution at P_2 can be obtained in two steps: (1) We construct the solution along $[0]b$ from the initial conditions along $[0]P_B$, and along $[0]a_2$ from the initial conditions along $[0]P_{A2}$; (2) we treat the problem in region $\bar{4}$ as a Cauchy problem of the second kind using the solutions along two different intersecting characteristics, namely, $[0]b$ and $[0]a_2$. It can be shown that the solution at P_2 obtained in this way is identical to that obtained by treating the problem as a Cauchy problem of the first kind over $P_{A2}P_B$.

B. The t Solution, in the Unfolded-Inverse Plane, Expressed as an Integral

Let us consider the evolution of the t solution along a line of slope $= -1$ in the unfolded (r, s) plane, as shown in Fig. 5. $t(P_1)$ in a $\bar{2}$ region is influenced by initial conditions along P_AP_B. At $P_{1,1}$, the corresponding point in the next $\bar{2}$ region, the solution $t(P_{1,1})$ is influenced by initial conditions along $P_{A,1}P_{B,1}$. Note that

$$x'_{A1}x'_{B1} = x'_Ax'_B + 4, \tag{5.2}$$

where the 4 in (5.2) corresponds to two full spatial periods. In general, the solution at $P_{1,m}$, a point in a $\bar{2}$ region having the same ξ and η as P_1, is given by

$$\begin{aligned} t(P_{1,m}) &= \tfrac{1}{2}\int_{x_{Am}}^{x_{Bm}} Gf^{-1/2}(x')\,dx' \\ &= \tfrac{1}{2}\left\{\int_{x_A}^{x_B}\cdots dx' + \int_{x_B}^{x_{Bm}}\cdots dx' + \int_{x_{Am}}^{x_A}\cdots dx'\right\}, \end{aligned} \tag{5.3}$$

or

$$t(P_{1,m}) = t(P_1) + mK(P_1), \tag{5.4}$$

where

$$K(P_1) = \int_{-2}^{+2} Gf^{-1/2}(x')\,dx'. \tag{5.5}$$

This follows because the integrand of (5.3) is periodic in x' with a period of 2.

$t(\xi, \eta)$ possesses certain symmetry properties in the (r, s) plane because the integrand in (5.3) is even in x' about the points $\ldots$ $[-1]$, $[0]$, $[1]$, $\ldots$ (see Fig. 4). Thus

$$t(P_1) = t(P'_1) \quad \text{or} \quad t_{\bar{2}}(\xi, \eta) = t_{\bar{3}}(\eta, \xi); \tag{5.6}$$

$$t(P_2) = t(P'_2) \quad \text{or} \quad t_{\bar{4}}(\xi, \eta) = t_{\bar{4}}(\eta, \xi); \tag{5.7}$$

$$t(P'_2) = t(P''_2) \quad \text{or} \quad t_{\bar{1}}(\xi, \eta) = t_{\bar{1}}(\eta, \xi). \tag{5.8}$$

The barred subscripts refer to the type of region.

C. The x Solution, in the Unfolded-Inverse Plane, Expressed as an Integral

Consider the evolution of the x solution along a line of slope $= -1$ in the unfolded plane. $x(P_1)$ is given directly by (4.28). The value of x at $P_{1,m}$ is given by

$$\begin{aligned} x(P_{1,m}) = {} & \tfrac{1}{2}C_1[(x_B + 2m)f^{-1/4}(x_B) \\ & + (x_A - 2m)f^{-1/4}(x_A)] \\ & - C_2\int_{(x_A-2m)}^{(x_B+2m)} x'\,\Pi(x')\,dx', \end{aligned} \tag{5.9}$$

where C_1, C_2, and $\Pi(x')$ are given in (4.29), (4.30), and (4.31). The integral in (5.9) is decomposed into three parts:

$$\int_{(x_A-2m)}^{(x_B+2m)} = \int_{(x_A-2m)}^{x_A} + \int_{x_A}^{x_B} + \int_{x_B}^{(x_B+2m)}.$$

This allows us to write (5.9) as

$$\begin{aligned} x(P_{1,m}) = {} & x(P_1) + mC_1[f^{-1/4}(x_B) - f^{-1/4}(x_A)] \\ & - C_2\left\{\int_{(x_A-2m)}^{x_A} x'\,\Pi(x')\,dx' \right. \\ & \left. + \int_{x_B}^{(x_B+2m)} x'\,\Pi(x')\,dx'\right\}. \end{aligned} \tag{5.10}$$

The x solution, (5.10), differs fundamentally from the t solution, not only in the presence of the non-integrated term, but also in the presence of the monotonic function in the integrand of the integrated term. If we introduce

$$\bar{\Pi}_A^{(i)} = \int_0^2 (x')^i\,\Pi(-x' + x_A)\,dx' \tag{5.11}$$

and

$$\bar{\Pi}_B^{(i)} = \int_0^2 (x')^i\,\Pi(+x' + x_B)\,dx', \tag{5.12}$$

then (5.10) takes the form

$$\begin{aligned} x(P_{1,m}) = {} & x(P_1) + mC_1[f^{-1/4}(x_A) - f^{-1/4}(x_B)] \\ & + mC_2\{\bar{\Pi}_A^{(1)} - \bar{\Pi}_B^{(1)}\}, \end{aligned} \tag{5.13}$$

since $\bar{\Pi}_A^{(0)} = \bar{\Pi}_B^{(0)} = 0$.

Similar considerations apply when P_1 lies in the lower part of a $\bar{4}$ or a $\bar{1}$ region.

D. Breakdown of the t Solution

It is a well-known fact that the solutions of the equations of a one-dimensional, polytropic, hydrodynamic fluid always exhibit a breakdown (the development of a discontinuity or singularity) if the initial conditions are arbitrary periodic func-

tions. We will show that the nonlinear string also develops a singularity in the *second* derivative after a "large" number of oscillations.[17]

Let us examine the relationship of t_r and r_x (or equivalently t_s and s_x).[18] Using Eq. (3.9) for t_r and the definition of j in (3.21), we obtain

$$t_r = \{2r_x[\beta(r+s)]^{1/3}\}^{-1}. \tag{5.14}$$

Similarly,

$$t_s = -\{2s_x[\beta(r+s)]^{1/3}\}^{-1}. \tag{5.15}$$

Thus, if t_r (or t_s) vanishes at some point in the (r, s) plane, this indicates that r_x (or s_x) has become infinite, since $(r+s)$ remains finite and nonzero in each square of the unfolded plane. However,

$$\begin{matrix} r_x \\ s_x \end{matrix} = \pm v_x + (1+\epsilon u)^{1/2}u_x$$

$$= \pm y_{tx} + (1+\epsilon y_x)^{1/2}y_{xx}. \tag{5.16}$$

Hence, a point along the string where r_x or s_x is infinite corresponds to an infinite second derivative, or a discontinuity in the first derivative. At the corresponding point in the unfolded (r, s) plane, j has become infinite and $J = 0$, indicating that the transformation has broken down.

Let us investigate t_ξ for a zero, using the integral representation given by (5.4).

$$2t_\xi = \int_{x_A}^{x_B} G_\xi f^{-1/2}(x')\,dx' + f^{-1/2}(x_B)G(x_B)[dx_B/d\xi] + m\int_{-2}^{2} G_\xi f^{-1/2}(x')\,dx', \tag{5.17}$$

where

$$\frac{dx_B}{d\xi} = \frac{1}{\xi_{x'}}\bigg|_{x'=x_B} = -\{a\pi^2[\sin \pi x_B f^{1/2}(x_B)]\}^{-1}, \tag{5.18}$$

where (5.18) is evaluated on the initial condition line $r = s$. The Riemann function and its derivatives on this line are

$$G = f^{1/4}(x')\{[(3\epsilon/4)(\xi+\eta)]^{-1/6}P_{-5/6}(q)\} = \mathcal{O}(\epsilon^0),$$
$$G_\xi = -(\epsilon/8)f^{1/4}(x')[(3\epsilon/4)(\xi+\eta)]^{-1/6} \times \{[(3\epsilon/4)(\xi+\eta)]^{-1}P_{-5/6} - (8/\epsilon)\,\partial P_{-5/6}/\partial\xi\} = \mathcal{O}(\epsilon^1). \tag{5.19}$$

Substituting (5.18) into (5.17) yields

$$2t_\xi = \epsilon\int_{x_A}^{x_B}(G_\xi/\epsilon)f^{-1/2}(x')\,dx' - G[a\pi^2(\sin \pi x_B)f(x_B)]^{-1} + \epsilon m\int_{-2}^{2}(G_\xi/\epsilon)f^{-1/2}(x')\,dx'. \tag{5.20}$$

If x_A and x_B are kept fixed and time (or m) allowed to increase, the last term, which is of $\mathcal{O}(\epsilon m)$, increases until it is of the *same order* as the second term and thereby cancels it, yielding a zero for t_ξ. For small ϵ, $G_\xi = -|\mathcal{O}(\epsilon)|$, and the zero of t_ξ will first occur for a value of x_B which makes $\sin \pi x_B < 0$, that is, $x_B = -\frac{1}{2}$.

We now estimate the *minimum* time required for the string to reach breakdown. We assume ϵ to be small and expand both G and G_ξ to terms of $\mathcal{O}(\epsilon)$. If (ϵm) is assumed to be $\mathcal{O}(\epsilon^0)$ and we neglect all terms of $\mathcal{O}(\epsilon)$ in (5.20), we obtain the relation

$$t_\xi = 0 = +1/(a\pi^2) - (m\epsilon/8)4 + \mathcal{O}(\epsilon),$$

or

$$m = 2/\epsilon a\pi^2. \tag{5.21}$$

We recall that m is the number of equivalent fundamental linear periods.[17] If one performs the same calculation for $t_\eta = 0$, one finds $x_A = +\frac{1}{2}$, and the same time-to-breakdown as given in (5.21).

In the FPU calculations illustrated in Fig. 1, the parameters were: $N = 32$, $\alpha = \frac{1}{4}$, and $a = 1.0$. Thus, $\epsilon = 1/64$ (Eq. 2.5) and $m = 12.95$. In Fig. 1, this corresponds, approximately, to the time when the energy in the second mode reaches its first maximum.

If the string is initially displaced in an nth mode,

$$y(x', 0) = a_n \sin \pi n x',$$

such that the energy is the same as in a first-mode displacement, then the *time*-to-breakdown is reduced by a factor of n. This results from two considerations. First, the energy invariance requires that $a_n = a_1/n$, where a_1 is the first-mode amplitude. Second, the smallness parameter in the continuum limit becomes $\epsilon_n = n\epsilon_1 = 2\alpha n/N$. This follows because an nth mode initial displacement over a discrete string of N particles looks like a first mode over a string of N/n particles. Thus, the number of *oscillations*-to-breakdown is an invariant quantity, since it is proportional to $1/(\epsilon_n a_n) = 1/(\epsilon_1 a_1)$. Because an nth mode makes one oscillation in $(1/n)$ the time required for a fundamental oscillation, then the *time*-to-breakdown is $(1/n)$ times as much.

[17] The oscillations referred to here and in the following are those of the equivalent linear string, namely, the one obtained by setting $\epsilon = 0$. For the normalization used, the period of one linear oscillation is 2.

[18] The methods used in Sec. 5D for studying the formation of discontinuities have been used in investigations in hydrodynamics. For example, see G. S. S. Ludford, Proc. Cambridge Phil. Soc. **48**, 499 (1952).

6. ANALYTIC REINVERSION OF THE SOLUTION

One can study the properties of the solution most conveniently when it is written in direct form, that is, in the physical plane where x and t are the independent variables. The procedure for accomplishing the reinversion is given below and is analogous to that used for reinverting the linear solution, as described at the end of Sec. 4. For the purposes of the example, we include only terms to $\mathcal{O}(\epsilon)$. A higher order analysis proceeds along similar lines.

As shown in Fig. 5, we examine the solution along a line of slope $= -1$ and in a $\bar{2}$ region of the (r, s) plane. The t solution, (4.27), is written as

$$t(\xi, \eta) = (2C_1)^{-1}\{(x_B - x_A) - (\epsilon a/4)[\sin \pi x_B - \sin \pi x_A] + \mathcal{O}(\epsilon^2)\} \quad (6.1)$$

because the Legendre function in the integrand of (4.27) contributes only its leading term $(= 1)$. Similarly, the x solution is written as

$$x(\xi, \eta) = (C_1/2)\{x_B[1 - (\epsilon a\pi/4) \cos \pi x_B] + x_A[1 - (\epsilon a\pi/4) \cos \pi x_A] + \mathcal{O}(\epsilon^2)\}. \quad (6.2)$$

C_1, (4.29), can be expanded in a power series in ϵ using the definitions of ξ (or r) and η (or s) as given in (4.15).

$$C_1 = 1 + (\epsilon a\pi/8)[\cos \pi x_A + \cos \pi x_B] + \mathcal{O}(\epsilon^2). \quad (6.3)$$

If (6.3) is substituted into (6.1) and (6.2) and the results rearranged, we obtain

$$t = \tfrac{1}{2}(x_B - x_A) - (\epsilon a/8)\ \Delta_s - \tfrac{1}{2}\epsilon\ \Delta_x \Sigma_c + \mathcal{O}(\epsilon^2) \quad (6.4)$$

$$x = \tfrac{1}{2}(x_B + x_A) - \tfrac{1}{2}\epsilon\ \Delta_x \Delta_c + \mathcal{O}(\epsilon^2), \quad (6.5)$$

where

$$\Delta_s = \sin \pi x_B - \sin \pi x_A; \quad \Delta_c = \cos \pi x_B - \cos \pi x_A; \quad (6.6)$$

$$\Sigma_s = \sin \pi x_B + \sin \pi x_A; \quad \Sigma_c = \cos \pi x_B + \cos \pi x_A;$$

$$\Delta_x = (a\pi/8)(x_B - x_A). \quad (6.7)$$

We now invert by adding and subtracting (6.4) and (6.5). Rearranging, we obtain the *implicit* form

$$x_A = (x - t) - \epsilon\ \Delta_x \cos \pi x_A - (\epsilon a/8)\ \Delta_s + \mathcal{O}(\epsilon^2) \quad (6.8)$$

$$x_B = (x + t) + \epsilon\ \Delta_x \cos \pi x_B + (\epsilon a/8)\ \Delta_s + \mathcal{O}(\epsilon^2). \quad (6.9)$$

One recognizes the leading terms of (6.8) and (6.9), $x_A = x - t$ and $x_B = x + t$, as the linear solutions, (4.32). To complete the inversion, one would have to express x_A and x_B explicitly as functions of x and t, rather than in the implicit form given in (6.8) and (6.9).

The direct solution is now recovered by forming $v = y_t$ from the difference of ξ and η as defined in (4.15).

$$v = \tfrac{1}{2}(\xi - \eta) = (3\epsilon)^{-1}[f^{3/2}(x_B) - f^{3/2}(x_A)] \quad (6.10)$$

or

$$v = \tfrac{1}{2}a\pi\ \{\Delta_c + (\epsilon a\pi/8)[\cos 2\pi x_B - \cos 2\pi x_A] + \mathcal{O}(\epsilon^2)\}, \quad (6.11)$$

or

$$v = -a\pi\{\sin (\pi/2)(x_A + x_B) \times \sin (\pi/2)(x_B - x_A) + (\epsilon a\pi/8) \times \sin \pi(x_A + x_B) \cdot \sin \pi(x_B - x_A)\} + \mathcal{O}(\epsilon^2). \quad (6.12)$$

Equation (6.12) is put into a more recognizable form by using the sums and differences of x_A and x_B in (6.4) and (6.5). We introduce the parameter τ which gives the temporal variation on a "slow" time scale:

$$\tau = (\epsilon a\pi/8)t. \quad (6.13)$$

This is obtained by substituting for $(1/2)(x_B - x_A)$ in Δ_x the leading term of (6.4), namely, t. Equation (6.12) then takes the form

$$v = -a\pi\{\sin \pi[x + \epsilon\Delta_c] \times \sin \pi[t + \tau\Sigma_c + (\epsilon a/8)\ \Delta_s] + (\epsilon a\pi/8) \sin 2\pi[x + \tau\Delta_c] \times \sin 2\pi[t + \tau\Sigma_c + (\epsilon a/8)\ \Delta_s]\} + \mathcal{O}(\epsilon^2). \quad (6.14)$$

The solution for u is recovered in similar fashion, namely, from

$$\tfrac{1}{2}(r + s) = \tfrac{1}{2}(\xi + \eta) \quad (6.15)$$

or

$$(2/3\epsilon)(1 + \epsilon u)^{3/2} = (1/3\epsilon)[f^{3/2}(x_A) + f^{3/2}(x_B)]. \quad (6.16)$$

Solving to first order in ϵ yields

$$u = y_x = \tfrac{1}{2}a\pi\Sigma_c + \epsilon(\tfrac{1}{4}a\pi\Delta_c)^2 + \mathcal{O}(\epsilon^2). \quad (6.17)$$

This can be expressed directly in terms of x, t, and τ by using the equations above.

The solutions for u and v given above are qualitatively of the form: a periodic function of a periodic function. Thus, a spatial Fourier decomposition of these functions will yield modal amplitudes which involve the Bessel functions of the argument τ.

7. CONCLUSIONS

We have shown that the standing vibrations of a continuous nonlinear string develop a discontinuity in the first derivative after elapsed times of $\mathcal{O}(1/\epsilon a)$. The computations of Fermi, Pasta, and Ulam indicate that the vibrations of a finite number N of coupled, nonlinear, equimass particles do not develop such a discontinuity. Thus, a continuous nonlinear system described by a partial differential equation of second order cannot describe the vibrations of the equivalent discrete system for "large" times.

To account for the FPU results by a continuum representation, one is led to include terms which measure the discreteness or "graininess" of the medium. These terms appear quite naturally if we retain quantities of $\mathcal{O}(1/N^2)$ that arise in the limiting process described in Sec. 2. These terms involve higher derivatives (for example, y_{tttt}, y_{xxxx}, etc.) and should affect the vibrations most at those points on the string where breakdown "tends" to occur. These terms are analogous to the viscosity-like terms that are added to the lowest order hydrodynamic equations to prevent a discontinuity from forming.

ACKNOWLEDGMENTS

The author wishes to thank Professor M. D. Kruskal for suggesting the problem of the continuous nonlinear string and for his constant encouragement, advice, and enthusiasm during the course of the investigation. Conversations with other members of the Plasma Physics Laboratory are gratefully acknowledged. The research was done under the auspices of the U. S. Atomic Energy Commission. The author gratefully acknowledges the advice and assistance rendered by members of the Bell Telephone Laboratories, where the final aspects of the research were completed and where the final version of the manuscript was prepared.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Proof of a Conjecture by Dyson

KENNETH G. WILSON*
Lyman Laboratory of Physics,
Harvard University, Cambridge, Massachusetts
(Received April 27, 1962)

We prove a mathematical conjecture by Dyson which he used in a study of the statistical distribution of energy levels in complex nuclei.

THE purpose of this paper is to prove a theorem that was proposed by Dyson in a recent paper.[1,2] This theorem (conjecture C of reference 1) may be stated as follows:

Let $z_1, \cdots, z_N$ be a set of N complex variables and $a_1, \cdots, a_N$ a set of N positive integers (≥ 0). Let y_j for $1 \leq j \leq N$ be

$$y_j = \prod_{k \neq j} \left(1 - \frac{z_j}{z_k}\right). \tag{1}$$

Let

$$P(z_1, \cdots, z_N) = \prod_j (y_j)^{a_j}, \tag{2}$$

which we shall write $P(z)$ for convenience. P can be expanded in positive and negative powers of the z_j; we are interested in the constant term F given by

$$F = (2\pi i)^{-N} \int z_1^{-1}\, dz_1 \cdots \int z_N^{-1}\, dz_N P(z), \tag{3}$$

the contours being the unit circle taken counterclockwise.

Theorem:

$$F = (\sum_j a_j)! \, \{\prod_j a_j!\}^{-1}. \tag{4}$$

The proof depends on three lemmas, which will be stated now and proved later.

Define

$$u_j = (y_j)^{-1}. \tag{5}$$

We compute F by making a change of variable from $z_1, \cdots, z_{N-1}$ to $u_2, \cdots, u_N$. Lemma 1 will state that this is possible. We note that because P is homogeneous in the z's, the integration over z_N becomes trivial after the other integrations have been performed, and that because the u's are homogeneous in the z's, only $N - 1$ are independent, which is why we make $N - 1$ instead of N changes of variable.

Lemma 1:

$$F = (2\pi i)^{-N+1} \int u_2^{-1}\, du_2 \cdots \int u_N^{-1}\, du_N \times [J(z)]^{-1} P(z) \tag{6}$$

where J is the Jacobian

$$J(z) = \frac{\partial(\ln u_2, \cdots, \ln u_N)}{\partial(\ln z_1, \cdots, \ln z_{N-1})} \tag{7}$$

and the z's are expressed in terms of the u's by Eq. (5). The paths of integration are the circles

$$|u_i| = R_i \tag{8}$$

taken $i - 1$ *times* counterclockwise, where the R_i are arbitrary except that they satisfy

$$R_{i+1} \ll R_i \ll 1 : 2 \leq i \leq N - 1. \tag{9}$$

Lemmas 2 and 3 will show that J and P are single-valued functions of the u's so we do not have to specify the branch of the solution of Eq. (5).

Lemma 2:

$$\sum_{j=1}^{N} u_j = 1. \tag{10}$$

This is an identity in the z's, when the u_j are regarded as functions of the z's through Eq. (5).

Lemma 3:

$$J(z) = (N - 1)!\, u_1. \tag{11}$$

See the note for Lemma 2.

As a result of these lemmas,

$$F = \frac{(2\pi i)^{-N+1}}{(N-1)!} \int du_2 \cdots \int du_N \prod_{j=1}^{N} (u_j)^{-a_j - 1}, \tag{12}$$

with u_1 given by Lemma 2. These integrals are elementary, but to save space we use a shortcut.

* Junior Fellow, Society of Fellows, Harvard University, now at CERN, Geneva, Switzerland.

[1] F. J. Dyson, J. Math. Phys., **3**, 140, 157, 166 (1962), this theorem has also been proved independently by J. Gunson, J. Math. Phys. (to be published).

[2] Conjectures A, B, and D of reference 1 are shown there to reduce to conjecture C; thus they will not be discussed here.

First we note that

$$(2\pi i)^{-1}\int u^{-n-1}e^{u}\,du = 1/n! \tag{13}$$

if the contour encloses the origin counterclockwise. Secondly, if $F(\lambda)$ is defined by Eq. (12), but with

$$u_1 = \lambda - \sum_{j=2}^{N} u_j, \tag{14}$$

where λ is on the unit circle, then by making a change of variable to $u_i' = u_i\lambda$ we see that

$$F(\lambda) = \lambda^{-a-1}F, \tag{15}$$

where

$$a = \sum_{j=1}^{N} a_j. \tag{16}$$

Hence, by Eq. (13),

$$F = (2\pi i)^{-1}a!\int e^{\lambda}F(\lambda)\,d\lambda. \tag{17}$$

Interchanging the order of integration so that we integrate over λ with $u_2, \cdots, u_N$ held fixed, we must evaluate the integral

$$\frac{1}{2\pi i}\int e^{\lambda}\left\{\lambda - \sum_{j=2}^{N} u_j\right\}^{-a_1-1} d\lambda = \frac{1}{a_1!}e^{u_2+\cdots+u_N} \tag{18}$$

[by Eqs. (8) and (9), the pole at $\lambda = u_2 + \cdots + u_N$ lies inside the unit circle].
Thus,

$$F = \frac{a!}{(2\pi i)^{N-1}(N-1)!\,a_1!}\int du_2 \cdots \int du_N \times \prod_{j=2}^{N} e^{u_j}u_j^{-a_j-1} = \frac{a!}{a_1!\cdots a_N!} \tag{19}$$

[the factor $(N-1)!$ is exactly compensated by the requirement of Lemma 1 that u_j execute a circle $j-1$ times].

Now we prove Lemmas 1–3. We start by stating another lemma. Lemmas 2–4 have probably all occurred in other work[3] but it is easier to prove them than to locate them in the literature.

Lemma 4:

Let $G(x_1, \cdots, x_M)$ be a function of M variables such that

1. G is a symmetric function of $x_1, \cdots, x_M$,
2. G is a ratio of two polynomials in the x's,
3. G is homogeneous of degree 0 in the x's,
4. The denominator of G is $\prod_{i<k}(x_i - x_k)$.

Then G is a constant. This is because, since the denominator changes sign when we interchange the values of any pair x_i, x_k, the numerator must also change sign under this exchange; thus, the numerator vanishes when $x_i = x_k$. Hence the numerator has $x_i - x_k$ as a factor (for any j and k), e.g., it has the entire denominator as a factor; thus G is a polynomial. Since it is of degree 0, it must be a constant.

Proof of Lemma 2:

Since $\sum_i u_i$ considered as a function of the z's satisfies the conditions of Lemma 4, it is constant. Putting $z_1 = 0$ we obtain $u_1 = 1$, $u_i = 0$ $(j > 1)$ so the constant is 1.

Proof of Lemma 3:

The Jacobian J is the determinant of the matrix

$$J_{ij} = \partial \ln u_i/\partial \ln z_j \tag{20}$$

(rows numbered $i = 2$ to N, columns $j = 1$ to $N-1$). Without changing the value of the determinant we may add columns $j = 2$ through $N-1$ to column 1; since $\ln u_i$ is homogeneous in the z's we now have

$$J_{i1} = -\partial \ln u_i/\partial \ln z_N. \tag{21}$$

Move this column to the right, calling it J_{iN}; thus,

$$J = (-1)^{N-2}\det|J_{ij}|,$$

$$\text{where}\quad 2 \le i \le N, \qquad 2 \le j \le N. \tag{22}$$

Now,

$$J_{ij} = -z_i(z_j - z_i)^{-1} \qquad (i \ne j), \tag{23}$$

$$J_{ii} = \sum_{k\ne i}\frac{z_i}{z_k - z_i} \tag{24}$$

Evidently J is the ratio of two polynomials in the z's the denominator being a product of factors $z_j - z_i$. No such factor occurs twice; for a denominator $z_j - z_i$ occurs only in the elements J_{ii}, J_{ij}, J_{ji}, and J_{jj} so that a term $(z_j - z_i)^2$ occurs in the denominator of J only if it occurs in the 2×2 determinant $J_{ii}J_{jj} - J_{ij}J_{ji}$. However the term in $J_{ii}J_{jj}$ containing the factor $(z_i - z_j)^{-2}$ cancels the corresponding term in $J_{ij}J_{ji}$. Furthermore, J has a factor $(z_2z_3\cdots z_N)$, it is symmetric in z_2 through z_N (but not in z_1), and it is homogeneous of degree 0 in the z's. Using the argument of Lemma 4, we must have

$$J = C\prod_{k=2}^{N} z_k(z_k - z_1)^{-1} = Cu_1, \tag{25}$$

where C is a constant. Since Eq. (4) is known to

[3] Lemma 2 appears in the theory of Lagrangian interpolation; see F. Hildebrand, *Introduction to Numerical Analysis* (McGraw-Hill Book Company, Inc., New York, 1956), p. 61, Eq. (3.2.5) with $x = 0$.

be true if all the a_i are 0 we shall have to have $C = (N-1)!$.

Proof of Lemma 1:

We shall prove Lemma 1 by introducing the new variables one at a time. For convenience we shall first change variables to t_j $(2 \leq j \leq N)$, where

$$(t_j)^{-j+1} = y_j. \tag{26}$$

Before making this change we change the paths of integration in Eq. (3) to be the circles

$$|z_j| = r_j \qquad (1 \leq j \leq N), \tag{27}$$

where

$$r_j \ll r_{j+1} \qquad (1 \leq j \leq N-1). \tag{28}$$

We now use mathematical induction. Suppose the following propositions are true for $m < n$:

1. Equation (26) can be solved for $2 \leq j \leq m+1$ to give z_1 through z_m as functions of $t_2, \cdots, t_{m+1}, z_{m+1}, \cdots, z_N$. We define functions ω_j for $1 \leq j \leq m$ by

$$z_j = z_{j+1} t_{j+1} (t_j)^{-(j-1)/j} e^{(i\pi/j)} \times \omega_j(t_2, \cdots, t_{j+1}, z_{j+1}, \cdots, z_N), \tag{29}$$

where we define $t_1 = 1$ and $0 \leq \arg t_j \leq 2\pi$. The solution can be chosen so that $\omega_j \approx 1$ in the region of interest. More specifically, ω_j is analytic as a function of t_{j+1} and z_{j+1} to z_N, and satisfies

$$|\omega_j - 1| < \tfrac{1}{2}\epsilon \tag{30}$$

when

$$|t_k| = R_k \qquad (2 \leq k \leq j), \tag{31}$$

$$\tfrac{1}{2}R_{j+1} < |t_{j+1}| < 2R_{j+1}, \tag{32}$$

$$\tfrac{1}{4}(1+\epsilon)^k r_k < |z_k| < 4(1-\epsilon)^k r_k \qquad (j+1 \leq k \leq N), \tag{33}$$

where

$$R_j = (r_1 r_2 \cdots r_{j-1})^{1/(j-1)} r_j^{-1} \tag{34}$$

and ϵ is a fixed number $\ll 1$.[4]

2.

$$F = \frac{1}{(2\pi i)^N} \int dt_2 \cdots \int dt_{m+1} \int \frac{dz_N}{z_N} \cdots \int \frac{dz_{m+1}}{z_{m+1}} \times \left\{ \frac{dt_2}{dz_1} \frac{dt_3}{dz_2} \cdots \frac{dt_{m+1}}{dz_m} \right\}^{-1} \{z_1 z_2 \cdots z_m\}^{-1} P(z), \tag{35}$$

where the integrations are carried out from right to left (e.g., z_{m+1} first, holding the other z's and t's fixed). The contours are the circles $|z_k| = r_k$, $|t_k| = R_k$. The symbol dt_{k+1}/dz_k stands for the partial derivative $\partial t_{k+1}/\partial z_k$ when $t_2, \cdots, t_k, z_{k+1}, \cdots z_N$ are held fixed.

We now prove these propositions for $m = n$. First, we must examine the dependence of t_{n+1} on z_n, when $t_2, \cdots, t_n, z_{n+1}, \cdots z_N$ are held fixed; z_1 to z_{n-1} are functions of these variables and z_n [through Eq. (26)], and satisfy the restrictions of proposition 1. We obtain

$$t_{n+1}(t_n)^{-(n-1)/n} = (y_n/y_{n+1})^{1/n} = (-1)^{1/n} z_n (z_{n+1})^{-1} g(z), \tag{36}$$

where

$$g(z) = \prod_{j=1}^{n-1} \left(1 - \frac{z_j}{z_n}\right)^{1/n} \left(1 - \frac{z_j}{z_{n+1}}\right)^{-1/n} \times \prod_{j=n+2}^{N} \left(1 - \frac{z_n}{z_j}\right)^{1/n} \left(1 - \frac{z_{n+1}}{z_j}\right)^{-1/n}. \tag{37}$$

If $t_2, \cdots, t_n$ and $z_n, \cdots, z_N$ satisfy the inequalities (31) and (33), one finds using Eqs. (29), (30), and (34) that Eq. (33) is satisfied also by $z_1, \cdots, z_{n-1}$. Thus (if the ratios r_j/r_{j+1} are sufficiently small) we may define the nth roots in $g(z)$ by requiring

$$|g(z) - 1| < \tfrac{1}{4}\epsilon; \tag{38}$$

the other nth roots are defined as in Eq. (29).

Now consider the equation $x = t_{n+1}(z_n)$ for values of x such that $\frac{1}{2}R_{n+1} < |x| < 2R_{n+1}$. This does not differ very much from the equation

$$x - (t_n)^{(n-1)/n}(-1)^{1/n} z_n (z_{n+1})^{-1} = 0, \tag{39}$$

which has a unique root z_n for a given x.

Now consider the functions

$$f(z_n) = x - t_{n+1}(z_n),$$
$$g(z_n) = x - (t_n)^{(n-1)/n}(-1)^{1/n} z_n (z_{n+1})^{-1}, \tag{40}$$

where $\frac{1}{2}R_{n+1} < |x| < 2R_{n+1}$. On the circles $|z_n| = r_n/3$ and $|z_n| = 3r_n$, f and g are almost equal so that

$$|g(z_n)[f(z_n)]^{-1} - 1| < 1. \tag{41}$$

Since $g(z_n)$ has a single root between the two circles, $f(z_n)$ does also, by Rouché's theorem.[5] From the explicit formula for the inverse function[5]

$$z_n(t_{n+1}) = \frac{1}{2\pi i} \int [t_{n+1}(z_n) - t_{n+1}]^{-1} \frac{dt_{n+1}}{dz_n} z_n \, dz_n, \tag{42}$$

where the contour is taken on the two circles, we

[4] To be precise, we should choose a sufficiently small value for ϵ, then choose the r_j with r_{j+1}/r_j sufficiently small.

[5] L. Ahlfors, *Complex Analysis* (McGraw-Hill Book Company, Inc., New York, 1953), p. 124.

see that the inverse function is analytic in t_{n+1} for $\frac{1}{2}R_{n+1} < |t_{n+1}| < 2R_n$. Since $t_{n+1}(z_n)$ is analytic in $z_{n+1}, \cdots, z_N$, the inverse function is also. Since $\omega_n = 1/g(z)$, it satisfies the inequality (30).

This proves proposition 1 for $m = n$. We now change variables in Eq. (35) (with $m = n - 1$) from z_n to t_{n+1}; the path of t_{n+1} is almost the circle $|t_{n+1}| = R_{n+1}$; we change it to be exactly this circle and then interchange the order of integration with the remaining z's. This proves proposition 2 for $m = n$. To complete the proof of propositions 1 and 2 they must be proved for $m = 1$; with some changes the above procedure can be used.

Now consider Eq. (35) with $m = N - 1$. The expression $I = dt_2/dz_1 \cdots, dt_N/dz_{N-1}$ is the Jacobian[6]

$$\frac{\partial(t_2, \cdots, t_N)}{\partial(z_1, \cdots, z_{N-1})};$$

by Lemma 3,

$$I = (t_2 \cdots t_N)(z_1 \cdots z_{N-1})^{-1}u_1^{-1}. \quad (43)$$

Since u_1 is a function only of the t's, by Lemma 2, the integral over z_N is trivial. Changing variables from t_n to $u_n = t_n^{n-1}$ we obtain Lemma 1 [Eq. (6)].

I am indebted to Dr. Paul Federbush for suggesting this problem.

[6] R. Courant, *Differential and Integral Calculus*, (Interscience Publishers, Inc., New York, 1936), Vol. II, pp. 247–256.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Analytic Functions of Continuant Matrices*

PHILLIP B. ABRAHAM AND GEORGE H. WEISS
Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, College Park, Maryland
(Received March 14, 1962)

It is shown that the Fourier method which was developed by us for the calculation of analytic functions of circulant matrices can also be applied to calculate analytic functions of continuant matrices. Similar calculations are developed for generalized continuant matrices which arise in connection with problems in greater than one dimension.

I. INTRODUCTION

In a recent paper,[1] we have discussed the problem of calculating analytic functions of circulant matrices and generalized circulant matrices, i.e., partitioned matrices whose blocks are successive cyclic permutations of those submatrices appearing in the first row. In this paper we shall use a similar method to that developed in our first paper to evaluate analytic functions of some types of continuant matrices. Cyclic matrices arise in many physical problems in statistical mechanics when cyclic boundary conditions are employed, that is, when the system is considered to be wrapped on a torus of the proper dimensionality. Continuant matrices arise from the same types of physical problems when different boundary conditions are employed. As an example, circulant matrices are used in lattice dynamical problems when the ends of the lattice are connected together. Continuant matrices are used for the same problem when the edge atoms of the lattice are held fixed. It is useful to be able to calculate analytic functions of both types of matrices since thermodynamic functions of the physical system are defined in terms of analytic functions. Analytic functions of circulant and continuant matrices are also useful in the perturbation theory of chemical systems.[2] The results for continuant matrices are somewhat less general than those for circulant matrices since we are only able to treat matrices with diagonal and first off-diagonal nonzero elements.

In addition to doing the calculation for ordinary continuant matrices, we shall introduce the notion of generalized continuant matrices which are used to solve physical problems in dimensions higher than one. The generalization to the calculation of analytic functions of these matrices is straightforward involving no more than the use of multiple Fourier series rather than Fourier series of a single variable.

To begin with, we shall recapitulate the principal ideas contained in I. Let $\mathbf{\Delta}$ be a circulant matrix of order $N+1$ of the form $\mathbf{\Delta} = (S_0, S_1, \cdots S_N)_{cyc}$ with eigenvalues λ_k given by

$$\lambda_k = \sum_{j=0}^{N} S_j \exp\left(\frac{2\pi ijk}{N+1}\right).$$

The unitary diagonalizing matrix $\mathbf{S}$ such that $\mathbf{S}^{-1}\,\mathbf{\Delta}\mathbf{S} = \mathbf{\Lambda}$ where $\mathbf{\Lambda}$ is diagonal has elements

$$(\mathbf{S})_{mk} = [1/(N+1)]^{1/2} \exp[2\pi imk/(N+1)], \quad (1)$$

where m and k run from 1 to N. If $F(x)$ is an analytic function with a convergent power series in some region about the origin, then we can write

$$F(\mathbf{\Delta}) = \sum_{n=0}^{\infty} a_n \mathbf{\Delta}^n \quad (2)$$

provided that the eigenvalues all lie within the circle of convergence. Hence we have

$$F(\mathbf{\Delta}) = \mathbf{S}F(\mathbf{\Lambda})\mathbf{S}^{-1} \quad (3)$$

or

$$\begin{aligned}[F(\mathbf{\Delta})]_{mn} &= \sum_{j,r=0}^{N} (\mathbf{S})_{mj}\{F(\mathbf{\Lambda})\}_{jr}(\mathbf{S}^{-1})_{rn} \\ &= \sum_{j=0}^{N} (\mathbf{S})_{mj}\{F(\mathbf{\Lambda})\}_{jj}(\mathbf{S}^{-1})_{jn} \\ &= \sum_{j=0}^{N} F(\lambda_j)(\mathbf{S})_{mj}(\mathbf{S}^{-1})_{jn} \\ &= \frac{1}{N+1}\sum_{j=0}^{N} F(\lambda_j)\exp\left\{\frac{2\pi i(m-n)j}{N+1}\right\}. \end{aligned} \quad (4)$$

* This research was supported by the U. S. Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 18(600)1315.

[1] P. Abraham and G. H. Weiss, J. Math. Phys. 3, 340 (1962), hereafter referred to as I.

[2] P. O. Löwdin, R. Pauncz, and J. de Heer, J. Math. Phys. 1, 461 (1960).

The next step in the calculation of $[F(\mathbf{\Delta})]_{mn}$ is to expand $F(\lambda_j)$ in a trigonometric series

$$F(\lambda_j) = \sum_{k=0}^{N} A_k^* \exp [2\pi ijk/(N+1)] \tag{5}$$

where the A_k^* are found in terms of the Fourier coefficients of the series for $F\{\sum_{i=0}^{N} S_i \exp (ij\theta)\}$. When Eq. (5) is substituted into Eq. (4) and the order of summations interchanged, it is found that the summation over j can be carried out explicitly, and the resulting matrix element is just one of the A_k.

It can be seen that the reason for the success of the method is the fact that the unitary matrix $\mathbf{S}$ has elements proportional to $\exp \{2\pi ijn/(N+1)\}$. Hence the same method will work, with minor modifications, for any matrix whose eigenvectors have components

$$(\mathbf{u}_j)_k = \exp (ijk\theta) \tag{6}$$

and by extension, any matrix whose eigenvectors have components

$$(\mathbf{u}_j)_k = \begin{matrix} \sin jk\theta \\ \cos jk\theta. \end{matrix} \tag{7}$$

II. FUNCTIONS OF SIMPLE CONTINUANT MATRICES

A particularly important class of matrices for which this holds is the symmetric Jacobi matrix

$$\mathbf{\Delta} = \begin{pmatrix} a & b & 0 & 0 & \cdots & 0 \\ b & a & b & 0 & \cdots & 0 \\ 0 & b & a & b & \cdots & 0 \\ \vdots & & & & & \\ 0 & \cdots & & b & a & \end{pmatrix}. \tag{8}$$

Some properties of this matrix have been elucidated by Rutherford,[3,4] but he did not discuss the calculation of analytic functions of $\mathbf{\Delta}$. We can actually carry out the calculations for a more general matrix of the form

$$\mathbf{\Gamma} = \begin{pmatrix} a & b_1 & 0 & \cdots & & 0 \\ c_1 & a & b_2 & \cdots & & 0 \\ 0 & c_2 & a & \cdots & & 0 \\ \vdots & & & \cdots & a & b_{N-1} \\ & \cdots & & & c_{N-1} & a \end{pmatrix} \tag{9}$$

where the condition

$$b_i c_i = \text{const} = b^2 \tag{10}$$

holds for all i, since any matrix having the same form as $\mathbf{\Gamma}$ is similar to a symmetric Jacobi matrix of the form of $\mathbf{\Delta}$. Consider the diagonal matrix $\mathbf{T}$ with elements

$$(\mathbf{T})_{ij} = (c_1 c_2 \cdots c_{j-1} b_j b_{j+1} \cdots b_{N-1})^{1/2}. \tag{11}$$

Then an easy calculation shows that

$$\mathbf{T}^{-1}\mathbf{\Gamma}\mathbf{T} = \begin{pmatrix} a & b & 0 & \cdots & 0 \\ b & a & b & \cdots & 0 \\ 0 & b & a & \cdots & 0 \\ \vdots & & & & \vdots \\ \cdots & & & b & a \end{pmatrix}. \tag{12}$$

Henceforth we will work with $\mathbf{\Delta}$ exclusively because of the property exhibited in Eq. (12).

The eigenvalues of $\mathbf{\Delta}$ are easily found to be

$$\lambda_j = a + 2b \cos [\pi j/(N+1)] \tag{13}$$

and the orthogonal matrix which brings $\mathbf{\Delta}$ to diagonal form has the elements

$$(\mathbf{S})_{mk} = \left(\frac{2}{N+1}\right)^{1/2} \sin \left(\frac{\pi mk}{N+1}\right), \qquad m, k = 1, 2, \cdots N. \tag{14}$$

In analogous fashion to the procedure of I, we define a continuous, periodic function

$$\lambda(\theta) = a + 2b \cos \theta. \tag{15}$$

We will now relate the elements of $F(\mathbf{\Delta})$ to the Fourier coefficients of $F\{\lambda(\theta)\}$. Inserting the matrix elements $(\mathbf{S})_{mk}$ into the relation

$$[F(\mathbf{\Delta})]_{mn} = \sum_j F(\lambda_j)(\mathbf{S})_{mj}(\mathbf{S}^{-1})_{jn}$$

we find

$$\begin{aligned} [F(\mathbf{\Delta})]_{mn} &= \frac{2}{N+1} \sum_{j=1}^{N} F(\lambda_j) \\ &\quad \times \sin \left(\frac{\pi mj}{N+1}\right) \sin \left(\frac{\pi nj}{N+1}\right) \\ &= \frac{1}{N+1} \sum_{j=1}^{N} F(\lambda_j) \left\{ \cos \left(\frac{\pi(m-n)j}{N+1}\right) \right. \\ &\quad \left. - \left(\cos \frac{\pi(m+n)j}{N+1}\right) \right\}. \end{aligned} \tag{16}$$

We shall assume that $F\{\lambda(\theta)\}$ can be represented

[3] D. E. Rutherford, Proc. Royal Soc. Edinburgh **62A**, 229 (1945).
[4] D. E. Rutherford, Proc. Royal Soc. Edinburgh **63A**, 232 (1951).

by a Fourier series

$$F\{\lambda(\theta)\} = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos n\theta \tag{17}$$

where

$$A_n = \frac{1}{\pi}\int_0^{2\pi} F\{\lambda(\theta)\} \cos n\theta \, d\theta. \tag{18}$$

and therefore that the $F(\lambda_j)$ result when θ is set equal to $[\pi j/(N+1)]$. Substituting the value of $F(\lambda_j)$ from Eq. (17) into Eq. (16), we have

$$\begin{aligned}[F(\mathbf{\Delta})]_{mn} = {} & \frac{A_0}{2(N+1)} \sum_{j=1}^{N} \left\{\cos\frac{\pi(m-n)j}{N+1} - \cos\frac{\pi(m+n)j}{N+1}\right\} \\ & + \frac{1}{2(N+1)} \sum_{k=1}^{\infty} A_k \\ & \times \sum_{j=1}^{N} \left\{\cos\frac{\pi(m-n-k)j}{N+1} + \cos\frac{\pi(m-n+k)j}{N+1} \right. \\ & \left. - \cos\frac{\pi(m+n-k)j}{N+1} - \cos\frac{\pi(m+n+k)j}{N+1}\right\}. \end{aligned} \tag{19}$$

To evaluate the finite sums appearing in this equation we note that they are all special cases of the prototype sum

$$\begin{aligned}\sum_{j=1}^{N} \cos\frac{\pi s j}{N+1} = {} & N\,\delta_{s,\pm 2r(N+1)} \\ & + \tfrac{1}{2}[(-1)^N - 1][1 - \delta_{s,\pm 2r(N+1)}]\,\delta_{s,\pm(2r+1)(N+1)} \\ & - (1 - \delta_{s,\pm 2r(N+1)} - \delta_{s,\pm(2r+1)(N+1)})\,\delta_{s,\pm 2r}, \end{aligned} \tag{20}$$

where $r = 0, 1, 2, \cdots$. The use of this formula in Eq. (19) results in

$$\begin{aligned}[F(\mathbf{\Delta})]_{mn} = {} & \tfrac{1}{2}(B_{m-n} + B_{-(m-n)} \\ & - B_{m+n} - B_{-(m+n)}) \qquad m \geq n \end{aligned} \tag{21}$$

where

$$B_i = \sum_{r=0}^{\infty} A_{i+2r(N+1)} \tag{22}$$

and a coefficient A_s is zero if s is negative.

The results for matrices of the form of $\mathbf{\Gamma}$ shown in Eq. (9) are similar. Assume that $F(\mathbf{\Delta})$ can be calculated. If the matrix which diagonalizes $\mathbf{\Delta}$ is denoted by $\mathbf{S}$ [cf. Eq. (14)] then the matrix which diagonalizes $\mathbf{\Gamma}$ is just $\mathbf{TS}$. When the proper substitutions are made in the relation

$$F(\mathbf{\Gamma}) = \mathbf{T}F(\mathbf{\Delta})\mathbf{T}^{-1} \tag{23}$$

it is found that

$$\begin{aligned}[F(\mathbf{\Gamma})]_{mn} = {} & [c_n c_{n+1} \cdots c_{m-1}/(b_n b_{n+1} \cdots b_{m-1})]^{1/2} \\ & \times [F(\mathbf{\Delta})]_{mn}, \qquad m > n+1 \\ [F(\mathbf{\Gamma})]_{mn} = {} & [b_m b_{m+1} \cdots b_{n-1}/(c_m c_{m+1} \cdots c_{n-1})]^{1/2} \\ & \times [F(\mathbf{\Delta})]_{mn}, \\ & \qquad n \geq m+1 \\ [F(\mathbf{\Gamma})]_{mm} = {} & [F(\mathbf{\Delta})]_{mm}. \end{aligned} \tag{24}$$

In the following paragraphs we shall give some specific applications of the formula of Eq. (21).

1. Calculation of $\mathbf{\Delta}^{-1}$

The Fourier coefficients of $\mathbf{\Delta}^{-1}$ are

$$\begin{aligned} A_n &= \frac{1}{\pi}\int_0^{\pi} \frac{\cos n\theta}{a + 2b\cos\theta}\, d\theta \\ &= \frac{2}{(a^2 - 4b^2)^{1/2}} \left(\frac{(a^2 - 4b^2)^{1/2} - a}{2b}\right)^n. \end{aligned} \tag{25}$$

When this expression is inserted into the definition of the B's, the resulting series are geometric series and the result can be written in closed form as

$$(\mathbf{\Delta}^{-1})_{mn} = \frac{u(v^{2n} - 1)}{v^{m+n}} \frac{(v^{2(N+1)} - v^{2m})}{1 - v^{2(N+1)}} \qquad m \geq n \tag{26}$$

where

$$u = 1/(a^2 - 4b^2)^{1/2}, \qquad v = \frac{(a^2 - 4b^2)^{1/2} - a}{2b} \tag{27}$$

2. Calculation of $\mathbf{\Delta}^{\nu}$, $\nu \neq 0, 1, 2, \cdots$

The Fourier coefficients can be written

$$\begin{aligned} A_n &= \frac{1}{\pi}\int_0^{2\pi} (a + 2b\cos\theta)^{\nu} \cos n\theta \, d\theta \\ &= \frac{a^{\nu}}{\pi(1+\beta^2)^{\nu}} \int_0^{-2\pi} (1 - 2\beta\cos\theta + \beta^{2\nu}) \cos n\theta \, d\theta \end{aligned} \tag{28}$$

where

$$\beta = -\frac{a - (a^2 + 4b^2)^{1/2}}{2b}$$

Hence the A's are

$$A_n = \frac{2a^{\nu}\beta^n \Gamma(n+\nu)}{(1+\beta^2)^{\nu}\Gamma(\nu)n!} F(\nu, \nu+n, n+1; \beta^2), \tag{29}$$

where $F(a, b, c; x)$ is a hypergeometric function. Thus we find

$$[\mathbf{\Delta}^\nu]_{mn} = \frac{a^\nu}{\Gamma(\nu)(1+\beta^2)^\nu}\left\{\sum_{r=0}^{\infty} F(\nu, \nu+m-n+2r(N+1), m-n+2r(N+1)+1; \beta^2)\right.$$
$$\times \frac{\Gamma(\nu+m-n+2r(N+1))}{\Gamma(m-n+2r(N+1)+1)}\beta^{m-n+2r(N+1)}$$
$$+\sum_{r=1}^{\infty} F(\nu, \nu+2r(N+1)-m+n, 2r(N+1)-m+n+1; \beta^2)$$
$$\times \frac{\Gamma(\nu-m+n+2r(N+1))}{\Gamma(n-m+2r(N+1)+1)}\beta^{-m+n+2r(N+1)}$$
$$-\sum_{r=0}^{\infty} F(\nu, \nu+m+n+2r(N+1), m+n+2r(N+1)+1; \beta^2)$$
$$\times \frac{\Gamma(\nu+m+n+2r(N+1))}{\Gamma(m+n+2r(N+1)+1)}\beta^{m+n+2r(N+1)}$$
$$-\sum_{r=1}^{\infty} F(\nu, \nu+2r(N+1)-m-n, 2r(N+1)-m-n+1; \beta^2)$$
$$\left.\times \frac{\Gamma(\nu+2r(N+1)-m-n)}{\Gamma(2r(N+1)-m-n+1)}\beta^{2r(N+1)-m-n}\right\}. \tag{30}$$

3. Calculation of exp $(t\mathbf{\Delta})$

The Fourier coefficients are

$$A_n = \frac{1}{\pi}\int_0^{2\pi} \exp([a+2b\cos\theta]t)\cos n\theta\, d\theta \tag{31}$$

so that

$$\begin{aligned} A_0 &= 2e^{at}[1+I_0(-2bt)] \\ A_n &= 2(-1)^n e^{at} I_n(-2bt) \end{aligned} \tag{32}$$

where $I_n(x)$ is the Bessel function of imaginary argument of the first kind and the B's can be calculated from Eq. (22).

III. FUNCTIONS OF GENERALIZED CONTINUANT MATRICES

In reference 1 we extended the definition of circulant matrices to a more general class, in order to handle physical problems which arise in two and three dimensions. We will now do the same for the case of continuant matrices, and show how to calculate analytic functions of generalized continuant matrices. We shall restrict ourselves to the generalization of symmetric continuant matrices.

The first-order continuant matrix is defined by its elements

$$\begin{aligned} \Delta_{ij} &= S_{|i-j|}, \qquad |i-j| = 0, 1 \\ &= 0 \qquad\qquad |i-j| > 1 \end{aligned} \tag{33}$$

where, for the particular case in Eq. (8), we have

$$S_0 = a, \qquad S_1 = b. \tag{34}$$

For the definition of the generalized continuant of order r we define a function $S_{\mathbf{k}}$ where $\mathbf{k}$ is an r-dimensional vector with integer components:

$$\mathbf{k} = (k_1, k_2, \cdots, k_r), \qquad k_j = 0, 1, 2, \cdots, N. \tag{35}$$

We will assume that $S_{\mathbf{k}}$ differs from zero only when $k_j = 0$ or 1 for $j = 1, 2, \cdots r$. Now define an integer vector

$$\mathbf{I} = (i_1, i_2, \cdots i_r) \tag{36}$$

where $i_j = 1, 2, \cdots N$. Then we define the generalized continuant matrix with elements $\mathbf{\Delta}(\mathbf{I}, \mathbf{J})$ by

$$\mathbf{\Delta}(\mathbf{I}, \mathbf{J}) = S_{|\mathbf{I}-\mathbf{J}|} \tag{37}$$

where the vector $|\mathbf{I} - \mathbf{J}|$ is defined to have components

$$|\mathbf{I}-\mathbf{J}| = (|i_1-j_1|, |i_2-j_2|, \cdots |i_r-j_r|). \tag{38}$$

In analogy to the one-dimensional case, we make the rule that there are no components of $\mathbf{\Delta}$ which have indices greater than N. An rth order continuant matrix is formed by replacing each scalar element in an $(r-1)$th order continuant, by a continuant matrix.

As an example of the application of this definition, the equations for the steady-state amplitudes of the displacements in an r-dimensional simple-cubic lattice with nearest-neighbor interactions only are:

$$M\omega^2 u_{\mathbf{m}} = \gamma_1 \Delta^2_{m_1} u_{\mathbf{m}} + \gamma_2 \Delta^2_{m_2} u_{\mathbf{m}} + \cdots + \gamma_r \Delta^2_{m_r} u_{\mathbf{m}} \quad (39)$$

where

$$\mathbf{m} = (m_1, m_2, \cdots, m_r),$$

and where, for example

$$\Delta^2_{m_j} u_{\mathbf{m}} = u_{m_1, \cdots m_j+1, \cdots m_r} + u_{m_1, \cdots m_j-1, \cdots m_r} - 2u_{\mathbf{m}}. \quad (40)$$

If the edges of the lattice are fixed, then $u = 0$ if any one of the indices is equal to 0 or $N + 1$ (where we have assumed that the crystal has n atoms on an edge). The solution to Eq. (39) is equivalent to finding the eigenvectors and eigenvalues of a generalized circulant matrix $\mathbf{\Delta}$

$$\begin{aligned} S_{(0,0,\cdots 0)} &= -2(\gamma_1 + \gamma_2 + \cdots + \gamma_r) \\ S_{(1,0,0,\cdots 0)} &= \gamma_1 \\ S_{(0,1,0,\cdots 0)} &= \gamma_2 \\ &\vdots \\ S_{(0,0,\cdots,1)} &= \gamma_r. \end{aligned} \quad (41)$$

In two dimensions the matrix $\mathbf{\Delta}$ can be written as

$$\mathbf{\Delta} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B} & \mathbf{A} & \mathbf{B} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{A} & \mathbf{B} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & \vdots & \vdots & & & \vdots \end{bmatrix}, \quad (42)$$

where

$$\mathbf{A} = \begin{bmatrix} -2(\gamma_1 + \gamma_2) & \gamma_1 & 0, & \cdots & 0 \\ \gamma_1 & -2(\gamma_1 + \gamma_2) & \gamma_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \end{bmatrix};$$

$$\mathbf{B} = \begin{bmatrix} 0 & \gamma_2 & 0 & \cdots \\ \gamma_2 & 0 & \gamma_2 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}. \quad (43)$$

It is a simple matter to find the eigenvalues and eigenvectors of the generalized continuant $\mathbf{\Delta}(\mathbf{I}, \mathbf{J})$. Let us write the eigenvalue equations as

$$\mathbf{\Delta}(\mathbf{I}, \mathbf{J})\mathbf{U} = \lambda \mathbf{U}, \quad (44)$$

where $\mathbf{U}$ is a vector with components $u(n_1, n_2, \cdots n_r)$ where $n_1, n_2, \cdots n_r = 1, 2, \cdots, N$. Assuming a solution of the form

$$u(n_1, n_2, \cdots n_r) = A \prod_{j=1}^{r} \sin n_j\theta, \quad (45)$$

we fit the boundary condition that u be equal to zero if any index is equal to $N + 1$ by setting

$$\theta = \pi(N + 1)^{-1} \quad (46)$$

and A will be chosen so that the vector $\mathbf{U}$ is normalized;

$$A = [2/(N + 1)]^{r/2}. \quad (47)$$

The eigenvalues of $\mathbf{\Delta}$ are readily found to be

$$\lambda(\mathbf{k}) = \sum_{\{\epsilon\}} 2^{\epsilon_1+\epsilon_2+\cdots+\epsilon_r} S_{(\epsilon_1,\epsilon_2,\cdots\epsilon_r)} \times \cos\frac{\pi k_1\epsilon_1}{N+1} \cos\frac{\pi k_2\epsilon_2}{N+1} \cdots \cos\frac{\pi k_r\epsilon_r}{N+1}. \quad (48)$$

The matrix which diagonalizes $\mathbf{\Delta}(\mathbf{I}, \mathbf{J})$ has the explicit representation

$$\mathbf{T}(\mathbf{n}, \mathbf{m}) = \left(\frac{2}{N+1}\right)^{r/2} \sin\frac{\pi n_1 m_1}{N+1} \times \sin\frac{\pi n_2 m_2}{N+1} \cdots \sin\frac{\pi n_r m_r}{N+1}. \quad (49)$$

Finally, we find as a representation for $F(\mathbf{\Delta})$

$$[F(\mathbf{\Delta})]_{\mathbf{mn}} = \left(\frac{2}{N+1}\right)^r \sum_{k_1} \cdots \sum_{k_r} F[\mathbf{\Lambda}(\mathbf{k})] \times \sin\frac{\pi m_1 k_1}{N+1} \sin\frac{\pi m_2 k_2}{N+1} \cdots \sin\frac{\pi m_r k_r}{N+1} \times \sin\frac{\pi n_1 k_1}{N+1} \sin\frac{\pi n_2 k_2}{N+1} \cdots \sin\frac{\pi n_r k_r}{N+1}. \quad (50)$$

where $\mathbf{\Lambda}(\mathbf{k})$ is the diagonal matrix of eigenvalues.

The following analysis is exactly analogous to the one dimensional case. We define the function of a continuous vector of variables $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots \theta_r)$:

$$\lambda(\boldsymbol{\theta}) = \sum_{\{\epsilon\}} 2^{\epsilon_1+\epsilon_2+\cdots+\epsilon_r} S_{(\epsilon_1,\cdots\epsilon_r)} \times \cos\pi\epsilon_1\theta_1 \cos\pi\epsilon_2\theta_2 \cdots \cos\pi\epsilon_r\theta_r \quad (51)$$

and assume that $F\{\lambda(\boldsymbol{\theta})\}$ has a Fourier expansion

$$F\{\lambda(\boldsymbol{\theta})\} = \sum_{\mathbf{n}} \frac{A_{\mathbf{n}}}{2^p} \cos n_1\theta_1 \cos n_2\theta_2 \cdots \cos n_r\theta_r, \quad n_1, n_2, \cdots n_r = 0, 1, 2, \cdots, \quad (52)$$

where p is the number of n's that are equal to zero in the particular term and

$$A_{\mathbf{n}} = \frac{1}{\pi^r} \int_0^{2\pi} \cdots \int_0^{2\pi} F\{\lambda(\boldsymbol{\theta})\} \cos n_1\theta_1 \times \cos n_2\theta_2 \cdots \cos n_r\theta_r \, d^r\boldsymbol{\theta}. \quad (53)$$

The value of $F\{\lambda(\mathbf{k})\}$ required for the general

formula of Eq. (52) is

$$F\{\lambda(\mathbf{k})\} = \sum_{n_1=1}^{N} \cdots \sum_{n_r=1}^{N} B^*_{\mathbf{n}} \times \cos\frac{\pi n_1 k_1}{N+1} \cdots \cos\frac{\pi n_r k_r}{N+1}. \quad (54)$$

The B^*'s are found from

$$B^*_{\mathbf{n}} = \sum_{l_1} \cdots \sum_{l_r} \frac{A_{\mathbf{n}+1}(N+1)}{2^p}, \quad (55)$$

where the vector $\mathbf{l} = (l_1, l_2, \cdots, l_r)$ has positive integer components which go from zero to infinity. When Eq. (55) is inserted into Eq. (52) we see that the sum over the k's is separable and can be carried out using the quantities

$$V_m = \sum_{j=1}^{N} \cos\frac{\pi m j}{N+1} = N\,\delta_{m,\pm 2s(N+1)} + \tfrac{1}{2}[(-1)^N - 1][1 - \delta_{m,\pm 2s(N+1)}]\,\delta_{m,\pm(2s+1)(N+1)} - (1 - \delta_{m,\pm 2s(N+1)} - \delta_{m,\pm(2s+1)(N+1)})\,\delta_{m,\pm 2s}, \quad (56)$$

where $s = 0, 1, 2, \cdots$. The final formula for the r-dimensional circulant matrix is

$$[F(\mathbf{\Delta})]_{\mathbf{m},\mathbf{n}} = \tfrac{1}{4} \sum_{\mathbf{p}} B^*_{\mathbf{p}} \prod_{s=1}^{r} [V_{m_s - n_s + p_s} + V_{m_s - n_s - p_s} - V_{m_s + n_s + p_s} - V_{m_s + n_s - p_s}]. \quad (57)$$

It does not seem possible to find a result analogous to Eq. (9) without additional assumptions being made on the commutativity of the submatrices which appear in the definition of the generalized continuant matrix.

It may be remarked at this stage that if one defines a lexicographical order for the vectors **I**, as was done in I, it is possible to write down explicitly the orthogonal matrix which brings the generalized continuant matrix to the form of a generalized diagonal matrix, regardless of dimensionality.

For instance if $\mathbf{\Delta}$ is the matrix

$$\mathbf{\Delta} = \begin{bmatrix} \mathbf{A} & \mathbf{B} & & & \mathbf{O} \\ \mathbf{B} & & \ddots & & \\ & \ddots & & \ddots & \\ & & \ddots & & \mathbf{B} \\ \mathbf{O} & & & \mathbf{B} & \mathbf{A} \end{bmatrix}, \quad (58)$$

where **A** and **B** may be generalized matrices themselves of the m'th order, then the diagonalizing orthogonal matrix **U** has the matrix elements:

$$U_{\mathbf{K},\mathbf{J}} = \left(\frac{2}{\tau^n + 1}\right)^{1/2} \sin\frac{\pi k j}{\tau^n + 1}\, \mathfrak{I}^{(m)}; \quad k, j = 1, \cdots, \tau^n, \quad (59)$$

where $\mathfrak{I}^{(m)}$ is the unit matrix of mth order, the vectors **K**, **J** are n-dimensional vectors whose components are taken from a set of τ numbers (not necessarily integers) and k, j are the positions of **K** and **J**, respectively, in the lexicographical order employed.

The generalized eigenvalues are

$$\mathfrak{\lambda}_j = \mathbf{A} + 2\mathbf{B}\cos\frac{\pi j}{\tau^n + 1}; \quad j = 1, \cdots, \tau^n. \quad (60)$$

Therefore, instead of having to diagonalize a matrix of order $m \times \tau^n$ we have to diagonalize one of order m only. Equation (60) is valid regardless of the commutability of **A** and **B**. If **A** and **B** are simple continuant (therefore commuting) we recover the formulas (48), the simple eigenvalues of $\mathbf{\Delta}$ being given by:

$$\lambda_{jk} = \lambda_k(\mathbf{A}) + 2\lambda_k(\mathbf{B})\cos\frac{\pi j}{\tau^n + 1}; \quad \begin{matrix} k = 1, \cdots, m \\ j = 1, \cdots, \tau^n \end{matrix} \quad (61)$$

In general if **A** and **B** commute and their eigenvectors and eigenvalues satisfy the periodicity requirements stated above, one can calculate analytic functions of $\mathbf{\Delta}$ using a single Fourier expansion.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Symmetry Restrictions on Field Dependent Tensors with Application to Galvanomagnetic Effects

LUDWIG GRABNER† AND JOHN A. SWANSON*
IBM Research Laboratory, Yorktown Heights, New York
(Received September 6, 1960)

The usual formalism for determining symmetry relations, due to macroscopic space symmetry, among the components of *constant* tensors, is shown to apply equally well to tensors that are a *function of the applied fields*. It differs only in two respects from the special case of constant tensors: (a) Only a subgroup, containing, in general, drastically fewer members than the entire group of symmetry operations, yields symmetry relations. (b) This subgroup contains elements other than the identity only for directions of the applied fields that are left invariant by the elements of macroscopic symmetry of the medium. Examples using first- and second-order tensors arising in electrical conductivity, with and without a magnetic field, are given and the even and odd parts of the tensor are separated.

I. INTRODUCTION

SYMMETRY relations in tensors are due to intrinsic symmetry, such as the Onsager relations,[1] for example, and macroscopic space symmetry. The latter is geometrical in nature whereas the former is derived from physical arguments independent of macroscopic space symmetry. We shall be concerned with the effects of macroscopic space symmetry and suppose the intrinsic symmetry given.

If a tensor is a function of the applied fields, it is the practice to expand it in a power series with respect to these fields. Symmetry considerations are then applied to the *constant* coefficients of the power series that define new tensors. The study of symmetry restrictions in constant tensors in material media with space symmetry has recently been formulated with great elegance.[2-10] The practical occasion for the study of higher rank constant tensors is the occurrence of interesting nonlinear phenomena, which one desires to describe approximately by a finite number of terms in a power series expansion. The number of independent tensor components appearing in the tensor coefficients of such a finite expansion is reduced by symmetry restrictions. Applications of this procedure to magnetoconductivity are given in various references[11-16].

A power series expansion is practical only to the first few terms and for some applications this may restrict its usefulness to a limited range of applied fields. We therefore divide the applied fields into "weak" and "strong" fields. Weak fields, by definition, are those with respect to which the effect is well described by an expansion of only a few terms; strong fields those for which this is not the case. If we expand the effect in terms of the weak fields only, the coefficients of the power series are tensors which are a function of the strong fields and one is led to inquire what restrictions space symmetry imposes on such "*field dependent tensors.*" We find that the symmetry restrictions upon field dependent tensors is determined by the intersection of the symmetry group of the medium and the symmetry group which leaves the strong applied fields invariant, in contrast to constant tensors where the much larger symmetry group of the medium imposes the symmetry restrictions on the tensor.

With respect to the physical basis for all symmetry relations it might be mentioned that it is founded in the principle of covariance; i.e., the assumption that the mathematical form in which physical laws are expressed must be identical in

† Present address: National Bureau of Standards, Washington, D.C.
* Deceased.
[1] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).
[2] F. G. Fumi, Acta Cryst. **5**, 44 (1952).
[3] F. G. Fumi, Acta Cryst. **5**, 691 (1952).
[4] F. G. Fumi, Nuovo cimento **9**, 739 (1952).
[5] R. Fieschi and F. G. Fumi, Nouvo cimento **10**, 865 (1953).
[6] H. Wondratschek, Neues Jahrb. Mineral. Monatsh. **85**, 217 (1952); **86**, 25 (1953).
[7] *Handbuch der Physik*, edited by S. Flügge (Springer-Verlag, Berlin, 1955), Vol. VII, Part 1, p. 40.
[8] C. S. Smith, *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic Press Inc., New York, 1958), Vol. 6.
[9] A. C. Pipkin and R. S. Rivlin, Arch. Ratl. Mech. Anal. **4**, 129 (1959).
[10] A. C. Pipkin and R. S. Rivlin, J. Math. Phys. **1**, 127 (1960).
[11] F. Seitz, Phys. Rev. **79**, 372 (1950).
[12] W. P. Mason, W. H. Hewitt, and R. F. Wick, J. Appl. Phys. **24**, 166 (1953).
[13] H. J. Juretschke, Acta Cryst. **8**, 716 (1955).
[14] T. Okada, Mem. Fac. Sci. Kyusyu Univ. **B1**, 157 (1955).
[15] J. R. Drabble and R. Wolfe, Proc. Phys. Soc. (London). **B69**, 1109 (1956).
[16] L. P. Kao and E. J. Katz, J. Phys. Chem. Solids **6**, 223 (1958).

equivalent coordinate systems. Thus, the method and point of view of this analysis are inherently identical with that used for constant tensors and the formalism established there can be carried over entirely to the present problem once it is recognized that only the elements common to the symmetry group of the medium and the symmetry group which leave the "strong" fields invariant are effective in producing symmetry restrictions on a field dependent tensor.

Our procedure will be to develop in Sec. II the mathematical formalism corresponding to these remarks and to give examples taken from electrical conductivity in Secs. III and IV.

II. CONVARIANCE APPLIED TO FIELD DEPENDENT TENSORS

In an orthogonal reference frame (k) consider the coupling relation between an effect E and n applied fields $F^{(1)} \cdots F^{(n)}$ through a tensor $E_{i\cdots j}$ which is a function of $F^{(1)} \cdots F^{(n)}$.[17]

$$E_{i\cdots j} = E_{i\cdots j}(\mathbf{F}^{(1)} \cdots \mathbf{F}^{(m)}; \mathbf{F}^{(m+1)} \cdots \mathbf{F}^{(n)}). \quad (1)$$

The fields F may be tensors of arbitrary rank. The superscripts identify the field. The symbol $\mathbf{F}$ stands for the ordered set of all components of F, thus if F is a vector, $\mathbf{F} = (F_1, F_2, F_3)$. Now, if for the particular application in mind, the fields $F^{(1)} \cdots F^{(m)}$ are "strong," whereas $F^{(m+1)} \cdots F^{(n)}$ are "weak," we expand (1) in terms of the "weak" fields. Thus, $E_{i\cdots j}$ will be a sum of expressions of the form:

$$E_{i\cdots j} = \phi_{i\cdots j;k\cdots l}(\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \cdots \mathbf{F}^{(m)}) \times F^{(m+1)}_{k\cdots} \cdots F^{(n)}_{\cdots l}, \quad (2)$$

where $\phi_{i\cdots j;k\cdots l}$ is a derivative of $E_{i\cdots j}$ with respect to the weak fields.

Examples of (1) and (2) are electrical conduction in a magnetic field: $E_i = E_i(\mathbf{H}, \mathbf{I})$ in which $\mathbf{E}$, $\mathbf{H}$, and $\mathbf{I}$ are the electric field, magnetic field, and current density. Expanding $E_i(\mathbf{H}, \mathbf{I})$ with respect to $\mathbf{H}$ or $\mathbf{I}$ depending on which of these fields is "weak," we get:

$$E_i = F_i(\mathbf{I}) + R_{ij}(\mathbf{I})H_j + \cdots \quad (3)$$

$$E_i = f_i(\mathbf{H}) + \rho_{ij}(\mathbf{H})I_j + \cdots \quad (4)$$

where $R_{ij}(\mathbf{I}) = \partial E_i/\partial H_j$, $\rho_{ij}(\mathbf{H}) = \partial E_i/\partial I_j$.

The connection between the geometrical properties of the medium and the laws of physics is provided by the principle of covariance according to which the laws of physics are form invariant with respect to geometrically equivalent frames; i.e., if another frame (k') is related to (k) by a symmetry operation of the medium, then in (k') the phenomenon is represented by

$$E'_{i\cdots j} = \phi_{i\cdots j;k\cdots l}(\mathbf{F}'^{(1)}, \mathbf{F}'^{(2)}, \cdots \cdots \mathbf{F}'^{(m)})F'^{(m+1)}_{k\cdots} \cdots F'^{(n)}_{\cdots l}, \quad (5)$$

where ϕ is the *same* function as in (2).

The simultaneous validity of Eqs. (2) and (5), together with the transformation rules for tensors, allows us to deduce the general nonintrinsic symmetry conditions on the functions $\phi_{i\cdots l}$.

In the interest of clarity of notation we shall show these for a tensor of arbitrary rank which is a function of a single applied field, an axial vector. This particular case shows all the essentials and the generalization to a tensor which is a function of any number of applied fields of arbitrary rank will be obvious. Given two Cartesian frames (k) and (k') and the orthogonal transformation connecting them $x'_i = s_{ij}x_j$, the components of ϕ in the primed system are determined by the usual laws of tensor transformation:

$$\phi'_{i\cdots l}(F'_1, F'_2, F'_3) = s_{ip} \cdots s_{lr}\phi_{p\cdots r}(F_1, F_2, F_3). \quad (6)$$

Equation (6) allows us to determine ϕ' as a function of $\mathbf{F}'$ merely by expressing (F_1, F_2, F_3) in terms of (F'_1, F'_2, F'_3) : $F_n = |s|s_{qn}F'_q \cdot |s|$ is the determinant of the coordinate transformation and must be included since $\mathbf{F}$ is assumed to be an axial vector.

However, if s is a symmetry operation of the medium without applied fields, the function $\phi'(F'_1, F'_2, F'_3)$ is known from the principle of covariance to be the original function ϕ, so that when either set of field components $\mathbf{F}$ is expressed in terms of the other, Eq. (6) becomes a set of functional relationships constituting the symmetry conditions on ϕ imposed by s:

$$\phi(|s|\, s_{1q}F_q, |s|\, s_{2q}F_q, |s|\, s_{3q}F_q) = s_{ip}\cdots s_{lr}\phi_{p\cdots r}(F_1, F_2, F_3). \quad (7)$$

The obvious generalization of (7) is:

$$\phi_{i\cdots l}(\mathbf{F}'^{(1)}, \mathbf{F}'^{(2)}, \cdots \mathbf{F}'^{(m)}) = s_{ip}\cdots s_{lr}\phi_{p\cdots r}(\mathbf{F}^{(1)}, \mathbf{F}^{(2)}, \cdots \mathbf{F}^{(m)}), \quad (8)$$

where the $\mathbf{F}$'s are tensors of arbitrary rank and in which the applied fields on the left-hand side of (8) are to be expressed in terms of those on the right-hand side as is done in the example (7), and where the polarity (axial or polar) of the various fields must be taken into account as in (7). Equation (8) is identical with Eq. (3.8) of reference 9.

[17] The notation of tensor analysis is used throughout. Summation over repeated indices is understood.

Equation (7) or the more general Eq. (8) is the point of departure for all discussions of symmetry restrictions. The complete set of spatial symmetry conditions is obtained by letting s run over all point group operations of the space group. That is, if

$$x_i' = s_{ij}x_j + v_i,$$

where v_i may be a whole or fractional lattice translation, is a symmetry operation of the lattice, then since translations do not affect our (constant) fields, the coordinate transformation

$$x_i' = s_{ij}x_j$$

is acceptable to us.

The experimental content of Eq. (8) is given by a reinterpretation of the transformed field components in the left member as the components of *differently oriented* fields in the *same* coordinate system. We are searching for symmetry *restrictions*; i.e., for numerical relationships among the tensor components in a *given* coordinate system for a *given* orientation of the "strong" fields. Clearly such exist whenever the "strong" fields are oriented such that

$$\mathbf{F}^{(h)} = \mathbf{F}'^{(h)}, \qquad 1 \leqq h \leqq m;$$

i.e., such that the transformation s leaves the field components invariant.

It should be noted that in the special case of constant tensors every symmetry operation of the medium yields, in general, a symmetry restriction. However, for field dependent tensors only those symmetry operations of the medium yield symmetry restrictions which leave the applied fields invariant; i.e., the intersection of the symmetry group of the medium with the symmetry group which leave the applied fields invariant. Thus, the general case of field dependent tensors differs in two respects from the special one of constant tensors: (a) Only a subgroup, containing in general drastically fewer members than the entire group of symmetry operations yields symmetry restrictions. (b) This subgroup contains elements other than the identity only for special directions of the "strong" fields.

III. APPLICATIONS

Intrinsic symmetry relations (Onsager relations, for example) usually state restrictions on a tensor under sign reversal of the strong fields. When only one strong field is involved, which is the case in our applications, the effect of intrinsic symmetry is more conveniently incorporated in the results by working with the "even" and "odd" tensors $\bar{\phi}$ and $\tilde{\phi}$ where $\phi = \bar{\phi} + \tilde{\phi}$ and

$$2\bar{\phi}_{i\ldots k}(\mathbf{F}) \equiv \phi_{i\ldots k}(\mathbf{F}) + \phi_{i\ldots k}(-\mathbf{F})$$

$$2\tilde{\phi}_{i\ldots k}(\mathbf{F}) \equiv \phi_{i\ldots k}(\mathbf{F}) - \phi_{i\ldots k}(-\mathbf{F}).$$

The tensors $\bar{\phi}$ and $\tilde{\phi}$ are obtained experimentally by reversing the direction of F. Equation (8) clearly applies to $\bar{\phi}$ and $\tilde{\phi}$. Furthermore, these tensors possess the properties $\bar{\phi}(\mathbf{F}) = \bar{\phi}(-\mathbf{F})$, $\tilde{\phi}(\mathbf{F}) = -\tilde{\phi}(-\mathbf{F})$ which allows us to extend the domain of allowable symmetry operations to those that reverse the large fields. In our applications we first apply this enlarged domain of spatial symmetry operations to the even and odd tensors. Further restrictions due to intrinsic symmetries are then easily applied to $\bar{\phi}$ and $\tilde{\phi}$.

We shall be concerned with the terms of the "partial" expansions (3) and (4). In (3), $F_i(\mathbf{I})$ describes nonlinear conduction, $R_{ij}(\mathbf{I})$ nonohmic conduction in a weak magnetic field. In (4), $\rho_{ij}(\mathbf{H})$ is the resistivity tensor in the approximation of Ohm's law, $f_i(\mathbf{H})$ vanishes because of time inversion symmetry.[18,19]

As the independent symmetry elements of the point group symmetries of crystals we choose the proper rotation $n = 1, 2, 3, 4, 6$, and the improper rotations $\bar{n} = \bar{1}, \bar{2}, \bar{4}$. $\bar{1}$ is an inversion, $\bar{2}$ a reflection with respect to a plane. With our choice of independent symmetry elements $\bar{3}$ and $\bar{6}$ are not independent. $\bar{3}$ is generated from 3 and $\bar{1}$, $\bar{6}$ from 3 and $\bar{2}$. Once a tensor is reduced for the independent symmetry elements, its form for dependent symmetry elements is easily obtained by virtue of the fact that if the symmetry operations s_1 and s_2 exist independently, a tensor satisfies $s_3 = s_1s_2$ if it satisfies s_1 and s_2. With respect to the details of obtaining the symmetry relations, the easily applied "direct inspection method"[2,3] can be used to advantage. Many examples using this method are given by Nye.[20] We shall make frequent use of a theorem due to Hermann[21] according to which a tensor of rank $r < n$ cannot distinguish an n-fold proper rotational symmetry axis from one for which $n = \infty$; i.e., cylindrical symmetry. In the following we use an orthogonal reference frame in which x_3 is parallel to the axis of rotation, x_1 and x_2 are arbitrary.

[18] H. Zocher and C. Török, Anais acad. brasil. cienc. **20**, 143 (1948).
[19] H. Zocher and C. Török, Proc. Natl. Acad. Sci. U. S. **39**, 681 (1953).
[20] J. F. Nye, *Physical Properties of Crystals* (Oxford University Press, New York, 1957).
[21] C. Hermann, Z. Krist. **89**, 32 (1934); see also reference 7, page 53.

A. Nonlinear Electrical Conduction: $E_i = F_i(I)$

$F_i(\mathbf{I})$ is a polar vector. Equation (8) for this case is:

$$F_i(s_{1q}I_q, s_{2q}I_q, s_{3q}I_q) = s_{ij}F_j(I_1, I_2, I_3) \qquad (9)$$

Applying (9) to $\tilde{F}_i(\mathbf{I})$ and $\tilde{F}_i(\mathbf{I})$ we find symmetry restrictions on $F_i(\mathbf{I})$ only for the following directions of $\mathbf{I}$:

1. $n = 1, 2, 3, 4, 6$

$\mathbf{I}$ parallel to an n axis or perpendicular to a $2n$ fold axis. In the latter case a twofold rotation is used to reverse the current. According to Hermann's theorem we need not distinguish axes with $n > 1$.

$$F_i(\mathbf{I} \parallel n > 1) = \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix} \qquad (10)$$

$$F_i(\mathbf{I} \perp 2n) = \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix} + \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \\ 0 \end{bmatrix} \qquad (11)$$

2. $\bar{n} = \bar{1}, \bar{2}, \bar{4}$

A center of inversion destroys $\tilde{F}_i$:

$$F_i(\bar{1}) = \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \\ \tilde{F}_3 \end{bmatrix} \qquad (12)$$

Thus, *if* an intrinsic symmetry relation exists in the form: $F_i(-\mathbf{I}) = \pm F_i(\mathbf{I})$, it must be odd to agree with the possibility of a center of inversion. For $\bar{2}$ and $\bar{4}$ we find:

$$F_i(\mathbf{I} \parallel \bar{2}) = \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix} \qquad (13)$$

$$F_i(\mathbf{I} \perp \bar{2}) = \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \\ 0 \end{bmatrix} + \begin{bmatrix} \tilde{F}_1 \\ \tilde{F}_2 \\ 0 \end{bmatrix} \qquad (14)$$

$$F_i(\mathbf{I} \parallel \bar{4}) = \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix} \qquad (15)$$

$F_i(\mathbf{I} \perp \bar{4})$ is covered by (11) since the group $\bar{4}$ contains the element $n = 2$. $F_i(\mathbf{I} \parallel \bar{3})$ or $F_i(\mathbf{I} \parallel \bar{6})$ is obtained by compounding (10) and (12), and (10) and (13), respectively.

For example,

$$F_i(\mathbf{I} \parallel \bar{6}) = \begin{bmatrix} 0 \\ 0 \\ \tilde{F}_3 \end{bmatrix}$$

The cases $F_i(\mathbf{I} \parallel \bar{3})$, $F_i(\mathbf{I} \parallel \bar{4})$ and $F_i(\mathbf{I} \parallel \bar{6})$ are found to give the same result.

B. Nonohmic Conduction in a Weak Magnetic Field: $E_i = R_{ij}(I)H_j$

We treat this case in the approximation to which the first power of the magnetic field is adequate. Since $\mathbf{E}$ is a polar vector, $\mathbf{H}$ an axial vector, $R_{ij}(\mathbf{I})$ must be an axial tensor, and Eq. (8) for this case is:

$$R_{ij}(s_{1q}I_q, s_{2q}I_q, s_{3q}I_q) = |s|\, s_{im}s_{jn}R_{mn}(I_1, I_2, I_3). \qquad (16)$$

Applying (16) to $\tilde{R}_{ij}(\mathbf{I})$ and $\tilde{R}_{ij}(\mathbf{I})$ we find symmetry restrictions only for the following directions of $\mathbf{I}$:

1. $n = 1, 2, 3, 4, 6$

According to Hermann's theorem we need not distinguish axes with $n > 2$.

$$R_{ij}(\mathbf{I} \parallel 2) = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ \tilde{R}_{21} & \tilde{R}_{22} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} + \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ \tilde{R}_{21} & \tilde{R}_{22} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} \qquad (17)$$

$$R_{ij}(\mathbf{I} \parallel n > 2) = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ -\tilde{R}_{12} & \tilde{R}_{11} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} + \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ -\tilde{R}_{12} & \tilde{R}_{11} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} \qquad (18)$$

$$R_{ij}(\mathbf{I} \perp 2n) = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ \tilde{R}_{21} & \tilde{R}_{22} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} + \begin{bmatrix} 0 & 0 & \tilde{R}_{13} \\ 0 & 0 & \tilde{R}_{23} \\ \tilde{R}_{31} & \tilde{R}_{32} & 0 \end{bmatrix} \qquad (19)$$

2. $n = \bar{1}, \bar{2}, \bar{4}$

A center of inversion destroys $\tilde{R}_{ij}(\mathbf{I})$:

$$R_{ij}(\bar{1}) = \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & \tilde{R}_{13} \\ \tilde{R}_{21} & \tilde{R}_{22} & \tilde{R}_{23} \\ \tilde{R}_{31} & \tilde{R}_{32} & \tilde{R}_{33} \end{bmatrix} \qquad (20)$$

For $\bar{2}$ and $\bar{4}$ we find:

$$R_{ij}(\mathbf{I} \parallel \bar{2}) = \begin{bmatrix} 0 & 0 & \tilde{R}_{13} \\ 0 & 0 & \tilde{R}_{23} \\ \tilde{R}_{31} & \tilde{R}_{32} & 0 \end{bmatrix} + \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ \tilde{R}_{21} & \tilde{R}_{22} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} \qquad (21)$$

$R_{ij}(\mathbf{I} \perp \bar{2})$

$$= \begin{bmatrix} 0 & 0 & \tilde{R}_{13} \\ 0 & 0 & \tilde{R}_{23} \\ \tilde{R}_{31} & \tilde{R}_{32} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \tilde{R}_{13} \\ 0 & 0 & \tilde{R}_{23} \\ \tilde{R}_{31} & \tilde{R}_{32} & 0 \end{bmatrix} \quad (22)$$

$R_{ij}(\mathbf{I} \parallel \bar{4})$

$$= \begin{bmatrix} \bar{R}_{11} & \bar{R}_{12} & 0 \\ \bar{R}_{12} & -\bar{R}_{11} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & 0 \\ -\tilde{R}_{12} & \tilde{R}_{11} & 0 \\ 0 & 0 & \tilde{R}_{33} \end{bmatrix} \quad (23)$$

$R_{ij}(\mathbf{I} \perp \bar{4})$, $R_{ij}(\mathbf{I} \parallel \bar{3})$, and $R_{ij}(\mathbf{I} \parallel \bar{6})$ are treated in the same way as the analogous cases for $F_i(\mathbf{I})$ in Sec. A. $R_{ij}(\mathbf{I} \perp \bar{4})$ gives, of course, the result (19) and $R_{ij}(\mathbf{I} \parallel \bar{3})$ and $R_{ij}(\mathbf{I} \parallel \bar{6})$ are found to give the same result:

$$R_{ij}(\mathbf{I} \parallel \bar{3} \text{ or } \bar{6}) = \begin{bmatrix} \bar{R}_{11} & \bar{R}_{12} & 0 \\ -\bar{R}_{12} & \bar{R}_{11} & 0 \\ 0 & 0 & \bar{R}_{33} \end{bmatrix}$$

C. Ohmic Conduction in a Strong Magnetic Field: $E_i = \varrho_{ij}(H) I_j$

This case is important in practice. $\rho_{ij}(\mathbf{H})$ is a polar tensor. Equation (8) for this case is

$$\rho_{ij}(|s|\, s_{1q} H_q,\ |s|\, s_{2q} H_q,\ |s|\, s_{3q} H_q) = s_{im} s_{jn} \rho_{mn}(H_1, H_2, H_3). \quad (24)$$

Since $\mathbf{H}$ is an axial vector, a center of inversion leaves (24) invariant and therefore may be added to the existing space symmetry as far as this phenomenon is concerned. Thus ohmic conduction in a magnetic field is unable to distinguish proper and improper symmetry operations.[22]

As an intrinsic symmetry relation we have the Onsager relation[1]: $\rho_{ij}(\mathbf{H}) = \rho_{ji}(-\mathbf{H})$ and as a result $\bar{\rho}_{ij}(\mathbf{H})$ is symmetric in the indices and $\tilde{\rho}_{ij}(\mathbf{H})$ antisymmetric in the indices.

It is necessary merely to apply these intrinsic conditions to the results listed in Eqs. (17) to (19), which apply now equally well to the corresponding improper rotations:

$\rho_{ij}(\mathbf{H} \parallel 2 \text{ or } \bar{2})$

$$= \begin{bmatrix} \bar{\rho}_{11} & \bar{\rho}_{12} & 0 \\ \bar{\rho}_{12} & \bar{\rho}_{22} & 0 \\ 0 & 0 & \bar{\rho}_{33} \end{bmatrix} + \begin{bmatrix} 0 & \tilde{\rho}_{12} & 0 \\ -\tilde{\rho}_{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (25)$$

$\rho_{ij}(\mathbf{H} \parallel n > 2 \text{ or } \bar{n} > \bar{2})$

$$= \begin{bmatrix} \bar{\rho}_{11} & 0 & 0 \\ 0 & \bar{\rho}_{11} & 0 \\ 0 & 0 & \bar{\rho}_{33} \end{bmatrix} + \begin{bmatrix} 0 & \tilde{\rho}_{12} & 0 \\ -\tilde{\rho}_{12} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (26)$$

$\rho_{ij}(\mathbf{H} \perp 2n \text{ or } 2\bar{n})$

$$= \begin{bmatrix} \bar{\rho}_{11} & \bar{\rho}_{12} & 0 \\ \bar{\rho}_{12} & \bar{\rho}_{22} & 0 \\ 0 & 0 & \bar{\rho}_{33} \end{bmatrix} + \begin{bmatrix} 0 & 0 & \tilde{\rho}_{13} \\ 0 & 0 & \tilde{\rho}_{23} \\ -\tilde{\rho}_{13} & -\tilde{\rho}_{23} & 0 \end{bmatrix} \quad (27)$$

IV. APPLICATIONS TO PARTICULAR MEDIA

In many cases there will exist more than one symmetry operation which leaves the "strong" fields invariant when they possess a particular direction. In such cases, just as with constant tensors, the tensor is to be reduced as far as possible by the simultaneous application of the different sets of symmetry operation. A complete characterization of the symmetry restrictions upon a field dependent tensor is obtained by listing the reduced tensors together with the field directions which occasion the reduction.

V. ACKNOWLEDGMENT

Discussions with Professor Fausto Fumi and his comments on this paper are gratefully acknowledged.

[22] D. Schoenberg, Proc. Cambridge Phil. Soc. **31**, 265, 271 (1935).

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Normal Forms of Complex Matrices*

BRUNO ZUMINO
Physics Department, New York University, New York, New York
(Received April 9, 1962)

Simple normal forms are given for symmetric, skew symmetric, and general complex matrices. These three cases can be combined into a single one, if one considers matrices with elements in a suitable ground field.

1. INTRODUCTION

IN this paper we state and prove several theorems on square complex matrices which do not seem to be described in the mathematical literature. Use of these theorems leads to simplifications in the treatment of some problems in quantum field theory and in the theory of superconductivity. The theorems also occur in the study of quantum mechanical time inversion. The physical applications, however, will be published separately.

Some of the results given in this paper have already been used by other authors. For instance, Lemma I is implicit in Wigner's work on time inversion.[1] Theorem I is suggested by some work of Valatin in the theory of superconductivity.[2] It was actually proven by Yang who used it in his discussion of long range order in Fermi systems.[3]

The three theorems given below as Theorems I, II, and III can be combined into one single very elegant theorem, stated below as Theorem IV. This fact was pointed out by F. J. Dyson in a letter to the author. Dyson's argument is reproduced in the Appendix. It is not difficult to give a direct proof of Theorem IV, but this shall not be done here.

2. NORMAL FORMS

We first prove two lemmas.

Lemma I. If S is unitary and skew symmetric, then[4]

$$S = UF\tilde{U}$$

where U is unitary and

$$F = \begin{vmatrix} 0 & -1 & & & 0 \\ 1 & 0 & & & \\ & & 0 & -1 & \\ & & 1 & 0 & \\ 0 & & & & \ddots \end{vmatrix}$$

Proof. Since $S\dagger = S^{-1}$ and $\tilde{S} = -S$, $S^* = -S^{-1}$. Therefore S and S^* commute. The matrices $A_1 = S + S^*$ and $A_2 = i(S - S^*)$ are real, skew symmetric, and commute with each other. By a known theorem,[5] they can be transformed simultaneously by means of a real orthogonal transformation O into the normal form

$$\begin{vmatrix} 0 & -\mu & & & 0 \\ \mu & 0 & & & \\ & & 0 & -\nu & \\ & & \nu & 0 & \\ 0 & & & & \ddots \end{vmatrix}, \tag{1}$$

where $\mu, \nu, \cdots$ are real numbers. The same is then true of S, however with $\mu, \nu, \cdots$ now complex. Since S is unitary, we have

$$S' = O^{-1}SO = \begin{vmatrix} 0 & -e^{i\alpha} & & & 0 \\ e^{i\alpha} & 0 & & & \\ & & 0 & -e^{i\beta} & \\ & & e^{i\beta} & 0 & \\ 0 & & & & \ddots \end{vmatrix}$$

Finally, with the unitary matrix

* This work was supported in part by the Office of Naval Research under Contract No. Nonr 285-40, and in part by the Army Research Office (Durham) under Contract No. DA-ARO-(D)-31-124-G133.

[1] E. P. Wigner, *Group Theory and its Application* (Academic Press Inc., New York, 1959).

[2] J. G. Valatin, Phys. Rev. **122**, 1012 (1961).

[3] C. N. Yang, Revs. Modern Phys. (to be published).

[4] The transposed of a complex matrix M is denoted by $\tilde{M}$, the complex conjugate by M^*, the Hermitian adjoint by $M\dagger = \tilde{M}^*$.

[5] See, e.g., F. R. Gantmacher, *The Theory of Matrices* (Chelsea Publishing Company, New York), Vol. I, p. 293.

$$V = \tilde{V} = \begin{vmatrix} e^{-i\alpha/2} & & & & 0 \\ & e^{-i\alpha/2} & & & \\ & & e^{-i\beta/2} & & \\ 0 & & & e^{-i\beta/2} & \\ & & & & \ddots \end{vmatrix},$$

we have

$$VS'\tilde{V} = F.$$

This proves the lemma.

An analogous lemma is valid for symmetric matrices.

Lemma II. If S is unitary and symmetric, then

$$S = UE\tilde{U} = U\tilde{U},$$

where U is unitary and E is the unit matrix.

Proof. One can always write $S = \exp i\eta$ where η is symmetric (and real).[6] Then, with $U = \tilde{U} = \exp i\eta/2$, the proof of the lemma is completed.

We now proceed to prove the following two theorems.

Theorem I. If M is a complex skew symmetric matrix, then

$$M = UX\tilde{U}, \tag{2}$$

where U is unitary and X has the normal form (1) with $\mu, \nu, \cdots$ *non-negative real numbers.*

Theorem II. If M is a complex symmetric matrix, then (2) applies, but X is now diagonal, real, and non-negative.

Proof. (upper signs refer to Theorem I, lower signs to Theorem II). Consider the Hermitian matrix $H = M^\dagger M = \mp M^* M$. Introduce M_1 through $M = VM_1\tilde{V}$ with V unitary. Then

$$H = V^* M_1^\dagger V^\dagger V M_1 \tilde{V} = V^* H_1 \tilde{V},$$

with

$$H_1 = M_1^\dagger M_1 = \mp M_1^* M_1.$$

Furthermore,

$$MH - H^* M = M_1 H_1 - H_1^* M_1 = 0. \tag{3}$$

Now V^* is also unitary, and it can be chosen so that H_1 is diagonal, since H is Hermitian. Let h_κ be the eigenvalues of H (and H_1). Clearly $h_\kappa \geq 0$. Equation (3) becomes, for the matrix elements $m_{\kappa\sigma}$ of M_1,

$$(h_\sigma - h_\kappa) m_{\kappa\sigma} = 0.$$

This shows that $m_{\kappa\sigma} = 0$ unless $h_\kappa = h_\sigma$. The matrix M_1 breaks up into submatrices corresponding to groups of h_κ's which are equal.

Consider one such submatrix Ψ. Dropping the index κ, we can write

$$\Psi^\dagger \Psi = hI \qquad h \geq 0,$$

where I is the corresponding subunit matrix. Clearly

$$\tilde{\Psi} = \mp \Psi.$$

Now, if $h = 0$, then $\Psi = 0$. If $h \neq 0$, then $(1/h^{1/2})\Psi$ is unitary. From Lemmas I and II, one can perform a further transformation of the type (2) and arrive at the normal form F or E. This transforms Ψ into $(h)^{1/2}F$ or $(h)^{1/2}E$, respectively, and completes the proof of the theorems.

Finally, we prove the following theorem.

Theorem III. If M is any complex matrix, then

$$M = UXV \tag{4}$$

where U and V are unitary and X is diagonal, real, non-negative.

Proof. First assume that M has an inverse. Then $M^\dagger M$ is Hermitian and positive, and $(M^\dagger M)^{-1/2} M^\dagger$ is unitary. The matrix $M(M^\dagger M)^{-1/2} M^\dagger$ is Hermitian and positive, and one can choose U unitary such that $U^\dagger M(M^\dagger M)^{-1/2} M^\dagger U$ is diagonal, real, and positive. If we identify

$$V^\dagger = (M^\dagger M)^{-1/2} M^\dagger U,$$

this proves our theorem.

If M is singular, one must perturb it infinitesimally so that the perturbed M has an inverse. One can then go through the proof just given. When the perturbation goes to zero, (4) is still valid. However, some of the elements of the real diagonal positive matrix X can tend to zero with the perturbation.

As explained in the introduction, Dyson has pointed out that Theorems I, II, and III are different aspects of a single more general theorem, Theorem IV below. Let ϕ be a ground field which can be real, complex, or quaternion. In all three cases, the coefficients are assumed to be complex numbers. Therefore ϕ real means that the numbers in ϕ are ordinary complex numbers, ϕ complex that they are complex numbers with complex coefficients, ϕ quaternion that they are quaternions with complex coefficients. Consider matrices Q which have elements in ϕ. Define the adjoint Q^D to be Q transposed with each element ϕ conjugated, but *not* coefficient conjugated. Define the Hermitian adjoint $Q^\dagger$ to be Q transposed, with the elements ϕ conjugated *and* coefficient conjugated. One has the following theorem.

[6] See, e.g., reference 5, Vol. II, p. 4.

Theorem IV. If $Q = Q^D$ is self-adjoint, then

$$Q = RXR^D$$

where R is unitary and X is diagonal, real, scalar and non-negative.

As we show in the Appendix, this Theorem IV is equivalent to Theorems I–III proven above.

3. APPENDIX

We consider separately the three cases.

Case 1: ϕ Real.

In this case ϕ consists of ordinary complex numbers and $Q^D \equiv \tilde{Q}$. Theorem IV becomes identical with Theorem II.

Case 2: ϕ Complex.

In this case the numbers in ϕ are of the form $a + jb$ with a and b complex numbers and $j^2 = -1$. We have, for any matrix Q with coefficients in ϕ, $Q = M_1 + jM_2$ where M_1 and M_2 are ordinary complex matrices. Similarly, $R = U_1 + jU_2$. Furthermore,

$$Q^D = \tilde{M}_1 - j\tilde{M}_2, \qquad R^\dagger = U_1^\dagger - jU_2^\dagger.$$

Theorem IV states that, if $Q = Q^D$, then

$$M_1 + jM_2 = (U_1 + jU_2)X(U_1^\dagger - jU_2^\dagger),$$

where

$$(U_1 + jU_2)(U_1^\dagger - jU_2^\dagger) = (U_1^\dagger - jU_2^\dagger)(U_1 + jU_2) = 1,$$

and X is diagonal, real, and non-negative. This is easily seen to be equivalent to Theorem III, either directly or making use of the representation

$$a + jb = \begin{vmatrix} a + ib & 0 \\ 0 & a - ib \end{vmatrix},$$

where i is the ordinary imaginary unit. The correspondence with the quantities occurring in the statement of Theorem III is as follows:

$$\begin{cases} M = M_1 + iM_2 & U = U_1 + iU_2 \\ \tilde{M} = M_1 - iM_2 & V = U_1 - iU_2. \end{cases}$$

Case 3: ϕ Quaternion.

In this case the numbers in ϕ are of the form $a + a_1\tau_1 + a_2\tau_2 + a_3\tau_3$ where a, a_1, a_2, a_3 are ordinary complex numbers and the quaternion units can be represented as

$$\tau_1 = \begin{vmatrix} 0 & i \\ i & 0 \end{vmatrix}, \qquad \tau_2 = \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix}, \qquad \tau_3 = \begin{vmatrix} i & 0 \\ 0 & -i \end{vmatrix}.$$

We observe that any $2n \times 2n$ complex matrix M can be written as an $n \times n$ quaternion matrix Q. The transposed of M corresponds to the quaternion matrix $-\tau_2 Q^D \tau_2$ and the Hermitian conjugate of M to the quaternion matrix $(Q^D)^*$, where the * denotes ordinary coefficient conjugation.

In quaternion notation, our Theorem I takes the following form:

If $Q = -\tau_2 Q^D \tau_2$, then

$$Q = RX\tau_2(-\tau_2 R^D \tau_2)$$

where X is diagonal, real, scalar, non-negative, and R is the unitary quaternion matrix corresponding to the unitary complex matrix U. Introducing $Q' = -Q\tau_2$, we see that

$$(Q')^D = Q' \quad \text{and} \quad Q' = RXR^D.$$

This shows that case 3 is equivalent to Theorem I.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 3, NUMBER 5 SEPTEMBER-OCTOBER 1962

Erratum: Variational Method for the Quantum Statistics of Interacting Particles.

[J. Math. Phys. **3**, 131 (1962)]

M. GIRARDEAU
Enrico Fermi Institute for Nuclear Studies, The University of Chicago, Chicago, Illinois

---

EQUATION (54) contains a misprint. The minus sign in front of the k' summation in the first line should be changed to a plus sign; the curly bracket in the first line of Eq. (54) should be identical with that in the first line of Eq. (52).