JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 8, NUMBER 2 FEBRUARY 1967

Induced and Spontaneous Emission*

SAUL M. BERGMANN†
Research Center, American Optical Company, Southbridge, Massachusetts
(Received 17 March 1966)

The problem of induced and spontaneous emission is investigated for an atomic two-level system with incident beams of radiation which are either in a coherent state or in a stationary state (contain a definite number of photons). The treatment is fully quantum-mechanical, and is confined to the case where the frequency spectrum of the incident beam is narrow compared to the natural linewidth of the system. It is shown that, under such conditions, the spontaneous emission for frequencies within the narrow band of the incident radiation is sharply reduced compared to the prediction of the natural lineshape. It is shown that a hole is burned in the natural lineshape within the narrow frequency band, thus effectively quenching the spontaneous emission at some frequency within the band. This effect is shown to occur both for the coherent and stationary beams. Quantities proportional to the induced and spontaneous probability amplitudes and the lifetimes are computed for times comparable to and long compared to the free lifetime of the state. An expression is found for the spectrum of the emergent radiation in terms of these quantities. Its physical meaning is briefly discussed. The density operator of the field for all times is given.

I. INTRODUCTION

INDUCED and spontaneous emission have been recognized as being the principal processes occurring in the operation of lasers, as has been shown by Schawlow and Townes.[1] These processes have been identified as first-order effects in ordinary perturbation theory.[2] On the other hand, spontaneous emission, as treated by Weisskopf and Wigner,[3] enabled them to find this emission as a function of frequency, and thus to determine the line shape. Their treatment differs from ordinary perturbation theory in that they invariably retain a unitary matrix $V(t)$ as will be shown in Sec. VIII. In ordinary perturbation theory (as is well known) where one expands the operator U in a power series and breaks off the series at the nth term, the operator U, as expressed by the first n terms, is no longer strictly unitary. In this paper, we wish to extend Weisskopf–Wigner's work of a two-level atomic system to the case where some incident radiation is present. We retain a unitary matrix throughout, and be able to obtain expressions as a function of frequency for both the induced and the spontaneous emissions. We choose a monochromatic beam in a coherent state discussed by Glauber and others[4-6] for one case of incident radiation and a beam in a stationary state (with a definite number of photons) for a wave packet which is narrow compared to the natural linewidth, for the other case to be treated. In both instances, it is found that a dip is occurring for the spontaneous emission as a function of frequency whose bottom, to a good

* This paper was sponsored by the U. S. Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract No. AF 19(628)-4156.
† Present address: Clarendon Laboratory, Oxford, England.
[1] A. L. Schawlow and C. H. Townes, Phys. Rev. 112, 1940 (1958).
[2] W. E. Lamb, Jr., in *Advances in Quantum Electronics*, J. Singer Ed. (Columbia University Press, New York, 1961), p. 370.
[3] V. F. Weisskopf and E. P. Wigner, Z. Physik 63, 54 (1930).
[4] R. J. Glauber, Phys. Rev. 131, 2766 (1963).
[5] S. S. Schweber, J. Math. Phys. 3, 831 (1962).
[6] V. Bargmann, Commun. Pure Appl. Math. 14, 187 (1961).

approximation, quenches the spontaneous emission at that frequency.

By an inducing field which is coherent for all modes k, we mean that it is of the form

$$|\{y_k\}\rangle = \prod_k \exp\left(-\tfrac{1}{2}|y_k|^2 + y_k a_k^\dagger\right)|0\rangle, \quad (1.1)$$

where $|0\rangle$ represents the state empty of all photons.

We are also interested in inducing fields of the form

$$|\{n_k\}\rangle = \prod_k \frac{a_k^{\dagger n_k}}{(n_k!)^{\frac{1}{2}}}|0\rangle, \quad (1.2)$$

which are the stationary states of the noninteracting field normally dealt with in quantum electrodynamics.

The atom is assumed to be in an excited state at $t = 0$. Since we are interested in the description of its decay, the equations of motion in the interaction picture are given,

$$i\hbar \frac{\partial}{\partial t} U(t) = H_1(t)U(t) \quad (1.3)$$

with the subsidiary condition $U(0) = 1$. (1.4)

The system is assumed to be in the state $|t = 0\rangle$ at time $t = 0$, and evolves with time into the state $|t\rangle$, governed by the unitary operator $U(t)$, thus

$$|t\rangle = U(t)\,|t = 0\rangle. \quad (1.5)$$

In Secs. II and III, we discuss the interaction Hamiltonian and the equations of motions, respectively. In Sec. IV, we discuss the Weisskopf–Wigner case expressed in our language. In Sec. V, we set up the integral equations governing the probability amplitude $a(t)$ for the system to remain in its original state when the incident beam is coherent. In Sec. VI, we discuss the stationary case. In Sec. VII, we discuss the meaning of the Weisskopf–Wigner, WW (no incident radiation present) and the Weisskopf–Wigner-like, WW-like (incident radiation present) approximations. In Secs. VIII and IX, we present the solutions of the integral equations set up in Secs. V and VI, respectively, and also present expressions for the quantities of physical interest. In Sec. X, we present a summary.

II. THE INTERACTION HAMILTONIAN

We choose as interaction Hamiltonian the form

$$H_1(t) = -\frac{1}{c}\int \mathbf{j}(\mathbf{r}, t)\cdot\mathbf{A}(\mathbf{r}, t)\,d\mathbf{r}, \quad (2.1)$$

where $\mathbf{A}$ is the transverse part of the vector potential which can be expanded as

$$\mathbf{A}(\mathbf{r}, t) = c\sum_k \left(\frac{\hbar}{2\omega_k}\right)^{\frac{1}{2}}(a_k\mathbf{u}_k(\mathbf{r})e^{-i\omega_k t} + a_k^\dagger\mathbf{u}_k^*(\mathbf{r})e^{i\omega_k t}). \quad (2.2)$$

The mode functions $\mathbf{u}_k(\mathbf{r})$ obey the orthonormality condition

$$\int \mathbf{u}_k^*(\mathbf{r})\cdot\mathbf{u}_l(\mathbf{r})\,d\mathbf{r} = \delta_{kl} \quad (2.3)$$

and the transversality condition

$$\nabla\cdot\mathbf{u}_k(\mathbf{r}) = 0. \quad (2.4a)$$

Subsequently, we are mostly concerned with plane-wave mode functions appropriate to a cubical volume V, such that Eq. (22) becomes

$$\mathbf{A}(\mathbf{r}, t) = c\sum_{\substack{k\\ \lambda=1,2}} \left(\frac{\hbar}{2\omega_k V}\right)^{\frac{1}{2}}\hat{\mathbf{e}}_k^{(\lambda)}[a_k^{(\lambda)}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)} + a_k^{\dagger(\lambda)}e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega_k t)}], \quad (2.4b)$$

where $\hat{\mathbf{e}}_k^{(\lambda)}$ is a unit polarization vector such that $\hat{\mathbf{e}}_k^{(1)}$, $\hat{\mathbf{e}}_k^{(2)}$, and $\mathbf{k}$ form a right-handed set, so that $\mathbf{k}\cdot\hat{\mathbf{e}}_k^{(\lambda)}$ and Eq. (2.4) is satisfied. a_k and $a_k^\dagger$ are the usual photon annihilation and creation operators.

The current density operator $\mathbf{j}(\mathbf{r}, t)$ can be written as

$$\mathbf{j}(\mathbf{r}, t) = ec\Psi^\dagger(\mathbf{r}, t)\boldsymbol{\alpha}\Psi(\mathbf{r}, t), \quad (2.5)$$

where $\Psi(\mathbf{r}, t)$ is the Dirac electron field operator and $\Psi^\dagger(\mathbf{r}, t)$ its Hermitian adjoint. There is a complete set of orthonormal energy eigenfunctions $w_i(n, \mathbf{r})$ and energy eigenvalues E_n for a single electron in a central field. The orthonormality condition is given by

$$\int \sum_j w_j^*(n, \mathbf{r})w_j(n', \mathbf{r})\,d\tau = \delta_{nn'}, \quad (2.6)$$

where the index j denotes the components of the Dirac spinor, and * stands for complex conjugate. We can now expand each component of the Dirac spinor electron field operator in terms of the eigenfunctions $w_i(n, \mathbf{r})$ thus

$$\Psi_i(\mathbf{r}, t) = \sum_n b_n e^{-i(E_n/\hbar)t}w_i(n, \mathbf{r}),$$
$$\Psi^\dagger(\mathbf{r}, t) = \sum_n b_n^\dagger e^{i(E_n/\hbar)t}w_i^*(n, r), \quad (2.7)$$

where the summation n extends over all energy states. The operators b_n and $b_n^\dagger$ are the usual electron destruction and creation operators. Equation (2.1) can be rewritten as

$$H_1(t) = -e\int \Psi^\dagger(\mathbf{r}, t)\boldsymbol{\alpha}\cdot\mathbf{A}(\mathbf{r}, t)\Psi(\mathbf{r}, t)\,d\mathbf{r}. \quad (2.8)$$

III. EQUATIONS OF MOTION

Assuming no diagonal matrix elements $\langle e|\, H_1(t)\, |e\rangle$ exist, Eq. (1.3) becomes

$$i\hbar \frac{\partial}{\partial t} \langle e|\, U(t)\, |e\rangle = \langle e|\, H_1(t)\, |g\rangle\langle g|\, U(t)\, |e\rangle, \tag{3.1a}$$

$$i\hbar \frac{\partial}{\partial t} \langle g|\, U(t)\, |e\rangle = \langle g|\, H_1(t)\, |e\rangle\langle e|\, U(t)\, |e\rangle. \tag{3.1b}$$

Satisfying Eq. (1.4), we integrate Eq. (3.1b) with respect to time,

$$\langle g|\, U(t)\, |e\rangle \equiv C(t) = \frac{1}{i\hbar} \int_0^t \langle g|\, H(t')\, |e\rangle\langle e|\, U(t')\, |e\rangle\, dt'. \tag{3.2}$$

Substituting Eq. (3.2) into (3.1a) and putting $B(t) \equiv \langle e|\, U(t)\, |e\rangle$, we get an integral equation in the radiation field operators only,

$$\begin{aligned} \frac{\partial}{\partial t} B(t) = -\frac{1}{\hbar^2 c^2} \int_0^t dt' \int d^3r \int d^3r' \\ \times \langle e|\, \mathbf{j}(\mathbf{r}, t)\, |g\rangle \cdot \mathbf{A}(\mathbf{r}, t) \\ \times \langle g|\, \mathbf{j}(\mathbf{r}', t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}', t') B(t'). \end{aligned} \tag{3.3}$$

Given the state vector $|f\rangle$ of the initial radiation field incident upon the atom, the state vector of the system at time t is

$$\begin{aligned} |t\rangle &= U(t)\, |e, f\rangle \\ &= \langle e|\, U(t)\, |e\rangle\, |e, f\rangle + \langle g|\, U(t)\, |e\rangle\, |g, f\rangle \\ &= B(t)\, |e, f\rangle + C(t)\, |g, f\rangle \\ &= B(t)\, |e, f\rangle - \frac{1}{i\hbar c} \int_0^t dt' \int d^3r \\ &\quad \times \langle g|\, \mathbf{j}(\mathbf{r}, t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}, t') B(t')\, |g, f\rangle. \end{aligned} \tag{3.4}$$

Normalization Condition

We wish to check the normalization condition requiring $U(t)$ to be unitary. This yields, using Eq. (3.4),

$$1 = U^\dagger(t)U(t) = B^\dagger(t)B(t) + C^\dagger(t)C(t). \tag{3.5}$$

To check the correctness of this equation for all times t, we first note that it is satisfied for $t = 0$. Therefore, if its differential with time is zero, Eq. (3.5) is satisfied.

By differentiating Eq. (3.5) with respect to time and equating it to zero, we are brought back to Eq. (3.3) for $B(t)$ and its Hermitian conjugate for $B^\dagger(t)$.

The probability of the atom being in the excited state is given by

$$W_e = \langle f, e|\, B^\dagger(t)B(t)\, |e, f\rangle. \tag{3.6}$$

The probability of the atom being in the ground state is given by

$$W_g = \langle f, g|\, C^\dagger(t)C(t)\, |g, f\rangle. \tag{3.7}$$

IV. SPONTANEOUS EMISSION

We assume first that there is no external radiation field present. In this case, the probability amplitude $\langle 0|\, B(t)\, |0\rangle$ for the atom remaining in the upper state is governed by Eq. (3.3) for the matrix element $\langle 0|\, B(t)\, |0\rangle$

$$\begin{aligned} \frac{\partial}{\partial t} \langle 0|\, B(t)\, |0\rangle = \\ -\frac{1}{\hbar^2 c^2} \int_0^t dt' \int d^3r \int d^3r' \\ \times \langle 0|\, [\langle e|\, \mathbf{j}(\mathbf{r}, t)\, |g\rangle \cdot A(r, t)] \\ \times [\langle g|\, j(\mathbf{r}', t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}', t')] B(t')\, |0\rangle. \end{aligned} \tag{4.1}$$

We now make the assumption that, in the radiation field operator product on the right-hand side of Eq. (4.1), we only retain the matrix element $\langle 0|\, B(t')\, |0\rangle$. Equation (4.1) then becomes

$$\begin{aligned} \frac{\partial}{\partial t} \langle 0|\, B(t)\, |0\rangle = -\frac{1}{\hbar^2 c^2} \int_0^t dt' \int d^3r \int d^3r' \\ \times \langle 0|\, [\langle e|\, \mathbf{j}(\mathbf{r}, t)\, |g\rangle \cdot \mathbf{A}(\mathbf{r}, t)] \\ \times [\langle g|\, \mathbf{j}(\mathbf{r}', t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}', t')]\, |0\rangle\langle 0|\, B(t')\, |0\rangle. \end{aligned} \tag{4.2}$$

Equation (4.2) is the Weisskopf–Wigner (WW) equation for the probability amplitude $a(t) \equiv \langle 0|\, B(t)\, |0\rangle$ of the atom remaining in the excited state.

Evaluating the matrix elements of the kernel of Eq. (4.2) [using Eqs. (2.8), (2.7a), and (2.7b)], replacing the Dirac single-electron wavefunctions $w_i(n, \mathbf{r})$ by the Schrödinger wavefunctions [$\Psi_g(\mathbf{r})$ and $\Psi_e(\mathbf{r})$ for the ground state and excited state respectively of the atomic system], and substituting the operator $c\boldsymbol{\alpha}$ by the operator $\mathbf{p}/m = -(i\hbar/m)$ **grad**, we get for Eq. (4.2)

$$\begin{aligned} \frac{\partial}{\partial t} a(t) = -\frac{1}{\hbar^2} \sum_{\lambda=1,2,k} \frac{\hbar}{2\omega_k V} |c_{k\lambda}|^2 \\ \times \int_0^t dt'\, e^{-i(\omega_k - \omega_0)(t-t')} a(t'), \end{aligned} \tag{4.3}$$

where

$$c_{k\lambda} = -\frac{ie\hbar}{m} \int d^3r\, \Psi_g^*(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} \mathbf{e}_k^{(\lambda)} \cdot \mathbf{grad}\, \Psi_e(\mathbf{r}), \tag{4.4}$$

and $\hbar\omega_0$ is the energy separation between the excited and ground states of the atom.

V. COHERENT INDUCING FIELD

We assume that the initial radiation field is of the form given by Eq. (1.1). We use again Eq. (3.3) in order to calculate the probability amplitude

$$a(t) \equiv \langle \{y_k\} | B(t) | \{y_k\} \rangle$$

for the atomic system remaining in the excited state. We thus have

$$\frac{\partial}{\partial t} \langle \{y_k\} | B(t) | \{y_k\} \rangle$$
$$= -\frac{1}{\hbar^2 c^2} \int_0^t dt' \int d^3r \int d^3r' \langle \{y_k\} |$$
$$\times [\langle e | \mathbf{j}(r, t) | g \rangle \cdot \mathbf{A}(\mathbf{r}, t)]$$
$$\times [\langle g | \mathbf{j}(\mathbf{r}', t') | e \rangle \cdot \mathbf{A}(\mathbf{r}', t')] B(t') | \{y_k\} \rangle. \quad (5.1)$$

We now wish to make a Weisskopf–Wigner-like (WW-like) approximation to Eq. (5.1), which consists in retaining only the matrix element $\langle \{y_k\} | B(t') | \{y_k\} \rangle$ in the operator product of the radiation operators on the right-hand side of Eq. (5.1). This then yields the integral equation

$$\frac{\partial}{\partial t} \langle \{y_k\} | B(t) | \{y_k\} \rangle$$
$$= -\frac{1}{\hbar^2 c^2} \int_0^t dt' \int d^3r \int d^3r' \langle \{y_k\} |$$
$$\times [\langle e | \mathbf{j}(\mathbf{r}, t) | g \rangle \cdot \mathbf{A}(\mathbf{r}, t)]$$
$$\times [\langle g | \mathbf{j}(\mathbf{r}', t') | e \rangle \cdot \mathbf{A}(\mathbf{r}', t')]$$
$$\times | \{y_k\} \rangle \langle \{y_k\} | B(t') | \{y_k\} \rangle. \quad (5.2)$$

The evaluation of the matrix element of the vector potential operators becomes

$$\langle \{y_k\} | \mathbf{A}(\mathbf{r}, t) \mathbf{A}(\mathbf{r}', t') | \{y_k\} \rangle$$
$$= \langle 0 | [\mathbf{A}(\mathbf{r}, t) + \mathbf{f}(\mathbf{r}, t)][\mathbf{A}(\mathbf{r}', t') + \mathbf{f}(\mathbf{r}', t')] | 0 \rangle$$
$$= \mathbf{f}(\mathbf{r}, t)\mathbf{f}(\mathbf{r}', t') + \langle 0 | \mathbf{A}(\mathbf{r}, t)\mathbf{A}(\mathbf{r}', t') | 0 \rangle \quad (5.3)$$

as a result of a well-known displacement property[4] of the state $|\{y_k\}\rangle$, with

$$\mathbf{f}(\mathbf{r}, t) \equiv c \sum_k \left(\frac{\hbar}{2\omega_k V}\right)^{\frac{1}{2}} \hat{\mathbf{e}}_k^{(\lambda)}$$
$$\times [y_k^{(\lambda)} e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_k t)} + y_k^{*(\lambda)} e^{-i(\mathbf{k}\cdot\mathbf{r} - \omega_k t)}]. \quad (5.4)$$

Using Eq. (5.3), evaluating all the other matrix elements of the kernel in Eq. (5.2), and going over to the nonrelativistic limit, we can, as with the previous spontaneous case, express Eq. (5.2) more explicitly as

$$\frac{\partial}{\partial t} a(t) = -\frac{1}{\hbar^2} \sum_{k_1} \left(\frac{\hbar}{2\omega_{k_1} V}\right) |c_{k_1\lambda}|^2$$
$$\times e^{-i(\omega_{k_1} - \omega_0)t} b_1(\omega_{k_1}, t)$$
$$- \frac{1}{\hbar} e^{i\omega_0 t} g^*(t) P_i, \quad (5.5)$$

where

$$d_{k\lambda} = -\frac{ie\hbar}{m} \int d^3r \, \Psi_g^*(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{e}_k^{(\lambda)} \cdot \mathbf{grad} \, \Psi_e(r), \quad (5.6)$$

$$g(t) = \sum_{k,\lambda} \left(\frac{\hbar}{2\omega_k V}\right)^{\frac{1}{2}} [c_{k\lambda} y_{k\lambda}^* e^{i\omega_k t} + d_{k\lambda} y_{k\lambda} e^{-i\omega_k t}], \quad (5.7)$$

$$b_1(\omega_{k_1}, t) = \int_0^t a(\tau) e^{-i(\omega_0 - \omega_{k_1})\tau} \, d\tau, \quad (5.8)$$

$$b_2(t) = \hbar P_i = \int_0^t g(\tau) a(\tau) e^{-i\omega_0 \tau} \, d\tau. \quad (5.9)$$

The summation over k_1 in Eq. (5.5) is to be carried out over all the modes of the field.

We shall invoke throughout the WW-like approximation and calculate other quantities of interest. From Eq. (3.7), we can calculate W_g, the probability of the atom being in the ground state

$$W_g = \frac{1}{\hbar^2 c^2} \int_0^t dt'' \int_0^t dt' \int d^3r' \int d^3r$$
$$\times \langle \{y_k\} | B^\dagger(t'') | \{y_k\} \rangle \langle \{y_k\} |$$
$$\times [\langle e | \mathbf{j}(\mathbf{r}', t'') | g \rangle \cdot \mathbf{A}(\mathbf{r}', t'')]$$
$$\times [\langle g | \mathbf{j}(\mathbf{r}, t') | e \rangle \cdot \mathbf{A}(\mathbf{r}, t')]$$
$$\times | \{y_k\} \rangle \langle \{y_k\} | B(t') | \{y_k\} \rangle, \quad (5.10)$$

which, after the evaluation of the matrix elements as before, becomes,

$$W_g = \frac{1}{\hbar^2} \sum_{k,\lambda} \frac{\hbar}{2\omega_k V} |c_{k\lambda}|^2 |b_1(\omega_k, t)|^2 + \frac{1}{\hbar^2} |b_2(t)|^2$$
$$= W_s + |P_i|^2, \quad (5.11)$$

where W_s represents the spontaneous emission transition probability and is given by the first term on the right-hand side of Eq. (5.11), and where P_i, defined in Eq. (5.9), represents the probability amplitude for induced emission. In a similar vein, the normalization condition given by Eq. (3.5) becomes

$$1 = \langle \{y_k\}, e | U^\dagger(t) U(t) | e, \{y_k\} \rangle$$
$$= |a(t)|^2 + W_s + |P_i|^2 = |a(t)|^2 + W_g. \quad (5.12)$$

For ease of notation, we henceforth omit the polarization index λ.

We wish to calculate a quantity that provides a measure of the frequency distribution of the outgoing radiation as a function of the frequency dis-

tribution of the incident radiation. To this effect, we calculate the expectation value of the operator $a_{k'}^{\dagger} a_{k'}$ for the state given by Eq. (3.4). This yields a quantity proportional to the expectation value of the energy in the mode k'. If applied to the spontaneous case, it simply yields the distribution of the natural linewidth in the WW theory. For the coherent case dealt with in this section, we have

$$\langle t|\, a_{k'}^{\dagger} a_{k'}\, |t\rangle = \langle \{y_k\}|\, B^{\dagger}(t) a_{k'}^{\dagger} a_{k'} B(t)\, |\{y_k\}\rangle + \langle \{y_k\}|\, C^{\dagger}(t) a_{k'}^{\dagger} a_{k'} C(t)\, |\{y_k\}\rangle. \tag{5.13}$$

After some laborious but routine manipulations in the evaluation of the matrix elements,[7] we get for Eq. (5.13)

$$\langle t|\, a_{k'}^{\dagger} a_{k'}\, |t\rangle = |y_{k'}|^2 + 2\,\mathrm{Re}\,(y_{k'} P_i P_{sk'}^{*}) + |P_{sk'}|^2, \tag{5.14}$$

where Re stands for "real value of" and where a probability amplitude $P_{sk'}$ for the spontaneous emission of a photon k'

$$P_{sk'} = \frac{1}{\hbar}\left(\frac{\hbar}{2\omega_{k'} V}\right)^{\frac{1}{2}} c_{k'} b_1(\omega_{k'}, t) \tag{5.15}$$

has been introduced. The first term on the right-hand side of Eq. (5.14) is a quantity proportional to the expectation value of the energy in mode k' of the incident radiation. The second term is an interference term which involves $y_{k'}$, the amplitude of the incident radiation of wavenumber k', P_i the probability amplitude for induced emission, and $P_{sk'}$ the probability amplitude for spontaneous emission of a photon k'. The last term can be interpreted as the transition probability for the emission of a photon k'.

VI. STATIONARY INDUCING FIELD

In this section, we deal with an initial radiation field of the form given by Eq. (1.2). We wish to transcribe Eqs. (5.2) and (5.5), Eqs. (5.13) and (5.14), Eqs. (5.10) and (5.11), and Eq. (5.12) with the appropriate matrix elements presently of interest. Instead of Eq. (5.2), we now have

$$\frac{\partial}{\partial t}\langle \{n_k\}|\, B(t)\, |\{n_k\}\rangle = -\frac{1}{\hbar^2 c^2}\int_0^t dt' \int d^3r \int d^3r' \times \langle \{n_k\}|\, [\langle e|\, \mathbf{j}(\mathbf{r}, t)\, |g\rangle \cdot \mathbf{A}(\mathbf{r}, t)] \times [\langle g|\, \mathbf{j}(\mathbf{r}', t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}', t')] \times |\{n_k\}\rangle\langle \{n_k\}|\, B(t')\, |\{n_k\}\rangle, \tag{6.1}$$

[7] S. M. Bergmann, "Theory of Induced and Spontaneous Emission," Final Report under Contract AF 19(628)-4156 AFCRL 65-505 AD 622-100 (1964-1965).

where again we have made the WW-like approximation and retained only matrix elements $a(t) \equiv \langle \{n_k\}|\, B(t)\, |\{n_k\}\rangle$.

The equation analogous to Eq. (5.5) becomes

$$\frac{\partial}{\partial t} a(t) = -\frac{1}{\hbar^2}\sum_k \frac{\hbar}{2\omega_k V} n_k\, |c_k|^2\, e^{-i(\omega_k - \omega_0)t} b_1(\omega_k, t) - \frac{1}{\hbar^2}\sum_k \frac{\hbar}{2\omega_k V} n_k\, |d_k|^2\, e^{i(\omega_k + \omega_0)t} b_2(\omega_k, t) - \frac{1}{\hbar^2}\sum_{k_1} \frac{\hbar}{2\omega_{k_1} V} |c_{k_1}|^2\, e^{-i(\omega_{k_1} - \omega_0)t} b_1(\omega_{k_1}, t), \tag{6.2}$$

where $b_1(\omega_k, t)$ is as defined by Eq. (5.8) and, in this section and Sec. IX,

$$b_2(\omega_k, t) = \int_0^t e^{-i(\omega + \omega_0)\tau} a(\tau)\, d\tau. \tag{6.3}$$

The summation over k_1 in Eq. (6.2) is to be carried out over all the modes of the field.

Similarly, instead of Eq. (5.10), we have

$$W_s = \frac{1}{\hbar^2 c^2}\int_0^t dt'' \int_0^t dt' \int d^3r' \int d^3r \times \langle \{n_k\}|\, B^{\dagger}(t'')\, |\{n_k\}\rangle\langle \{n_k\}| \times [\langle e|\, \mathbf{j}(\mathbf{r}', t'')\, |g\rangle \cdot \mathbf{A}(\mathbf{r}', t'')] \times [\langle g|\, \mathbf{j}(\mathbf{r}, t')\, |e\rangle \cdot \mathbf{A}(\mathbf{r}, t')]\, |\{n_k\}\rangle\langle \{n_k\}|\, B(t')\, |\{n_k\}\rangle, \tag{6.4}$$

which finally leads to[7]:

$$W_s = \frac{1}{\hbar^2}\sum_k \frac{\hbar}{2\omega_k V} n_k\, |c_k|^2\, |b_1(\omega_k, t)|^2 + \frac{1}{\hbar^2}\sum_k \frac{\hbar}{2\omega_k V} n_k\, |d_k|^2\, |b_2(\omega_k, t)|^2 + \frac{1}{\hbar^2}\sum_{k_1} \frac{\hbar}{2\omega_{k_1} V} |c_{k_1}|^2\, |b_1(\omega_{k_1}, t)|^2. \tag{6.5}$$

The first term on the right-hand side is the transition probability for induced emission, the second term is the transition probability for absorption, and the third term is the transition probability for spontaneous emission. The normalization condition analogous to Eq. (5.12) becomes

$$1 = |a(t)|^2 + W_s \tag{6.6}$$

with $a(t)$ and W_s as given in this section. We now wish to calculate

$$\langle t|\, a_{k'}^{\dagger} a_{k'}\, |t\rangle = \langle \{n_k\}|\, B^{\dagger}(t) a_{k'}^{\dagger} a_{k'} B(t)\, |\{n_k\}\rangle + \langle \{n_k\}|\, C^{\dagger}(t) a_{k'}^{\dagger} a_{k'} C(t)\, |\{n_k\}\rangle. \tag{6.7}$$

After some straightforward computation,[7] Eq. (6.7) becomes

$$\langle t|\, a_{k'}^{\dagger} a_{k'}\, |t\rangle = n_{k'} + |b_{k'}(t)_{\text{em}}|^2 - |b_{k'}(t')_{\text{abs}}|^2, \quad (6.8)$$

where

$$|b_{k'}(t)_{\text{em}}|^2 = \frac{1}{\hbar^2}\frac{\hbar}{2\omega_{k'}V}(n_{k'}+1)\,|c_{k'}|^2\,|b_1(\omega_{k'}, t)|^2 \quad (6.9)$$

and

$$|b_{k'}(t)_{\text{abs}}|^2 = \frac{1}{\hbar^2}\frac{\hbar}{2\omega_{k'}V} n_{k'}\,|d_{k'}|^2\,|b_2(\omega_{k'}, t)|^2 \quad (6.10)$$

are the transition probabilities, respectively, of emission and absorption of a single photon k'. Their explicit evaluation is dependent on $b_1(\omega_{k'}, t)$ and $b_2(\omega_{k'}, t)$, which in turn are dependent on the solution of the integral equation (6.2). This is attempted in Sec. IX. The first term on the right-hand side of Eq. (6.8) is the number of photons of wave-number k' in the incident beam. As to the remaining two terms on the right-hand side of Eq. (6.8), a comparison with Eq. (5.14) indicates that here we do not have an interference term.

VII. THE WW AND WW-LIKE APPROXIMATIONS

In the WW and WW-like approximations made in the previous sections, we retained only the diagonal matrix element $\langle i|\, B\, |i\rangle \equiv \langle i, e|\, U\, |e, i\rangle$, where $|i\rangle$ is the initial state of the radiation field considered in Secs. IV, V, and VI. This means that, for probability amplitudes for which the atomic system is in the excited state, we are ignoring effects with finite $\langle f, e|\, U\, |e, i\rangle$ for $\langle f| \neq \langle i|$. By so doing, we are describing induced and spontaneous emission. The description is indicated by the integral equations (5.2) and (5.5) for the coherent case, and by the integral equations (6.1) and (6.2) for the stationary case. A glance at the right-hand side of Eq. (6.2) indicates that there is a competition between three terms, the first being induced emission, the second absorption from the exciting beam, and the third representing spontaneous emission. This last term is the only one present in the WW formulation as given by Eq. (4.3), and leads to the natural line shape through a quantity proportional to $b_1(\omega_{k_1}, t)$. But, because of the presence of the other two terms in Eq. (6.2), it is unlikely that this term still indicates an undistorted line shape for $t \to \infty$. In fact, we may anticipate our result and state that a dip occurs at the inducing frequency, if it falls within the linewidth, as indicated by Eq. (9.22). Similar considerations apply to the integral equations, (5.2) and (5.5), pertaining to the coherent case. In Eq. (5.5), the first term on the right-hand side denotes spontaneous emission. We interpret the second term on the right-hand side as induced emission.

Another way of looking at the WW-like approximation is to say that we replace the operator B wherever it occurs in the previous sections by $|i\rangle\langle i|\, B$. By making this substitution, we hereby indicate which part of the total operator B is retained in the WW and WW-like approximations. In the particular case of $|i\rangle = |0\rangle$, we have to make the substitution $B \to |0\rangle\langle 0|\, B$ which corresponds to the WW approximation. We wish to make the substitution $B \to |i\rangle\langle i|\, B$ in Eq. (3.4). Defining a new unitary operator $V(t)$, we have for Eq. (3.4)

$$\begin{aligned}|t\rangle &= V(t)\,|e, i\rangle \\ &= |i\rangle\langle i|\, B(t)\,|e, i\rangle - \frac{1}{i\hbar c}\int_0^t dt' \int d^3r \\ &\quad \times \langle g|\, \mathbf{j}(\mathbf{r}, t')\,|e\rangle\cdot\mathbf{A}(r, t')\,|i\rangle\langle i|\, B(t')\,|g, i\rangle \\ &\equiv |i\rangle\langle i|\, B(t)\,|e, i\rangle \\ &\quad + \int_0^t dt'\, D(t')\,|i\rangle\langle i|\, B(t')\,|g, i\rangle. \end{aligned} \quad (7.1)$$

The proof of the normalization condition

$$\begin{aligned}1 = V^{\dagger}(t)V(t) &= B^{\dagger}(t)\,|i\rangle\langle i|\, B(t) \\ &\quad + \int_0^t dt' \int_0^t dt''\, B^{\dagger}(t'')\,|i\rangle \\ &\quad \times \langle i|\, D^{\dagger}(t'')D(t')\,|i\rangle\langle i|\, B(t') \end{aligned} \quad (7.2)$$

corresponding to Eq. (3.5) proceeds as before, except that we are now brought back to Eq. (3.3) with $B(t) \to |i\rangle\langle i|\, B(t)$.

Reducing the density operator $|t\rangle\langle t|$ as constructed from Eq. (7.1), we find, for the density operator $\rho(t)$ of the field alone,

$$\begin{aligned}\rho(t) &= |\langle i|\, B(t)\,|i\rangle|^2\,|i\rangle\langle i| \\ &\quad + \int_0^t dt' \int_0^t dt''\,\langle i|\, B(t')\,|i\rangle D(t')\,|i\rangle \\ &\quad \times \langle i|\, D^{\dagger}(t'')\langle i|\, B^{\dagger}(t'')\,|i\rangle. \end{aligned} \quad (7.3)$$

VIII. SOLUTION OF THE INTEGRAL EQUATION (5.5)

Quantities of Physical Interest

A solution of Eq. (5.5) would yield an expression for $a(t)$ and the quantities $b_1(\omega_k, t)$ and $b_2(t)$. They occur in Eq. (5.11) for W_{σ}, in Eq. (5.15) for $P_{\sigma k'}$, and in Eq. (5.9) for P_i. These quantities evaluated

for all time and especially for $t \to \infty$ are of physical interest.

We can rewrite Eq. (5.5) in the form

$$\frac{\partial}{\partial t} a(t) + \frac{1}{2\pi} \int_0^\infty f\left(\frac{\omega}{2\pi}\right) e^{i(\omega_0 - \omega)t} b_1(\omega, t)\, d\omega + G e^{i\omega_0 t} b_2(t) = 0, \quad (8.1a)$$

which is to be solved with the subsidiary condition

$$a(0) = 1, \quad (8.1b)$$

where

$$G(t) \equiv \hbar^{-1} g^*(t) \quad (8.2a)$$

and the substitution

$$\sum_k \to \frac{V}{(2\pi)^3} \int d^3k \to \int_0^\infty f\left(\frac{\omega}{2\pi}\right) d\omega \quad (8.2b)$$

has been made.

Introduction of Dimensionless Quantities

The integral equation (8.1) can be solved by using asymptotic techniques.[7] It is convenient to introduce dimensionless and scaled variables and determine the small parameters. Our final results are also to be expressed in this form. Let

$$\bar{t} = \omega_0 t, \quad (8.3a) \qquad G = \epsilon\omega_0 \bar{G}, \quad (8.3d)$$

$$\bar{\omega} = \omega/\omega_0, \quad (8.3b) \qquad B_1 = \omega_0 b_1, \quad (8.3e)$$

$$f(\omega/2\pi) = 2\lambda\omega_0 \bar{f}(\bar{\omega}), \quad (8.3c) \qquad \epsilon B_2 = b_2. \quad (8.3f)$$

We have defined the new quantities given by Eqs. (8.3a)–(8.3f). We can now, without danger of ambiguity, drop all bars and rewrite Eqs. (8.1), (5.8), and (5.9) in dimensionless form.

Dimensionless Equations

Equation (8.1) becomes in dimensionless form

$$\frac{\partial}{\partial t} a + \frac{\lambda}{\pi} \int_0^\infty f(\omega) B_1(\omega, t) e^{i(1-\omega)t}\, d\omega + \epsilon^2 G e^{it} B_2(t) = 0 \quad (8.4)$$

with

$$B_1(\omega, t) = \int_0^t a(\tau) e^{-i(1-\omega)\tau}\, d\tau, \quad (8.5a)$$

$$B_2(t) = \int_0^t G^*(\tau) a(\tau) e^{-i\tau}\, d\tau. \quad (8.5b)$$

The Small Parameters λ and ϵ

We choose λ so as to make $f_{\max} = 0(1)$ on the right-hand side of Eq. (8.3c).

A glance of Eq. (8.4) indicates that, for $\epsilon = 0$ (that is, when no external radiation is present), Eq. (8.4) reduces to the WW equation for a and corresponds to spontaneous emission. It is then easily seen that $f(\omega_0/2\pi)$ in Eq. (8.3c) is the natural line width and is the transition probability per unit time for spontaneous emission by ordinary perturbation theory. It is of the order of 10^8 sec^{-1} and very small compared to ω_0 at optical frequencies. Thus, the parameter λ will always be small. The factor of 2 in Eq. (8.3c) is introduced in order to facilitate notation of subsequent expressions.

As for ϵ, we choose it to make $|G|_{\max} = 0(1)$ on the right-hand side of Eq. (8.3d). In the analysis of this section, we are only concerned with the case of a single monochromatic incident wave. Thus, we choose a single term out of the summation in Eq. (5.7).

We have, for the expectation value of the number of photon operator,

$$|y_k|^2 = \langle y_k |\, a_k^\dagger a_k \,| y_k \rangle. \quad (8.6)$$

We wish to specify the single monochromatic wave, such that

$$\lim_{V\to\infty} \frac{\hbar}{2\omega_k V} |y_k|^2 = b \neq 0. \quad (8.7)$$

Thus, in nondimensional quantities for a single monochromatic wave, we have

$$G(t) = \beta_{-1} e^{-i\Omega t} + \beta_1 e^{i\Omega t}, \quad (8.8)$$

where $|\beta_{-1}|$ and $|\beta_1|$ are chosen to be $O(1)$ as indicated in Eq. (8.3d).

We henceforth limit this investigation to incident radiation intensities for which ϵ always remain small, and for which our chosen model of a two-level atomic system remains valid.

Solution of Eq. (8.4) for the Case Ω near 1

In the following, we present very briefly the solution of Eq. (8.4) and the other quantities related to it. Since it can be shown that, when the frequency Ω of the monochromatic incident beam is not near 1, the solution reduces to the WW case (spontaneous emission), we only concentrate our attention to the case of Ω being near 1, which means within the frequency of the natural linewidth. Let

$$\Omega = 1 + \alpha, \quad (8.9)$$

where α is small and represents the deviation from the atomic resonant frequency. We restrict this treatment to such intensities of the incident radiation for which

$$\epsilon = 0(\lambda) \text{ or less.} \quad (8.10)$$

The expression for $a(t)$ is given by

$$a(t) = \frac{1}{\mu_0 - \mu_1}[(\alpha + \mu_0)e^{i\mu_0 t} - (\alpha + \mu_1)e^{i\mu_1 t}]. \tag{8.11}$$

Here, μ_0 and μ_1 are the roots, $\mu = \mu_0$ (+ sign) and $\mu = \mu_1$ (− sign) of the equation

$$\mu = \lambda g(-1) - \frac{\epsilon^2 |\beta_{-1}|^2}{\mu + \alpha} = 0, \tag{8.12}$$

yielding

$$\mu = \tfrac{1}{2}\{\lambda g_0 - \alpha \pm [(\lambda g_0 + \alpha)^2 + 4\epsilon^2 |\beta_{-1}|^2]^{\frac{1}{2}}\}. \tag{8.13}$$

In Eq. (8.13), we have put

$$\lambda g(-1) \equiv \lambda g_0(-1) = \Delta + i\Gamma, \tag{8.14}$$

where $\Gamma > 0$ and is given by $\Gamma = \lambda f(1)$, which is the damping factor for spontaneous emission. The quantity Δ is the frequency shift, also in the spontaneous emission case. Since, in this treatment, this quantity diverges, we can replace it by its correct value arrived at by more sophisticated means or we can ignore it altogether without imparing the validity of the results of our work.

The quantity B_1 in Eq. (8.3e) for $t \to \infty$ is given by

$$B_1(\omega, \Omega, \infty) = \frac{-i(\Omega - \omega)}{(\omega - 1 + \mu_0)(\omega - 1 + \mu_1)}. \tag{8.15}$$

It can be readily verified that $B_1(\omega, \Omega, \infty)$ in Eq. (8.15) reduces, in the absence of incident radiation, to the ordinary WW value of the probability amplitude for spontaneous emission of a single photon and yields the natural line shape. From Eq. (8.15), we have

$$|B_1(\omega, \Omega, \infty)|^2 = \frac{(\Omega - \omega)^2}{|\omega - 1 + \mu_0|^2 |\omega - 1 + \mu_1|^2} = \frac{(\Omega - \omega)^2}{D(\omega)\, D^*(\omega)}, \tag{8.16}$$

where μ_0, μ_1, $D(\omega)$, and $|D(\omega)|^2$ are given in Appendix I by Eqs. (A12), (A13), (A14), and (A15), respectively. We notice that $|B_1(\omega, \Omega, \infty)|^2 = 0$ for $\Omega = \omega$, indicating that there is a dip at that frequency. This result is interesting. Since $|B_1|^2$ is proportional to the spontaneous emission which is omnidirectional, this would indicate that a quenching of spontaneous emission is predicted to be taking place at the frequency Ω of the incident plane wave of radiation, and that it is present for all directions. However, since in the forward direction we also have induced emission, the effect would be observable for directions other than the forward one. The quantity $B_1(\Omega, \omega, t)$ which is proportional to the spontaneous emission is given by

$$B_1(\Omega, \omega, t) = B_1(\Omega, \omega, \infty) - i\frac{e^{i(\omega-1)t}}{\mu_0 - \mu_1}\left\{\frac{\Omega - 1 + \mu_0}{\omega - 1 + \mu_0} e^{i\mu_0 t} - \frac{\Omega - 1 + \mu_1}{\omega - 1 + \mu_1} e^{i\mu_1 t}\right\}. \tag{8.17}$$

The quantity $B_2(\Omega, t)$ of Eq. (8.3f) which is proportional to the induced emission is given by

$$B_2(\Omega, t) = -i\frac{\beta_{-1}^*}{\mu_0 - \mu_1} e^{i\alpha t}(e^{i\mu_0 t} - e^{i\mu_1 t}) \tag{8.18}$$

and its absolute value squared becomes

$$|B_2(\Omega, t)|^2 = \frac{2|\beta_{-1}|^2}{r_1 r_2} e^{-\Gamma t}\{\cosh[t(r_1 r_2)^{\frac{1}{2}} \sin \tfrac{1}{2}(\theta_1 - \theta_2)] - \cos[t(r_1 r_2)^{\frac{1}{2}} \cos \tfrac{1}{2}(\theta_1 - \theta_2)]\} \tag{8.19}$$

with r_1, r_2, θ_1, and θ_2 defined in Appendix I. We note that as $t \to \infty$ the cosh term blows up but it can be shown[7] that the factor multiplying t in the argument is smaller than Γ, so that $B_2 \to 0$ as $t \to \infty$. For $\Omega = 1$, $\alpha + \Delta = 0$, $2\epsilon |\beta_{-1}| > \Gamma$, Eq. (8.19) becomes,

$$|B_2(\Omega, t)|^2 = \frac{2|\beta_{-1}|^2}{4\epsilon^2 |\beta_{-1}|^2 - \Gamma^2} e^{-\Gamma t} \times \{1 - \cos[t(4\epsilon^2 |\beta_{-1}|^2 - \Gamma^2)^{\frac{1}{2}}]\}. \tag{8.20}$$

For $\Omega = 1$, $\alpha + \Delta = 0$, $2\epsilon |\beta_{-1}| < \Gamma$, Eq. (8.19) becomes

$$|B_2(\Omega, t)|^2 = \frac{2|\beta_{-1}|^2}{\Gamma^2 - 4\epsilon^2 |\beta_{-1}|^2} e^{-\Gamma t} \times \{\cosh[t(\Gamma^2 - 4\epsilon^2 |\beta_{-1}|^2)^{\frac{1}{2}}] - 1\}, \tag{8.21}$$

and for long times, Eq. (8.19) becomes

$$|B_2(\Omega, t)|^2 = \frac{|\beta_{-1}|^2}{r_1 r_2} \exp[(r_1 r_2)^{\frac{1}{2}} \sin \tfrac{1}{2}|\theta_1 - \theta_2| - \Gamma]t. \tag{8.22}$$

Equations (8.11)–(8.22) have one important feature in common—the presence of the damping factor Γ. This factor, in turn, is a direct result of the vacuum term $\langle 0|\, \mathbf{A}(\mathbf{r}, t)\mathbf{A}(\mathbf{r}', t')\, |0\rangle$ of Eq. (5.3) which contributes to the first term on the right-hand side of Eq. (5.5). The quantity proportional to the spontaneous emission is given by Eqs. (8.15)–(8.17). It is seen from Eq. (8.16) that spontaneous emission can be quenched despite the presence of Γ in the

denominator, despite the full contribution of the vacuum term. Thus, the physical contribution of the vacuum state, as expressed by Γ, does not necessarily imply spontaneous emission if incident radiation is present. In other words, in the presence of incident radiation, the concepts of damping and decay (associated with the physical contribution of the vacuum state) are distinct from the concept of spontaneous emission.

IX. SOLUTION OF THE INTEGRAL EQUATION (6.2)

A solution of Eq. (6.2) would yield an expression for $a(t)$ and the quantities $b_1(\omega_k, t)$ (emission) and $b_2(\omega_k, t)$ (absorption). They occur in Eq. (6.5) for W_s, in Eq. (6.9) for $b_{k_1}(t)_{\text{em}}$, and in Eq. (6.10) for $b_{k_1}(t)_{\text{abs}}$. These quantities evaluated for all time and especially for $t \to \infty$ are of physical interest.

We can rewrite Eq. (6.2) in the form

$$\frac{\partial}{\partial t} a(t) + \frac{1}{2\pi} \int_0^\infty f\left(\frac{\omega}{2\pi}\right) e^{i(\omega_0 - \omega)t} b_1(\omega, t)\, d\omega + \frac{1}{2\pi} \int_\alpha^\beta h_1\left(\frac{\omega}{2\pi}\right) e^{i(\omega_0 - \omega)t} b_1(\omega, t)\, d\omega + \frac{1}{2\pi} \int_\alpha^\beta h_2\left(\frac{\omega}{2\pi}\right) e^{i(\omega_0 + \omega)t} b_2(\omega, t)\, d\omega = 0, \tag{9.1}$$

where

$$a(0) = 1, \tag{9.2}$$

and we have made the substitutions

$$-\frac{1}{\hbar^2} \sum_{k_1} \frac{\hbar}{2\omega_{k_1} V} |c_{k_1}|^2 \to \frac{1}{2\pi} \int_0^\infty f\left(\frac{\omega}{2\pi}\right) d\omega, \tag{9.3}$$

$$-\frac{1}{\hbar^2} \sum_{k} \frac{\hbar}{2\omega_k V} n_k |c_k|^2 \to \frac{1}{2\pi} \int_\alpha^\beta h_1\left(\frac{\omega}{2\pi}\right) d\omega, \tag{9.4}$$

$$-\frac{1}{\hbar^2} \sum_{k} \frac{\hbar}{2\omega_k V} n_k |d_k|^2 \to \frac{1}{2\pi} \int_\alpha^\beta h_2\left(\frac{\omega}{2\pi}\right) d\omega. \tag{9.5}$$

The limits α and β denote the frequency band limits of the incident radiation.

Introduction of Dimensionless Quantities

The integral equation (9.1) can be solved exactly. In order to be able to make analogies with the coherent wave case treated in the previous section, however, we introduce dimensionless and scaled variables so as to bring out the small parameters describing the physical situation. Let

$$t = \frac{\bar{t}}{\omega_0}, \tag{9.6}$$

$$\omega = \omega_0 \bar{\omega}, \tag{9.7}$$

$$f\left(\frac{\omega}{2\pi}\right) = 2\lambda\omega_0 \bar{f}(\bar{\omega}), \tag{9.8}$$

$$h_1\left(\frac{\omega}{2\pi}\right) = \epsilon_1 \omega_0 \bar{h}_1(\bar{\omega}), \tag{9.9}$$

$$h_2\left(\frac{\omega}{2\pi}\right) = \epsilon_2 \omega_0 \bar{h}_2(\bar{\omega}), \tag{9.10}$$

$$B_1 = \omega_0 b_1, \tag{9.11}$$

$$B_2 = \omega_0 b_2. \tag{9.12}$$

We henceforth drop all bars.

The Small Parameters λ and ϵ

In order to determine $a(t)$ for this case and draw an analogy with Eq. (8.11), we make the following assumptions:

(1) We choose the bandwidth of the incident radiation to be very narrow compared to the natural linewidth, so that

$$\beta - \alpha \ll \lambda. \tag{9.13}$$

(All that has been said about the size of λ in the previous section remains valid here.)

(2) We suppose that h_1 has been scaled in such a way that

$$\int_\alpha^\beta h_1\, d\nu = 0(\beta - \alpha). \tag{9.14}$$

(This is no real restriction since this can always be done.)

(3) We assume that the intensity of the incident radiation is such that

$$\epsilon_1(\beta - \alpha) = 0(\lambda^2). \tag{9.15}$$

(This puts an upper limit to the intensity of radiation compatible with our chosen model.)

(4) We assume that

$$\epsilon_2 \int_\alpha^\beta h_2(\nu)\, d\nu = 0(\lambda^2). \tag{9.16}$$

Solution of Eq. (9.1) in Dimensionless Form

It is convenient to introduce the quantities

$$\Omega = \tfrac{1}{2}(\alpha + \beta) = 1 + \sigma, \tag{9.17}$$

$$k^2 = \frac{1}{2\pi} \epsilon_1 \int_\alpha^\beta h_1(\nu)\, d\nu. \tag{9.18}$$

In the following, we are only concerned with the situation where Ω is near 1 which means that α and β are within the natural linewidth. This is the interesting situation, since it can be shown that the expressions reduce to those of the WW case when

the frequency Ω is not near 1. Thus, σ is a small quantity in Eq. (9.17). The expression for $a(t)$ is given by Eq. (8.11), where, instead of Eq. (8.12), we have

$$\eta - \lambda g_0 - [k^2/(\eta + \sigma)] = 0, \tag{9.19}$$

yielding the roots

$$\eta = \tfrac{1}{2}\{\lambda g_0 - \sigma \pm [(\lambda g_0 + \sigma)^2 + 4k^2]^{\frac{1}{2}}\}. \tag{9.20}$$

The roots $\eta = \eta_0$ (+ sign) and $\eta = \eta_1$ (− sign) become identical with μ_0 and μ_1 for the substitution

$$k^2 \to \epsilon^2 |\beta_{-1}|^2. \tag{9.21}$$

This indicates that the probability amplitude $a(t)$ remains the same if the quantities proportional to the expectation value of the number of photon operator are the same.

Evaluation of $B_1(\omega, \Omega, \infty)$

The exact expression for $B_1(\omega, \Omega, \infty)$ is given by

$$B_1(\omega, \Omega, \infty) = -i[1 - \omega - \lambda g(-\omega) + \epsilon_1 d_1(-\omega) + \epsilon_2 d_2(\omega)]^{-1}, \tag{9.22}$$

where

$$g(-\omega) = if(\omega), \tag{9.23}$$

$$d_1(-\omega) = \frac{i}{2} h_1(\omega) + \frac{1}{2\pi} P \int_\alpha^\beta \frac{h_1(\omega')\, d\omega'}{\omega' - \omega}, \tag{9.24a}$$

and

$$d_2(\omega) = \frac{1}{2\pi} \int_\alpha^\beta \frac{h_2(\omega')\, d\omega'}{\omega' + \omega}, \tag{9.24b}$$

and P stands for "principal value" in Eq. (9.24). Here again, as in the previous section, the quantity $|B_1(\omega, \Omega, \infty)|^2$ indicates that there is a dip somewhere between α and β. We interpret this effect as quenching of spontaneous emission, since the value of spontaneous emission at the dip is below that which would prevail if there were no incident radiation. This is not inconsistent with the contribution of the vacuum term which is given by the second term on the left-hand side of Eq. (9.1) and corresponds to $\lambda g(-\omega)$ in Eq. (9.22). Thus, here again, the omnipresence of the vacuum state expressed in Eq. (9.22) by $\lambda g(-\omega)$ does not necessarily imply spontaneous radiation, when incident radiation is present.

Evaluation of $B_2(\omega, \Omega, \infty)$

The exact expression for $B_2(\omega, \Omega, \infty)$ is given by

$$B_2(\omega, \Omega, \infty) = -i[1 + \omega - \lambda g(\omega) - \epsilon_1 d_1(\omega) + \epsilon_2 d_2(-\omega)]^{-1}, \tag{9.25}$$

where

$$g(\omega) = \frac{1}{\pi} \int_0^\infty \frac{f(\omega')}{\omega' + \omega} d\omega', \tag{9.26}$$

$$d_1(\omega) = \frac{1}{2\pi} \int_\alpha^\beta \frac{h_1(\omega')}{\omega' + \omega} d\omega', \tag{9.27}$$

and

$$d_2(-\omega) = -\frac{i}{2} h_2(\omega) + \frac{1}{2\pi} P \int_\alpha^\beta \frac{h_2(\omega')}{\omega' - \omega} d\omega'. \tag{9.28}$$

In connection with Eq. (9.26) it should be remembered that, since this quantity diverges, no physical meaning can be attached to it. This is closely related to our remarks made in connection with the frequency shift. For our purposes, this quantity is small and can be ignored. A glance at Eq. (9.25) indicates that $|B_2(\omega, \Omega, \infty)|^2$ is small owing to the plus sign in front of ω, the second term in the denominator.

X. SUMMARY

In this work, we have calculated quantities proportional to the induced and the spontaneous emission, for times comparable to, and long compared to, the free lifetime of the excited state of a two-level atomic system. Of physical interest is the prediction indicated by Eqs. (8.16) and (9.22) that, for $t \to \infty$, there is a partial quenching of spontaneous emission within the frequency band of the incident radiation. This result is valid for both the coherent and the stationary beams, despite the difference in the nature of the governing integral equations. In the case of an incident plane wave, the effect ought to be observed for any direction other than the forward one. In this work, we have also clarified the relation between the physical contribution of the vacuum state and the occurrence of spontaneous emission when incident radiation is present. It is indicated that spontaneous emission can be quenched despite the full contribution of the vacuum state. The latter contribution expresses itself physically through damping. We found an expression for the spectrum of the emergent radiation both for coherent and stationary incident beams given respectively by Eq. (5.14) and Eq. (6.8). In the coherent case, an interference term between the incident wave and the probability amplitudes for induced and spontaneous emission has been found. In the stationary case, only transition probabilities are involved. The density operator of the field for all times is given by Eq. (7.3).

ACKNOWLEDGMENTS

The author wishes to thank Professor L. J. Sanders for assistance with mathematical problems and Professor P. C. Martin for reading the manuscript and offering valuable comments.

APPENDIX

More Explicit Expressions for μ_0 and μ_1

We wish to derive more convenient expressions for μ_0 and μ_1 in the case $\alpha = 0(\lambda)$. We have

$$\mu_0 = \tfrac{1}{2}\{\lambda g_0 - \alpha + [(\lambda g_0 + \alpha)^2 + 4\epsilon^2 |\beta_{-1}|^2]^{\frac{1}{2}}\}, \quad \text{(A1)}$$

$$\mu_1 = \tfrac{1}{2}\{\lambda g_0 - \alpha - [(\lambda g_0 + \alpha)^2 + 4\epsilon^2 |\beta_{-1}|^2]^{\frac{1}{2}}\}. \quad \text{(A2)}$$

Let us call $\lambda g_0(-1) = \Delta + i\Gamma$, where $\Gamma > 0$, $\Gamma = \lambda f(1)$ (the damping factor), and Δ is the frequency shift in the spontaneous emission case. Since in this treatment this quantity diverges, we can replace it by its correct value arrived at by more sophisticated means without impairing the validity of the other results of our work. We have

$$[(\lambda g_0 + \alpha)^2 + 4\epsilon^2 |\beta_{-1}|^2]^{\frac{1}{2}} = \{[\alpha + \Delta + i(2\epsilon |\beta_{-1}| + \Gamma)] \times [\alpha + \Delta - i(2\epsilon |\beta_{-1}| - \Gamma)]\}^{\frac{1}{2}}. \quad \text{(A3)}$$

Case $2\epsilon_1 |\beta_{-1}| - \Gamma \geq 0$

We draw a $\alpha + \Delta$ plane as shown in Fig. 1.

$$x = 2\epsilon |\beta_{-1}| - \Gamma \geq 0, \quad \text{(A4)}$$

$$y = 2\epsilon |\beta_{-1}| + \Gamma > 0, \quad \text{(A5)}$$

$$r_1^2 = (\alpha + \Delta)^2 + x^2, \quad \text{(A6)}$$

$$r_2^2 = (\alpha + \Delta)^2 + y^2, \quad \text{(A7)}$$

$$x, y > 0, \quad \text{(A8a)}$$

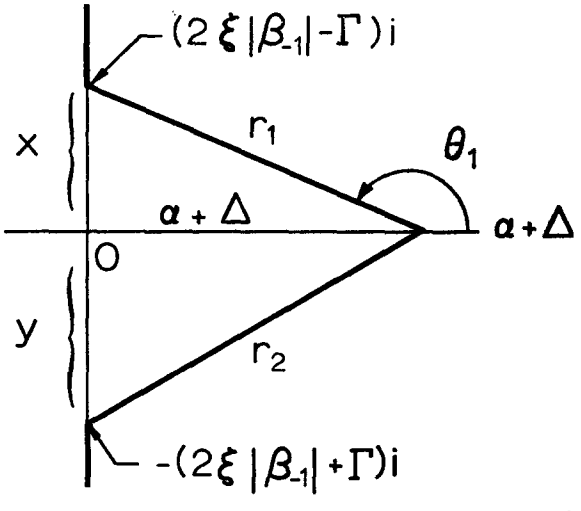

FIG. 1. $\alpha + \Delta$ plane for the evaluation of μ_0 and μ_1 as given by Eqs. (A12) and (A13), case $2\epsilon|\beta_{-1}| - \Gamma \geq 0$.

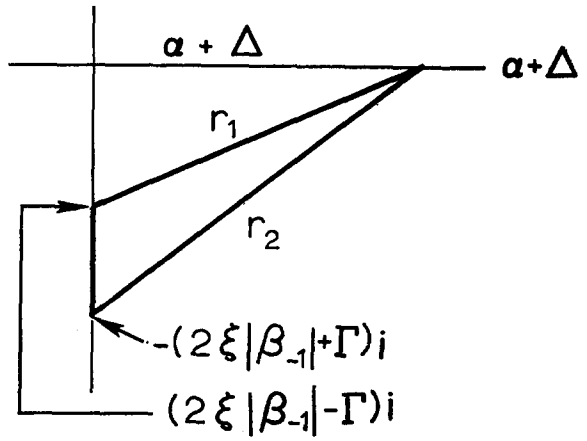

FIG. 2. $\alpha + \Delta$ plane for the evaluation of μ_0 and μ_1 in the case $2\epsilon|\beta_{-1}| - \Gamma = x < 0$.

$$r_1, r_2 \geq 0, \quad \text{by definition,} \quad \text{(A8b)}$$

$$\alpha + \Delta - ix = -r_1 e^{i\theta_1}, \quad \text{(A9)}$$

$$\alpha + \Delta + iy = -r_2 e^{-i\theta_2}. \quad \text{(A10)}$$

With these definitions, Eq. (A3) becomes

$$[(\lambda g_0 + \alpha)^2 + 4\epsilon^2 |\beta_{-1}|^2]^{\frac{1}{2}} = (r_1 r_2)^{\frac{1}{2}} e^{\frac{1}{2}i(\theta_1 - \theta_2)} = (r_1 r_2)^{\frac{1}{2}}[\cos \tfrac{1}{2}(\theta_1 - \theta_2) + i \sin \tfrac{1}{2}(\theta_1 - \theta_2)] \quad \text{(A11)}$$

and Eqs. (A1) and (A2) become, respectively,

$$\mu_0 = \tfrac{1}{2}[\Delta - \alpha + i\Gamma + (r_1 r_2)^{\frac{1}{2}} e^{\frac{1}{2}i(\theta_1 - \theta_2)}], \quad \text{(A12)}$$

$$\mu_1 = \tfrac{1}{2}[\Delta - \alpha + i\Gamma - (r_1 r_2)^{\frac{1}{2}} e^{\frac{1}{2}i(\theta_1 - \theta_2)}]. \quad \text{(A13)}$$

Case $2\epsilon |\beta_{-1}| - \Gamma = x < 0$

The foregoing results apply also to this case but instead of Fig. 1 we have Fig. 2.

The Quantity D (ω)

Finally, we wish to introduce the quantity

$$D(\omega) = [\omega - 1 + \tfrac{1}{2}(\Delta - \alpha) + \tfrac{1}{2}i\Gamma + \tfrac{1}{2}(r_1 r_2)^{\frac{1}{2}} e^{\frac{1}{2}i(\theta_1 - \theta_2)}] \times [\omega - 1 + \tfrac{1}{2}(\Delta - \alpha) + \tfrac{1}{2}i\Gamma - \tfrac{1}{2}(r_1 r_2)^{\frac{1}{2}} e^{\frac{1}{2}i(\theta_1 - \theta_2)}], \quad \text{(A14)}$$

which yields

$$DD^* = \{[\omega - 1 + \tfrac{1}{2}(\Delta - \alpha) + \tfrac{1}{2}(r_1 r_2)^{\frac{1}{2}} \cos \tfrac{1}{2}(\theta_1 - \theta_2)]^2 + \tfrac{1}{4}[\Gamma + (r_1 r_2)^{\frac{1}{2}} \sin \tfrac{1}{2}(\theta_1 - \theta_2)]^2\} \times \{[\omega - 1 + \tfrac{1}{2}(\Delta - \alpha) - \tfrac{1}{2}(r_1 r_2)^{\frac{1}{2}} \cos \tfrac{1}{2}(\theta_1 - \theta_2)]^2 + \tfrac{1}{4}[\Gamma - (r_1 r_2)^{\frac{1}{2}} \sin \tfrac{1}{2}(\theta_1 - \theta_2)]^2\}. \quad \text{(A15)}$$

Discrete Series for the Universal Covering Group of the 3 + 2 de Sitter Group*

N. T. EVANS
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England
(Received 21 February 1966)

A classification is given of the irreducible unitary representations of the universal covering group of the 3 + 2 de Sitter group which contract to the usual physical representations of the Poincaré group. These representations include the discrete series for the 3 + 2 de Sitter group. The classification problem is reduced from one for the group to the corresponding one for the Lie algebra. The method used by Thomas for the representations of the 4 + 1 de Sitter group is then followed, except that a representation is reduced out with respect to the irreducible unitary representations of a noncompact 2 + 2 subalgebra. It is conjectured that each representation of this subalgebra occurs at most once. The action on the representation spaces of a basis for the Lie algebra is given. The contractions of the representations to those of the Poincaré, oscillator and the Galilei groups are briefly considered.

1. INTRODUCTION

THE 4 + 1 and 3 + 2 de Sitter groups are the only simple Lie groups which can be contracted, in the Inönü–Wigner sense,[1] to the Poincaré group.[2] The relationship of the Poincaré group to the Galilei group through contraction,[2] suggests that any attempt to base a theory of elementary particles on a space–time group other than the Poincaré group should at least begin with a study of the de Sitter groups.[3] In the spirit of Wigner, this means that we should first classify all irreducible unitary ray representations of the de Sitter groups. Bargmann[4] has shown that this is the same as classifying all continuous unitary irreducible representations of their universal covering groups.

If we pass to the Lie algebras of the de Sitter groups and consider the element iM_{05} whose image under contraction is represented by the energy operator, then we find a decisive difference between the two de Sitter spaces.[5] In the Lie algebra of the 4 + 1 de Sitter group, iM_{05} belongs to a noncompact Cartan subalgebra. If we diagonalize its representing operator on an irreducible unitary representation space for the group, its "energy" eigenvalues are continuous and symmetrically distributed about the origin. Thus, it seems to be impossible to define a positive-definite "energy" in any simple way in such a space and this casts serious doubts on the use of the 4 + 1 de Sitter group as a space–time symmetry group for elementary particles. However, iM_{05} belongs to a compact Cartan subalgebra of the 3 + 2 Lie algebra and its "energy" eigenvalues are discrete. Furthermore, there exist unitary irreducible representations of the 3 + 2 de Sitter group which possess positive-definite or negative-definite "energies," and it is precisely these representations which contract to the physical representations of the Poincaré group. These representations belong to the discrete series for the 3 + 2 de Sitter group.

The purpose of this paper is to classify these representations. Because we are restricting the classification to the discrete series, we show that we can formally follow a method used by Thomas[6] for the unitary representations of the 4 + 1 de Sitter group. We first reduce the problem to one for the Lie algebra and reduce out the representations with respect to the representations of a semisimple noncompact Lie algebra of order 6.

Previously, Ehrman[7] has classified unitary irreducible representations of the universal covering group of the 3 + 2 de Sitter group using a method developed by Harish-Chandra. However, this method is not suitable for the discrete series and, except for certain "singleton" representations classified by the

* The research reported in this document has been sponsored in part by the Air Force Office of Scientific Research, OAR, under Grant AF EOAR 65-36, through the European Office of Aerospace Research, United States Air Force.

[1] E. Inönü and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. 39, 510 (1953).

[2] W. T. Sharp, *Racah Algebra and the Contraction of Groups*, CRT-935 (Atomic Energy of Canada Ltd., Chalk River, Ontario, 1960), Chap. 10.

[3] The first such study was made by P.A.M. Dirac, Ann. Math. 36, 657 (1935). For a recent summary of the situation see F. Gursey, *Group Theoretical Concepts and Methods in Elementary Particle Physics* (Gordon and Breach Science Publishers, Inc., New York, 1964).

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

[5] T. Phillips, Ph.D. Thesis, Princeton University (1962).

[6] L. H. Thomas, Ann. Math. 42, 113 (1941).

[7] J. B. Ehrman, Proc. Cambridge Phil. Soc. 53, 290 (1956); and Ph.D. Thesis, Princeton University (1954).

Thomas[6] method, his representations contract to the imaginary mass representations of the Poincaré group.

In Sec. 2 we discuss the universal covering group of the 3 + 2 de Sitter group and several of its important subgroups. We also introduce its Lie algebra. In Sec. 3 we investigate the relationship between unitary irreducible representations of the universal covering group of the 3 + 2 de Sitter group and the corresponding irreducible representations of the Lie algebra. We are able to reduce the classification problem for the discrete series for the group, to an explicit problem for the irreducible "physical" representations of the Lie algebra. The matrix elements of a basis for the Lie algebra are computed in Sec. 4 and in Sec. 5, we use these to complete the classification. Section 6 contains a brief account of the contractions of these representations to the physical ray representations of the Galilei group, via the representations of both the Poincaré group and an oscillator group. The Appendix summarizes some relevant details of the discrete series for the 2 + 1 Lorentz group.

2. UNIVERSAL COVERING GROUP OF THE 3 + 2 DE SITTER GROUP AND ITS LIE ALGEBRA

Let $x_5^2 + x_0^2 - x_1^2 - x_2^2 - x_3^2$ be a nonsingular quadratic form in the Euclidean space $\mathbf{R}^5$. In this coordinate system we can introduce the diagonal metric tensor

$$g^{ij} = g_{ij};\quad g_{00} = g_{55} = 1,\quad g_{11} = g_{22} = g_{33} = -1,$$
$$g_{ij} = 0 (i \neq j). \tag{1}$$

The quadratic form is $g^{ij}x_ix_j$, where the usual summation convention for tensor indices is observed. The metric tensor is also used for raising and lowering indices.

The group of automorphisms of $\mathbf{R}^5$ which leaves invariant the quadratic form has four connected components.[4] Let G' be the connected component of the identity. We call G' the 3 + 2 de Sitter group (but we may often omit the label 3 + 2). The center of G' is the identity element only. The group manifold of G' is homeomorphic to the direct product of a sphere, of a circle, and of a six-dimensional Euclidean space.[8] Let G be the universal covering group of G' and let ϕ be the natural homomorphism of G onto G'. The kernel of ϕ is the center Z of G. It is one-dimensional and isomorphic to the fundamental group of G', i.e., to the direct product of a cyclic group of order 2 and an infinite cyclic group. G' is isomorphic to G/Z, the adjoint group of G. G is a connected, simply connected, real simple Lie group and is, of course, noncompact.

G has an infinite center and therefore has no faithful finite-dimensional representations. But we can easily find a faithful, finite dimensional (and thus nonunitary) representation of G' by identifying the linear transformations of $\mathbf{R}^5$ with their matrices with respect to the natural basis for $\mathbf{R}^5$. A matrix preserving $g^{ij}x_ix_j$ belongs to G' if it is unimodular and if the second-order subdeterminant formed from its first two rows and columns is positive.

Identify the Lie algebra g of G with that of G'. g is a simple, real Lie algebra. It is well known[4] that a 5×5 matrix $A = (A_j^i)(i, j = 5, 0, 1, 2, 3)$ belongs to the corresponding representation of g if and only if

$$A_{ij} + A_{ji} = 0 \qquad (A_{ij} = g_{ik}A_j^k).$$

A basis for g over $\mathbf{R}$ is given by the ten linearly independent matrices M_{ij}, with elements

$$(M_{ij})_l^k = g_{jk}\,\delta_i^l - g_{ik}\,\delta_j^l \quad (i, j, k, l = 5, 0, 1, 2, 3).$$

They satisfy the commutation relations

$$[M_{ij}, M_{kl}] = g_{il}M_{jk} + g_{jk}M_{il} - g_{ik}M_{jl} - g_{jl}M_{ik}. \tag{2}$$

Define the following elements of g:

$$H = M_{21} + M_{50}\quad N_1 = M_{25} + M_{10}\quad N_2 = M_{51} + M_{20}, \tag{3}$$
$$H' = M_{21} - M_{50}\quad N_1' = M_{25} - M_{10}\quad N_2' = M_{51} - M_{20}.$$

The linear envelopes

$$\begin{aligned} f_1 &= ((M_{23}, M_{31}, M_{12})), \\ f_2 &= ((M_{50}, M_{03}, M_{35})), \\ f_3 &= ((H, N_1, N_2)), \\ f_4 &= ((N', N_1', N_2')), \end{aligned} \tag{4}$$

are simple, rank-1, subalgebras of g. By examining their respective multiplication tables, we can verify that each of f_2, f_3, and f_4 is isomorphic to the Lie algebra of the group $SO(2, 1)$. However, f_1 is compact and isomorphic to the Lie algebra of $SO(3)$. Its image, F_1' in the de Sitter group G' is the group of three-dimensional rotations leaving the x_5 and x_0 axes fixed. $F_1 = \phi^{-1}(F_1')$ is the spinor covering group of $SO(3)$.

Let f be the subalgebra of g generated by $M_{ij}(i, j \neq 3)$. f is semisimple but not simple. f_3 and f_4 are simple ideals of f, and $[f_3, f_4] = 0$, $f_3 \cap f_4 = 0$. Therefore $f = f_3 + f_4$.

[8] E. P. Wigner, Proc. Natl. Acad. Sci. U. S. 36, 184 (1950).

Let F^*, F_3^*, and F_4^* be the connected and simply connected Lie groups determined by f, f_3, and f_4, respectively. Then $F^* = F_3^* \otimes F_4^*$ where F_3^* is isomorphic to F_4^*, and F^*, F_3^* are isomorphic to the universal covering groups of $SO(2, 2)$ and $SO(2, 1)$, respectively. Let F, F' be the closed, connected subgroups of G, G', respectively, determined by f. Neither F nor F' is simply connected. F' is isomorphic to $SO(2, 2)$, the connected component of the identity in the group of linear transformations in $\mathbf{R}^4$ preserving the quadratic form $x_5^2 + x_0^2 - x_1^2 - x_2^2$. It has a center of order 2 consisting of the identity and the element $\exp(\pi H) = \exp(\pi H')$, i.e., the 5×5 matrix diag $(-1, -1, -1, -1, 1)$. The kernel of ϕ restricted to F is an infinite cyclic group. Any connected group with Lie algebra f (F in particular) is isomorphic to F^*/D, where D is a discrete central subgroup of F^*.

The subalgebra $h = ((H, H'))$ is a compact Cartan subalgebra of both f and g. We denote the complexification of a real Lie algebra by a subscript c. Introduce the following elements of g_c:

$$E_{\pm\alpha} = N_1 \pm iN_2, \qquad E_{\pm\beta} = N_1' \pm iN_2', \tag{5}$$
$$E_{\pm\gamma} = M_{35} \pm iM_{30}, \qquad E_{\pm\delta} = \pm M_{23} + iM_{31}.$$

Together with H, H' they form a Weyl basis for g_c. The other nonisomorphic real forms of g_c are the Lie algebras of $SO(5)$ and the $4 + 1$ de Sitter group $SO(4, 1)$.

We give the multiplication table for g_c in full, since we need it later.

$$[H, H'] = 0, \tag{6}$$
$$[H, E_{\pm\alpha}] = \pm 2iE_{\pm\alpha}, \qquad [H, E_{\pm\beta}] = 0, \tag{7}$$
$$[H', E_{\pm\alpha}] = 0, \qquad [H', E_{\pm\beta}] = \pm 2iE_{\pm\beta}, \tag{8}$$
$$[E_\alpha, E_{-\alpha}] = -4iH, \qquad [E_\beta, E_{-\beta}] = -4iH', \tag{9}$$
$$[E_{\pm\alpha}, E_{\pm\beta}] = 0 \quad \text{(all combinations)},$$
$$[H, E_{\pm\gamma}] = \pm iE_{\pm\gamma}, \qquad [H, E_{\pm\delta}] = \pm iE_{\pm\delta}, \tag{10}$$

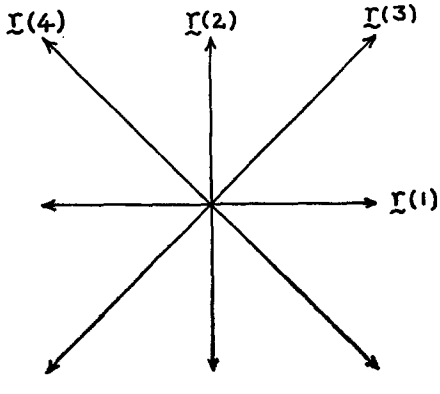

FIG. 1. The root diagram of g_c.

$$[H', E_{\pm\gamma}] = \mp iE_{\pm\gamma}, \qquad [H', E_{\pm\delta}] = \pm iE_{\pm\delta}, \tag{11}$$
$$[E_\alpha, E_\gamma] = 0, \qquad [E_{-\alpha}, E_{-\delta}] = 0, \tag{12}$$
$$[E_\alpha, E_{-\delta}] = 2E_\gamma, \qquad [E_{-\alpha}, E_\gamma] = 2E_{-\delta},$$
$$[E_\alpha, E_\delta] = 0, \qquad [E_{-\alpha}, E_{-\gamma}] = 0, \tag{13}$$
$$[E_\alpha, E_{-\gamma}] = -2E_\delta, \qquad [E_{-\alpha}, E_\delta] = -2E_{-\gamma},$$
$$[E_\beta, E_\delta] = 0, \qquad [E_{-\beta}, E_\gamma] = 0, \tag{14}$$
$$[E_\beta, E_\gamma] = -2E_\delta, \qquad [E_{-\beta}, E_\delta] = -2E_\gamma,$$
$$[E_\beta, E_{-\gamma}] = 0, \qquad [E_{-\beta}, E_{-\delta}] = 0, \tag{15}$$
$$[E_\beta, E_{-\delta}] = 2E_{-\gamma}, \qquad [E_{-\beta}, E_{-\gamma}] = 2E_{-\delta},$$
$$[E_\gamma, E_{-\gamma}] = -i(H - H'),$$
$$[E_\delta, E_{-\delta}] = -i(H + H'),$$
$$[E_\gamma, E_\delta] = E_\alpha, \qquad [E_\gamma, E_{-\delta}] = -E_{-\beta}, \tag{16}$$
$$[E_{-\gamma}, E_{-\delta}] = -E_{-\alpha}, \qquad [E_{-\gamma}, E_\delta] = E_\beta.$$

If we introduce a new basis for h_c and normalize the root vectors

$$2H_1 = -i(H + H'), \quad 2H_2 = -i(H - H'),$$
$$\sqrt{2}E_{\pm 1} = -E_{\pm\delta}, \qquad \sqrt{2}E_{\pm 2} = -E_{\pm\gamma}, \tag{17}$$
$$2E_{\pm 3} = -E_{\pm\alpha}, \qquad 2E_{\pm 4} = -E_{\mp\beta},$$

we can summarize the multiplication table in the familiar form:

$$[H_i, H_j] = 0,$$
$$[H_i, E_{\pm\lambda}] = \pm r_i(\lambda)E_{\pm\lambda}, \qquad [E_\lambda, E_{-\lambda}] = \sum_{i=1}^{2} r_i(\lambda)H_i,$$
$$[E_\lambda, E_\mu] = \begin{cases} 0 & \text{if } \mathbf{r}(\lambda) + \mathbf{r}(\mu) \text{ is not a root,} \\ N_{\lambda\mu}E_\nu & \text{if } \mathbf{r}(\lambda) + \mathbf{r}(\mu) = \mathbf{r}(\nu) \text{ is a root,} \end{cases}$$
$$N_{\lambda\mu} = N_{-\mu,-\lambda} = -N_{\mu\lambda},$$
$$N_{12} = N_{1,-2} = N_{-1,3} = N_{2,-3} = N_{41} = 1,$$
$$\mathbf{r}(1) = (1, 0), \quad \mathbf{r}(2) = (0, 1), \quad \mathbf{r}(-\mu) = -\mathbf{r}(\mu),$$
$$\mathbf{r}(3) = (1, 1), \quad \mathbf{r}(4) = (-1, 1),$$
$$(i, j = 1, 2; \quad \pm\lambda, \pm\mu, \pm\nu = 1, 2, 3, 4). \tag{18}$$

From this table we can easily construct the root diagram for $g_c = B_2$ in Cartan's notation.

Four distinct A_1 subalgebras can be selected from a Weyl basis for g_c[9]:

$$f_{1c} = ((H_1, E_{\pm 1})), \qquad f_{2c} = ((H_2, E_{\pm 2})), \tag{19}$$
$$f_{3c} = ((2H_3 = (H_1 + H_2), E_{\pm 3})),$$
$$f_{4c} = ((2H_4 = (H_2 - H_1), E_{\pm 4})).$$

[9] In general, this number is equal to the number of positive roots, i.e., $\frac{1}{2}$ (order–rank), of the simple Lie algebra.

As the notation indicates, they are the complexifications of the corresponding real forms (4). From $f = f_3 + f_4$ we have

$$f_c = f_{3c} + f_{4c}. \tag{20}$$

Let $\mathfrak{B}$ be the universal enveloping algebra of g.[10] Because g has rank 2, the center of $\mathfrak{B}$ has two algebraically independent generators. We take for these the Casimir operator Π and the fourth-order invariant Ω,

$$\begin{aligned}\Pi &= \tfrac{1}{2}M_{ij}M^{ij} \\ &= \tfrac{1}{2}(H^2 - N_1^2 - N_2^2 + H'^2 - N_1'^2 - N_2'^2) \\ &\qquad - M_{35}^2 - M_{30}^2 + M_{31}^2 + M_{32}^2, \\ \Omega &= W_iW^i,\end{aligned}$$

with $W_i = \frac{1}{8}\epsilon_{ijklm}M^{jk}M^{lm}$, where ϵ_{ijklm} is antisymmetric in all indices and normalized, $\epsilon_{50123} = +1$. Let Π_a be the Casimir operator for the subalgebra $f_a (a = 1, 2, 3, 4)$, defined thus:

$$\begin{aligned}\Pi_1 &= -(M_{23}^2 + M_{31}^2 + M_{12}^2), \\ \Pi_2 &= M_{50}^2 - M_{03}^2 - M_{35}^2, \\ \Pi_3 &= \tfrac{1}{4}(H^2 - N_1^2 - N_2^2), \\ \Pi_4 &= \tfrac{1}{4}(H'^2 - N_1'^2 - N_2'^2).\end{aligned} \tag{21}$$

Now, Π, Ω can be written more simply as

$$\Pi = 2(\Pi_3 + \Pi_4) - M_{35}^2 - M_{30}^2 + M_{31}^2 + M_{32}^2, \tag{22}$$

$$\Omega = [M_{35}, \Phi]^2 + [M_{30}, \Phi]^2 - [M_{31}, \Phi]^2 - [M_{32}, \Phi]^2 - \Phi^2, \tag{23}$$

with

$$\Phi = \Pi_3 - \Pi_4.$$

3. RELATION BETWEEN REPRESENTATIONS OF G AND g

Let K' be the maximal compact subgroup of the de Sitter group, G'. It is isomorphic to $SO(3) \otimes SO(2) = F_1' \otimes SO(2)$, in previous notation. The $SO(2)$ subgroup leaves invariant the x_1, x_2, and x_3 axes in R^5. Its infinitesimal generator is iH_2. $\phi^{-1}(K') = K$ is the universal covering group of K'. $K = F_1 \otimes T$, where

$$T = \{T_\tau = \exp(i\tau H_2) \mid -\infty < \tau < \infty\}. \tag{24}$$

T is a closed subgroup of G isomorphic to the additive group of reals. Neither T nor K is compact. Z is the center of both G and K, and $K' \cong K/Z$. The element $T_{2\pi}$ generates the infinite cyclic subgroup of Z. We call K the maximal essentially compact subgroup of G, i.e., its image in the adjoint group G' is compact.

Irreducible unitary representations of K are finite-dimensional, even though K is noncompact, and corresponding to the decomposition $K = F_1 \otimes T$, they are a tensor product of an irreducible representation of F_1 and an irreducible representation of T. They are specified by a pair, (j, ϵ), where $2j$ is a nonnegative integer and ϵ is any real number, and have dimensionality $2j + 1$.

Much work has been done by Gårding, Harish-Chandra, Nelson, and others on the relation between representations of a semisimple Lie group and those of its Lie algebra. In the following lemmas and remarks we apply several of their results to the present case. For the 4 + 1 de Sitter group this has been done by Dixmier,[11] who also indicates that his conclusions carry over to an arbitrary real semisimple Lie group.

Lemma 1: Let ρ' be an irreducible unitary representation of G on a separable complex Hilbert space, $\mathfrak{H}$. Then $\rho' \mid K$, the restriction of ρ' to K, is completely reducible; that is, $\mathfrak{H}$ decomposes into a direct sum of finite-dimensional subspaces which are invariant under $\rho'(K)$ and irreducible for $\rho' \mid K$.

Proof: $T_{2\pi} \in Z$ and hence $\rho'(T_{2\pi})$ commutes with $\rho'(G)$ and is therefore a multiple of the identity I. Let $\rho'(T_{2\pi}) = \exp(i2\pi\epsilon^0)\mathrm{I}$, with $0 \leq \epsilon^0 < 1$. Now, define ν' to be the unitary representation of $K = F_1 \otimes T$ on $\mathfrak{H}$:

$$\nu' : F_1(x)T_\tau \to e^{-i\epsilon^0\tau}\rho'[F_1(x)T_\tau] \quad (F_1(x) \in F_1, T_\tau \in T). \tag{25}$$

Then $\nu'[F_1(x)T_{\tau+2\pi}] = \nu'[F_1(x)T_\tau]$, and thus ν' is a unitary representation on $\mathfrak{H}$ of the compact group $SU(2) \otimes SO(2)$. Hence ν' is completely reducible by the well-known theorem for compact groups. This shows that $\rho' \mid K$ is completely reducible also.

Let Λ_0 be the set of all inequivalent irreducible unitary representations of K. Elements of Λ_0 are specified by (j, ϵ). Let $\Lambda \subset \Lambda_0$ be the set of all distinct (j, ϵ) which occur in the complete reduction of $\rho' \mid K$. Then it is known that each (j, ϵ) in Λ occurs at most $2j + 1$ times in the decomposition.[12] In fact, Ehrman[7] has shown that this upper bound can be reduced to $j + \frac{1}{2}$ or $j + 1$ for $2j$ an odd or even integer, respectively. Only for special ρ' does each (j, ϵ) in Λ has multiplicity one. Following Ehrman, we call such representations singletons.

[10] Harish-Chandra, Ann. Math. 50, 900 (1949).

[11] J. Dixmier, Bull Soc. Math. (France) 89, 9 (1961).

[12] Harish-Chandra, Trans. Am. Math Soc. 75, 185 (1953), p. 243.

For $(j, \epsilon) \in \Lambda$, let $V_{j,\epsilon}$ be the (finite) direct sum of all irreducible subspaces of $\mathfrak{H}$ of type (j, ϵ). Then $\mathfrak{H}$ is the direct sum

$$\mathfrak{H} = \bigoplus_{(j,\epsilon)\in\Lambda} V_{j,\epsilon},$$

and the algebraic sum

$$V = \sum_{(j,\epsilon)\in\Lambda} V_{j,\epsilon} \tag{26}$$

is dense in $\mathfrak{H}$.

Let $\mathfrak{A} \subset \mathfrak{H}$ be the set of vectors analytic[13] for ρ'. Harish-Chandra[14] has shown (i) $V \subset \mathfrak{A}$, (ii) $\rho'(g)V \subset V$. Let ρ be this representation of g (or $\mathfrak{G}$) on V; then ρ is algebraically irreducible.[15] For a class of X in $\mathfrak{G}$, the closure of $i\rho(X)$ is self-adjoint. This class includes all elements of g and the skew-symmetric elements belonging to the center of $\mathfrak{G}$ or to the center of a subalgebra of $\mathfrak{G}$.[16] In particular, the operators $\rho(\Omega)$, $\rho(\Pi)$, and $\rho(\Pi_a)$ $(a = 1, 2, 3, 4)$ are essentially self-adjoint; i.e., their closures are self-adjoint.

Take the simultaneous eigenvectors $|jm\epsilon;\ \alpha\rangle$ of the commuting operators $\rho(\Pi_1)$, $\rho(H_1)$, $\rho(H_2)$ as an orthonormal basis for V. The eigenvalue equations are

$$\begin{aligned}
\rho(\Pi_1)\ |jm\epsilon;\alpha\rangle &= j(j+1)\ |jm\epsilon;\alpha\rangle,\\
\rho(H_1)\ |jm\epsilon;\alpha\rangle &= m\ |jm\epsilon;\alpha\rangle,\\
\rho(H_2)\ |jm\epsilon;\alpha\rangle &= \epsilon\ |jm\epsilon;\alpha\rangle.
\end{aligned} \tag{27}$$

For fixed $(j, \epsilon) \in \Lambda$, the $|jm\epsilon;\ \alpha\rangle$ span $V_{j,\epsilon}$. The discrete parameter α lifts the (j, m, ϵ) degeneracy. We have seen that the range of α is finite and, in general, dependent on both j and ϵ. It is only for singleton representations that α becomes redundant.

Lemma 2: In the algebraically irreducible representation ρ of $\mathfrak{G}$ on V, the eigenvalues ϵ and $j(j+1)$ of operators $\rho(H_2)$ and $\rho(\Pi_1)$ obey the restrictions:

(i) $\epsilon \equiv \epsilon^0 \pmod 1$ with $0 \leq \epsilon^0 < 1$; that is ϵ^0, the fractional part of ϵ, is constant.

(ii) $2j$ takes either even integer or odd integer values.

Proof:

(i) With the notation of Lemma 1, let

$$\rho'(T_\tau) = \exp\,[i\tau\rho(H_2)], \qquad \nu'(T_\tau) = \exp\,[i\tau\nu(H_2)].$$

From Lemma 1, ν' is a (continuous) unitary representation of $SO(2)$ and therefore $\nu(H_2)$ has only integer eigenvalues. But, from the definitions of ν' and ν, $\rho(H_2) = \epsilon^0 I + \nu(H_2)$ $(0 \leq \epsilon^0 < 1)$, and hence (i) follows.

(ii) It is well known that, under a rotation through 2π in the $x_1x_2x_3$ plane, $|jm\epsilon;\ \alpha\rangle$ becomes multiplied by $(-1)^{2j}$. Since this operation is in the center of G, the parity of $2j$ must be constant.

For physical applications, it is essential that we find a set of commuting, essentially self-adjoint operators in $\rho(\mathfrak{G})$, whose eigenvalues completely parametrize a basis for V. Racah[17] has shown that we need at least $\frac{1}{2}$ (order g − 3 rank g) = 2 such operators, in addition to $\rho(\Pi)$, $\rho(\Omega)$, $\rho(H_1)$, and $\rho(H_2)$. We take $\rho(\Pi_3)$ and $\rho(\Pi_4)$ as the extra commuting operators. Only for the special case of singleton representations is $\rho(\Pi_1)$ an appropriate choice as one of the extra pair; for the general ρ we cannot find a commuting self-adjoint operator with eigenvalues α.

With $\rho(\Pi_3)$ and $\rho(\Pi_4)$ in the set of commuting operators, it is more convenient to label the basis with eigenvalues λ and λ' of the operators representing H_3 and H_4, (19):

$$\begin{aligned}
\rho(H_3) &= \tfrac{1}{2}(\rho(H_1) + \rho(H_2)), \qquad \lambda = \tfrac{1}{2}(m+\epsilon),\\
\rho(H_4) &= \tfrac{1}{2}(\rho(H_2) - \rho(H_1)), \qquad \lambda' = \tfrac{1}{2}(\epsilon - m).
\end{aligned} \tag{28}$$

From Lemma 2 (ii), $\lambda - \lambda'$ is an integer or half-odd integer throughout ρ.

We can generalize the idea of the weight diagram of a representation of a compact group to the non-compact case. The weight space of ρ is two dimensional. Its points (weights) are defined either by the coordinates (m, ϵ) or (λ, λ'), corresponding to the two different bases for the Cartan subalgebra, h_c. These bases are related by a rotation through $\frac{1}{4}\pi$ and a change of scale on the weight diagram. The multiplicity of a weight is equal to the dimensionality of the associated eigenspace. It is important to notice that a weight may have an infinite multiplicity corresponding to the occurrence of j values up to infinity. We now prove a theorem which rules out this possibility for the representations of physical interest.

Theorem 1[18]: Let ρ be an algebraically irreducible

[13] E. Nelson, Ann. Math. 70, 572 (1959). [A vector $|\psi>$ in $\mathfrak{H}$ is analytic for ρ' if for every vector $|\phi>$ in $\mathfrak{H}$, $<\phi|\ \rho'(\mathfrak{N})\ |\psi>$ is analytic in some neighborhood $\mathfrak{N}$ of the identity of G.]

[14] See Ref. 12, especially Lemma 34.

[15] Harish-Chandra, Am. J. Math. 78, 564 (1956).

[16] I. E. Segal, Proc. Am. Math. Soc. 3, 13 (1952). E. Nelson and W. F. Stinespring, Am. J. Math. 81, 547 (1959).

[17] G. Racah, in *Ergebnisse der exakten Naturwissenschaften* (Springer-Verlag, Berlin, 1965).

[18] C. Fronsdal, Rev. Mod. Phys. 37, 221 (1965) gives a similar theorem.

representation of g on $V = \sum_{(j,\epsilon)\in\Lambda} V_{j,\epsilon}$. Then if ϵ is bounded below (above):

(i) λ and λ' are bounded below (above),

(ii) ϵ, λ, and λ' are positive-definite (negative-definite),

(iii) given (j, ϵ) in Λ, for fixed ϵ, $j < |\epsilon|$.

Proof: (i) First suppose ϵ is bounded below. Let $\epsilon_0 \in \Lambda$ be a lower bound. Then $|m\epsilon_0\rangle$ is a basis vector for V, for some finite m. From (18), it is easy to check that $\rho(E_{-a})(a = 2, 3, 4)$ are lowering operators for ϵ. Therefore $\rho(E_{-a})\ |m\epsilon_0\rangle$ must vanish. Now,[19] $V = \rho(\mathfrak{B})\ |m\epsilon_0\rangle$. Let X be an element of $\mathfrak{B}$ of degree n. Again using table (18), we find that $\rho(E_{-a}^{n+1})\rho(X)\ |m\epsilon_0\rangle$ vanishes. Thus every vector in V is annihilated by a finite product of the $\rho(E_{-a})$. But $\rho(E_{-3})$, $\rho(E_{-4})$ are lowering operators for λ, λ', respectively. It follows that both λ and λ' are bounded below. For ϵ bounded above, the result follows using $\rho(E_3)$ and $\rho(E_4)$.

(ii) It follows from (i) and (A3) that in the reduction of $\rho \mid f_a(a = 2, 3, 4)$, the irreducible components must all be of type σ_l^+ or σ_l^- according as ϵ, λ, λ' are bounded below or above, respectively. The result (ii) is a consequence of (A5).

(iii) Fix a weight $(m, \epsilon) = (\lambda, \lambda')$. For $\epsilon > 0$; $\lambda > 0$, $\lambda' > 0 \Rightarrow |m| < \epsilon$ from (28). Similarly, with $\epsilon < 0$; $\lambda < 0$, $\lambda' < 0 \Rightarrow |m| < -\epsilon$. For (j, ϵ) in Λ, $j \leqq \max |m| < |\epsilon|$. This completes the proof.

Corollary 1: For ϵ bounded below or above, each weight has a finite multiplicity.

Proof: This is immediate from (iii) and the finite dimensionality of the $V_{j,\epsilon}$ subspaces.

Corollary 2[18]*:* Suppose ϵ, λ, λ' are bounded below (above), with lower (upper) bounds ϵ_0, λ_0, λ_0', respectively. Then the basis vectors $|m\epsilon_0\rangle$, $|\lambda_0\lambda'\rangle$, $|\lambda\lambda_0'\rangle$ transform under $\rho(f_1)$, $\rho(f_4)$, $\rho(f_3)$ as irreducible unitary representations of type $\mathfrak{D}_{j_0}$, σ^+, $\sigma^+(\mathfrak{D}_{j_0}, \sigma^-, \sigma^-)$; that is, the weights (m, ϵ_0), (λ_0, λ'), and (λ, λ_0') are simple.

Proof: Consider ϵ, λ, λ' bounded below. Let $\mathfrak{B}_1$ be the subalgebra of $\mathfrak{B}$ generated by $\{1, f_1\}$. $\mathfrak{B}_1$ is the enveloping algebra of f_1 and we can identify the representations of f_1 and $\mathfrak{B}_1$. Subalgebras $\mathfrak{B}_2$, $\mathfrak{B}_3$, and $\mathfrak{B}_4$ are defined analogously. Choose a nonzero basis vector $|m\epsilon_0\rangle$, then $V = \rho(\mathfrak{B})\ |m\epsilon_0\rangle$. Suppose $|m\epsilon_0\rangle' = \rho(X)\ |m\epsilon_0\rangle$ for some X in $\mathfrak{B}$ has weight (m, ϵ_0). We may commute through and thus remove all $E_{-a}(a = 2, 3, 4)$ in X, since these are lowering operators for ϵ. Obviously all $E_a(a = 2, 3, 4)$ must also have disappeared. Further, $[H_2, \mathfrak{B}_1] = 0$ and thus we may take X in $\mathfrak{B}_1$. But $\rho(\mathfrak{B}_1)$ generates an irreducible representation, $\mathfrak{D}_{j_0}$ say, of f_1 with $j_0 < \epsilon_0$. In $\mathfrak{D}_{j_0}$, each weight m is simple. Therefore $|m\epsilon_0\rangle'$ is a multiple of $|m\epsilon_0\rangle$. The rest of the proof follows analogously.

[19] See Ref. 12, Lemma 33.

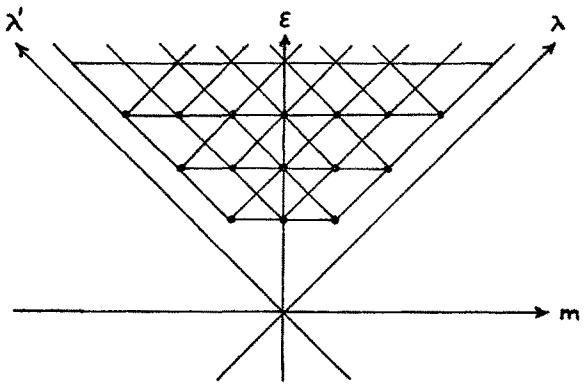

FIG. 2. The weight diagram of a typical representation of type ρ^+.

Let ρ^+, ρ^- denote the irreducible representations ρ with ϵ bounded below and above, respectively. Theorem 1 allows us to construct the form of the weight diagram of ρ^+. Horizontal lines connect members of (f_1, H_2) multiplets and oblique lines connect $f = f_3 + f_4$ multiplets. Corollaries 1 and 2 show that weights on the perimeter are simple and interior weights have finite multiplicities. For ρ^- we reflect the diagram in the m axis.

Decomposition of $\rho|f$

Let $\sigma_{ll'}^+$ be the algebraically irreducible representation $\sigma_l^+ \otimes \sigma_{l'}^+$ of $f = f_3 + f_4$, defined on the Kronecker product space $V_{ll'} = V_l \otimes V_{l'}$(cf., Appendix). For $X = X_3 + X_4$ in f and $|ll'\rangle = |l\rangle \otimes |l'\rangle$ in $V_{ll'}$:

$$\sigma_{ll'}^+(X)\ |ll'\rangle = (\sigma_l^+(X_3)\ |l\rangle) \otimes |l'\rangle + |l\rangle \otimes (\sigma_{l'}^+(X_4)\ |l'\rangle). \quad (29)$$

We can now deduce several results concerning the decomposition of $\rho^+|\ f$. (For $\rho^-|\ f$ the modifications are obvious, with $\sigma_{ll'}^-$ defined analogously.)

(A) The irreducible components are of type $\sigma_{ll'}^+$, by Theorem 1. (The identity representation of f occurs if and only if ρ^+ is the identity representation and we agree to exclude this case from the class $\rho^\pm$.)

(B) Each $\sigma_{ll'}^+$ occurs at most with finite multiplicity, by Theorem 1, Corollary 1.

(C) The $\sigma_{ll'}^+$ are subrepresentations of $\rho^+ \mid f$; that is, V decomposes into the algebraic sum of subspaces $V_{ll'}$ on which $\rho^+ \mid f$ acts irreducibly according

to $\sigma^+_{ll'}$. This is a consequence of (B), Lemma 2(i) and (A7).

(D) Let $(m_0, \epsilon_0)(m_0 = -j_0, -j_0+1, \cdots, j_0)$ be the $2j_0+1$ simple weights of Theorem 1, Corollary 2. Each (m_0, ϵ_0) is a "terminal" weight of the irreducible representation $\sigma^+_{ll'}$ with

$$l = -\tfrac{1}{2}(m_0 + \epsilon_0), \qquad l' = -\tfrac{1}{2}(\epsilon_0 - m_0) \tag{30}$$

and these $2j_0+1$ distinct representations occur once and only once in the decomposition of $\rho^+ \mid f$. We conjecture that, in fact, each $\sigma^+_{ll'}$ occurs at most once in $\rho^+ \mid f$; or equivalently, the operators representing Π, Ω, Π_3, Π_4, H_3, and H_4 form a complete commuting set. While we have no proof of this, we can make it plausible. We have observed that, by counting the parameters in g, we can establish the need for at least six elements in the commuting set.[17] If a further commuting operator was required, then it would depend on f, since we have found $2j_0+1$ of the $\sigma^+_{ll'}$ which occur only once. Such an operator is, necessarily, a function of the previous six operators.[20] Dixmier[21] has proved the analogous result for the irreducible unitary representations of $SO(4, 1)$ decomposed with respect to its maximal compact subgroup $SO(4)$. His proof involves the Iwasawa decomposition of $SO(4, 1)$, in which the maximal compact subgroup occurs explicitly. It does not carry over to our case where the subgroup with respect to which we decompose ρ is a covering group of $SO(2, 2)$. Hereafter we consider only those $\rho^\pm$ for which the conjecture holds. The reasons given above suggest that this is not a restriction at all.

(E) Let $\Gamma = \{(l, l') \mid \sigma^+_{ll'} \text{ occur in } \rho^+ \mid f\}$. Then from (C)

$$V = \sum_{(l,l')\in\Gamma} V_{ll'}. \tag{31}$$

From (17), (29), and (A6), we can choose an orthonormal basis for $V_{ll'}$ so that

$$\begin{aligned}
&\sigma^+_{ll'}(H)\ |l\lambda l'\lambda'\rangle = 2i\lambda\ |l\lambda l'\lambda'\rangle,\\
&\sigma^+_{ll'}(H')\ |l\lambda l'\lambda'\rangle = -2i\lambda'\ |l\lambda l'\lambda'\rangle,\\
&\sigma^+_{ll'}(E_{\pm\alpha})\ |l\lambda l'\lambda'\rangle\\
&\qquad = 2i[(\lambda \mp l)(\lambda \pm l \pm 1)]^{\frac{1}{2}}\ |l\lambda \pm 1 l'\lambda'\rangle,\\
&\sigma^+_{ll'}(E_{\pm\beta})\ |l\lambda l'\lambda'\rangle\\
&\qquad = 2i[(\lambda' \pm l')(\lambda' \mp l' \mp 1)]^{\frac{1}{2}}\ |l\lambda l'\lambda' \mp 1\rangle,
\end{aligned} \tag{32}$$

with

$$\lambda + l = 0, 1, 2, \cdots; \qquad \lambda' + l' = 0, 1, 2, \cdots.$$

These equations are formally identical with those for the case f is the real compact form of f_c.[22]

(F) If we discuss the representation ρ^+ of g on V in terms of the basis $|l\lambda l'\lambda'\rangle$ in which f is diagonal, then we can make a unitary transformation to the basis $|jm\epsilon;\ \alpha\rangle$:

$$|jm\epsilon;\alpha\rangle = \sum_{ll'} \langle l\lambda l'\lambda' \mid jm\epsilon;\alpha\rangle\ |l\lambda l'\lambda'\rangle.$$

The transformation functions vanish for all save a finite number of $(l, l') \in \Gamma$. This diagonalizes the operators $\rho^+(\mathrm{II}_1)$ and $\rho^+(H_2)$. V is dense in $\mathfrak{H}$, and we can use arguments similar to those given by Dixmier[11] to show that there exists a unique irreducible unitary representation, ρ'^+ of G on $\mathfrak{H}$, whose infinitesimal representation is just ρ^+.

Thus, we have completely reduced the problem of constructing the irreducible representations of type $\rho'^\pm$ to a similar one for its Lie algebra g. Now, it is a trivial modification of the proof of Theorem 1 to show that, if at least one irreducible representation of type $\sigma^+(\sigma^-)$ occurs in the decomposition of one of the reducible unitary representations $\rho \mid f_a (a = 2, 3, 4)$, then ρ is of type $\rho^+(\rho^-)$. We know from the work of Harish-Chandra[15] that G possesses a discrete series of unitary irreducible representations ρ' and that on the restriction of ρ' to a subgroup of G which also has a discrete series,[23] all irreducible components belong to its discrete series. (This follows from the square-integrability of members of the discrete series.[15]) Therefore, a classification of representations of type $\rho^\pm$ includes a classification of the discrete series for G. As has already been observed for the universal covering groups of $SO(2, 1)$[24] and $SO(4, 1)$,[11] we must expect that a certain subset of the $\rho^\pm$ will not be square-integrable, and hence will not appear in the Plancherel formula for G.

4. MATRIX ELEMENTS OF THE LIE ALGEBRA g

Hereafter, the representation ρ^+ of g is written simply as ρ. It also denotes the uniquely determined unitary irreducible representation of G. We introduce an orthonormal basis, $|l\lambda l'\lambda'; PW\rangle[(l, l') \in \Gamma; \lambda + l = 0, 1, 2, \cdots; \lambda' + l' = 0, 1, 2, \cdots]$ for the representation space V, corresponding to the decomposition (31), with

[20] There is the possibility that we may not be able to find a complete set of commuting symmetric elements in $\mathfrak{B}$, or that our choice is an inconvenient one.

[21] J. Dixmier, Compt. Rend. **250**, 3263 (1960). The proof is given for the decomposition of irreducible unitary representations of $SO(n, 1)$ with respect to $SO(n)$.

[22] See Ref. 11, Eqs. (13) and (14).

[23] These are the subgroups which possess essentially compact Cartan subgroups, viz. F, F_2, F_3, and F_4.

[24] L. Pukánszky, Math. Ann. **156**, 96 (1964).

$$\rho(\Pi)\ |l\lambda l'\lambda';PW\rangle = P\ |l\lambda l'\lambda';PW\rangle,$$
$$\rho(\Omega)\ |l\lambda l'\lambda';PW\rangle = W\ |l\lambda l'\lambda';PW\rangle, \tag{33}$$

and satisfying (32) with respect to $\rho(f)$. For simplicity, we omit the labels P, W.

The commutations relations (6) to (9) of the subalgebra f_c are automatically satisfied. We use the method of Thomas[6] to determine the matrix elements of the operators $\rho(E_{\pm\gamma})$, $\rho(E_{\pm\delta})$. The end results are given by Dixmier[11] for the case that g_c arises from the Lie algebra of the group $SO(4, 1)$. We sketch a derivation as the range of our eigenvalue is different from that of Thomas and Dixmier. From Eqs. (10) and (11), we find

$$\langle k\nu k'\nu'|\ \rho(E_{\pm\gamma})\ |l\lambda l'\lambda'\rangle = 0, \quad \text{for } \nu \neq \lambda \pm \tfrac{1}{2},\ \nu' \neq \lambda' \pm \tfrac{1}{2},$$
$$\langle k\nu k'\nu'|\ \rho(E_{\pm\delta})\ |l\lambda l'\lambda'\rangle = 0, \quad \text{for } \nu \neq \lambda \pm \tfrac{1}{2},\ \nu' \neq \lambda' \mp \tfrac{1}{2}. \tag{34}$$

This also places a restriction on the (l, k), (l', k') pairs involved in nonzero matrix elements. In fact, we easily deduce

$$l - k \equiv \tfrac{1}{2}(\mathrm{mod}\ 1), \qquad l' - k' \equiv \tfrac{1}{2}(\mathrm{mod}\ 1) \qquad (l, l', k, k' < 0) \tag{35}$$

for the (l, k) and (l', k') pairs occurring in a nonzero matrix element.

The two vanishing commutators in (12) enable us to define reduced matrix elements with respect to λ for $\rho(E_\gamma)$ and $\rho(E_{-\delta})$. Using (32) and (34) we find

$$\langle k\lambda + \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_\gamma)\ |l\lambda l'\lambda'\rangle = \left[\frac{(\lambda + k + \frac{1}{2})!\,(\lambda - k - \frac{1}{2})!}{(\lambda + l)!\,(\lambda - l - 1)!}\right]^{\frac{1}{2}} \times \langle kk'\lambda' + \tfrac{1}{2}||\ \rho(E_\gamma)\ ||ll'\lambda'\rangle,$$
$$\langle k\lambda - \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_{-\delta})\ |l\lambda l'\lambda'\rangle = \left[\frac{(\lambda + l)!\,(\lambda - l - 1)!}{(\lambda + k - \frac{1}{2})!\,(\lambda - k - \frac{3}{2})!}\right]^{\frac{1}{2}} \times \langle kk'\lambda' + \tfrac{1}{2}||\ \rho(E_{-\delta})\ ||ll'\lambda'\rangle, \tag{36}$$

with $x! = (x - 1)!x$.

We show that the left-hand sides of (36) vanish for all k which make the factorials negative. For these values of k, we define the reduced matrix elements to be zero. From each of the remaining two relations in (12) we can determine the ratio of the reduced matrix elements in (36). Then for consistency we require

$$[(l - k)^2 - \tfrac{1}{4}][(l + k + 1)^2 - \tfrac{1}{4}] = 0, \tag{37}$$

that is, the matrix elements (36) vanish unless

$$k = l \pm \tfrac{1}{2},\ -l - \tfrac{1}{2},\ -l - \tfrac{3}{2} \qquad (l < 0, k < 0).$$

Together with (35), it can be seen that $k = l \pm \frac{1}{2}$ gives all allowed k values. Substituting these values of k in (36), we have

$$\langle l - \tfrac{1}{2}\lambda + \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_\gamma)\ |l\lambda l'\lambda'\rangle = (\lambda - l)^{\frac{1}{2}}\langle l - \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}||\ \rho(E_\gamma)\ ||l l'\lambda'\rangle \quad (l < 0),$$
$$\langle l + \tfrac{1}{2}\lambda + \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_\gamma)\ |l\lambda l'\lambda'\rangle = (\lambda + l + 1)^{\frac{1}{2}}\langle l + \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}||\ \rho(E_\gamma)\ ||l l'\lambda'\rangle \quad (l < -\tfrac{1}{2}),$$
$$\langle l - \tfrac{1}{2}\lambda - \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_{-\delta})\ |l\lambda l'\lambda'\rangle = (\lambda + l)^{\frac{1}{2}}\langle l - \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}||\ \rho(E_{-\delta})\ ||l l'\lambda'\rangle \quad (l < 0),$$
$$\langle l + \tfrac{1}{2}\lambda - \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}|\ \rho(E_{-\delta})\ |l\lambda l'\lambda'\rangle = (\lambda - l - 1)^{\frac{1}{2}}\langle l + \tfrac{1}{2}k'\lambda' + \tfrac{1}{2}||\ \rho(E_{-\delta})\ ||l l'\lambda'\rangle \quad (l < -\tfrac{1}{2}). \tag{38}$$

Note that the left-hand side of the third equation in (38) is not defined for $\lambda = -l$ with $-\frac{1}{2} \leqq l < 0$. However, the right-hand side vanishes and we choose to regard the left-hand side as being defined but zero.

The commutation relations (13) lead to similar definitions of reduced matrix elements with respect to λ for $\rho(E_{-\gamma})$ and $\rho(E_\delta)$. Similarly, (14) and (15) allow us to define reduced matrix elements with respect to λ'. We can now combine the four sets of equations analogous to (38) and define four reduced matrix elements with respect to both λ and λ'. Because we have completely parametrized our basis, these final four reduced matrix elements are complex numbers $a_{ll'}$, $b_{ll'}$, $c_{ll'}$, and $d_{ll'}$. In terms of these, we have

$$\begin{aligned}\rho(E_{\pm\gamma})\ |l\lambda l'\lambda'\rangle = {} & \pm[(\lambda \mp l)(\lambda' \mp l')]^{\frac{1}{2}} \\ & \times a_{ll'}\ |l - \tfrac{1}{2}\lambda \pm \tfrac{1}{2}\ l' - \tfrac{1}{2}\lambda' \pm \tfrac{1}{2}\rangle \\ & \pm [(\lambda \pm l \pm 1)(\lambda' \pm l' \pm 1)]^{\frac{1}{2}} \\ & \times b_{ll'}\ |l + \tfrac{1}{2}\lambda \pm \tfrac{1}{2}\ l' + \tfrac{1}{2}\lambda' \pm \tfrac{1}{2}\rangle \\ & \pm [(\lambda \pm l \pm 1)(\lambda' \mp l')]^{\frac{1}{2}} \\ & \times c_{ll'}\ |l + \tfrac{1}{2}\lambda \pm \tfrac{1}{2}\ l' - \tfrac{1}{2}\lambda' \pm \tfrac{1}{2}\rangle \\ & \pm [(\lambda \mp l)(\lambda' \pm l' \pm 1)]^{\frac{1}{2}} \\ & \times d_{ll'}\ |l - \tfrac{1}{2}\lambda \pm \tfrac{1}{2}\ l' + \tfrac{1}{2}\lambda' \pm \tfrac{1}{2}\rangle,\end{aligned} \tag{39}$$

$$\begin{aligned}\rho(E_{\pm\delta})\ |l\ \lambda\ l'\ \lambda'\rangle &= \mp i[(\lambda \mp l)(\lambda' \pm l')]^{\frac{1}{2}} \\ &\times a_{ll'}\ |l - \tfrac{1}{2}\ \lambda \pm \tfrac{1}{2}\ l' - \tfrac{1}{2}\ \lambda' \mp \tfrac{1}{2}\rangle \\ &\mp i[(\lambda \pm l \pm 1)(\lambda' \mp l' \mp 1)]^{\frac{1}{2}} \\ &\times b_{ll'}\ |l + \tfrac{1}{2}\ \lambda \pm \tfrac{1}{2}\ l' + \tfrac{1}{2}\ \lambda' \mp \tfrac{1}{2}\rangle \\ &\mp i[(\lambda \pm l \pm 1)(\lambda' \pm l')]^{\frac{1}{2}} \\ &\times c_{ll'}\ |l + \tfrac{1}{2}\ \lambda \pm \tfrac{1}{2}\ l' - \tfrac{1}{2}\ \lambda' \mp \tfrac{1}{2}\rangle \\ &\mp i[(\lambda \mp l)(\lambda' \mp l' \mp 1)]^{\frac{1}{2}} \\ &\times d_{ll'}\ |l - \tfrac{1}{2}\ \lambda \pm \tfrac{1}{2}\ l' + \tfrac{1}{2}\ \lambda' \mp \tfrac{1}{2}\rangle. \qquad (40)\end{aligned}$$

Remarks:

(1) For $\rho = \rho^-$, (39) remains valid but we must change the over-all sign of the right-hand sides of (40).

(2) The vectors on the right-hand sides may become undefined:

(a) If $(l - \frac{1}{2}, l' - \frac{1}{2}) \notin \Gamma$; then we agree to define $a_{ll'} = 0$. Similarly, if $(l + \frac{1}{2}, l' + \frac{1}{2}) \notin \Gamma$, define $b_{ll'} = 0$, if $(l + \frac{1}{2}, l' - \frac{1}{2}) \notin \Gamma$, define $c_{ll'} = 0$ and if $(l - \frac{1}{2}, l' + \frac{1}{2}) \notin \Gamma$, define $d_{ll'} = 0$.

(b) If $\lambda - \frac{1}{2}$ becomes negative or zero. This can happen only when $\lambda = -l$ with $-\frac{1}{2} \leqq l < 0$. In these cases either the coefficient contains $(\lambda + l)^{\frac{1}{2}}$ and is zero or the vector contains the label $(l + \frac{1}{2}, l' \pm \frac{1}{2})$ not in Γ, in which case $b_{ll'} = c_{ll'} = 0$ by (a). Similarly for $\lambda' - \frac{1}{2} \leqq 0$.

(3) For nonzero reduced matrix elements, the expressions under the square roots are nonnegative.

(4) $2(l - l')$ is either even integer or odd integer throughout ρ. This is equivalent to the statement of Lemma 2 (ii).

For the representation $\rho(G)$ to be unitary, we require that the operators $i\rho(g)$ be essentially self-adjoint on V. From (5), we see that we must apply this condition to the linear combinations $i\rho(E_\gamma + E_{-\gamma})$, $\rho(E_\gamma - E_{-\gamma})$, $\rho(E_\delta + E_{-\delta})$, and $i\rho(E_\delta - E_{-\delta})$ of (39) and (40). We obtain the reality conditions

$$a_{ll'} = \bar{b}_{l-\frac{1}{2},l'-\frac{1}{2}}, \qquad c_{ll'} = \bar{d}_{l+\frac{1}{2},l'-\frac{1}{2}}, \qquad (41)$$

where the bar denotes complex conjugation.

All the commutation relations are satisfied save those of (16), and it is easy to check that only one of these six is now independent. We choose for this

$$[E_\gamma, E_{-\gamma}] = -i(H - H'). \qquad (42)$$

Given $(l, l') \in \Gamma$, we can take diagonal matrix elements of (42), insert a complete set of intermediate states and use (41) to deduce

$$\begin{aligned}&(\lambda l' + \lambda' l)\ |a_{ll'}|^2 - (\lambda(l' + 1) - \lambda'(l + 1))\ |b_{ll'}|^2 \\ &\quad + (\lambda l' - \lambda'(l+1))\ |c_{ll'}|^2 - (\lambda(l' + 1) - \lambda' l)\ |d_{ll'}|^2 \\ &= -(\lambda + \lambda') \\ &(\lambda + l = 0, 1, 2, \cdots\ ;\ \lambda' + l' = 0, 1, 2, \cdots). \qquad (43)\end{aligned}$$

We obtain two further equations involving the quantities $|a_{ll'}|^2, \cdots, |d_{ll'}|^2$ by taking (diagonal) matrix elements of $\rho(\Pi)$ and $\rho(\Omega)$ in (22) and (23). We use (33) to express the irreducibility of ρ and we need (5), (39), (40), and (A6) in the evaluation. This gives the first pair of equations (44), and (43) gives the second pair. For $(l, l') \in \Gamma$,

$$\begin{aligned}&ll'\ |a_{ll'}|^2 + (l + 1)(l' + 1)\ |b_{ll'}|^2 \\ &\qquad - (l + 1)l'\ |c_{ll'}|^2 \\ &\qquad - l(l' + 1)\ |d_{ll'}|^2 \\ &\quad = P/2 + l(l + 1) + l'(l' + 1), \\ &ll'(l - l')^2\ |a_{ll'}|^2 + (l + 1)(l' + 1)(l - l')^2\ |b_{ll'}|^2 \\ &\qquad - (l + 1)l'(l + l' + 1)^2\ |c_{ll'}|^2 \\ &\qquad - l(l' + 1)(l + l' + 1)^2\ |d_{ll'}|^2 \\ &\quad = W/2 + \tfrac{1}{2}(l(l + 1) - l'(l' + 1))^2, \qquad (44) \\ &l\ |a_{ll'}|^2 - (l + 1)\ |b_{ll'}|^2 \\ &\qquad - (l + 1)\ |c_{ll'}|^2 + l\ |d_{ll'}|^2 = -1, \\ &l'\ |a_{ll'}|^2 - (l' + 1)\ |b_{ll'}|^2 \\ &\qquad + l'\ |c_{ll'}|^2 - (l' + 1)\ |d_{ll'}|^2 = -1.\end{aligned}$$

We can write these equations in the matrix form (symbolically)

$$A\mathbf{x} = \mathbf{y} \qquad (45)$$

with $\det A = 2ll'(l + 1)(l' + 1)(2l + 1)^2(2l' + 1)^2$. Since l, l' are negative-definite, Eqs. (44) have a unique solution for all $(l, l') \in \Gamma$, except when

$$(2l + 1)(l + 1)(2l' + 1)(l' + 1) = 0. \qquad (46)$$

When (46) is not satisfied, we can invert (45) to obtain

$$\begin{aligned}|b_{ll'}|^2 &= \frac{(l + l' + 1)(l + l' + 2)[P + (l + l')(l + l' + 3)] - W}{4(l + 1)(l' + 1)(2l + 1)(2l' + 1)}, \\ |d_{ll'}|^2 &= \frac{(l - l' - 1)(l - l')[P + (l + l' - 2)(l - l' + 1)] - W}{4l(l' + 1)(2l + 1)(2l' + 1)}, \qquad (47) \\ |a_{ll'}| &= |b_{l-\frac{1}{2},l'-\frac{1}{2}}|, \qquad |c_{ll'}| = |d_{l+\frac{1}{2},l'-\frac{1}{2}}|.\end{aligned}$$

Equation (42) has four off-diagonal matrix elements but, because of (41), only two are independent. For $(l, l') \in \Gamma$, we find

$$\begin{aligned} &lb_{l-\frac{1}{2},l'+\frac{1}{2}}d_{ll'} \\ &\quad = (l+1)d_{l+\frac{1}{2},l'+\frac{1}{2}}b_{ll'} \quad \text{if} \quad (l, l'+1) \in \Gamma, \\ &l'b_{l+\frac{1}{2},l'-\frac{1}{2}}\bar{d}_{l+\frac{1}{2},l'-\frac{1}{2}} \\ &\quad = (l'+1)\bar{d}_{l+1,l'}b_{ll'} \quad \text{if} \quad (l+1, l') \in \Gamma. \end{aligned} \tag{48}$$

It is seen that these equations will be satisfied by any reasonable choice of phases in (47). Here we choose the positive square roots, which make all the $a_{ll'}$, $b_{ll'}$, $c_{ll'}$, and $d_{ll'}$ real and nonnegative. The form of (47) suggests that we change to the new parameters

$$\begin{aligned} \sigma &= -(l+l'), \\ \delta &= l' - l, \end{aligned} \tag{49}$$

and we say that $(\sigma, \delta) \in \Delta$ if and only if $(l, l') \in \Gamma$. We write the decomposition (31):

$$V = \sum_{(\sigma,\delta)\in\Delta} V_{\sigma\delta}. \tag{50}$$

From remark (4) above and with $l, l' < 0$, we deduce:

$$\begin{cases} \sigma > |\delta|, \\ 2\delta, \quad \text{an even integer throughout} \\ \qquad \text{or odd integer throughout.} \end{cases} \tag{51}$$

Write $a_{\sigma\delta} \equiv a_{ll'}, \cdots, d_{\sigma\delta} \equiv d_{ll'}$ and then $a_{\sigma,-\delta} = a_{l'l}, \cdots, d_{\sigma,-\delta} = d_{l'l}$. With the above choice of phases, (47) and (49) give

$$\begin{aligned} b_{\sigma\delta} &= \left[\frac{(\sigma-1)(\sigma-2)(P+\sigma(\sigma-3))-W}{(\sigma+\delta-1)(\sigma+\delta-2)(\sigma-\delta-1)(\sigma-\delta-2)}\right]^{\frac{1}{2}}, \\ d_{\sigma\delta} &= \left[\frac{\delta(\delta+1)(P+(\delta-1)(\delta+2))-W}{(\sigma+\delta)(\sigma+\delta-1)(\sigma-\delta-1)(\sigma-\delta-2)}\right]^{\frac{1}{2}}, \\ a_{\sigma\delta} &= b_{\sigma+1,\delta}, \qquad c_{\sigma\delta} = d_{\sigma,\delta-1} \end{aligned} \tag{52}$$

as a solution of (44) for all $(\sigma, \delta) \in \Delta'$, where [see (46)]

$$\Delta' = \{(\sigma, \delta) \in \Delta \mid \sigma \pm \delta \neq 1, 2\}. \tag{53}$$

It will be convenient to consider the matrix elements for the singular cases in the next section.

5. CLASSIFICATION OF THE DISCRETE SERIES FOR G

Definitions: A pair $(\sigma, \delta) \in \Delta$ is called a Δ-weight of ρ, and the two-dimensional vector space spanned by the set of Δ-weights is called the Δ-weight space of ρ. The set of all Δ-weights forms the Δ-weight diagram of ρ. It is a consequence of Sec. 3 (D) that every Δ-weight is simple.

With our choice of phases, we can use (39) and (40) to formulate the following criterion. The pair (σ, δ) belongs to Δ only if Eqs. (44) determine nonnegative $a_{\sigma\delta}^2, \cdots, d_{\sigma\delta}^2$, of which at least one is nonzero.

By interchanging l and l' in (44) and (48), and using (41), we obtain

$$a_{\sigma\delta} = a_{\sigma,-\delta}, \qquad b_{\sigma\delta} = b_{\sigma,-\delta}, \qquad c_{\sigma\delta} = d_{\sigma,-\delta} \tag{54}$$

for all $(\sigma, \delta) \in \Delta$. Therefore the Δ-weight diagram is symmetric about the σ axis, and for classification purposes we need consider only $\delta \geq 0$. Also, with our phases, (41) reads

$$a_{\sigma\delta} = b_{\sigma+1,\delta}, \qquad c_{\sigma\delta} = d_{\sigma,\delta-1}. \tag{55}$$

We have previously determined $2j_0 + 1$ Δ-weights (ϵ_0, m_0) of ρ [see (30)] corresponding to the $2j_0 + 1$ values of m_0. Furthermore, the weight diagram of Fig. 1, together with (39) and (40), implies

$$d_{\epsilon_0 j_0} = b_{\epsilon_0 m_0} = 0 \quad (m_0 = j_0, j_0 - 1, \cdots, -j_0). \tag{56}$$

This motivates our next definition.

Definition: $(\sigma, \delta) \in \Delta$ is called an extreme Δ-weight of ρ if and only if

$$b_{\sigma\delta} = d_{\sigma\delta} = 0. \tag{57}$$

One extreme Δ-weight of ρ is (ϵ_0, j_0). Furthermore, we show that this is the only extreme Δ-weight of ρ and that it determines ρ uniquely up to equivalence.

Firstly, let $(q, s) \in \Delta'$ be an extreme Δ-weight of ρ. Applying (57) to (52), we require

$$\begin{aligned} W &= (q-1)(q-2)(P+q(q-3)), \\ W &= s(s+1)(P+(s-1)(s+2)), \end{aligned} \tag{58}$$

that is

$$\begin{aligned} P &= -(s(s+1)+q(q-3)), \\ W &= -s(s+1)(q-1)(q-2), \end{aligned} \tag{59}$$

and [using (51) with $\delta \geq 0$],

$$q > s; \qquad s = 0, \tfrac{1}{2}, 1, \cdots. \tag{60}$$

Substituting (59) in (52),

$$\begin{aligned} a_{\sigma\delta} &= \left[\frac{(\sigma+s)(\sigma-s-1)(\sigma-q+1)(\sigma+q-2)}{(\sigma+\delta)(\sigma+\delta-1)(\sigma-\delta)(\sigma-\delta-1)}\right]^{\frac{1}{2}}, \\ c_{\sigma\delta} &= \left[\frac{(s+\delta)(s-\delta+1)(q-\delta-1)(q+\delta-2)}{(\sigma+\delta-1)(\sigma+\delta-2)(\sigma-\delta)(\sigma-\delta-1)}\right]^{\frac{1}{2}}, \\ b_{\sigma\delta} &= a_{\sigma-1,\delta}, \qquad d_{\sigma\delta} = c_{\sigma,\delta+1}. \end{aligned} \tag{61}$$

Of course, the conditions (60) are not sufficient for (q, s) to belong to Δ', since we must also guarantee that $a^2_{qs} \geq 0$, $c^2_{qs} \geq 0$ and that at least one is nonzero. However, before giving a set of sufficient conditions (Lemma 4), it is useful to have the solutions of (44) for the singular cases $\sigma \pm \delta = 1, 2$, (46). Using (51) with $\delta \geq 0$, we find that only $\sigma - \delta = 1, 2$ are independent, and further that $(\delta + 1, \delta) \in \Delta$ is necessarily an extreme Δ-weight and that $(\delta + 2, \delta)$ is either itself extreme or occurs in the same Δ-weight diagram as an extreme Δ-weight of the first type.

Case 1. $q = s + 1$ $(s = 0, \frac{1}{2}, 1, \cdots)$

With $\sigma = q = s + 1$, $\delta = s$ in (44) and (45), rank (A) = 1 or 2 according as s is zero or not. In either case the consistency condition, rank (A) = rank $(A \mid y)$, is

$$P = -2(s^2 - 1), \qquad W = -s^2(s^2 - 1), \tag{62}$$

and (57) gives

$$b_{qs} = d_{qs} = 0 \quad (q = s + 1; s = 0, \tfrac{1}{2}, \cdots). \tag{63}$$

Note that (62) also follows formally from (59) with $q = s + 1$. For $s = 0$, (54) gives $c_{10} = 0$ and then (44) reduces to

$$a_{10} = (2)^{\frac{1}{2}}. \tag{64}$$

For rank (A) = 2, the two independent equations in (44) give

$$a_{qs} = c_{qs} = 1 \quad (q = s + 1; s = \tfrac{1}{2}, 1, \cdots). \tag{65}$$

Case 2. $\sigma = \delta + 2$ $(\delta = 0, \frac{1}{2}, 1, \cdots)$

Here, rank (A) = 2 or 3 according as δ is zero or not. For $\delta = 0$, the consistency conditions are

$$P = 2, \qquad W = 0, \tag{66}$$

and then from (44) and (54) we obtain

$$a_{20} = 1, \qquad c_{20} = d_{20} = 0, \qquad b_{20} \text{ arbitrary}. \tag{67}$$

For $\delta = \frac{1}{2}, 1, \cdots$, the consistency condition is

$$W = \delta(\delta + 1)(P + (\delta - 1)(\delta + 2)), \tag{68}$$

which is (58) with $s = \delta$, and three of Eqs. (44) are independent.

Suppose, firstly, that $(\sigma = q = s + 2, \delta = s)$ is an extreme Δ-weight. Then, using (57) and the above equations, we can solve (44) uniquely to obtain

$$\begin{aligned} &a^2_{qs} = (s + 1)/(s + 2), \qquad c^2_{qs} = s/(2s + 1), \\ &b_{qs} = d_{qs} = 0, \\ &P = -2(s^2 + s - 1), \\ &W = -s^2(s + 1)^2 \quad (q = s + 2; s = 0, \tfrac{1}{2}, 1, \cdots). \end{aligned} \tag{69}$$

These P, W values are also given by (59), with $q = s + 2$. Thus (59) is true for all extreme Δ-weights (q, s).

Secondly, when $(\delta + 2, \delta)$ is not extreme, it is associated with an extreme Δ-weight $(s + 1, s)$ by (55), with $\delta = s$ or $s - 1$. With P, W given by (62), (68) is satisfied for both these values of δ. Using (55), (64), and (65), we obtain from (44)

$$\begin{aligned} &a_{20} = b_{20} = (2)^{\frac{1}{2}}, \quad c_{20} = d_{20} = 0 \quad (q = 1, s = 0), \\ &a_{q+1,s} = b_{q+1,s} = 1, \qquad c_{q+1,s} = d_{q+1,s} = 0, \\ &a_{q,s-1} = b_{q,s-1} = 0, \qquad c_{q,s-1} = d_{q,s-1} = 1 \\ &\qquad\qquad (q = s + 1; s = \tfrac{1}{2}, 1, \cdots). \end{aligned} \tag{70}$$

This completes the solution of (44) for the singular Δ-weights.

Lemma 3: Let (q, s) be an extreme Δ-weight of ρ. Then

$$\begin{aligned} &b_{q\delta} = 0 \quad \text{for} \quad s - \delta = 0, 1, \cdots, [s] \qquad (\delta \geq 0), \\ &d_{\sigma s} = 0 \quad \text{for} \quad \sigma - q = 0, 1, 2, \cdots, \end{aligned}$$

where $q > s$, $s = 0, \frac{1}{2}, 1, \cdots$ and $[s]$ denotes the greatest integer not exceeding s.

Proof: For the singular case $q = s + 1 (s = 0, \frac{1}{2}, 1, \cdots)$, $b_{q,s-1} = d_{q+1,s} = 0$ from (70). The rest follows immediately from (60) and (61).

Corollary 1: Let (σ, δ) belong to the Δ-weight diagram containing an extreme Δ-weight (q, s). Then the possible (σ, δ) values are

$$\begin{aligned} &\sigma - q = 0, 1, 2, \cdots, \\ &\delta + s = 0, 1, \cdots, 2s. \end{aligned}$$

Proof: This follows from the lemma, using (39), (40), and the symmetry about the σ axis, (54).

Corollary 2: There is one and only one extreme Δ-weight of ρ.

Proof: We know from (56) that (ϵ_0, j_0) is an extreme Δ-weight of ρ. Denote this by (q, s). Suppose $(q', s') \in \Delta$ and is also extreme. We deduce from Corollary 1 that $q' = q$, $s' = s$.

It can be seen from the above results and equations (39), (40), and (54) that all Δ-weight diagrams for g can now be found by determining the signs of $|a_{\sigma\delta}|^2$ and $|c_{\sigma\delta}|^2 (\delta \geq 0)$, for each allowed extreme Δ-weight (q, s). These properties are summarized in the next lemma.

Lemma 4: Given the real number pairs (q, s) $(q > s;\ s = 0, \frac{1}{2}, 1, \cdots)$ and $(\sigma, \delta)(\sigma - q =$

$0, 1, 2, \cdots ; s - \delta = 0, 1, \cdots, [s]$), the solutions $|a_{\sigma\delta}|^2$ and $|c_{\sigma\delta}|^2$ of Eqs. (44) and (59) satisfy the inequalities

(a) $s + 1 > q > s + \frac{1}{2} \quad (s = 0, \frac{1}{2})$,

$q > s + 1 \quad (s = 0, \frac{1}{2}, 1, \cdots)$,

$|a_{\sigma\delta}|^2 > 0$,

$$|c_{\sigma\delta}|^2 = \begin{cases} 0 & (s = 0) \\ >0 & (s = \frac{1}{2}, 1, \frac{3}{2}, \cdots). \end{cases}$$

(b) $q = s + 1$,

$|a_{qs}|^2 > 0$,

$$|c_{qs}|^2 = \begin{cases} 0 & (s = 0) \\ >0 & (s = \frac{1}{2}, 1, \frac{3}{2}, \cdots), \end{cases}$$

and $a_{q\delta} = 0, |c_{q\delta}|^2 > 0, |a_{\sigma s}|^2 > 0, c_{\sigma s} = 0$, otherwise.

(c) $q = s + \frac{1}{2} \quad (s = 0)$.

$|a_{\frac{1}{2}0}|^2 > 0, \quad a_{\frac{3}{2}0} = c_{\frac{1}{2}0} = c_{\frac{3}{2}0} = 0.$

(d) $q = s + \frac{1}{2} \quad (s = \frac{1}{2})$.

$a_{1\frac{1}{2}} = 0, \quad |c_{1\frac{1}{2}}|^2 > 0.$

(e) $s + \frac{1}{2} > q > s \quad (s = 0, \frac{1}{2})$,

$s + 1 > q > s \quad (s = 1, \frac{3}{2}, 2, \cdots)$.

$|a_{q0}|^2 > 0, \quad |a_{q+1,0}|^2 < 0,$

$|a_{q\frac{1}{2}}|^2 < 0; \quad |c_{qs}|^2 > 0, \quad |a_{q,s-1}|^2 < 0.$

Proof: For all nonsingular (σ, δ), we can use (60) and the given conditions to deduce

$$|a_{\sigma\delta}|^2 > 0 \Leftrightarrow (\sigma - s - 1)(\sigma + q - 2) \times (\sigma + \delta - 1)(\sigma - \delta - 1) > 0,$$

$$|c_{\sigma\delta}|^2 > 0 \Leftrightarrow (q - \delta - 1)(q + \delta - 2)(\sigma - \delta - 1) \times (\sigma + \delta - 1)(\sigma + \delta - 2) > 0 \quad (s \neq 0).$$

We consider $s = 0$, and $s = \frac{1}{2}$ separately.

$s = 0$. Then $\delta = 0, c_{\sigma 0} = 0$ and

$$|a_{\sigma 0}|^2 > 0 \Leftrightarrow (\sigma + q - 2)(\sigma - 1) > 0$$

$$\Leftrightarrow q > 1, \quad 1 > q > \tfrac{1}{2}$$

except $\{\sigma = q \mid q = s + 2 = 2\}$.

Also $|a_{\frac{1}{2}0}|^2 > 0, a_{\frac{1}{2}0} = 0$ for $q = s + \frac{1}{2} = \frac{1}{2}$.

$s = \frac{1}{2}$. Then $\delta = \frac{1}{2}$ and

$$\left.\begin{matrix} |a_{\sigma\frac{1}{2}}|^2 > 0 \\ |c_{\sigma\frac{1}{2}}|^2 > 0 \end{matrix}\right\} \Leftrightarrow \left\{\begin{matrix} (\sigma + q - 2)(\sigma - \frac{1}{2}) > 0 \\ (\sigma - \frac{1}{2}) > 0 \end{matrix}\right\}$$

$$\Leftrightarrow q > 1$$

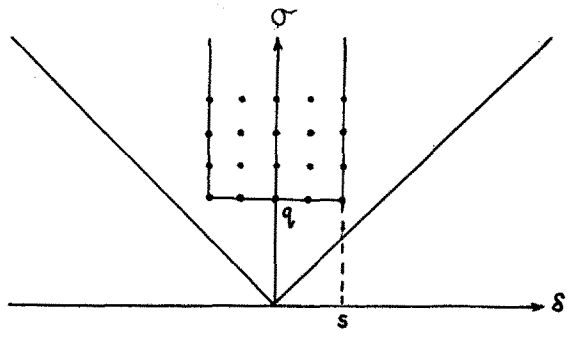

FIG. 3. The Δ-weight diagram of ρ_{qs}^{+} with $q > s + \frac{1}{2}$ $(s = 0, \frac{1}{2})$, $q > s + 1$ $(s = 1, 3/2, 2, \ldots)$.

except $\{\sigma = q, q + 1 \mid q = s + 1 = \frac{3}{2}\}$

and $\{\sigma = q \mid q = s + 2 = \frac{5}{2}\}$.

Also $a_{1\frac{1}{2}} = 0, |c_{1\frac{1}{2}}|^2 > 0$, for $q = s + \frac{1}{2} = 1$.

$s = 1, \frac{3}{2}, 2, \cdots$. Here $(\sigma + \delta - 1) > 0$, $(\sigma + q - 2) > 0$ and therefore

$$|a_{\sigma\delta}|^2 > 0 \Leftrightarrow (\sigma - s - 1)(\sigma - \delta - 1) > 0$$

$$\Leftrightarrow q > s + 1 \text{ except } \{\sigma = q \mid q = s + 2\},$$

and this also guarantees

$$|c_{\sigma\delta}|^2 > 0.$$

Equations (69) extend these results to the singular cases, $\{\sigma = q \mid q = s + 2\}$ excepted above. This already proves (a), (c), and (d). Result (b) follows immediately from (64) and (65) for $(\sigma = q, \delta = s)$, from (70) for $(\sigma = q + 1, \delta = s)$ and $(\sigma = q, \delta = s - 1)$, and from (61) otherwise. Finally, (e) is easily verified by using (61).

Using Lemmas 3 and 4, we can now draw a set of typical Δ-weight diagrams of unitary irreducible representations of G of type ρ^{+}. Figures 3, 4, 5, and 6 correspond, respectively, to (a), (b), (c), and (d) of Lemma 4.

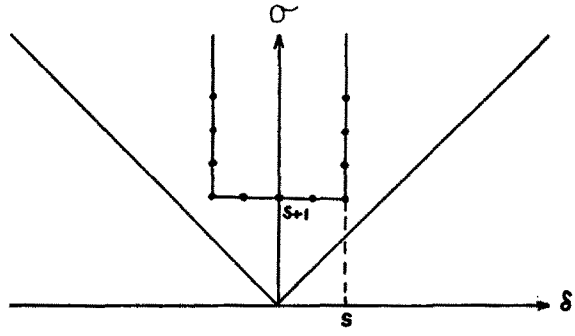

FIG. 4. The Δ-weight diagram of ρ_{qs}^{+} with $q = s + 1$ $(s = 1, 3/2, 2, \ldots)$.

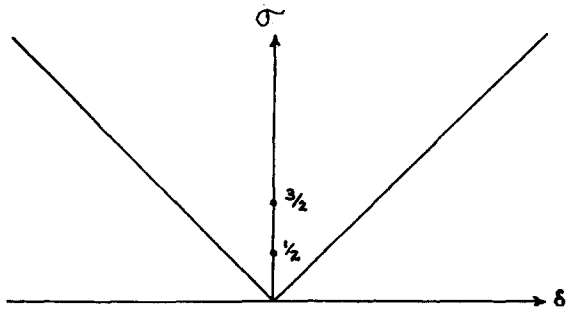

FIG. 5. The Δ-weight diagram of $\rho_{qs}^{\pm}$ with $q = s + \frac{1}{2}$ $(s = 0)$.

We can follow through the same classification procedure for the representations of type ρ^- and we arrive at an identical set of Δ-weight diagrams. However, the weight diagrams of ρ^+ and ρ^- are related by reflection in the "m" axis. We assemble these results in the next theorem.

Theorem 2: Each pair $(q, s)[q \geq s + \frac{1}{2}(s = 0, \frac{1}{2}), q \geq s + 1(s = 1, \frac{3}{2}, 2, \cdots)]$ determines two inequivalent unitary irreducible representations $\rho_{qs}^{\pm}$ of G. To each (σ, δ) in Δ_{qs}, there corresponds a unitary irreducible subrepresentation $\sigma_{ll'}^{\pm}$ of $\rho_{qs}^{\pm} \mid F$ with $2l = -(\sigma + \delta)$, $2l' = \delta - \sigma$. We distinguish four cases:

(a) $q > s + \frac{1}{2}$ $(s = 0, \frac{1}{2})$,

$q > s + 1$ $(s = 1, \frac{3}{2}, 2, \cdots)$

(cf. Fig. 3),

$$(\sigma, \delta) \in \Delta_{qs} \Leftrightarrow \begin{cases} \sigma - q = 0, 1, 2, \cdots, \\ \delta + s = 0, 1, \cdots, 2s \quad \text{for each} \quad \sigma. \end{cases}$$

(b) $q = s + 1$ $(s = 1, \frac{3}{2}, 2, \cdots)$

(cf. Fig. 4),

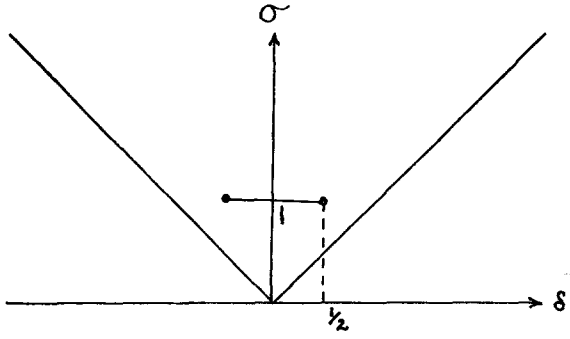

FIG. 6. The Δ-weight diagram of $\rho_{qs}^{\pm}$ with $q = s + \frac{1}{2}$ $(s = \frac{1}{2})$.

$$(\sigma, \delta) \in \Delta_{qs} \Leftrightarrow \begin{cases} \sigma - q = 0, \pm 1, \pm 2, \cdots, \\ \delta + s = \begin{cases} 0, \pm 1, \cdots, 2s & \text{for} \quad \sigma = q, \\ 0, 2s & \text{for each} \quad \sigma \neq q. \end{cases} \end{cases}$$

(c) $q = s + \frac{1}{2}$ $(s = 0)$ (cf. Fig. 5),

$$(\sigma, \delta) \in \Delta_{qs} \Leftrightarrow \begin{cases} \sigma = \frac{1}{2}, \frac{3}{2}, \\ \delta = 0. \end{cases}$$

(d) $q = s + \frac{1}{2}$ $(s = \frac{1}{2})$ (cf. Fig. 6),

$$(\sigma, \delta) \in \Delta_{qs} \Leftrightarrow \begin{cases} \sigma = 1, \\ \delta = \frac{1}{2}, -\frac{1}{2}. \end{cases}$$

In each case,

$$\begin{aligned} \rho_{qs}^{\pm}(\Pi) &= -(s(s+1) + q(q-3))1, \\ \rho_{qs}^{\pm}(\Omega) &= -s(s+1)(q-1)(q-2)1. \end{aligned} \tag{59}$$

This exhausts the unitary irreducible representations of G in which ϵ is bounded below or above.

Corollary: The unitary irreducible representations ρ_{qs}^{+}, $\rho_{q's'}^{+}$ of G are equivalent if and only if $q = q'$, $s = s'$. Similarly, for ρ_{qs}^{-}, $\rho_{q's'}^{-}$.

It is also useful to know the composition of Λ for the decomposition of the representation spaces of $\rho_{qs}^{\pm}$ into subspaces invariant under the maximal essentially compact subgroup K. From our knowledge of the Δ_{qs}-weight diagram of $\rho_{qs}^{\pm}$ and the weight diagrams of $\sigma_{ll'}^{\pm}$, we can easily construct the weight diagrams of $\rho_{qs}^{\pm}$, including the multiplicities of the degenerate weights. Let us define the multiplicity of $(j, \epsilon) \in \Lambda$ as $(2j + 1)^{-1} \dim V_{j,\epsilon}$ [see (26)]. The (j, ϵ) content of $\rho_{qs}^{\pm}$, together with the multiplicities can be read from the weight diagrams. The cases (a)–(d) below correspond to (a)–(d) of Theorem 2, and the $\pm\epsilon$ correspond to $\rho_{qs}^{\pm}$.

(j, ϵ) content of $\rho_{qs}^{\pm}$.

(a) $2s = 1, 3, 5, \cdots$.

$$j - k = 0, 1, 2, \cdots \quad (2k = 1, 3, \cdots, 2s)$$

$$\pm\epsilon - (q + j + s - 2k) = 0, 1, 2, \cdots \text{ for each } j, k.$$

$$s = 0, 1, 2, \cdots$$

$$j - k = 0, 1, 2, \cdots \quad (k = 1, 2, \cdots, s)$$

$$\pm\epsilon - (q + j + s - 2k) = 0, 1, 2, \cdots \text{ for each } j, k,$$

$$\text{and } j = 0, 1, 2, \cdots$$

$$\pm\epsilon - (q + j + s) = 0, 2, 4, \cdots \text{ for each } j.$$

(b) $2s = 2, 3, 4, \cdots$

$$j - s = 0, 1, 2, \cdots,$$

$$\pm\epsilon - j = 1, 2, 3, \cdots \text{ for each } j.$$

(c) $s = 0$

$$j = 0, 1, 2, \cdots,$$

$$\pm\epsilon = j + \tfrac{1}{2} \text{ for each } j.$$

(d) $s = \tfrac{1}{2}$

$$2j = 1, 3, 5, \cdots,$$

$$\pm\epsilon = j + \tfrac{1}{2} \text{ for each } j.$$

Remarks: For the $\rho^{\pm}_{qs}$ in (a), the greatest (j, ϵ) multiplicity is $[s] + 1$, i.e., they are singletons if and only if $s = 0, \frac{1}{2}$. The representations in (b), (c), and (d) are also singletons. The singletons of type $\rho^{\pm}$ given in Sec. 7, (3) of Ehrman[7] are precisely those given above. Our Eqs. (59) can be obtained from his Eqs. (20) and (21) by substituting $\nu_1 = q$, $\nu_2 = s + 1$, $R = P$, and $P_1 = -W$. It can then be checked that the $\rho^{\pm}_{qs}$ indeed satisfy 2, 3, or 4 of his reducibility conditions, (24) or (25). His conjecture that singletons are the only unitary irreducible representations of G in which the (j, ϵ) multiplicity has an upper bound independent of both j and ϵ is contradicted by our result for $\rho^{\pm}_{qs}$ in (a), with $2s = 2, 3, 4, \cdots$. Explicit realizations of the representations (c) and (d) have been given by Dirac.[25]

6. INÖNÜ-WIGNER CONTRACTIONS OF G

In order to discuss such contractions of G, it is useful to introduce the following symmetric operators on the representation space, V:

$$\begin{aligned} P_0 &= i\omega M_{05}, \\ \mathbf{P} = (P_i) &= \frac{i\omega}{c}(M_{i5}), \\ \mathbf{L} = (L_i) &= \frac{i}{c}(M_{i0}), \\ \mathbf{J} = (J_i) &= \frac{i}{2}(\epsilon_{ijk}M^{jk}), \\ & (i = 1, 2, 3), \end{aligned} \tag{71}$$

where c, ω are real parameters.

Hereafter, we let the same symbol denote an element of g_c and its representing operator on V. The multiplication table for g_c now reads

$$\begin{aligned} [J_i, J_j] &= i\epsilon_{ijk}J_k, & [P_i, P_j] &= -i\omega^2/c^2\ \epsilon_{ijk}J_k, \\ [J_i, L_j] &= i\epsilon_{ijk}L_k, & [P_0, P_i] &= i\omega^2 L_i, \\ [J_i, P_j] &= i\epsilon_{ijk}P_k, & [L_i, P_j] &= i/c^2\ \delta_{ij}P_0, \\ [L_i, L_j] &= -i/c^2\ \epsilon_{ijk}J_k, & [L_i, P_0] &= iP_i \\ & & & (i, j, k = 1, 2, 3), \end{aligned} \tag{72}$$

with all other commutators zero.

The contraction of the de Sitter group to the Galilei group is a two-step process associated with the limits $\omega \to 0$ and $c \to \infty$. The intermediate group in this chain will obviously depend on the order in which we take these limits.

Usually we first let $\omega \to 0$, corresponding to the curvature of the de Sitter space tending to zero, and we obtain the Lie algebra of the Poincaré group. In the standard way,[1] we obtain faithful unitary representations of the Poincaré group by considering the limit of a sequence of representations ρ^{+}_{qs}. (We consider only the positive energy representations throughout.) Substituting from (71) into the expressions (22) and (23) for Π and Ω, we find the Poincaré invariants $\tilde{P}$ and $\tilde{W}$ given by

$$\begin{aligned} \tilde{P} &= -\lim_{\omega\to 0}(\omega^2\Pi), \\ \tilde{W} &= \lim_{\omega\to 0}(\omega^2\Omega). \end{aligned} \tag{73}$$

For the representation ρ^{+}_{qs}, we have from (59):

$$\begin{aligned} \tilde{P} &= \lim_{\omega\to 0}\{\omega^2(s(s+1) + q(q-3))\}, \\ \tilde{W} &= -\lim_{\omega\to 0}\{\omega^2 s(s+1)(q-1)(q-2)\}. \end{aligned} \tag{74}$$

With q, s always positive, we see that we never obtain the nonphysical representations of the Poincaré group with $\tilde{P} < 0$. The real, nonzero mass representations P_s are obtained from a sequence of ρ^{+}_{qs} of type (a) by taking $q \to \infty$ such that $\omega q \to M$, the mass of the representation, and with s remaining fixed and determining the spin of P_s. The zero mass representation O_s, with $2s$ an integer, occurs as an irreducible component of the contraction of $\rho^{+}_{q|s|}$. Finally, we can obtain the classes $O(\Xi)$, $O'(\Xi)$ of mass zero representations by taking $q \to \infty$, $s \to \infty$ such that $\omega q \to 0$, $\omega s \to 0$ while $\omega qs \to \Xi$. These limiting procedures definitely associate the class O_s, not the class $O^{(\prime)}(\Xi)$, with the zero mass limit of the class P_s. No other unitary irreducible representations of G contract to P_s. The imaginary mass class is obtained by contracting irreducible representations classified by Ehrman.[7] We can also give explicitly the matrix elements of the Lie algebra of the Poincaré group by contracting the matrix

[25] P. A. M. Dirac, J. Math. Phys. 4, 901 (1963).

elements of $\rho^{+}_{a^2}(g)$.[26] However, the basis for the contracted representation space turns out to be most unusual. It is given by the simultaneous eigenvectors of the two invariants $\tilde{P}$, $\tilde{W}$ and of the operators representing P_0, P_3, J_3, and $\tilde{W}_3$, the third component of the intrinsic angular momentum $\tilde{W}_\mu$. This undesirable basis arises because we have reduced V out with respect to subspaces invariant under transformations which leave the quadratic form $x_5^2 + x_0^2 - x_1^2 - x_2^2$ invariant. Further contraction to the physical (ray) representations of the Galilei group, by taking $c \to \infty$, is well known.[1]

Contracting in the reverse order by first taking $c \to \infty$, we arrive at a transformation group of a curved space which stands in the same relation to the Galilei group as does the 3 + 2 de Sitter group to the Poincaré group. Because it differs from the Galilei group only through the commutators $[P_0, P_i] = i\omega^2 L_i (i = 1, 2, 3)$, we call this nonrelativistic de Sitter group the oscillator group. It has many features in common with the Galilei group, e.g., it possesses a one-dimensional infinity of equivalence classes of unitary ray representations. The representations $\rho^{+}_{a^2}$ contract to these ray representations, provided we first renormalize the operator P_0. The further contraction, $\omega \to 0$, to (ray) representations of the Galilei group is of the normal Inönü–Wigner type.

ACKNOWLEDGMENTS

The author wishes to thank Dr. J. Goldstone for his interest in this work and Unilever (N. Z.) Ltd. for the award of a scholarship.

APPENDIX. UNITARY IRREDUCIBLE REPRESENTATIONS OF THE UNIVERSAL COVERING GROUP OF $SO(2,1)$

It is well known that there exists a 1–1 correspondence between these representations and algebraically irreducible representations of the Lie algebra in which every element of the algebra is represented by a skew-Hermitian operator.[24,27] f_3[cf.(4)] is such an algebra and we use it to demonstrate some of the properties of these representations.[24,27,28]

The multiplication table for f_{3c} is

$$[H_3, E_{\pm 3}] = \pm E_{\pm 3}, \qquad [E_3, E_{-3}] = 2H_3. \tag{A1}$$

Let σ be a representation of f_3 of the above type and let V_3 be the representation space. In σ, the Casimir operator takes a fixed real value. A basis for V_3 is formed by the simultaneous eigenvectors $|Q, \lambda\rangle$ of $\sigma(\Pi_3)$ and $\sigma(H_3)$

$$\begin{aligned} \sigma(\Pi_3)\ |Q, \lambda\rangle &= Q\ |Q, \lambda\rangle, \\ \sigma(H_3)\ |Q, \lambda\rangle &= \lambda\ |Q, \lambda\rangle. \end{aligned} \tag{A2}$$

The irreducible representations σ are specified by Q and the range of λ. They fall into a continuous class and a discrete class. Within the continuous class there exist two distinct series, the principal series and the supplementary series.

Continuous Class

If λ is an eigenvalue of $\sigma(H_3)$, then so is $\lambda + n$ for n integer; that is, the spectrum of $\sigma(H_3)$ is unbounded above and below. Each eigenvalue is nondegenerate. (A3)

Q can take a continuous range of real values and is not determined by λ, except that λ determines the upper bound of Q for the supplementary series. (A4)

Discrete Class

It is composed of two inequivalent series σ^{+}_{l} and σ^{-}_{l}, $Q = -l(l+1)$, l real and negative definite. The two series are distinguished by the ranges of λ:

$$\begin{aligned} &\sigma^{+}_{l};\ \ \lambda + l = 0, 1, 2, \cdots, \\ &\sigma^{-}_{l};\ \ \lambda - l = 0, -1, -2, \cdots. \end{aligned} \tag{A5}$$

Thus the eigenvalues of $\sigma^{+}_{l}(H_3)$ are bounded below and positive and reversely for $\sigma^{-}_{l}(H_3)$.

By a choice of phase, we have

$$\begin{aligned} \sigma^{\pm}_{l}(H_3)\ |l, \lambda\rangle &= \lambda\ |l, \lambda\rangle, \\ \sigma^{\pm}_{l}(iE_{\pm 3})\ |l, \lambda\rangle &= [(\lambda \mp l)(\lambda \pm l \pm 1)]^{\frac{1}{2}}\ |l, \lambda \pm 1\rangle, \\ \sigma^{\pm}_{l}(\Pi_3)\ |l, \lambda\rangle &= -l(l+1)\ |l, \lambda\rangle. \end{aligned} \tag{A6}$$

Remarks

For σ^{+}_{l} (respectively, σ^{-}_{l}), $-l = \min \lambda$ (respectively, $l = \max \lambda$). (A7)

The factors under the square root are (separately for σ^{+}_{l}, collectively for σ^{-}_{l}) nonnegative, and we mean the positive root. Thus $\sigma^{\pm}_{l}(E_3)^{\dagger} = -\sigma^{\pm}_{l}(E_{-3})$. (A8)

$\sigma^{\pm}_{l}$ are square integrable in the sense of Bargmann[27] for $l \leqq -\frac{1}{2}$. They occur in the direct integral reduction of the regular representation. (A9)

There is a formal similarity of (A6) to the case in which H_3, $E_{\pm 3}$ arise from an irreducible representation of the compact form. (A10)

[26] See S. Ström, Arkiv Fysik **30**, 455 (1965), for similar results for the 4 + 1 de Sitter group.
[27] V. Bargmann, Ann. Math. **48**, 568 (1947).
[28] A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) **A287**, 532 (1965).

Structure of the Crossing Matrix for Arbitrary Internal Symmetry Groups. II. Matrices in $SU(n)$

D. B. FAIRLIE
Department of Mathematics, University of Durham, Durham, England
(Received 4 April 1966)

It is shown that in $SU(n)$ two distinct processes can give rise to identical crossing matrices in general. In $SU(2)$ and $SU(3)$, however, the crossing matrix corresponds to a unique scattering process, and thus the dynamical equations in the elastic channel are sufficient to discriminate between different processes. Some $SU(n)$ crossing matrices, which generalize the work of Cook, Muturza, and Rashid, are given.

1. INTRODUCTION

In the last few years the fundamental role of the crossing matrix, as the intermediary through which properties of the internal symmetry group of elementary particles affect dynamical calculations, has been recognized more fully[1-3] and attention has shifted from utilitarian calculations of numerical values of specific crossing matrices to the study of more general features of their structure.[4-7] Recently Cook, Muturza, and Rashid[6] and Mani, Mohan, Pande, and Singh[7] have obtained several crossing matrices for elastic scattering of particles transforming under totally symmetric, totally antisymmetric, or regular representations of $SU(n)$, and their results have a somewhat surprising simplicity. The existence of these calculations in $SU(n)$ prompts the question: is there any distinguishing feature of the results for $n = 2, 3$ which might give some clue as to why $SU(2)$ and $SU(3)$ appear to be the symmetry groups favored in nature? In this paper it is shown by an explicit calculation that in $SU(n)$ for $n > 3$, two different scattering processes may have exactly the same crossing matrix; and hence share the same Low equation. Explicitly the elastic scattering of a particle transforming as a representation described by the Young diagram $[1^h]$ (i.e., h boxes in a vertical line) off a particle transforming as a totally symmetric representation with k boxes in a line $[k]$, is found to give the same crossing matrix as that of a quark scattering off a representation described by a Young diagram consisting of a rectangle of length k boxes, height h boxes in $SU(n)$. In $SU(2)$ and $SU(3)$ these two processes are identical. In these groups alone, therefore, the Low equations are sufficient to determine the process uniquely. (The internal symmetries considered here exclude spin, which couples directly to the angular momentum.)

In the remainder of this paper some calculations are presented which generalize the crossing matrix for [3] off [3] in $SU(n)$ given by Cook, Muturza, and Rashid.[6] Matrices are given for the elastic scattering of a representation of width $r (r = 1, 2, 3)$ boxes, height h off a completely symmetric representation [k]. These matrices possess remarkable symmetry properties, and show much more clearly than before the pattern of these calculations with increasing r. Finally a method of obtaining the crossing matrix to the inelastic t channel for these processes is given.

2. CROSSING MATRIX PROPERTIES

In the case of elastic processes with only two amplitudes, the crossing matrix is completely determined by the dimensions of the irreducible representations in the direct channel N_i, say, and the crossed channel M_i, say.[3] Then the crossing matrix is

$$\frac{1}{M_1 + M_2}\begin{bmatrix} M_1 - \rho N_1^{-1} M_2 + \rho N_1^{-1} \\ M_1 + \rho N_2^{-1} M_2 - \rho N_2^{-1} \end{bmatrix} \tag{1}$$

with $\rho^2 = M_1 M_2 N_1 N_2$.

In the scattering

$$[1^h] + [k] \to [1^h] + [k] \tag{2}$$

the irreducible representations occurring in the S channel are $[k + 1, 1^{h-1}]$ and $[k, 1^h]$, while in the U channel they are $[k + 1, 1^{n-h-1}]$ and $[k, 1^{n-h}]$, respectively. The crossing matrix for the process is given in Table I, while the dimensionalities of

[1] J. Rotheitner, Z. Physik 177, 287 (1964).
[2] A. W. Martin and W. D. McGlinn, Phys. Rev. 136, B1515 (1964).
[3] D. B. Fairlie, Nuovo Cimento 43, 859 (1966).
[4] D. B. Fairlie, J. Math. Phys. 7, 811 (1966).
[5] D. E. Neville, Phys. Rev. 132, 844 (1963).
[6] C. L. Cook, G. Muturza, and M. A. Rashid, Nuovo Cimento 41, 122 (1966).
[7] H. S. Mani, G. Mohan, L. K. Pande, and V. Singh, Ann. Phys. (N. Y.) 36, 285 (1966).

TABLE I. $[1^h] + [k] \to [1^h] + [k]$.

	$[k+1,1^{n-h-1}]$	$[k,1^{n-h}]$
$[k+1,1^{h-1}]$	$\frac{n-h}{n+k-h}$	$\frac{k}{n+k-h}$
$[k,1^h]$	$\frac{n+k}{n+k-h}$	$\frac{-h}{n+k-h}$

the irreducible representations are listed in Table IV. Similarly, the irreducible representations in the S channel for the process

$$[1] + [k^h] \to [1] + [k^h] \tag{3}$$

are $[k+1, k^{h-1}]$ and $[k^h, 1]$, respectively, while the U channel representations are $[(k+1)^h, 1^{n-h-1}]$ and $[k^{h-1}, k-1]$. The dimensionalities are given in Table IV, and the crossing matrix turns out to be identical with that for process (2). This circumstance serves to distinguish the cases when $n = 2$ and $n = 3$, from that of general n, for in $SU(2)$ the only nontrivial value of h permitted is 1, and in $SU(3)$, h can be 1 or 2, but $[k^2]$ is the conjugate representation to $[k]$. In both cases processes (2) and (3) are identical, and are the crossed channel processes for

$$[1] + [k] \to [1] + [k], \tag{4}$$

where in (2) the quark has been crossed, and in (3) the representation $[k]$ is the one transferred.

Since the crossing matrix, and not the particle multiplicities enters into Low equations for scattering, nothing distinguishes the scattering of the generally different processes (2) and (3) in this approximation. This is a theoretical objection, though obviously not a crucial one, to the application of special unitary groups with $n > 3$, such as $SU(4)$, favored by many authors[8,9], to the internal symmetry structure of elementary particles. Of course, one must emphasize here that spin is not regarded as an internal symmetry in this context. It is easy to conjecture an obvious generalization of the above example by considering the scattering of two representations with rectangular Young diagrams, again the possible processes are equivalent when $n \leq 3$.

The writer has obtained the solution of the Low equation[3] for process (3) in the case $h = 1$, using a two parameter crossing matrix. It turns out that if k is an integral multiple of n, the solution for the s matrix elements takes the form of rational fractions of two polynomials in a single function, analytic in the cut energy plane. Putting the parameters from the crossing matrix into Eq. (17) of Ref. 3 it is found that for the more general process (3) with arbitrary k the criterion for a solution of the above type is unaltered, i.e., does not depend upon k.

3. SPECIFIC $SU(n)$ CROSSING MATRICES

Crossing matrices for processes

$$[2^h] + [k] \to [2^h] + [k] \tag{5}$$

and

$$[3^h] + [k] \to [3^h] + [k] \tag{6}$$

are given in Tables II and III. They are constructed by combining the matrix for [3] + [3] given by Cook and Muturza[6] with the matrices for (5) and (6) for the case $n = 2$, $h = 1$, k arbitrary, which are obtainable from tables of Racah coefficients,[10] since the isotopic spin matrices are simply related to Racah coefficients for $SU(2)$.

The generalization to arbitrary h is achieved by requiring that the matrix for process (5) with $k = 1$ is identical with that for process (3) with $k = 2$, and by insisting that the matrices satisfy the following necessary requirements[4,6]:

$$\sum_i C_{ij} = 1, \tag{7}$$

$$C_{ij}^{-1} = (N_i/M_j)C_{ij} \qquad (\text{all } i, j), \tag{8}$$

$$C_{ij}^{-1}(h) = C_{ij}(n-h). \tag{9}$$

Equation (10) follows from the observation[1,4,11] that any matrix C may be written as

$$C_{ij} = N_i^{-\frac{1}{2}} O_{ij} M_j^{\frac{1}{2}}, \tag{10}$$

where N_i and M_j are the dimensionalities of the irreducible representations in the direct and crossed channels, respectively, and O_{ij} is an orthogonal matrix. In Eq. (11) $C_{ij}(h)$ stands for the matrix for processes (2), (5), (6) while $C_{ij}(n-h)$ stands for the same processes with h replaced by $n-h$. But $[3^n]$ and $[3^{n-h}]$ are conjugate representations, hence (11) follows. These requirements are all met, and the equivalence of matrix IV with $k = 2$ with that of III with $k = 3$ gives another example of

[8] I. S. Gerstein and M. L. Whippman, Phys. Rev. **134**, B1123 (1964).
[9] P. Tarjanne and V. Teplitz, Phys. Rev. Letters **11**, 447 (1963); J. P. Antoine, D. Speiser, and R. J. Oakes, Phys. Rev. **141**, 1542 (1966).
[10] L. C. Biedenharn, J. M. Blatt, and M. E. Rose, Rev. Mod. Phys. **24**, 249 (1952).
[11] L. L. Foldy and R. F. Peierls, Phys. Rev. **130**, 1585 (1963).

TABLE II. $[2^h] + [k] \to [2^h] + [k]$.

	$[k+2,2^{n-h-1}]$	$[k+1,2^{n-h-1},1]$	$[k,2^{n-h}]$
$[k+2,2^{h-1}]$	$\frac{(n-h)(n-h+1)}{(n+k-h)(n+k-h+1)}$	$\frac{2k(n-h)}{(n+k-h+1)(n+k-h-1)}$	$\frac{k(k-1)}{(n+k-h)(n+k-h-1)}$
$[k+1,2^{h-1},1]$	$\frac{(n-h+1)(n+k+1)}{(n+k-h)(n+k-h+1)}$	$\frac{k(k+n)-(h+1)(n-h+1)}{(n+k-h+1)(n+k-h-1)}$	$\frac{(h+1)(k-1)}{(n+k-h)(n+k-h-1)}$
$[k\ 2^h]$	$\frac{(n+k+1)(n+k)}{(n+k-h)(n+k-h+1)}$	$\frac{-2h(n+k)}{(n+k-h+1)(n+k-h-1)}$	$\frac{h(h+1)}{(n+k-h)(n+k-h-1)}$

the general conjecture of the equivalence of crossing matrices for the processes

$$[k^h] + [k'^{h'}] \to [k^h] + [k'^{h'}], \tag{11}$$

$$[k'^{h}] + [k^{h'}] \to [k'^{h}] + [k^{h'}]. \tag{12}$$

In this example $k' = 3$, $h' = 1$, $k = 2$.

In these matrices the pattern of dependence of the elements on n, k, and h for more general processes begins to be apparent. The matrices I, II, and III satisfy some curious symmetry properties. If the order of the matrix is denoted by N, they all satisfy the properties

$$C_{iN-i}(n, k, h) = C_{ii}[n, -(n-h), (n+k)], \tag{13}$$

$$C_{N-i,i}(n, k, h) = C_{ii}(n, -h, -k). \tag{14}$$

These relations were obtained empirically.

4. CROSSING MATRICES CONNECTING THE INELASTIC CHANNEL

So far the matrices given relate the two elastic channels, say the S and the U channels in an elastic

TABLE III. $[3^h] + [k] \to [3^h] + [k]$.

	$[k+3,3^{n-h-1}]$	$[k+2,3^{n-h-1}]$
$[k+3,3^{h-1}]$	$\frac{(n-h)(n-h+1)(n-h+2)}{(n+k-h)(n+k-h+1)(n+k-h+2)}$	$\frac{3k(n-h)(n-h+1)}{(n+k-h-1)(n+k-h)(n+k-h+2)}$
$[k+2,3^{h-1},1]$	$\frac{(n-h+1)(n-h+2)(n+k+2)}{(n+k-h)(n+k-h+1)(n+k-h+2)}$	$\frac{(n-h+1)[2k(k+n+1)-(h+2)(n-h+2)]}{(n+k-h-1)(n+k-h)(n+k-h+2)}$
$[k+1,3^{h-1},2]$	$\frac{(n-h+2)(n+k+2)(n+k+1)}{(n+k-h)(n+k-h+1)(n+k-h+2)}$	$\frac{(n+k+1)[(k-2)(n+k-2)-2h(n-h+1)]}{(n+k-h-1)(n+k-h)(n+k-h+2)}$
$[k,3^h]$	$\frac{(n+k)(n+k+1)(n+k+2)}{(n+k-h)(n+k-h+1)(n+k-h+2)}$	$\frac{-3h(n+k)(n+k+1)}{(n+k-h-1)(n+k-h)(n+k-h+2)}$
	$[k+1,3^{n-h-1},2]$	$[k,3^{n-h}]$
$[k+3,3^{h-1}]$	$\frac{3k(k-1)(n-h)}{(n+k-h+1)(n+k-h)(n+k-h-2)}$	$\frac{k(k-1)(k-2)}{(n+k-h)(n+k-h-1)(n+k-h-2)}$
$[k+2,3^{h-1},1]$	$\frac{-(k-1)[2(h+2)(n-h+1)-k(k+n)]}{(n+k-h+1)(n+k-h)(n+k-h-2)}$	$\frac{-(h+2)(k-1)(k-2)}{(n+k-h)(n+k-h-1)(n+k-h-2)}$
$[k+1,3^{h-1},2]$	$\frac{-(h+1)[2(k-2)(n+k+1)-h(n-h)]}{(n+h-k+1)(n+k-h)(n+k-h-2)}$	$\frac{(h+2)(h+1)(k-2)}{(n+k-h)(n+k-h-1)(n+k-h-2)}$
$[k,3^h]$	$\frac{3h(h+1)(n+k)}{(n+h-k+1)(n+k-h)(n+k-h-2)}$	$\frac{h(h+1)(h+2)}{(n+k-h)(n+k-h-1)(n+k-h-2)}$

scattering process. However, the crossing matrices for the T to the S channel C^{st}, and C^{tu} are related to the elastic crossing matrix C^{su} by the following considerations, which hold also in the case where all channels are inelastic. Defining the T channel for the process $a + b \to c + d$ as

$$a + \bar{c} \to d + \bar{b} \tag{15}$$

cross to the S channel

$$a + b \to c + d, \tag{16}$$

thence to the U channel

$$a + \bar{d} \to c + \bar{b}, \tag{17}$$

TABLE IV. Dimensions of irreducible representations occurring for the processes $[1^h]+[k]\to[1^h]+[k]$ and $[1]+[k^h]\to[1]+[k^h]$.

Representation	Dimensionality
$[k+1,1^{h-1}]$	$\frac{(n+k)h}{h+k}\rho$
$[k,1^n]$	$\frac{(n-h)k}{h+k}\rho$
$[k+1,1^{n-h-1}]$	$\frac{(n+k)(n-h)}{n+k-h}\rho$
$[k,1^{n-h}]$	$\frac{kh}{n+k-h}\rho$
$[k+1,k^{h-1}]$	$\frac{(n+k)h}{h+k}\rho$
$[k^h,1]$	$\frac{(n-h)k}{h+k}\sigma$
$[(k+1)^h,1^{n-h-1}]$	$\frac{(n+k)(n-h)}{n+k-h}\sigma$
$[k^h,k-1]$	$\frac{kh}{n+k-h}\sigma$

with $\rho = \dfrac{(n+k-1)!}{(n-h)!k!h!}$

and $\sigma = \prod_{r=1}^{h} \dfrac{(n+k-r)!(r-1)!}{(n-r)!(h+k-r)!}$

thence back to the T channel

$$b + \bar{d} \to \bar{a} + c. \tag{18}$$

Since the only difference between processes (15) and (18) is that (18) is obtained from (15) by conjugating the particles in (15) the S-matrix elements for the two processes differ only by real phase factors; hence

$$C^{tu}C^{us}C^{st} = D, \tag{19}$$

where D is a diagonal matrix with diagonal elements ± 1. This property was mentioned briefly in Ref. 4, and is the equivalent for crossing matrices of the well known Racah sum rule.[10] Since the crossing matrix may be written in the form

$$C^{su}_{ij} = N_i^{-\frac{1}{2}}(s)O^t_{ij}N_j(u^{\frac{1}{2}}), \tag{20}$$

where $N_i(s)$ are the dimensionalities of the irreducible representations occurring in the S channel, and O^t_{ij} is an orthogonal matrix which is known to be symmetric in the case where C^{su} connects two elastic channels,[4,11] Eq. (19) gives

$$O^sO^tO^u = D. \tag{21}$$

Also the eigenvalues of O^t are the nonzero elements of D.[4] Hence (21) has a solution $O^u = \tilde{O}^s$, and is the orthogonal matrix which diagonalizes the symmetric matric matrix O^t. Thus the matrices C^{st} and C^{tu} are determined in terms of (20) by the matrix C^{su} through (21). Applying this result to the processes (1) and (2) and using the property that the irreducible representations occurring in the t channel for both processes are the scalar (associated with element $+1$ of D) and the regular representation $N(t) = n^2 - 1$ (associated with element -1 of D), the C^{st} matrices are completely determined. In the notation of Table IV the inelastic crossing matrices for the two processes are related by a multiplicative factor as

$$C^{st}\ (\text{process 1}) = (\rho/\sigma)^{\frac{1}{2}}C^{st}\ (\text{process 2}). \tag{22}$$

Thus the inelastic crossing matrices for the two processes are proportional and the T-channel forces have a greater strength in one case than the other.

Note on the Robinson–Trautman Solutions*

J. FOSTER†‡
King's College, University of London, London, England
AND
E. T. NEWMAN
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received 6 June 1966)

A procedure is given which enables one to construct all the type N Robinson–Trautman solutions and an infinite class of type III solutions. Some approximate solutions, consistent with the Bondi–Sachs radiation conditions for bounded source fields, are also given. These approximate solutions, obtained by perturbing the Schwarzschild solution, are Schwarzschild in the asymptotic future. It is also shown that there exists an infinity of exact solutions which are Schwarzschild in the asymptotic future.

I. INTRODUCTION

THE solutions of the Einstein field equations for empty space given by Robinson and Trautman[1] are characterized by the fact that they admit a congruence of null rays which is geodesic, shear-free, hypersurface-orthogonal, and diverging. These solutions are also easily obtained by using the spin coefficient formalism[2] as was shown in a paper[3] on empty space metrics containing hypersurface-orthogonal geodesic rays. This paper together with a paper[4] on the behavior of asymptotically flat empty spaces are the starting point for the present one; the notation of these papers is used throughout with the exception that lower case Greek indices now take the values 0, 1, 2, 3 and Latin indices 2, 3.

In Sec. II we set out the equations to be solved in spin coefficient formalism, and give some of the known solutions. An algorithm is given which enables one to construct all the null RT solutions and an infinite class of type III solutions. These two classes of solutions are generated from two special solutions, one corresponding to flat space, the other to a type III solution given in RT.

In Sec. III we give some approximate RT solutions which can be regarded as small perturbations of the Schwarzschild solution subject to the shear-free condition of Robinson and Trautman. By using a coordinate transformation developed in NU, it is shown that these approximate solutions satisfy the boundary conditions suggested by Bondi[5] and Sachs[6] as suitable for defining gravitational radiation from bounded sources, but there are technical difficulties involved in solving the field equations and in transforming to Bondi–Sachs coordinates which force us to consider approximate solutions only.

II. EXACT SOLUTIONS

We consider a space–time in which, by hypothesis, there exists a family of null hypersurfaces containing shear-free null geodesics; these are used to construct a coordinate system as described in RT. This coordinate system is also used in NU (where the shear-free condition is relaxed) together with a null tetrad system, and the Newman–Penrose equations for empty space then take a simpler form. The coordinate $u = x^0$ labels the hypersurfaces, $r = x^1$ is an affine parameter along the null geodesics lying in the hypersurfaces, and the coordinates x^i $(i = 2, 3)$ pick out a particular null geodesic in a hypersurface $u = \text{const}$.

In this coordinate system the radial equations have been solved explicitly[3]; the notation employed here for the functions independent of r introduced by this integration (α^0 etc.) agrees with that of NU. The nonradial equations yield relations between these functions, and one is eventually led to the metric[3]

$$\begin{aligned} g^{01} &= 1, \qquad g^{00} = g^{0i} = 0, \\ g^{11} &= a_{-1}r + a_0 + a_1 r^{-1}, \\ g^{1i} &= 0, \\ g^{ij} &= -2P^2\,\delta^{ij} r^{-2}, \end{aligned} \tag{1}$$

* The research reported in this document has been sponsored by the Aerospace Research Laboratories under Contract AF 61(052)-457 through the U. S. Air Force European Office of Aerospace Research (OAR).
† In receipt of a grant from the Science Research Council of Great Britain.
‡ Present address: University of Sussex, Birghton, England.
[1] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962), subsequently referred to as RT.
[2] E. T. Newman and R. Penrose J. Math. Phys. **3**, 566 (1962).
[3] E. T. Newman and L. A. Tamburino, J. Math. Phys. **3**, 902 (1962).
[4] E. T. Newman and T. W. J. Unti, J. Math. Phys. **3**, 891 (1962), subsequently referred to as NU.
[5] H. Bondi, M. G. J. van der Burg, and A. W. K. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962).
[6] R. K. Sachs, Proc. Roy. Soc. (London) **A270**, 103 (1962).

where

$$a_{-1} = -4\gamma^0 = 2(\log P)_{,0},$$
$$a_0 = 2U^0 = -2P^2\nabla\bar{\nabla}\log P,$$
$$a_1 = -2\Psi_2^0,$$

and P and Ψ_2^0 (real functions of u, x^i)[7] satisfy

$$\nabla\Psi_2^0 = 0, \tag{2}$$

$$\Psi_2{}^0{}_{,0} - 3\Psi_2^0(\log P)_{,0} + P^2\nabla\bar{\nabla}(P^2\nabla\bar{\nabla}\log P) = 0. \tag{3}$$

In the above $_{,0}$ denotes differentiation with respect to u, and $\nabla = \partial/\partial x^2 + i\partial/\partial x^3$. Introducing the complex coordinate $\zeta = x^2 + ix^3$ (so that $\nabla = 2\partial/\partial\bar{\zeta}$), we may write the remaining coordinate freedom as

$$u' = \gamma(u), \qquad r' = r/\dot{\gamma}, \tag{4}$$

$\zeta' = f(\zeta)$, which induces $P'^2 = (\partial f/\partial\zeta)(\partial\bar{f}/\partial\bar{\zeta})P^2$. (5)

Equation (4) is a relabeling of hypersurfaces, and (5) a relabeling of geodesics within the hypersurfaces.

The 2-spaces $V_2(u, r)$: u = const, r = const, have line elements r^2dl^2, where dl^2 is the line element of a 2-space V_2^0, given by

$$V_2^0: dl^2 = \tfrac{1}{2}P^{-2}[(dx^2)^2 + (dx^3)^2] = \tfrac{1}{2}P^{-2}\,d\zeta\,d\bar{\zeta}; \tag{6}$$

the Gaussian curvature of a 2-space with metric (6) is

$$K = 2P^2\nabla\bar{\nabla}\log P, \tag{7}$$

so that (3) may be written

$$\Psi_2{}^0{}_{,0} - 3\Psi_2^0(\log P)_{,0} + \tfrac{1}{2}P^2\nabla\bar{\nabla}K = 0. \tag{8}$$

There are now two cases to consider: (a) $\Psi_2^0 = 0$, giving solutions of type III, N or flat.

Equation (2) is satisfied, and (3) reduces to $\nabla\bar{\nabla}K = 0$. The condition that these solutions be type N (possibly flat) is $\nabla K = \bar{\nabla}K = 0$, implying $K = K(u)$ a function of u alone. In any region where K is positive or negative it can be reduced to $+1$ or -1, respectively, by a coordinate transformation (4) with $\dot{\gamma}(u) = (\pm K)^{\frac{1}{2}}$. This can be seen from the transformation under (4) of g^{11}, whose second term is $a_0 = -K$.

In order to obtain all the null solutions we require the most general solution of

$$P^2\nabla\bar{\nabla}\log P = \tfrac{1}{2}K, \qquad K = -1, 0, 1. \tag{9}$$

[7] The p and m of RT are here denoted $\sqrt{2}P$ and $-\Psi_2{}^0$, respectively.

Since each of the spaces V_2^0 now has constant curvature, a particular solution of (9) is $P = 2^{-\frac{1}{2}}(1 + \frac{1}{4}K\zeta\bar{\zeta})$. Although this solution yields flat space, it may nevertheless be used to generate the most general solution of (9). This is achieved by defining a new coordinate ζ' and a new function $P'(\zeta', u)$ by

$$\zeta' = \zeta'(\zeta, u),$$
$$P'^2 = (\partial\zeta'/\partial\zeta)(\partial\bar{\zeta}'/\partial\bar{\zeta})P^2. \tag{10}$$

For each value of u this represents a coordinate transformation of V_2^0, under which (9) remains invariant, but it is not a coordinate transformation of the four-dimensional space.

This method of constructing null solutions also gives flat-space solutions; for example the choice

$$\zeta' = [a(u)\zeta + b(u)]/[c(u)\zeta + d(u)], \qquad ad - bc = 1 \tag{11}$$

in (10) yields a flat-space metric based on a family of null cones whose vertices lie on an arbitrary spacelike, null or timelike curve, given by $K = -1, 0, +1$, respectively. These solutions are generalizations of some flat-space solutions[8] given previously.

A particular type III solution given in RT is

$$P = (\zeta + \bar{\zeta})^{\frac{1}{2}},$$
$$K = -12(\zeta + \bar{\zeta}). \tag{12}$$

From this particular solution one can generate an infinite class of type III solutions by defining a new coordinate ζ' and a new function P' by Eq. (10). This does not lead to all the type III solutions, as the expression for K in (12) is not the most general solution of $\nabla\bar{\nabla}K = 0$, $\nabla K \neq 0$.

(b) $\Psi_2^0 \neq 0$, giving solutions of type II or D.

From (2), Ψ_2^0 is a function of u only, therefore by a coordinate transformation (4) it may be made equal to -1.[9] The remaining equation to be solved is (3), which now takes the form

$$3P_{,0} + P^3\nabla\bar{\nabla}(P^2\nabla\bar{\nabla}\log P) = 0. \tag{13}$$

Various solutions of this equation are given in RT, but they cannot be used to generate new solutions because (13) is not invariant under the substitution (10).

[8] E. T. Newman and T. W. J. Unti, J. Math. Phys. 4, 1467 (1963).

[9] In Ref. 3 it is made equal to $+1$, but here the minus sign is more convenient.

The Schwarzschild solution (with mass $K^{-\frac{1}{2}}$) is given by

$$K = 2P^2\nabla\bar{\nabla}\log P = \text{const} > 0, \qquad P = (1/\sqrt{2})(1 + \tfrac{1}{4}K\zeta\bar{\zeta}). \tag{14}$$

For the next section it is convenient to introduce a new coordinate $\zeta' = V + i\phi$ defined by

$$\zeta = 2K^{-\frac{1}{2}}e^{V+i\phi}; \tag{15}$$

then from Eq. (5)

$$P' = (1/\sqrt{2})K^{\frac{1}{2}}\cosh V. \tag{16}$$

The transformation to the usual null polar coordinate system is then

$$r' = K^{-\frac{1}{2}}r, \qquad u' = K^{\frac{1}{2}}u, \qquad \tanh V = \cos\theta. \tag{17}$$

III. APPROXIMATE SOLUTIONS

Using the coordinates V, ϕ defined above, the line element (6) of V_2^0 for the Schwarzschild metric in the form (16) is

$$dl^2 = K^{-1}\operatorname{sech}^2 V(dV^2 + d\phi^2), \tag{18}$$

and on putting $\tanh V = \cos\theta$ this becomes

$$dl^2 = K^{-1}(d\theta^2 + \sin^2\theta\, d\phi^2), \tag{19}$$

which is the metric of a sphere of radius $A = K^{-\frac{1}{2}}$.

The approximate solutions to be presented here are obtained by considering 2-spaces V_2^0 which are only approximately spherical, and neglecting the square of a small parameter which measures their deviation from spherical form. They can thus be regarded as perturbations of the Schwarzschild solution subject to the RT shear-free condition. More precisely we consider the surfaces formed by revolving the curve, given in polar coordinates (R, θ) by

$$R = A[1 + \epsilon P_n(\cos\theta)], \tag{20}$$

about the polar axis, where ϵ is a small quantity whose square can be neglected. One could, of course, take a more complicated surface given by a linear combination of spherical harmonics, but without any significant loss of generality we confine ourselves to the simple axially symmetric surface represented by Eq. (20) so as to avoid cumbersome expressions later. The cases $n = 0$ and $n = 1$ both represent spheres, that in the latter case having its center displaced a distance $A\epsilon$ along the axis of revolution, and so to achieve nontrivial results we take $n \geq 2$.

The line element (6) for the surface given by (20) is

$$dl^2 = A^2[1 + 2\epsilon P_n]\operatorname{sech}^2 V(dV^2 + d\phi^2), \tag{21}$$

where $P_n = P_n(\cos\theta) = P_n(\tanh V) = P_n(\mu)$ say, so the function P is

$$P = A^{-1}2^{-\frac{1}{2}}(1 - \epsilon P_n)\cosh V. \tag{22}$$

Substitution in Eq. (13) shows that this is an approximate RT solution if

$$\epsilon = \epsilon_0 \exp\left[-2\binom{n+2}{4}A^{-4}u\right], \tag{23}$$

where ϵ_0 is the value of ϵ on the hypersurface $u = 0$. As $u \to \infty$, $\epsilon \to 0$ and the 2-spaces tend to spheres. The approximation is not valid when carried indefinitely into the past, but this difficulty can be avoided by relaxing the shear-free condition prior to $u = 0$ say, and taking ϵ to be given on $u = 0$ as a small quantity; then Eq. (23) shows that at later times ϵ is still small and the approximation is valid.

The metric (1) now has the form

$$g^{01} = 1, \qquad g^{00} = g^{0i} = 0, \qquad g^{1i} = 0,$$
$$g^{11} = a_{-1}r + a_0 + a_1r^{-1},$$
$$g^{ij} = -2P^2\,\delta^{ij}r^{-2},$$

where

$$a_{-1} = 4\binom{n+2}{4}A^{-4}\epsilon P_n,$$
$$A_0 = -A^{-2}[1 + \epsilon(n+2)(n-1)P_n], \tag{24}$$
$$a_1 = 2,$$
$$P = A^{-1}2^{-\frac{1}{2}}(1 - \epsilon P_n)\cosh V.$$

We wish to interpret this solution as representing radiation from a bounded source, and to do this we must transform to a coordinate system in which it is apparent that the boundary conditions suggested by Bondi[5] and Sachs[6] are satisfied. The first task is the removal of the term a_{-1} in the metric (24), which is accomplished by using a coordinate transformation developed in NU.[4] In general this transformation is given by infinite series in r^{-1}, of which only the first few terms are readily calculated, and the transformed metic is not expressible in closed form. However, in our case the term to be transformed away is of first order in ϵ, and so an infinitesimal coordinate transformation is sufficient. Quoting from NU, Sec. IV, we see that it is of the form $x^{\mu'} = x^\mu + \zeta^\mu(x^\nu)$ where in the axially sym-

metric case which we are considering the infinitesimal quantities are given by

$$\begin{aligned}\zeta^0 &= \zeta^{00}(u, V),\\ \zeta^1 &= -\zeta^{00}{}_{,0}r + \zeta^{01}(u, V),\\ \zeta^2 &= \zeta^{02}(V) - 2P^2r^{-1}\zeta^{00}{}_{,2},\\ \zeta^3 &= 0,\end{aligned} \tag{25}$$

and satisfy

$$\begin{aligned}\zeta^{02}{}_{,2} &= \zeta^{02}{}_{,3} = 0,\\ \zeta^{01} &= P^2\zeta^{00}{}_{,22}.\end{aligned} \tag{26}$$

Considering the effect on g^{11}, we have

$$\zeta^{00} = \frac{1}{2}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} A^4\epsilon P_n,$$

and taking $\zeta^{02} = 0$, this gives the transformation

$$\begin{aligned}\zeta^0 &= \frac{1}{2}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} A^4\epsilon P_n,\\ \zeta^1 &= \epsilon P_n r - [6/(n-1)(n+2)]A^2\epsilon P_n,\\ \zeta^2 &= -\frac{1}{2}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} A^2\epsilon\frac{dP_n}{d\mu}r^{-1},\\ \zeta^3 &= 0, \quad \text{where} \quad \mu = \tanh V.\end{aligned} \tag{27}$$

The transformation (25) is similar to an infinitesimal Bondi–Metzner transformation,[5] but is more general as ζ^{00} is a function of u as well as of V. Taking $\zeta^{02} = 0$ is analogous to using a supertranslation rather than a general BM transformation.

The transformed metric has components

$$\begin{aligned}g^{0'0'} &= g^{0'i'} = 0,\\ g^{0'1'} &= 1,\\ g^{1'1'} &= a'_0 + a'_1r'^{-1} + a'_2r'^{-2},\\ g^{1'2'} &= b'^2{}_2r'^{-2} + b'^2{}_3r'^{-3},\\ g^{1'3'} &= 0,\\ g^{i'j'} &= -2P'^2\,\delta^{ij}r'^{-2} + d'^{ij}{}_3r'^{-3},\end{aligned} \tag{28}$$

where

$$\begin{aligned}a'_0 &= -A^{-2},\\ a'_1 &= 2 + 6\epsilon' P'_n,\\ a'_2 &= \frac{-12}{(n-1)(n+2)}A^2\epsilon' P'_n,\\ b'^2{}_2 &= \frac{6}{n(n+1)}\epsilon'\frac{dP'_n}{d\mu'},\\ b'^2{}_3 &= \begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} A^2\epsilon'\frac{dP'_n}{d\mu'},\\ 2P'^2 &= A^{-2}\cosh^2 V',\\ d'^{23}{}_3 &= 0,\\ d'^{22}{}_3 &= -d'^{33}{}_3 = \frac{1}{2}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1}\epsilon' P'^2{}_n\cosh^2 V',\end{aligned}$$

and here

$$\begin{aligned}\epsilon' &= \epsilon_0\exp\left[-2\begin{bmatrix} n+2 \\ 4 \end{bmatrix}A^{-4}u'\right],\\ P'_n &= P_n(\tanh V') = P_n(\cos\theta'),\\ \mu' &= \tanh V' = \cos\theta'.\end{aligned}$$

Finally by a transformation to new coordinates u, r, V [not to be confused with the original coordinates used in (24)]

$$\begin{aligned}u &= A^{-1}u', \qquad r = Ar',\\ V &= V', \qquad \phi = \phi',\end{aligned} \tag{29}$$

and putting $A^3 = m_0$, one obtains the metric

$$\begin{aligned}g^{00} &= g^{01} = 0,\\ g^{01} &= 1,\\ g^{11} &= a_0 + a_1r^{-1} + a_2r^{-2},\\ g^{12} &= b_2^2r^{-2} + b_3^2r^{-3},\\ g^{13} &= 0,\\ g^{ij} &= -2P^2\,\delta^{ij}r^{-2} + d_3^{ij}r^{-3},\end{aligned} \tag{30}$$

where

$$\begin{aligned}a_0 &= -1,\\ a_1 &= 2m_0 + 6m_0\epsilon P_n,\\ a_2 &= [-12/(n-1)(n+2)]m_0^2\epsilon P_n,\\ b_2^2 &= \frac{6}{n(n+1)}m_0\epsilon\frac{dP_n}{d\mu},\\ b_3^2 &= \begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} m_0^2\epsilon\frac{dP_n}{d\mu},\\ 2P^2 &= \cosh^2 V, \qquad d_3^{23} = 0,\\ d_3^{22} &= -d_3^{33} = \frac{1}{2}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} m_0\epsilon P_n^2\cosh^2 V,\end{aligned}$$

and here

$$\begin{aligned}\epsilon &= \epsilon_0\exp\left[-2\begin{bmatrix} n+2 \\ 4 \end{bmatrix}m_0^{-1}u\right],\\ P_n &= P_n(\tanh V) = P_n(\cos\theta) = P_n(\mu).\end{aligned}$$

The offensive term a_{-1} has now been removed (at the expense of introducing shear into the coordinate hypersurfaces), and as $r \to \infty$ the 2×2 metric associated with the surfaces of constant u and r tends to

$$dl^2 = \tfrac{1}{2}P^{-2}(dV^2 + d\phi^2)$$
$$= \operatorname{sech}^2 V(dV^2 + d\phi^2) = d\theta^2 + \sin^2\theta\, d\phi^2,$$

which is that of a sphere. Thus apart from using an affine parameter r instead of a luminosity distance we have made a transformation (of the region $u > 0$) to Bondi–Sachs coordinates, and the approximate solution is now clearly seen to represent radiation from a bounded source.

Making use of Eqs. (32) and (25a)–(25d) of NU we can pick out from the metric (30) the following leading terms of the spin coefficients and tetrad components of the curvature tensor:

$$\sigma^0 = \frac{1}{4}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} m_0 \epsilon P_n^2, \tag{31a}$$

$$\omega^0 = -[3\sqrt{2}/n(n+1)]m_0\epsilon P_n^1, \tag{31b}$$

$$\lambda^0 = -\tfrac{1}{2}\epsilon P_n^2, \tag{31c}$$

$$\Psi_0^0 = 0, \tag{31d}$$

$$\Psi_1^0 = \frac{3\sqrt{2}}{4}\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} m_0^2 \epsilon P_n^1, \tag{31e}$$

$$\Psi_2^0 = -m_0, \tag{31f}$$

$$\Psi_3^0 = -[(n-1)(n+2)/2\sqrt{2}]\epsilon P_n^1, \tag{31g}$$

$$\Psi_4^0 = -\begin{bmatrix} n+2 \\ 4 \end{bmatrix}^{-1} m_0^{-1} \epsilon P_n^2. \tag{31h}$$

These are of course correct to first order only in ϵ.

In Eq. (31a) σ^0 is Bondi's news function, and corresponds to 2^n-pole radiation. As $n \geq 2$ quadrupole is the lowest pole radiation which occurs, as was to be expected.

The mass of the system at any time u as defined by Janis and Newman[10] is

$$m(u) = -\frac{1}{4}\int_{-1}^{1} (\Psi_2^0 + \bar{\Psi}_2^0) P_0(\mu)\, d\mu = m_0.$$

The mass $m_B(u)$ as defined by Bondi differs from $m(u)$ by terms quadratic in σ^0, and so when ϵ^2 is ignored $m_B(u) = m(u)$, and neither mass changes to first order in ϵ. The change in either mass depends only on terms quadratic in σ^0,[11] and so both $\partial m/\partial u$ and $\partial m_B/\partial u$ can in fact be calculated. Quoting from Newman and Unti,[11]

$$\frac{\partial m}{\partial u} = -\frac{1}{2}\int_{-1}^{1} \left|\frac{\partial \sigma^0}{\partial u}\right|^2 d\mu + \frac{1}{4}\int_{-1}^{1} \frac{\partial^2}{\partial u^2}(\sigma^0\bar{\sigma}^0)\, d\mu$$
$$= \frac{1}{4(2n+1)}\frac{(n+2)!}{(n-2)!}\epsilon^2,$$

which is a positive quantity, while

$$\frac{\partial m_B}{\partial u} = -\frac{1}{2}\int_{-1}^{1} \left|\frac{\partial \sigma^0}{\partial u}\right|^2 d\mu$$
$$= -\frac{1}{4(2n+1)}\frac{(n+2)!}{(n-2)!}\epsilon^2,$$

which is of necessity a negative quantity. The difference in sign of the two rates of change of mass occurs whenever the news function depends exponentially on u, but it is not really significant as the system does not start from a static configuration. It is only in the static case that the mass of the system can be positively identified, and then $m(u) = m_B(u)$; in a system which has an initial static configuration and a final static configuration, but is active in between, both definitions give the same net mass loss.

We end this section with a few remarks about the solution of the propagation equation (13), which may be written

$$6P_{,0} + P^3\nabla\bar{\nabla}K = 0, \tag{32}$$

where K is the Gaussian curvature of V_2^0,

$$K = 2P^2\nabla\bar{\nabla}\log P.$$

Suppose $P(u, \zeta, \bar{\zeta})$ is given by a Taylor series in u, i.e.,

$$P(u) = P_0 + u\dot{P}_0 + (1/2!)u^2\ddot{P}_0 + \cdots, \tag{33}$$

where the suffix zero indicates that the function is evaluated at $u = 0$, and a dot denotes differentiation with respect to u. If P_0 is given (i.e., if a 2-surface V_2^0 is given at $u = 0$) then $P(u)$ is determined uniquely at other times u by Eq. (32). For

$$\dot{P}_0 = -\tfrac{1}{6}P_0^3\nabla\bar{\nabla}K_0 = f_1(P_0),$$
$$\ddot{P}_0 = -\tfrac{1}{2}P_0^2\dot{P}_0\nabla\bar{\nabla}K_0 - \tfrac{1}{6}P_0^3\nabla\bar{\nabla}\dot{K}_0$$
$$= g_2(\dot{P}_0, P_0) = f_2(P_0),$$

and in general

$$\overset{(n)}{P}_0 = g_n(\overset{(n-1)}{P}_0, \cdots, P_0) = f_n(P_0).$$

As a trivial example one could take

$$P_0 = (1/\sqrt{2})(1 + \tfrac{1}{4}a^{-2}\zeta\bar{\zeta}),$$

[10] A. I. Janis and E. T. Newman, J. Math. Phys. **6**, 902 (1965).

[11] E. T. Newman and T. W. J. Unti, J. Math. Phys. **6**, 1806 (1965).

corresponding to a sphere of radius a. Then $K_0 = a^{-2} = \text{const}$, $\dot{P}_0 = \ddot{P}_0 = \cdots = 0$, and the 2-surfaces V_2^0 remain spherical. By taking more complicated initial surfaces and neglecting higher powers of u one can in this way obtain approximate RT solutions valid for small u.

Alternatively one can ask for solutions which take a prescribed form in the infinite future. Putting $w = u^{-1}$, Eq. (32) becomes

$$6w^2P' - P^3\nabla\bar{\nabla}K = 0, \tag{34}$$

and analogous to (33)

$$P(w) = P_0 + wP_0' + \frac{1}{2!}w^2P_0'' + \cdots, \tag{35}$$

where a prime denotes differentiation with respect to w. Repeated differentiation of (34) yields

$$\nabla\bar{\nabla}K_0 = 0, \tag{36a}$$

$$\nabla\bar{\nabla}K_0' = 0, \tag{36b}$$

$$\nabla\bar{\nabla}K_0'' = 12P_0^{-3}P_0', \tag{36c}$$

and in general

$$\nabla\bar{\nabla}K_0^{(n)} = f(P_0^{(n-1)}, P_0^{(n-2)}, \cdots, P_0). \tag{36d}$$

Equation (36a) puts a restriction on the Gaussian curvature of the 2-surfaces which can be prescribed in the infinite future. Even if a function P_0 compatible with (36a) is given, $P(w)$ is not uniquely determined. Equation (36b) is a differential equation for P'_0, and its solution involves arbitrary functions; P'_0 can then be substituted in (36c) to give a differential equation for P''_0, and again arbitrary functions are introduced by the integration, and so on, arbitrary functions being introduced as each $P_0^{(n)}$ is determined.

A possible form for P_0 which is compatible with (36a) is that corresponding to a sphere. Thus we infer that there is an infinite number of RT solutions which are Schwarzschild in the asymptotic future.

ACKNOWLEDGMENTS

One of the authors (E.T.N.) wishes to express his appreciation to Professor H. Bondi and King's College for their hospitality during the period of his stay.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 2 FEBRUARY 1967

Point-Loop Renormalization in Regularized Field Theories. I

WILLIAM M. FRANK*

Department of Nuclear Physics, The Weizmann Institute of Science, Rehovoth, Israel

(Received 3 May 1965; final manuscript received 15 June 1966)

The relation between the augmented and the standard vacuum amplitudes is derived (for a finite-valued amplitude for the point fermion–antifermion loop) by diagrammatic methods. This relation is shown to consist of a mass renormalization, whereby the point fermion–antifermion loop is absorbed into the fermion propagator, plus some additional factors, and the analytical properties of the amplitudes are greatly modified. The effects of the renormalization are the increase of the density of zeros, the increase of the exponential order in the variable z, and the change of the singularities in η from simple poles to essential singularities. The significance of these for massless fermions or bosons is briefly discussed.

I. INTRODUCTION

IN three earlier papers[1-3] (denoted hereafter by CT, AP, and EO, respectively), an investigation of the analytic properties of coupling constant expansions in Regularized Field Theories (RFT) was initiated. The field theories studied in those articles were restricted by two major "finitizing" conditions (i.e., conditions imposed to make the amplitudes of all diagrams finite numbers); restriction to a finite four-dimensional periodicity volume V (in the Euclidean metric),[4] and regularization of the boson and fermion propagators.[5] While we presently con-

* Present Address: U. S. Naval Ordnance Laboratory, Silver Spring, Maryland.

[1] W. M. Frank, Ann. Phys. (N. Y.) **29**, 175 (1964).
[2] W. M. Frank, Ann. Phys. (N. Y.) **29**, 217 (1964).
[3] W. M. Frank, Ann. Phys. (N. Y.) **30**, 328 (1964).

[4] In an alternative approach to the "existence of solutions" problem in quantum field theory by K. Symanzik, the analysis proceeds from the same formulation, an Euclidean metric, a finite four-dimensional periodicity volume, and regularized propagators. The relation between field theory in the Lorentz and Euclidean metrics is presented there in terms of the Wightman function formalism. See K. Symanzik, "A Model of Euclidean Quantum Theory," NYU Report IMM-NYU327, June 1964.

[5] Details of the regularization are discussed in CT, Sec. IIB, and briefly in Sec. II of this paper.

corresponding to a sphere of radius a. Then $K_0 = a^{-2} = \text{const}$, $\dot{P}_0 = \ddot{P}_0 = \cdots = 0$, and the 2-surfaces V_2^0 remain spherical. By taking more complicated initial surfaces and neglecting higher powers of u one can in this way obtain approximate RT solutions valid for small u.

Alternatively one can ask for solutions which take a prescribed form in the infinite future. Putting $w = u^{-1}$, Eq. (32) becomes

$$6w^2P' - P^3\nabla\bar{\nabla}K = 0, \tag{34}$$

and analogous to (33)

$$P(w) = P_0 + wP_0' + \frac{1}{2!}w^2P_0'' + \cdots, \tag{35}$$

where a prime denotes differentiation with respect to w. Repeated differentiation of (34) yields

$$\nabla\bar{\nabla}K_0 = 0, \tag{36a}$$

$$\nabla\bar{\nabla}K_0' = 0, \tag{36b}$$

$$\nabla\bar{\nabla}K_0'' = 12P_0^{-3}P_0', \tag{36c}$$

and in general

$$\nabla\bar{\nabla}K_0^{(n)} = f(P_0^{(n-1)}, P_0^{(n-2)}, \cdots, P_0). \tag{36d}$$

Equation (36a) puts a restriction on the Gaussian curvature of the 2-surfaces which can be prescribed in the infinite future. Even if a function P_0 compatible with (36a) is given, $P(w)$ is not uniquely determined. Equation (36b) is a differential equation for P'_0, and its solution involves arbitrary functions; P'_0 can then be substituted in (36c) to give a differential equation for P''_0, and again arbitrary functions are introduced by the integration, and so on, arbitrary functions being introduced as each $P_0^{(n)}$ is determined.

A possible form for P_0 which is compatible with (36a) is that corresponding to a sphere. Thus we infer that there is an infinite number of RT solutions which are Schwarzschild in the asymptotic future.

ACKNOWLEDGMENTS

One of the authors (E.T.N.) wishes to express his appreciation to Professor H. Bondi and King's College for their hospitality during the period of his stay.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 2 FEBRUARY 1967

Point-Loop Renormalization in Regularized Field Theories. I

WILLIAM M. FRANK*

Department of Nuclear Physics, The Weizmann Institute of Science, Rehovoth, Israel

(Received 3 May 1965; final manuscript received 15 June 1966)

The relation between the augmented and the standard vacuum amplitudes is derived (for a finite-valued amplitude for the point fermion–antifermion loop) by diagrammatic methods. This relation is shown to consist of a mass renormalization, whereby the point fermion–antifermion loop is absorbed into the fermion propagator, plus some additional factors, and the analytical properties of the amplitudes are greatly modified. The effects of the renormalization are the increase of the density of zeros, the increase of the exponential order in the variable z, and the change of the singularities in η from simple poles to essential singularities. The significance of these for massless fermions or bosons is briefly discussed.

I. INTRODUCTION

IN three earlier papers[1-3] (denoted hereafter by CT, AP, and EO, respectively), an investigation of the analytic properties of coupling constant expansions in Regularized Field Theories (RFT) was initiated. The field theories studied in those articles were restricted by two major "finitizing" conditions (i.e., conditions imposed to make the amplitudes of all diagrams finite numbers); restriction to a finite four-dimensional periodicity volume V (in the Euclidean metric),[4] and regularization of the boson and fermion propagators.[5] While we presently con-

* Present Address: U. S. Naval Ordnance Laboratory, Silver Spring, Maryland.

[1] W. M. Frank, Ann. Phys. (N. Y.) **29**, 175 (1964).
[2] W. M. Frank, Ann. Phys. (N. Y.) **29**, 217 (1964).
[3] W. M. Frank, Ann. Phys. (N. Y.) **30**, 328 (1964).

[4] In an alternative approach to the "existence of solutions" problem in quantum field theory by K. Symanzik, the analysis proceeds from the same formulation, an Euclidean metric, a finite four-dimensional periodicity volume, and regularized propagators. The relation between field theory in the Lorentz and Euclidean metrics is presented there in terms of the Wightman function formalism. See K. Symanzik, "A Model of Euclidean Quantum Theory," NYU Report IMM-NYU327, June 1964.

[5] Details of the regularization are discussed in CT, Sec. IIB, and briefly in Sec. II of this paper.

sider only the vacuum-to-vacuum amplitude; these results are crucial to the study of the physical amplitudes. For simplicity, only spinless fermions are considered; however, all results are essentially applied to fermions with spin.

This investigation is being carried out in anticipation of eventual consideration of the physical limit; the limit as the regularized series approaches the renormalized series (to be known as the "renormalization limit"), and the limit as $V \to \infty$. In this paper and in a paper to follow, the first step of the renormalization limit will be studied in the form of renormalizing the augmented[6] (blind) S matrix by including the point fermion–antifermion loop in the nonaugmented[6] S matrix via a mass renormalization of the fermion propagator. The case of a finite point-loop amplitude is studied in this paper, while the circumstance of an infinite point-loop amplitude is investigated in a subsequent paper. From the physical point of view, this renormalization is really a "straw man" in that the point fermion–antifermion loop is not part of the physical amplitude, but was invented for mathematical convenience. Mathematically, however, this simple renormalization is extremely illuminating. A brief review of definitions and earlier results is presented in Sec. II. The Point-Loop Renormalization Relation (PLRR) is derived by diagrammatic methods in Sec. III. In Sec. IV, the structure and some interesting consequences of the PLRR are derived. The results are discussed in Sec. V. In the subsequent article, convergence of the nonaugmented S matrix is demonstrated for the case that the point loop is infinite provided that other diagrams have finite amplitudes, and the implications of these results as well as the general considerations concerning renormalization are discussed.

II. DEFINITIONS AND EARLIER RESULTS

The vacuum-to-vacuum amplitude was constructed in Ref. 1, in terms of two functions $\Delta(x)$ and $S(x)$ of a Euclidean four-vector variable x, which have the periodicity of a four-dimensional Euclidean periodicity volume V. Their Fourier coefficients are

$$\hat{\Delta}(k) = \int_V dx\ \Delta(x)e^{-ikx}, \tag{1}$$

$$\hat{S}(k) = \int_V dx\ S(x)e^{-ikx}, \tag{2}$$

where k is a vector of the four-dimensional momentum lattice $\mathfrak{L}$ and

$$\Delta(x) = \frac{1}{V}\sum_{k\in\mathfrak{L}} \hat{\Delta}(k)e^{ikx}, \tag{3}$$

$$S(x) = \frac{1}{V}\sum_{k\in\mathfrak{L}} \hat{S}(k)e^{ikx}. \tag{4}$$

They correspond, respectively, to the bare boson and fermion propagators, after the transformation to the Euclidean metric described in Ref. 1 has been performed. The latter function is to be considered the $\eta = 0$ value of the function $S_\eta(x)$, defined by

$$S_\eta(x) = \frac{1}{V}\sum_{k\in\mathfrak{L}} \hat{S}_\eta(k)e^{ikx} \equiv \frac{1}{V}\sum_{k\in\mathfrak{L}} \frac{e^{ikx}}{\hat{S}(k)^{-1} - \eta}, \tag{5}$$

where η is a complex variable. In the case of fermions with spin, Eqs. (2), (4), and (5) would be replaced by more elaborate expressions [Eqs. (84), (B33), (B34) in CT] for nongradient scalar coupling. The expansions to be studied are [Eqs. (4), (9) of AP]

$$\mathcal{S}(z, \eta) \equiv \sum_{n=0}^{\infty} \frac{(2z)^n}{(2n)!} \int_V dx_1 \cdots \int_V dx_{2n} \times S^{(2n)}(x_1, \cdots, x_{2n}; \eta)\ \Delta^{(2n)}(x_1, \cdots, x_{2n}), \tag{6}$$

$$\mathring{\mathcal{S}}(z, \eta) \equiv \sum_{n=0}^{\infty} \frac{(2z)^n}{(2n)!} \int_V dx_1 \cdots \int_V dx_{2n} \times S_0^{(2n)}(x_1, \cdots, x_{2n}; \eta)\ \Delta^{(2n)}(x_1, \cdots, x_{2n}), \tag{7}$$

known, respectively, as the "augmented (blind)" and the "standard (blind)" (vacuum) amplitudes. These are the amplitudes expressed in the Euclidean metric after the removal of the i's (hence "blind"). The quantity

$$S^{(2n)}(x_1, \cdots, x_{2n}; \eta) \equiv S^{(2n)}(12, \cdots, 2n; \eta)$$

is the determinant of the matrix $\mathbf{S}^{(2n)}(x_1, \cdots, x_{2n}; \eta)$ (we adhere to the notation of representing the matrix by a boldface symbol, and its determinant by the corresponding plain symbol) whose i, j matrix element is

$$[\mathbf{S}^{(2n)}(x_1, \cdots, x_{2n}; \eta)]_{ij} = S_\eta(x_i - x_j), \tag{8}$$

and the elements of the matrix $\mathbf{S}_0^{(2n)}(x_1, \cdots, x_{2n}\eta)$ whose determinant is

$$S_0^{(2n)}(x_1, \cdots, x_{2n}; \eta) \equiv S_0^{(2n)}(12 \cdots 2n; \eta)$$

differ from those of $\mathbf{S}^{(2n)}(x_1, \cdots, x_{2n}; \eta)$ only in that its diagonal elements are zero. $\Delta^{(2n)}(x_1, \cdots, x_{2n})$

[6] These terms are defined in CT Sec. IIG, and in Sec. II of this paper, where the epithet "nonaugmented" is replaced by "standard". The adjective "blind" refers to the removal of all the (imaginary unit) i's. "Augmented" refers to the inclusion of the point fermion–antifermion loop.

is the sum of n-fold products of the form

$$\Delta^{(2n)}(x_1, \cdots, x_{2n}) \equiv \Delta^{(2n)}(12 \cdots 2n) = \sum_{\substack{1 \le i_1 < i_3 \cdots < i_{2n-1} \\ \text{all } i_k \text{ distinct}}} \Delta(x_{i_1} - x_{i_2}) \times \Delta(x_{i_3} - x_{i_4}) \cdots \Delta(x_{i_{2n-1}} - x_{i_{2n}})$$

corresponding to all ways of forming distinct pairwise boson connections between the $2n$ points. Certain "regularization conditions" are demanded of the functions $\hat{\Delta}(k)$ and $\hat{S}_\eta(k)$ so that all integrals appearing in $\mathcal{S}(z, \eta)$ and $\mathring{\mathcal{S}}(z, \eta)$ are finite. They are to be specified in the course of this article. The variable z corresponds to $\frac{1}{2}g^2$, where g would be the coupling constant for a Yukawa-type coupling of the fermions to the bosons.

In AP and EO the regularization conditions

$$\Delta_1 \equiv \int_V dx\ |\Delta(x)| < \infty \tag{9}$$

and

$$S(0) \equiv \frac{1}{V} \sum_{k \in \mathfrak{L}} \hat{S}(k) < \infty \tag{10}$$

were imposed.

The latter condition was somewhat strengthened to the condition on the large momentum behavior of $\hat{S}(k)$

$$\overline{\lim_{|k| \to \infty}}\ |k|^t\ |\hat{S}(k)| < \infty \tag{11}$$

for some $t > 4$. Then, it was shown in AP that $\mathring{\mathcal{S}}(z, \eta)$ had exponential order of at most 1 in z,[7] and in EO it was shown that $\mathcal{S}(z, 0)$ had exponential order exactly $(\frac{1}{2}t - 1)^{-1}$. For $\mathcal{S}(z, \eta)$ $(\eta \neq 0)$, it was shown that, except for a discrete set of points $[\eta = \hat{S}(k)^{-1},\ k \in \mathfrak{L}]$, where $\mathcal{S}(z, \eta)$ had poles in η, $\mathcal{S}(z, \eta)$ had exponential order of at most $(\frac{1}{2}t - 1)^{-1}$ in z. In this paper, the case $t > 4$ corresponding to the regularization condition Eq. (10) is demanded. In a following paper, the range $2 < t \le 4$ will be studied, where $\mathcal{S}(z, \eta)$ is no longer defined since $S(0)$ of Eq. (10) is a divergent sum. The condition Eq. (9) will be maintained.

Our ultimate concern is with the quantities $U(z, \eta)$ and $\mathring{U}(z, \eta)$ defined by

$$\mathcal{S}(z, \eta) \equiv \exp\,[-VU(z, \eta)], \qquad \mathring{\mathcal{S}}(z, \eta) \equiv \exp\,[-V\mathring{U}(z, \eta)]. \tag{12}$$

These are "observed" quantities in the sense that, in the $V \to \infty$ limit, they are finite. Physically, they represent the vacuum energy density. They are given as the sum of the amplitudes of all connected diagrams calculated by the proper prescription (see CT Sec. IIC).

III. THE POINT-LOOP RENORMALIZATION RELATION

In the Appendix B of CT, relations [Eqs. (B18) and (B19)] were derived connecting the functions $\mathcal{S}(z, \eta)$ and $\mathring{\mathcal{S}}(z, \eta)$. These relations will now be rederived by diagrammatic methods, and their physical content will be made clearer. The relation to be derived[8] (in a somewhat different notation than in CT)

$$\mathring{\mathcal{S}}(z, \eta) = \mathcal{S}[z, \eta + 2\Delta_0 z S_\eta(0)] \times \exp V\left(\Delta_0 z S_\eta^2(0) - \int_\eta^{\eta + 2\Delta_0 z S_\eta(0)} d\rho\ S_\rho(0)\right) \tag{13}$$

is termed, for reasons to become clear later, the "Point-Loop Renormalization Relation" (PLRR). We have defined $\Delta_0 \equiv \Delta(0) = \int_V dx\ \Delta(x)$.

It should be recalled that $\hat{S}_\eta(k)$ obeys the relation

$$(\partial/\partial\eta)\hat{S}_\eta(k) = \hat{S}_\eta(k)^2, \tag{14}$$

which allows the interpretation that the operation $\partial/\partial\eta$ has the effect of attaching a vertex (of unit strength) of an external line carrying four-momentum zero into the propagator $\hat{S}_\eta(k)$ (to be termed a "zero vertex"). Equation (13) is thus a generalized Ward identity.[9] Consequently, $\partial/\partial\eta$, acting on any diagrammatic amplitude, produces the sum of the amplitudes of all diagrams obtained from the original one by the insertion of the zero vertex into all possible internal fermion propagators. Iteration of $\partial/\partial\eta$ leads correspondingly to the attachment of further zero vertices into the original diagram.

[7] In AP only an upper bound of unity was established for the exponential order of the standard vacuum S matrix. That the order is exactly unity follows immediately from the observations at the end of Sec. IV of this paper, or by arguments similar to those used in EO based on Eq. (28) of this paper.

[8] Equations (B18) and (B19) of CT contain a number of errors in sign [a consequence of an error in Eqs. (B16) and (B25) of CT]. In Eqs. (B16), (B25), and the rightmost side of Eq. (B24), replace the explicit V by $-V$ in both places, and place an additional factor of $(-)^q$ on the right side. The origin of this error is the omission of a minus in the first term on the right side of Eqs. (B21), (B22), and (B24). Equation (B17) is corrected by an additional $(-)^q$ on the right side as well as the replacement of V by $-V$ (except for the V multiplying $\lambda^2\Delta_0 z$). Equation (B18) is corrected by changing λ to $-\lambda$ and interchanging the upper and lower limits on the integral. Equation (B19) is corrected by preceding the integral with a $-$ rather than a $+$, and setting $\eta = 2S(0)\Delta_0 z$ in the rightmost subscript. In this paper, $-\lambda$ has been used everywhere in place of the λ of CT [see Eq. (16)].

[9] These generalized Ward identities can be put to the same use as the original Ward identity in simplifying the analysis of overlapping divergences in the perturbation theoretic version of renormalization.

Let us consider now the power series

$$\mathfrak{S}_\lambda(z,\eta) \equiv \sum_{n=0}^{\infty} \frac{(2z)^n}{(2n)!} \int_V dx_1 \cdots \int_V dx_{2n} \times S_\lambda^{(2n)}(x_1 \cdots x_{2n};\eta)\,\Delta^{(2n)}(x_1 \cdots x_{2n}), \quad (15)$$

where

$$[\mathbf{S}_\lambda^{(2n)}(12\cdots 2n;\eta)]_{ij} \equiv S_\eta(x_i - x_j), \quad \text{for} \quad i \neq j,$$
$$[\mathbf{S}_\lambda^{(2n)}(12\cdots 2n;\eta)]_{ii} = S_\eta(0) - \lambda. \quad (16)$$

Diagrammatically, the new terms appearing in $\mathfrak{S}_\lambda(z,\eta)$, which would vanish when $\lambda = 0$, arise from the inclusion of point loop of strength λ attached by a zero-momentum boson line to fermion lines. This contribution can be determined by considering the λ contributions via $U_\lambda(z,\eta)$, the total amplitude for all connected diagrams. There are two types of λ-dependent diagrams appearing in $U_\lambda(z,\eta)$: (a) those connected diagrams where removal of the "λ-polyps" (i.e., λ point loop plus attached boson line) leads to a connected diagram already in $U(z,\eta)$, (b) those not of type (a), (see Figs. 1, 2). The contribution of type (a) [where we include $U_0(z,\eta)$ in the sum] is readily evaluated by considering the fact that the attachment of k λ-polyps is achieved by acting on $U(z,\eta)$ with

$$(1/k!)(2z\Delta_0\lambda)^k(\partial^k/\partial\eta^k). \quad (17)$$

The operator $\partial/\partial\eta$ attaches a zero-momentum boson line to all possible fermion lines in $U(z,\eta)$; $\Delta(0) \equiv \Delta_0$ is the propagator for the zero-momentum boson line; $2z\lambda = g^2\lambda$ is the strength factor arising from the two ends of the boson line and the strength λ of the point loop, $1/k!$ removes the multiplicity

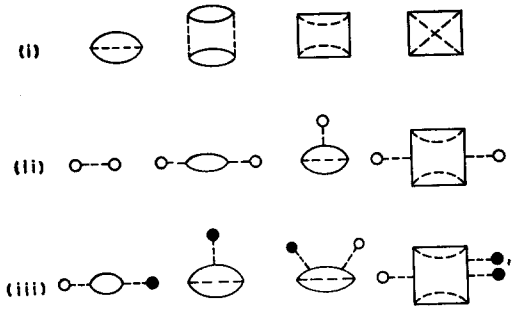

FIG. 1. Type-(a) diagrams. The hollow circle denotes a point fermion–antifermion loop of strength $S_\eta(0)$. The black circle denotes the point loop of strength λ (λ-polyp). On line (i) are some typical connected vacuum diagrams corresponding to the standard vacuum S matrix. Line (ii) includes some of the connected diagrams associated with the augmented vacuum S matrix. Line (iii) includes some of the vacuum diagrams arising from the inclusion of the λ-polyp, which, upon detachment of the λ-polyps, are part of the augmented vacuum S matrix.

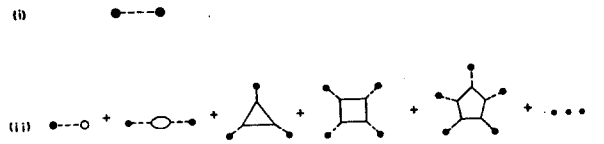

FIG. 2. Type-(b) diagrams. The diagram on line (i), upon removal of the λ-polyps, reduces to the null diagram. The diagram (ii), upon removal of λ-polyps, reduce to the point loop $S_\eta(0)$ which is not part of the augmented amplitude.

due to the different orders of attaching the polyps to a diagram and all terms appear with the same sign. The total contribution for (a)-type diagrams is

$$\mathring{U}_\lambda(z,\eta) = \sum_{k=0}^{\infty} \frac{(\lambda\nu)^k}{k!} \frac{\partial^k}{\partial\eta^k} U(z,\eta) = U(z,\eta+\lambda\nu), \quad (18)$$

where

$$\nu \equiv 2\Delta_0 z.$$

Another way of obtaining this result is by considering

$$\mathring{\mathfrak{S}}_\lambda(z,\eta) \equiv \exp[-V\mathring{U}_\lambda(z,\eta)]. \quad (19)$$

The net effect of including the λ-polyps in $\mathring{\mathfrak{S}}(z,\eta)$ can be obtained by summing the contribution of all λ-polyps attached to each fermion line. This is just a mass renormalization of the fermion propagator $\hat{S}_\eta(k)$, and the renormalized propagator is [see Fig. 3]

$$\hat{S}_\eta^R(k) = \hat{S}_\eta(k) + \lambda\nu\hat{S}_\eta(k)^2 + \lambda^2\nu^2\hat{S}_\eta(k)^3 + \cdots = [\hat{S}_\eta(k)^{-1} - \lambda\nu]^{-1} = [\hat{S}(k)^{-1} - \eta - \lambda\nu]^{-1} = \hat{S}_{\eta+\lambda\nu}(k). \quad (20)$$

Thus, η is essentially the mass, i.e., the quantity to be augmented by mass renormalization in the case of regularized propagators. From Eq. (20), one immediately finds

$$\mathring{\mathfrak{S}}_\lambda(z,\eta) = \mathfrak{S}(z,\eta+\lambda\nu) \quad (21)$$

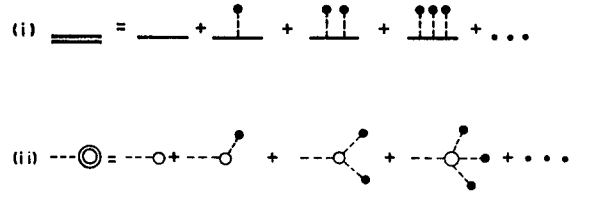

FIG. 3. The double heavy line in (i) denotes the renormalized fermion line obtained by including any number of λ-polyp attachments. The same renormalization of the point loop of strength $S_\eta(0)$ is shown on line (ii). This renormalization in the augmented amplitude gives rise to the augmented S matrix with all λ-polyp attachments whose removal leads back to diagrams in the augmented S matrix [viz. type (a)].

or

$$\overset{a}{U}_\lambda(z, \eta) = U(z, \eta + \lambda\nu)$$

as in Eq. (18).

The (b)-type diagrams (as mentioned), upon detachment of the λ polyps, reduce to diagrams not contained in $U(z, \eta)$. Since any connected diagram with internal meson lines is included in $U(z, \eta)$, the (λ-polyp) detached (b)-type diagrams consist exactly of two diagrams, the "null" diagram containing no lines, and a single, closed fermion loop with strength $S_\eta(0)$. From the null diagram, by attachment of λ polyps, one can form only one diagram (see Fig. 2) (remembering that a λ-point loop can connect to only one boson line). From the point fermion loop, one can form simply the set of all polygons with λ polyps connected at each vertex (see Fig. 2). The sum of contributions from all of these is readily evaluated in closed form. One finds

$$-\overset{b}{U}_\lambda(z, \eta) = \tfrac{1}{2}\nu\lambda^2 - \sum_{m=1}^{\infty} \frac{(\nu\lambda)^m}{m} \frac{1}{V} \sum_{k\in\mathfrak{L}} [\hat{S}_\eta(k)]^m$$

$$= \tfrac{1}{2}\nu\lambda^2 + \frac{1}{V} \sum_{k\in\mathfrak{L}} \ln [1 - \lambda\nu\hat{S}_\eta(k)]. \qquad (22)$$

In Eq. (22), the first term is the contribution from the attached null diagram, while the remaining sum is over the attached polygons. The $1/m$ denominator in the polygon sum is an enumerative factor.[10] An alternative expression for the sum in Eq. (22) gives

$$-\tfrac{1}{2}\nu\lambda^2 - \overset{b}{U}_\lambda(z, \eta) = \frac{1}{V} \sum_k \int_0^{\nu\lambda} d\rho \, \frac{1}{\hat{S}_\eta(k)^{-1} - \rho}$$

$$= \frac{1}{V} \sum_k \int_0^{\nu\lambda} d\rho \, \hat{S}_{\eta+\rho}(k) = \int_\eta^{\eta+\nu\lambda} d\rho \, \mathcal{S}_\rho(0). \qquad (23)$$

Thus, from Eqs. (21), (22), and (23),

$$\mathcal{S}_\lambda(z, \eta) = \exp\{-V[\overset{a}{U}_\lambda(z, \eta) + \overset{b}{U}_\lambda(z, \eta)]\}$$

$$= \exp(V\Delta_0\lambda^2 z)\mathcal{S}(z, \eta + \lambda\nu) \prod_{k\in\mathfrak{L}} [1 - \lambda\nu\hat{S}_\eta(k)]$$

$$= \mathcal{S}(z, \eta + \lambda\nu) \exp V\left[\tfrac{1}{2}\nu\lambda^2 - \int_\eta^{\eta+\lambda\nu} d\rho \, S_\rho(0)\right]. \qquad (24)$$

The last form corresponds to Eq. B(18) in CT. Setting $\lambda = S_\eta(0)$, one obtains

$$\mathcal{S}_{-s_\eta(0)}(z, \eta) = \mathring{\mathcal{S}}(z, \eta)$$

$$= \exp(\tfrac{1}{2}S_\eta^2(0)\nu V) \prod_{k\in\mathfrak{L}} [1 - \nu S_\eta(0)\hat{S}_\eta(k)]$$

$$\times \mathcal{S}(z, \eta + \nu S_\eta(0)) = \mathcal{S}[z, \eta + 2\Delta_0 z S_\eta(0)]$$

$$\times \exp V\left[\Delta_0 z S_\eta^2(0) - \int_\eta^{\eta+2\Delta_0 z S_\eta(0)} d\rho \, S_\rho(0)\right] \qquad (25)$$

in accordance with the corrected Eq. B(19) of CT.

IV. ANALYTIC PROPERTIES OF THE PLRR

There is an interesting analytical structure in the relations Eqs. (24) and (25). In AP some of the analytical properties of $\mathcal{S}(z, \eta)$ and $\mathring{\mathcal{S}}(z, \eta)$ in the variable η were discussed, though the discussion there emphasized the branch cut in η obtained in the $V = \infty$ limit. For simplicity, and with no loss of generality from a physical point of view, let us assume that the function $\hat{S}(k)$ repeats no value for $k \in \mathfrak{L}$. For finite V, the functions $\mathcal{S}(z, \eta)$ and $\mathring{\mathcal{S}}(z, \eta)$ have singularities in η at the points

$$\eta_i \equiv [\hat{S}(k_i)]^{-1}, \qquad k_i \in \mathfrak{L}. \qquad (26)$$

These originate in the function $S_\eta(x)$ which is a meromorphic function in η having simple poles at the η_i. The set of η_i is denoted by $\{\eta\}$. The function $\mathcal{S}(z, \eta)$ has simple poles at the points of $\{\eta\}$, while $\mathring{\mathcal{S}}(z, \eta)$ has an essential singularity at these points. This is readily recognized from the relations:

$$\mathcal{S}(z, \eta) = \sum_{n=0}^{\infty} \frac{z^n}{n!\, V^{2n}} \sum_{0\le i_1<i_2\cdots<i_{2n}} \prod_{j=1}^{2n} \hat{S}_\eta(k_{i_j})$$

$$\times \int_V dx_1 \cdots \int_V dx_{2n} \, |D(k_{i_1} \cdots k_{i_{2n}}; x_1 \cdots x_{2n})|^2$$

$$\times \Delta(1, 2)\, \Delta(3, 4) \cdots \Delta(2n - 1, 2n), \qquad (27)$$

where $\Delta(i, j) \equiv \Delta(x_i - x_j)$, and

$$\mathring{\mathcal{S}}(z, \eta) = \sum_{n=0}^{\infty} \frac{z^n}{n!\, V^{2n}} \sum_{i_1, i_2\cdots i_{2n}} \hat{S}_\eta(k_{i_j})$$

$$\times \int_V dx_1 \cdots \int_V dx_{2n} \exp\left(-i \sum_{j=1}^{2n} k_{i_j} x_j\right)$$

$$\times D_0(k_{i_1} \cdots k_{i_{2n}}; x_1 \cdots x_{2n})$$

$$\times \Delta(1, 2) \cdots \Delta(2n - 1, 2n). \qquad (28)$$

Equation (27) is the same as Eq. (24) of EO, and Eq. (25) is readily established by the same type of reasoning as Eq. (27). For any $k_i \in \mathfrak{L}$, each term in the expansion in Eq. (27) either has a simple pole at $\eta = \eta_i$ or none at all, while in Eq. (28), a factor $\hat{S}_\eta(k_i)$ appears to arbitrarily high powers. That an essential singularity in η is compounded in this way is to be seen from the PLRR.

What about the analytic properties in z? It has already been shown in AP and EO that, for $t > 4$ [i.e., $S_\eta(0)$ finite], both $\mathcal{S}(z, \eta)$ and $\mathring{\mathcal{S}}(z, \eta)$ are entire

[10] A detailed discussion of this enumerative factor can be found in W. M. Frank, J. Math. Phys. 5, 363 (1964), Sec. IV.

functions in z. The PLRR is very illuminating in this connection. The function $\mathcal{S}(z, \eta + 2\lambda\Delta_0 z)$ appearing in Eqs. (24) and (25) has a functional dependence on z arising both from the "explicit" dependence on z (i.e., the first argument variable), and the dependence on $\eta + 2\lambda\Delta_0 z$ (i.e., the second argument variable). It will have simple poles in z arising from this second dependence at points

$$z_i = (2\Delta_0\lambda)^{-1}[\hat{S}(k_i)^{-1} - \eta] = [2\lambda\Delta_0\hat{S}_\eta(k_i)]^{-1}. \quad (29)$$

However, these are exactly cancelled by factors in

$$F_\lambda(z, \eta) \equiv \prod_{k\in\mathfrak{L}} [1 - 2\lambda\Delta_0 z\hat{S}_\eta(k)], \quad (30)$$

leaving $\mathcal{S}_\lambda(z, \eta)$ entire in z. $\mathring{\mathcal{S}}(z, \eta)$ is, of course, a special case of this. A further comment on analytic properties in z will be made later.

Essential Singularity in η

Let $\mathfrak{L}[k_1, \cdots, k_l]$ denote the momentum lattice *excluding* the points $k_1, k_2, \cdots, k_l$, and let $\{\eta[k_1, \cdots, k_l]\}$ denote the set $\{\eta\}$ excluding the points $\hat{S}(k_1)^{-1}$, $\hat{S}(k_2)^{-1}$, $\cdots$, $\hat{S}(k_l)^{-1}$ ($\equiv \eta_1, \eta_2, \cdots, \eta_l$, respectively). Let

$$F_\lambda(z, \eta; k_1, \cdots, k_l) \equiv \prod_{k\in\mathfrak{L}(k_1,\cdots,k_l)} [1 - 2\lambda\Delta_0 z\hat{S}_\eta(k)]. \quad (31)$$

This function has a simple pole in η at each point of $\{\eta[k_1, \cdots, k_l]\}$. From Eqs. (24) and (27),

$$\begin{aligned} Q(z, \eta; \lambda) &\equiv \exp(-V\lambda^2\Delta_0 z)\mathcal{S}_\lambda(z, \eta) \\ &= \mathcal{S}(z, \eta + 2\lambda\Delta_0 z)F_\lambda(z, \eta) \\ &= \sum_{n=0}^{\infty} \frac{z^n}{n!V^n} \sum_{0\le i_1<i_2\cdots<i_{2n}} F_\lambda(z, \eta; k_{i_1}, k_{i_2}, \cdots, k_{i_{2n}}) \\ &\quad \times \prod_{j=1}^{2n} \hat{S}_\eta(k_{i_j})\mathfrak{D}(k_{i_1}, \cdots, k_{i_{2n}}), \end{aligned} \quad (32)$$

where

$$\begin{aligned} \mathfrak{D}(k_1, \cdots, k_{2n}) &\equiv \frac{1}{V^n}\int_V dx_1 \cdots \int_V dx_{2n} \\ &\quad \times |D(k_1, \cdots, k_{2n}; x_1, \cdots, x_{2n})|^2 \\ &\quad \times \Delta(1, 2)\,\Delta(3, 4)\cdots\Delta(2n-1, 2n), \end{aligned} \quad (33)$$

and the relation

$$\hat{S}_{\eta+\rho}(k) = \hat{S}_\eta(k)/[1 - \rho\hat{S}_\eta(k)] \quad (34)$$

has been used. In Eq. (32) it is clear that each term in the last expression has a simple pole in η at each point of $\{\eta\}$. Moreover, the arguments in AP make clear that the only singularities of $\mathring{\mathcal{S}}(z, \eta)$ in η are for $\eta \in \{\eta\}$. The sum of the series in Eq. (32) has at worst a simple pole at each η of $\{\eta\}$ and the same is true for $\mathcal{S}_\lambda(z, \eta)$ for any fixed λ. When $\lambda = S_\eta(0)$ is chosen, as in

$$\begin{aligned} \mathring{\mathcal{S}}(z, \eta) &= \exp[VS_\eta^2(0)\,\Delta_0 z]F_{s_\eta(0)}(z, \eta) \\ &\quad \times \mathcal{S}[z, \eta + 2\Delta_0 z S_\eta(0)], \end{aligned} \quad (35)$$

new singularities in η appear as a result of the simple poles of $S_\eta(0)$ for $\eta \in \{\eta\}$. Then, both the exponential factor and the remaining factors in Eq. (35) both involve poles of infinite order, i.e., essential singularities for $\eta \in \{\eta\}$. We show that no systematic cancellation between these singularities takes place, so that the poles of infinite order appearing in $\mathring{\mathcal{S}}(z, \eta)$ are really there and represent an essential singularity in η. Let us assume that the function $\mathring{\mathcal{S}}(z, \eta)$ has poles of a maximum finite order p at $\eta = \eta_i$ (and therefore for all $\eta \in \{\eta\}$ because of the symmetric dependence on the lattice points). The residue $R_i^{(p+1)}(z)$ of the pole of order $p + 1$ at $\eta = \eta_i$ would have to be identically zero. This residue is expressible as

$$R_i^{(p+1)}(z) = \frac{1}{2\pi i}\int_{C_i} d\eta'\,(\eta' - \eta_i)^p\mathring{\mathcal{S}}(z, \eta') \quad (36)$$

with C_i a closed contour in the η-plane enclosing only η_i from among the singular points in $\{\eta\}$. From the fact that the location of the singularities in η is independent of z and that $\mathring{\mathcal{S}}(z, \eta)$ is analytic everywhere in z for all η at which $\mathring{\mathcal{S}}(z, \eta)$ is analytic (see AP Sec. B), one can readily establish that $R_i^{(p+1)}(z)$ defined by Eq. (36) is an analytic function of z, which by hypothesis should be identically zero. This is contradicted by the nonvanishing value of

$$\frac{\partial^{p+1}}{\partial z^{p+1}} R_i^{(p+1)}(z)|_{z=0} \quad (37)$$

as obtained explicitly from the power series expression in Eq. (28).

We conclude that $\mathring{S}(z, \eta)$ has an essential singularity in η for all $\eta \in \{\eta\}$. The significance of this result will be discussed in Sec. V.

Exponential Order of $Q(z, \eta; \lambda)$

While it was shown in AP that $\mathring{\mathcal{S}}(z, \eta)$ has exponential order unity,[7] it will now be shown that this growth arises solely from the $\exp(\frac{1}{2}\lambda^2\nu V)$ factor in Eq. (25), and that, in fact, $Q(z, \eta; \lambda)$ of Eq. (32) has exponential order $\rho_Q = 4/t$ which is taken to be less than unity in this paper. The result is a relatively straightforward consequence of the rightmost expression in Eq. (32). From the fact that,

for each $0 < \alpha \leq 1$, there is a constant C_α such that, for all complex z

$$|1 + z| \leq \exp(C_\alpha |z|^\alpha), \tag{38}$$

it follows that if $\alpha > 4/t$

$$|F_\lambda(z, \eta; k_1, \cdots, k_l)| = \prod_{k \in \mathcal{L}[k_1, \cdots k_l]} |1 - 2\lambda\Delta_0 z S_\eta(k)| \leq \exp(\mu |z|^\alpha), \tag{39}$$

where

$$\mu = C_\alpha |2\lambda\Delta_0|^\alpha \sum_{k \in \mathcal{L}} |\hat{S}_\eta(k)|^\alpha < \infty, \tag{40}$$

since the numbers $\hat{S}_\eta(k)$ have exponent of convergence $4/t$.[11] Then, in the notation, and by the same arguments of Sec. II in EO,

$$|Q(z, \eta; \lambda)| \leq e^{\mu|z|^\alpha} \sum_{n=0}^{\infty} \frac{|z|^n}{n!\, V^n} (2n)^n \Delta_1^n \times \sum_{0 \leq i_1 < i_2 < \cdots < i_{2n}} \prod_{j=1}^{2n} |\hat{S}_\eta(k_{i_j})|. \tag{41}$$

The summation on the right is precisely of the form of the upper bound in Eq. (26) of EO, whose growth for large $|z|$ was shown to correspond to an exponential order $\rho_* = (\frac{1}{2}t - 1)^{-1}$. The growth of $|Q(z, \eta; \lambda)|$ in z therefore corresponds to α, which is the larger of the two exponential orders α and ρ_* if $t > 4$. The greatest lower bound of α is $4/t$ subject to the convergence of the sum in Eq. (40).

V. DISCUSSION

The renormalization relation in this specialized case has been seen to cause some interesting and quite marked modifications in the analytic properties of the perturbation expansion. The renormalized expansion (i.e., the sum over the more restricted class of renormalized diagrams, namely, those without the point fermion–antifermion loop) has a denser distribution of zeros and a higher exponential growth in z. A particularly striking result is the emergence of an essential singularity in η from the renormalization. In the case of a fermion of mass zero, the point $\eta = 0$ becomes one of the singular points of $S(z, \eta)$ and $\mathring{S}(z, \eta)$, and, in the latter case, it is an essential singularity. The same thing happens if the boson has mass zero in which case Δ_0 is unbounded. In these cases the contributions of large numbers of external zero momentum lines no longer fall off in the manner characteristic of the high terms in a convergent power series. This corresponds to the fact that the contribution of processes involving arbitrarily large numbers of photons or neutrinos have a qualitative effect on the total behavior of a process. In the $V \to \infty$ limit, the branch cut in η discussed in AP runs into the origin $\eta = 0$ for both $U(z, \eta)$ and $\mathring{U}(z, \eta)$. These cuts in $U(z, \eta)$ and $\mathring{U}(z, \eta)$ differ presumably in some very characteristic way (since one is an accumulation point of simple poles, and the other of essential singularities). These matters are deferred for a later investigation. In a forthcoming paper, we shall discuss the interesting situation when an infinite "subtraction" is necessary for renormalization, as well as the program for tackling the nontrivial renormalizations.

ACKNOWLEDGMENT

The author is grateful to Dr. A. Katz for very valuable criticism.

[11] The exponent of convergence of a set of numbers $\{\alpha_n\}$ is the greatest lower bound of the set of numbers r for which $\sum_n |\alpha_n|^r < \infty$. The exponent of convergence of the set $\{\hat{S}_\eta(k)\}$ was established in Sec. I of EO.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 2 FEBRUARY 1967

Instantaneous Action-at-a-Distance in Classical Relativistic Mechanics

ROBERT NYDEN HILL
Department of Physics, University of Delaware, Newark, Delaware
(Received 29 November 1965)

The possibility of describing orbits in classical relativistic mechanics in instantaneous action-at-a-distance fashion by second-order differential equations (as in Newton's gravitational theory) is investigated with particular emphasis on the two-body problem of classical relativistic electrodynamics. Differential conditions are stated to guarantee world-line invariance and form-invariance of the equations of motion under Lorentz transformation for such a description of an N-particle system in three dimensions. A pair of integrodifferential equations for the equations of motion are derived to provide an explicit means of passing from a description via direct interaction along light cones to an instantaneous action-at-a-distance description for a two-body problem. These integrodifferential equations are applicable to the two-body problem of classical electrodynamics with either retarded interactions and radiation damping or with half-advanced plus half-retarded interactions.

I. INTRODUCTION

A BASIC difficulty of the field-theoretic description of interparticle interaction in relativistic mechanics is the necessity of introducing an infinite number of degrees of freedom to describe the world lines of a finite number of particles. As a consequence of this, the simplest problem in the relativistic electrodynamics of interacting particles, namely the two-body problem, has never been solved. The elimination of the fields by making use of the Lienard–Wiechert potentials does not eliminate the difficulty; the resulting equations of motion have a differential-difference structure and the two-body problem still defies solution. If one attempts to escape the difficulty by expanding these differential-difference equations in a Taylor series about some observer's present time, one obtains infinite-order differential equations—again a description involving an infinite number of degrees of freedom.

We will begin our inquiry by focusing on the two-body problem, which already contains all of the aforementioned difficulties. With the exception of the circular-orbit solutions found by Schild,[1] solutions to the two-body problem are known only in the nonrelativistic limit, in which one obtains the usual Kepler problem discussed in undergraduate mechanics texts, or in the limit in which the mass of one of the particles goes to infinity,[2] which is no longer a two-body problem.

"Why can one solve two-body problems in nonrelativistic mechanics?" is a logical first question to ask in seeking to solve a two-body problem in relativistic mechanics. The answer is essentially the following: Invariance under space and time translation provides four constants of the motion[3] (momentum and energy conservation), rotational invariance provides three (angular momentum conservation), and invariance under Galilean velocity transformations provides three (center-of-mass theorem) for a total of ten constants of the motion. These conservation laws permit the reduction of the primitive equations of motion, whose solution contains twelve arbitrary constants, to a second-order ordinary differential equation which can be solved easily for many interaction potentials. In the relativistic situation, the Galilean velocity transformations are replaced by the pure Lorentz transformations; the geometric invariances should still provide ten constants of the motion. However, the fact that the interaction is viewed as occurring along light cones rather than instantaneously prevents the reduction to a single second-order differential equation.

The two situations are contrasted in Fig. 1. If instantaneous action-at-a-distance is employed, the acceleration at the world point P_0 is written in terms of coordinates and velocities at P_0 and Q_0; the acceleration at Q_0 is also written in terms of quantities at P_0 and at Q_0, and one obtains a closed set of equations. However, if one describes the world lines by means of retarded interactions along light cones, the acceleration at P_0 is written in terms of quantities at P_0 and Q_1, the acceleration at Q_1 is written in terms of quantities at Q_1 and P_1, and so on. One is caught in an infinite regression and does not obtain a closed set of equations. The same difficulty arises

[1] A. Schild, Phys. Rev. 131, 2762 (1963).
[2] A. Sommerfeld, Ann. Physik 51, 1 (1916).
[3] This follows immediately from Noether's theorem if an action principle exists. A clear and concise discussion of Noether's theorem is given by E. L. Hill, Rev. Mod. Phys. 23, 253 (1951).

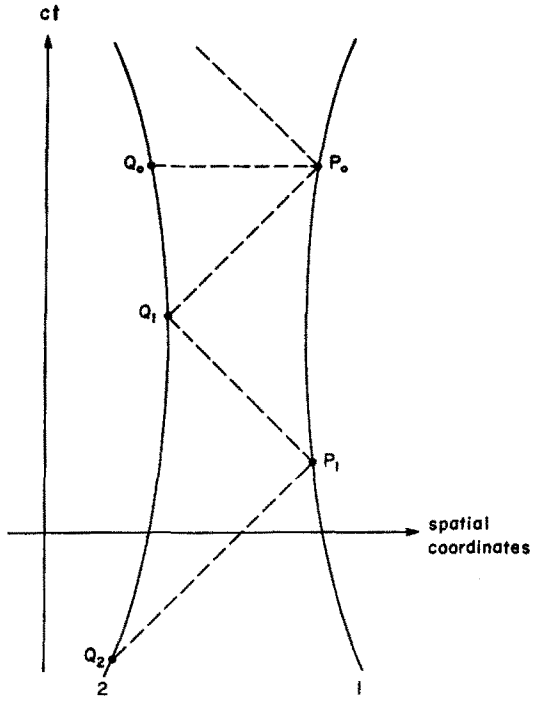

FIG. 1. Instantaneous action-at-a-distance compared with direct interaction along light cones.

if some combination of retarded and advanced interactions along light cones is employed, as in Wheeler–Feynman[4] electrodynamics and in the generalization of Wheeler–Feynman electrodynamics given by Rohrlich.[5]

A way out of the difficulty in the relativistic situation has been shown by Kerner.[6] Kerner expands the differential-difference equations which one obtains in the case of two particles interacting via half-advanced plus half-retarded potentials in a Taylor series about some observer's present time and obtains a set of infinite-order differential equations. He then proceeds to systematically depress the order of the equations by means of an expansion in powers of e^2, thus obtaining an instantaneous action-at-a-distance description of the orbits via second-order differential equations.

The use of Taylor expansion followed by order depression to obtain successive approximations to an instantaneous action-at-a-distance description is a quite general procedure, being in principle limited only by the question of convergence of the expansions involved. Section II provides a relatively simple example of such a "brute force" perturbative calculation for a one-dimensional electrodynamic two-body problem. The possibility of using such a successive approximation procedure implies that an instantaneous action-at-a-distance description is in principle always available, even if energy and momentum are carried away (as radiation) by a field; all that is required *a priori* is a decision to confine attention to the particle's world lines, as is done in hydrodynamics when the equations for the velocity field of the fluid are solved (with appropriate boundary conditions) to obtain the viscous drag force (of which Stoke's law is the first term of the low Reynolds number expansion for the force on a sphere).

In Sec. III, the compatibility of instantaneous action-at-distance with special relativity is investigated. Differential conditions guaranteeing this compatibility for a one-dimensional two-body problem are derived and used to show that summing to all orders in c^{-1} for each order in e^2 of the expansion of Sec. II is not sufficient: all powers of e^2 as well as all powers of c^{-1} must appear for compatibility with special relativity.

In Sec. IV, the results of Sec. II and the insight of Sec. III into the role played by special relativity are used as stepping stones to a more powerful method of dealing with the problem of reducing the one-dimensional electrodynamic two-body problem of Sec. II to instantaneous action-at-a-distance form. A pair of integrodifferential equations, which formally sum the approximation method of Sec. II, are derived; their solutions (the instantaneous action-at-a-distance equations of motion) satisfy the conditions of Sec. III for compatibility with special relativity. These integrodifferential equations are restricted in their application to the reduction of interactions along light cones to instantaneous interaction form; they are thus less general than the method of Sec. II.

The basic ideas and methods developed in Secs. III and IV are generalized to three dimensions in Sec. V. The resulting equations can be used to sum Kerner's scheme. The development, although considerably more complicated in detail, parallels that given in the one-dimensional case. The inclusion of radiation damping is also considered. In Sec. VI, the differential conditions expressing compatibility of special relativity with instantaneous action-at-a-distance for N particles in three dimensions are derived and used to show that the principle of relativity demands many-body forces in an instantaneous action-at-a-distance description.

[4] J. A. Wheeler and R. P. Feynman, Rev. Mod. Phys. 21, 425 (1949).
[5] F. Rohrlich, Phys. Rev. Letters 12, 375 (1964); F. Rohrlich, *Classical Charged Particles* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1965), Sec. 7-1.
[6] E. H. Kerner, J. Math. Phys. 6, 1218 (1965).

II. A ONE-DIMENSIONAL ELECTRODYNAMIC TWO-BODY PROBLEM

The basic difficulty of the finite propagation time required for the interaction is already present in the problem of two three-dimensional point particles, whose motion is confined to one dimension as a consequence of the initial data. Inasmuch as both the radiation fields and the magnetic field due to one particle then vanish at the location of the other particle, we will examine this simpler problem first before attacking the complexity of the full three-dimensional problem. As a further simplification, we assume that the two particles have the same charge and the same mass: $e_1 = e_2 = e$ and $m_1 = m_2 = m$. We employ the Lienard–Wiechert expressions[7] for the fields of a point charge and assume that particle 1 is always to the right of particle 2 so that $x_{12} = x_1(t_0) - x_2(t_0)$ is always positive. The retarded and advanced fields $(E_1)_{\text{ret}}$ and $(E_1)_{\text{adv}}$ felt by particle 1 due to the motion of 2 are

$$(E_1)_{\text{ret}} = \left[\frac{e}{x_{12}^2}\left(\frac{1 + v_2/c}{1 - v_2/c}\right)\right]_{\text{ret}}, \qquad (E_1)_{\text{adv}} = \left[\frac{e}{x_{12}^2}\left(\frac{1 - v_2/c}{1 + v_2/c}\right)\right]_{\text{adv}}. \tag{1}$$

The fields $(E_2)_{\text{ret}}$ and $(E_2)_{\text{adv}}$ felt by particle 2 due to the motion of 1 are

$$(E_2)_{\text{ret}} = -\left[\frac{e}{x_{12}^2}\left(\frac{1 - v_1/c}{1 + v_1/c}\right)\right]_{\text{ret}}, \qquad (E_2)_{\text{adv}} = -\left[\frac{e}{x_{12}^2}\left(\frac{1 + v_1/c}{1 - v_1/c}\right)\right]_{\text{adv}}. \tag{2}$$

The equations of motion are

$$m\dot{v}_1 = (1 - v_1^2)^{\frac{3}{2}} eE_1, \qquad m\dot{v}_2 = (1 - v_2^2)^{\frac{3}{2}} eE_2. \tag{3}$$

We will systematically rewrite the equations of motion in instantaneous action-at-a-distance form by means of an expansion in powers of the coupling constant e^2. We will do this for both purely retarded and half-advanced plus half-retarded interactions, but will include no radiation damping, since the radiation damping term as given by Dirac[8] has the form of a finite self-interaction term and is not involved in the basic difficulty of a finite propagation time for the interaction.

Before proceeding, we note that the mass and charge appear only in the combination e^2/m after the fields have been introduced into the equations of motion. Thus, the only length which can be formed from the available parameters is the classical charge radius e^2/mc^2; the only time is e^2/mc^3. If this length and this time are used to rewrite the equations of motion in dimensionless form, it becomes clear that *an expansion in powers of the coupling constant e^2 is necessarily also an expansion in inverse powers of the spatial distance x_{12}*. Inasmuch as an expansion in inverse powers of a spatial distance is typically not Lorentz invariant, we cannot expect any approximate instantaneous action-at-a-distance electrodynamics (obtained by truncating the expansion in powers of e^2 at some finite stage) to be Lorentz invariant. *Only the full theory obtained by summing the expansion will be compatible with the principle of relativity*, as will be shown in the next section.

We now proceed to construct the first few terms of the expansion. These first few terms will give us something to work with in later sections. We set $e^2/m = c = 1$ and write instantaneous action-at-a-distance equations of Newtonian order in the form

$$\dot{v}_1 = (1 - v_1^2)^{\frac{3}{2}} \sum_{n=0}^{\infty} e_n(v_1, v_2)/x_{12}^{n+2}, \qquad \dot{v}_2 = -(1 - v_2^2)^{\frac{3}{2}} \sum_{n=0}^{\infty} \bar{e}_n(v_1, v_2)/x_{12}^{n+2}, \tag{4}$$

where x_1, x_2, v_1, and v_2 are all to be evaluated at the present time. Since the particles are identical, $\bar{e}_n(v_1, v_2) = e_n(v_2, v_1)$, the terms $e_n(v_1, v_2)/x_{12}^{n+2}$ are successive approximations to the field E_1 rewritten in terms of present positions and velocities. The retarded and advanced times are computed from

$$x_1(t_0) - x_2(t_{\text{ret}}) = t_0 - t_{\text{ret}}, \qquad x_1(t_0) - x_2(t_{\text{adv}}) = t_{\text{adv}} - t_0, \tag{5}$$

by making Taylor series expansions about the present time t_0. We obtain

$$t_0 - t_{\text{ret}} = \frac{x_{12}}{1 - v_2} + \frac{1}{1 - v_2} \times \sum_{m=1}^{\infty} \frac{(-1)^m}{(m+1)!} \frac{d^m v_2}{dt^m} (t_0 - t_{\text{ret}})^{m+1} \tag{6}$$

for the retardation.

The infinite series, involving as it does the acceleration and all higher derivatives (to be computed by differentiating the equations of motion), is a

[7] L. Landau and E. Lifshitz, *The Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1951), p. 177.

[8] P. A. M. Dirac, Proc. Roy. Soc. (London) A167, 148 (1938). See also F. Rohrlich, *loc. cit.* The *mathematical* consistency of using purely retarded interactions while neglecting radiation damping has been emphasized by J. L. Synge, Proc. Roy. Soc. (London) A177, 118 (1940).

correction to the first term, which gives $t_0 - t_{ret}$ when particle 2 is in straight-line motion. The approximation scheme proceeds by computing better approximations to the series; the first n terms of (4) are sufficient to compute $t_0 - t_{ret}$ accurately enough for the computation of the $(n + 1)$st term of (4). Since $t_0 - t_{ret} = (x_{12})_{ret}$, a knowledge of $t_0 - t_{ret}$ to a given order of approximation gives $(x_{12})_{ret}$ to that order of approximation. $(v_2)_{ret}$ is computed from the series

$$(v_2)_{ret} = v_2 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \frac{d^n v_2}{dt^n} (t_0 - t_{ret})^n, \tag{7}$$

which is similarly evaluated by successive approximations. To lowest order, $(v_2)_{ret} \approx v_2$ and $(x_{12})_{ret} \approx x_{12}(1 - v_2)$. Hence, denoting retarded quantities by a subscript r and advanced quantities by a subscript a, the zeroth-order retarded field is $e_{0r}/x_{12}^2 = (1 - v_2^2)/x_{12}^2$. The zeroth-order advanced field is the same: $e_{0a} = e_{0r}$. The details of computing the next two approximations are given in Appendix A; for retarded interactions without radiation damping the results are

$$e_{0r}(v_1, v_2) = 1 - v_2^2, \tag{8a}$$

$$e_{1r}(v_1, v_2) = -\frac{2(1 - v_1^2)(1 - v_2^2)^{5/2}}{(v_2 - v_1)^2} \times \left[\frac{(v_2 - v_1)(1 - v_1 v_2)}{(1 - v_1)(1 - v_2^2)} + \ln\left(\frac{1 - v_2}{1 - v_1}\right) \right], \tag{8b}$$

$$\begin{aligned} e_{2r}(v_1, v_2) &= \frac{(1 - v_1^2)^{3/2}(1 - v_2^2)^{5/2}}{(v_2 - v_1)^4} \ln\left(\frac{1 - v_2}{1 - v_1}\right) \xi_1(v_1, v_2) \\ &+ \frac{(1 - v_1^2)^{3/2}(1 - v_2^2)^{5/2}}{(v_2 - v_1)^3(1 - v_2)} \xi_2(v_1, v_2) \\ &+ \frac{3(1 - v_1^2)^2(1 - v_2^2)^4}{(v_2 - v_1)^4} \left[\ln\left(\frac{1 - v_2}{1 - v_1}\right) \right]^2 \\ &+ \frac{(1 - v_1^2)^2(1 - v_2^2)^3}{(v_2 - v_1)^4(1 - v_1)} \ln\left(\frac{1 - v_2}{1 - v_1}\right) \xi_3(v_1, v_2) \\ &+ \frac{(1 - v_1^2)^2(1 - v_2^2)^2}{(v_2 - v_1)^3(1 - v_1)^2} \xi_4(v_1, v_2), \end{aligned} \tag{8c}$$

where

$$\begin{aligned} \xi_1(v_1, v_2) &= 6 - 2v_1^2 - 2v_1 v_2 - 8v_2^2 \\ &\quad - 2v_1^3 v_2 + 4v_1^2 v_2^2 + 4v_1 v_2^3, \\ \xi_2(v_1, v_2) &= 6 + 3v_1 - 3v_2 - 3v_1 v_2 \\ &\quad - 9v_2^2 + v_1^2 v_2 - 5v_1 v_2^2 + 4v_2^3 - 3v_1^2 v_2^2 + 9v_1 v_2^3, \end{aligned} \tag{8d}$$

$$\begin{aligned} \xi_3(v_1, v_2) &= 6 - 12v_1 + 6v_2 \\ &\quad - 4v_1 v_2 - 2v_2^2 + 10v_1^2 v_2 - 4v_1 v_2^2, \\ \xi_4(v_1, v_2) &= 6 - 12v_1 + 6v_2 + 2v_1^2 \\ &\quad - 8v_1 v_2 - 6v_2^2 + v_1^3 + 11v_1^2 v_2 + 3v_1 v_2^2 \\ &\quad - 3v_2^3 - 2v_1^2 v_2^2 + 8v_1 v_2^3 - 4v_1^3 v_2^2 - 2v_1^2 v_2^3. \end{aligned}$$

If half-advanced plus half-retarded fields act, the zeroth (8a) and first (8b) approximations to the retarded fields remain the same, but the e_{2r} of (8c) must be replaced by $e_{2r} + \Delta e_{2r}$, where

$$\Delta e_{2r} = \frac{(1 - 2v_2 + v_1 v_2)(1 - v_1^2)^{5/2}(1 - v_2^2)^{5/2}}{(1 - v_1)^2(v_2 - v_1)^2} \times \left[\ln\left(\frac{1 + v_2}{1 + v_1}\right) - \ln\left(\frac{1 - v_2}{1 - v_1}\right) - \frac{2(v_2 - v_1)(1 - v_1 v_2)}{(1 - v_1^2)(1 - v_2^2)} \right]. \tag{9}$$

The successive approximations to the advanced fields e_{na} may be obtained from the e_{nr} by reversing the signs of the velocities. The apparent pole at $v_1 = v_2$ is spurious, since the numerators contain compensating zeros of the necessary order. The only singularities of the e_{na} and e_{nr} through $n = 2$ are logarithmic and square-root branch points at the velocity of light.

The most obvious conclusion to be drawn from the results (8) and (9) is that this is not the most efficient way to solve the problem. More powerful methods are needed if the general term of the series (4) is to be surveyed. It will nevertheless be reassuring to find that the more powerful group-theoretic methods of Secs. III and IV, which do make it possible to survey the general term, are in agreement with the more pedestrian calculations of this section.

III. THE PRINCIPLE OF SPECIAL RELATIVITY IN INSTANTANEOUS ACTION-AT-A-DISTANCE THEORIES

Better computational methods frequently follow from a better over-all understanding. We therefore seek to illuminate the role played by the (special) principle of relativity in the present problem.

The principle of relativity places two requirements on a theory of interacting point particles. The first is that all observers in inertial frames shall calculate with the same equations (form invariance of the equations of motion). The second is that the family of particle trajectories is the same no matter who calculates them: a pair of trajectories which form a solution of the equations of motion in one inertial

frame also forms a solution of the equations of motion when described by the coordinates appropriate to another inertial frame (world line invariance). For an instantaneous action-at-a-distance theory, these requirements can be satisfied by insisting that the fact that simultaneity is different for different observers be compensated by the differences (given by the usual Lorentz transformation formulas) in the sets of numbers assigned by different observers to kinematical quantities at the same world point.

For simplicity, we continue to work in one dimension. In the notation of Fig. 2, Q_0 and P_0 are simultaneous in the inertial frame S; Q_0' and P_0 are simultaneous in S'. Thus, an observer in S computes the acceleration at P_0 in terms of coordinates and velocities at Q_0 and P_0, while an observer is S' computes the acceleration at P_0 in terms of coordinates and velocities at Q_0' and P_0. We require that the two observers' expressions for the acceleration at P_0, written as

$$a_1(P_0) = f_1[x_1(P_0) - x_2(Q_0), v_1(P_0), v_2(Q_0)] \tag{10}$$

in S and as

$$a_1'(P_0) = f_1[x_1'(P_0) - x_2'(Q_0'), v_1'(P_0), v_2'(Q_0')] \tag{11}$$

in S' (same f_1 in both S and S'), agree when one makes use of the Lorentz transformation formula for accelerations

$$a_1'(P_0) = (1 - z^2)^{3/2}[1 - zv_1(P_0)]^{-3}a_1(P_0) \tag{12}$$

to relate the numbers assigned to the acceleration at P_0 by the two observers.

This requirement is most easily enforced by imposing it upon the infinitesimal Lorentz transformation. Inasmuch as it is (from the viewpoint of group theory) the requirement that the infinitesimal transformation on the form of the equations of motion vanish, the group property of the Lorentz transformations then guarantees that it holds for all proper Lorentz transformations. We now think of S' as moving with respect to S with an infinitesimal velocity $z = \delta\beta$. Then $x' = x - t\delta\beta$ and $t' = t - x\delta\beta$ are the kinematical transformations; $t(Q_0') - t(Q_0) = -[x_1(P_0) - x_2(Q_0)]\delta\beta$ takes account of the change in simultaneity. Using these,

$$\begin{aligned} x_1'(P_0) &= x_1(P_0) - t_0\,\delta\beta, \\ x_2'(Q_0') &= x_2(Q_0) - v_2(Q_0) \\ &\quad \times [x_1(P_0) - x_2(Q_0)]\,\delta\beta - t_0\,\delta\beta, \\ v_1'(P_0) &= v_1(P_0) - [1 - v_1^2(P_0)]\,\delta\beta, \end{aligned} \tag{13}$$

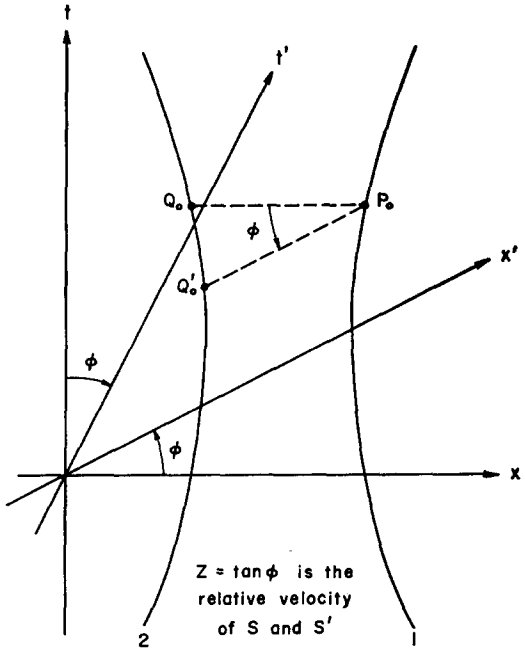

FIG. 2. Instantaneous action-at-a-distance in different Lorentz frames.

$$\begin{aligned} v_2'(Q_0') &= v_2(Q_0) - a_2(Q_0) \\ &\quad \times [x_1(P_0) - x_2(Q_0)]\,\delta\beta - [1 - v_2^2(Q_0)]\,\delta\beta, \\ a_1'(P_0) &= a_1(P_0) + 3a_1(P_0)v_1(P_0)\,\delta\beta. \end{aligned}$$

If we insert the transformations (13) into (11), expand to first order in $\delta\beta$, and demand agreement with (10) to first order, we obtain the condition

$$\begin{aligned} 3a_1v_1 = x_{12}v_2(\partial f_1/\partial x_{12}) &- (1 - v_1^2)(\partial f_1/\partial v_1) \\ &- (1 - v_2^2 + x_{12}a_2)(\partial f_1/\partial v_2). \end{aligned}$$

We now use $a_1 = f_1$ and assume that the acceleration of particle 2 has been written in the instantaneous action-at-a-distance form, $a_2 = f_2(x_{12}, v_1, v_2)$. We then arrive at the differential statements of the Lorentz invariance of a one-dimensional two-body instantaneous action-at-a-distance theory[9]

$$3v_1f_1 + Lf_1 = -x_{12}f_2(\partial f_1/\partial v_2), \tag{14a}$$

$$3v_2f_2 + \tilde{L}f_2 = x_{12}f_1(\partial f_2/\partial v_1), \tag{14b}$$

where the linear first-order differential operators L and $\tilde{L}$ are defined by

$$\begin{aligned} L \equiv -x_{12}v_2(\partial/\partial x_{12}) &+ (1 - v_1^2)(\partial/\partial v_1) \\ &+ (1 - v_2^2)(\partial/\partial v_2), \end{aligned} \tag{15a}$$

[9] Conditions equivalent to these have been dervied independently by D. G. Currie, Phys, Rev. 142, 817 (1966).

$$\tilde{L} \equiv -x_{12}v_1(\partial/\partial x_{12}) + (1 - v_1^2)(\partial/\partial v_1) + (1 - v_2^2)(\partial/\partial v_2). \quad (15b)$$

The conditions (14) are quite general, and apply to any two-body one-dimensional instantaneous action-at-a-distance theory.[10] The corresponding conditions in three dimensions are written down in Sec. V and generalized to N particles in Sec. VI.

We are now in a position to confirm the conclusion drawn from dimensional analysis at the beginning of Sec. II. We note that the operators $3v_1 + L$ and $3v_2 + \tilde{L}$ which appear in (14) are homogeneous of degree zero as far as x_{12} is concerned [e.g., $(3v_1 + L)$ operating on $x_{12}^m\lambda(v_1, v_2)$ with λ an arbitrary function of v_1, v_2 produces something of the form $x_{12}^m\mu(v_1, v_2)$]. Thus, a series solution to (14) in inverse powers of x_{12} which begins with x_{12}^{-2} must contain all inverse powers: any truncation of the series (4) cannot satisfy (14). We can, however, verify that the conditions (14) are satisfied to the order to which we have computed (4).

It is convenient to remove the factor $(1 - v_1^2)^{\frac{1}{2}}$, which arises from the relativistic variation of mass with velocity, and work directly with the electric field E_1, which is a scalar under Lorentz transformations in one dimension. We assume that E_1 has been rewritten as a function of instantaneous positions and velocities; then $f_1 = (1 - v_1^2)^{\frac{1}{2}}E_1$ and E_1 satisfies

$$LE_1 = -x_{12}f_2(\partial E_1/\partial v_2). \quad (16)$$

Direct computation now shows that

$$L(x_{12}^{-2}e_0) = 0,$$

$$L(x_{12}^{-3}e_1) = x_{12}^{-3}(1 - v_2^2)^{\frac{1}{2}}\bar{e}_0(\partial e_0/\partial v_2),$$

$$L(x_{12}^{-4}e_2) = x_{12}^{-4}(1 - v_2^2)^{\frac{1}{2}}[\bar{e}_0(\partial e_1/\partial v_2) + \bar{e}_1(\partial e_0/\partial v_2)]$$

hold for both the purely retarded and half-advanced plus half-retarded cases considered in Sec. II, so that the results (8) and (9) satisfy the differential expressions (14) of the principle of relativity to the accuracy of the computation.

IV. THE INTEGRODIFFERENTIAL EQUATION FOR THE FIELD IN ONE DIMENSION

The differential conditions (14) form a system of first-order partial differential equations; as such they have an infinity of solutions corresponding to different one-dimensional two-body problems. What singles out the solution to the electrodynamic two-body problem with which we began? An answer to this question may well provide the more efficient computational method sought after; we now address ourselves to this task.

The fact that the operator L in (16) operating on $x_{12}^m\lambda(v_1, v_2)$ (with λ arbitrary) produces something of the form $x_{12}^m\mu(v_1, v_2)$ for any m suggests computing the series in inverse powers of x_{12} by treating the right-hand side of (16) as a perturbation, i.e., inverting L and rewriting (16) as

$$E_1 = \frac{1 - v_2^2}{x_{12}^2} - L^{-1}\left(x_{12}f_2\frac{\partial E_1}{\partial v_2}\right). \quad (17)$$

Unfortunately, the infinity of possible solutions to (14) manifests itself in the infinity of possible inverses L^{-1}. There seems to be no obvious way to find L^{-1} other than by trial and error: one looks for an inverse L^{-1} which reproduces the e_1 and e_2 of (8) and (9) without the need for arbitrarily adding on solutions φ of the homogeneous equation $L\varphi = 0$.

After some experimentation, it was found useful to define the following pair of (formal) inverses[11] for L:

$$L_{\text{ret}}^{-1} \equiv \int_0^\infty e^{-sL}\,ds, \quad (18a)$$

$$L_{\text{adv}}^{-1} \equiv -\int_0^\infty e^{sL}\,ds. \quad (18b)$$

The effect of the operators (18a) and (18b) on a function $\varphi = \varphi(x_{12}, v_1, v_2)$ can be obtained by introducing the new variables $\xi = (1 - v_2^2)^{\frac{1}{2}}/x_{12}$, $\eta = \tanh^{-1}v_1$, and $\zeta = \tanh^{-1}v_2$. Then $L = \partial/\partial\eta + \partial/\partial\zeta$ and

$$e^{\pm sL}\psi(\xi, \eta, \zeta) = \psi(\xi, \eta \pm s, \zeta \pm s).$$

If we now set $z = \tanh s$ and transform back to the original variables, we find that

$$L_{\text{ret}}^{-1} = \int_0^1 \frac{dz}{1 - z^2}\,\mathfrak{R}(z), \quad (19a)$$

$$L_{\text{adv}}^{-1} = \int_0^{-1} \frac{dz}{1 - z^2}\,\mathfrak{R}(z), \quad (19b)$$

where the operator $\mathfrak{R}(z)$ is defined by

$$\mathfrak{R}(z)\varphi(x_{12}, v_1, v_2) = \varphi\left(\frac{x_{12}(1 - z^2)^{\frac{1}{2}}}{1 - v_2z}, \frac{v_1 - z}{1 - v_1z}, \frac{v_2 - z}{1 - v_2z}\right) \quad (20)$$

for φ any function of the three indicated arguments. The L_{ret}^{-1} and L_{adv}^{-1} are only inverses for L when applied to a class of functions which vanishes fast

[10] Some examples of such theories obtained by solving the Eqs. (14) have been given by E. H. Kerner, Phys. Rev. Letters **16**, 667 (1966).

[11] The consideration of a formal inverse of the type (18a) the suggested to the author by E. H. Kerner.

enough as $|v_1|$ or $|v_2| \to 1$ for the contributions at the upper limits in (18a) and (18b) to be discarded. In our problem, the fact that the electric fields E_1 and E_2 are bounded and of bounded variation, while $f_2 = (1 - v_2^2)^{\frac{1}{2}}E_2$ contains a factor $(1 - v_2^2)^{\frac{1}{2}}$ ensures that $[-x_{12}f_2(\partial E_1/\partial v_2)]$ has this property.

Direct computation and comparison with the results (8) and (9) show that

$$e_{1r} = x_{12}^3 L_{\text{ret}}^{-1} x_{12}^{-3}(1 - v_2^2)^{\frac{1}{2}}\tilde{e}_0(\partial e_0/\partial v_2),$$

$$e_{2r} = x_{12}^4 L_{\text{ret}}^{-1} x_{12}^{-4}(1 - v_2^2)^{\frac{1}{2}}[\tilde{e}_0(\partial e_{1r}/\partial v_2) + \tilde{e}_{1r}(\partial e_0/\partial v_2)],$$

$$\Delta e_{2r} = \tfrac{1}{2}x_{12}^4 L_{\text{ret}}^{-1} x_{12}^{-4}(1 - v_2^2)^{\frac{1}{2}}$$

$$\times [\tilde{e}_0(\partial e_{1a}/\partial v_2 - \partial e_{1r}/\partial v_2) + (\tilde{e}_{1a} - \tilde{e}_{1r})(\partial e_0/\partial v_2)].$$

These results suggest the conjecture

$$E_{1r} = \frac{1 - v_2^2}{x_{12}^2} - L_{\text{ret}}^{-1}\left(x_{12}f_2\frac{\partial E_{1r}}{\partial v_2}\right), \qquad (21a)$$

$$E_{1a} = \frac{1 - v_2^2}{x_{12}^2} - L_{\text{adv}}^{-1}\left(x_{12}f_2\frac{\partial E_{1a}}{\partial v_2}\right). \qquad (21b)$$

We note that the full f_2 (not just the advanced or retarded contribution to it) enters in (21a) and (21b). This is to be expected, since f_2 has implicit in it the second particle's orbit, which is just what is needed to obtain the fields. We now proceed to prove the correctness of (21).

The fact that $\mathcal{R}(z)$, operating on the velocities v_1 and v_2, produces the corresponding velocity in a Lorentz frame moving with velocity z down the positive x axis suggests that the conjecture (21) produces the retardation (or advancement) by parametrizing the portion of the second particle's trajectory between the present time t_0 and the retarded time t_{ret} with the velocity of the Lorentz frame in which the world point $x_2(t_s)$, $t_{\text{ret}} < t_s < t_0$, is simultaneous with the world point $x_1(t_0)$ as indicated in Fig. 3. When $z = 0$, $x_2 = x_2(t_0)$; when $z = 1$, $x_2 = x_2(t_{\text{ret}})$. The suspicion that a proof can be constructed along these lines is strengthened by the recognition that $\mathcal{R}(z)x_{12} = (1 - z^2)^{\frac{1}{2}}[x_{12}/(1 - v_2z)]$ is just the number assigned to $x_1'(t_0) - x_2'(t_s)$ in the S' frame if $x_2(t_s)$ is reckoned in straight-line approximation: $x_2(t_s) \approx x_2(t_0) + (t_s - t_0)v_2(t_0)$. This is easily seen by using $t_0 - t_s = [x_1(t_0) - x_2(t_s)]$ tan φ, which follows from the geometry of Fig. 3, and $z = \tan\varphi$ to obtain

$$t_s - t_0 = -z[x_1(t_0) - x_2(t_s)]. \qquad (22)$$

Following this line of reasoning, the first step is the production of a "correction formula," which can be used to write the quantity to be calculated;

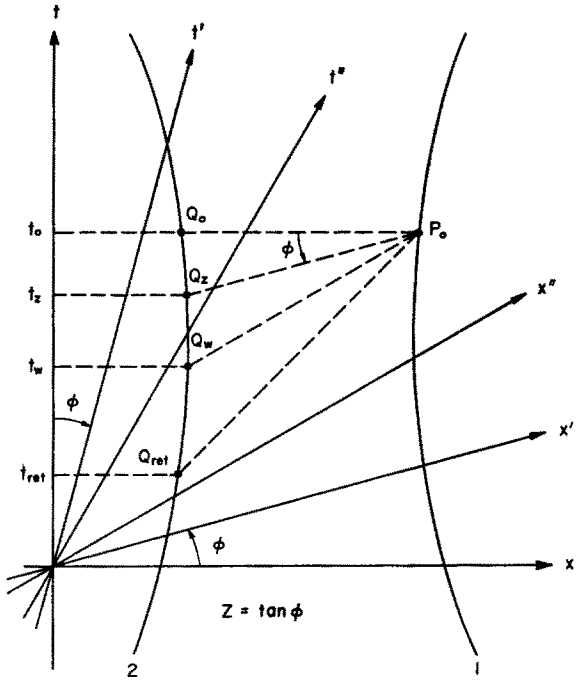

FIG. 3. Parametrization of an integral to the retarded time on the velocity of Lorentz frames.

$$E_1 = [1 + v_2(t_{\text{ret}})]/\{[x_1(t_0) - x_2(t_{\text{ret}})]^2[1 - v_2(t_{\text{ret}})]\},$$

as this quantity reckoned by a straight-line approximation to the second particle's motion plus a correction term in the form of an integral from t_0 to t_{ret} which involves the acceleration of the second particle linearly. The formula

$$\varphi[x_1(t_0) - x_2(t_w) + (t_w - t_0)v_2(t_w), v_1(t_0), v_2(t_w)]$$
$$= \varphi[x_1(t_0) - x_2(t_0), v_1(t_0), v_2(t_0)]$$
$$+ \int_{t_0}^{t_w} dt_s\, a_2(t_s)\frac{\partial}{\partial v_2(t_s)}$$
$$\times \varphi[x_1(t_0) - x_2(t_s) + (t_s - t_0)v_2(t_s), v_1(t_0), v_2(t_s)], \qquad (23)$$

which follows from the fact that the integrand in (23) is just

$$(d/dt_s)\varphi[x_1(t_0) - x_2(t_s) + (t_s - t_0)v_2(t_s), v_1(t_0), v_2(t_s)],$$

serves our purpose: if φ is a straight-line approximation, the $(t_w - t_0)v_2(t_w)$ on the left-hand side removes the effect of the straight-line approximation and leaves the quantity to be calculated.

We now change from the time variable t_s to the

velocity variable z and integrate $0 \leq z \leq w$. From (22),

$$dt_z = -\frac{x_1(t_0) - x_2(t_z)}{1 - zv_2(t_z)}\,dz. \tag{24}$$

$a_2(t_z)$ may be obtained by recognizing that the overall Lorentz invariance of the theory permits writing the acceleration of particle 2 at the world point Q_z as seen by an observer in the frame S', for whom P_0 and Q_z are simultaneous, as

$$a_2'(Q_z) = f_2[x_1'(P_0) - x_2'(Q_z), v_1'(P_0), v_2'(Q_z)]. \tag{25}$$

If one now employs the usual Lorentz transformation formulas for position, velocity, and acceleration to re-express (25) in the units appropriate to S,

$$a_2(t_z) = \frac{[1 - zv_2(t_z)]^3}{(1 - z^2)^{\frac{1}{2}}}\,\mathfrak{R}'(z)f_2. \tag{26}$$

The operator $\mathfrak{R}'(z)$ is defined by

$$\begin{aligned}\mathfrak{R}'(z)&\varphi(x_{12}, v_1, v_2)\\ &= \varphi\{[x_1(t_0) - x_2(t_z)](1 - z^2)^{\frac{1}{2}},\\ &\qquad [v_1(t_0) - z]/[1 - zv_1(t_0)],\\ &\qquad [v_2(t_z) - z]/[1 - zv_2(t_z)]\}\end{aligned} \tag{27}$$

for φ any function of x_{12}, v_1, v_2. The $\mathfrak{R}(z)$ defined in (20) is the straight-line approximation to $\mathfrak{R}'(z)$. We now use (24) and (26), and let φ be the straight-line approximation to some function ψ, i.e.,

$$\varphi(x_{12}, v_1, v_2) = \psi(x_{12}/(1 - wv_2), v_1, v_2).$$

The correction formula (23) then takes the form

$$\begin{aligned}&\psi[x_1(t_0) - x_2(t_w), v_1(t_0), v_2(t_w)]\\ &= \psi[x_{12}/(1 - wv_2), v_1, v_2]\\ &\quad - \int_0^w \frac{dz}{1 - z^2}\,[\mathfrak{R}'(z)x_{12}f_2]\,\frac{[1 - zv_2(t_z)]^2}{1 - z^2}\,\frac{\partial}{\partial v_2(t_z)}\\ &\quad \times \psi\left\{\frac{[x_1(t_0) - x_2(t_z)][1 - zv_2(t_z)]}{1 - wv_2(t_z)}, v_1(t_0), v_2(t_z)\right\}.\end{aligned} \tag{28}$$

We now apply (28) to the computation of the retarded field by setting $z = 1$ and $\psi = x_{12}^{-2}(1 + v_2)/(1 - v_2)$. We use

$$\begin{aligned}\mathfrak{R}'(z)&\psi[x_{12}/(1 - v_2), v_1, v_2] = \psi\{[x_1(t_0) - x_2(t_z)]\\ &\qquad \times [1 - zv_2(t_z)]/[1 - v_2(t_z)], v_1(t_0), v_2(t_z)\}\end{aligned}$$

for this choice of ψ (which can be verified directly) and the operator relation

$$\frac{[1 - zv_2(t_z)]^2}{1 - z^2}\,\frac{\partial}{\partial v_2(t_z)}\,\mathfrak{R}'(z) = \mathfrak{R}'(z)\,\frac{\partial}{\partial v_2} \tag{29}$$

to obtain

$$\begin{aligned}\left\{\frac{1}{x_{12}^2}\left(\frac{1 + v_2}{1 - v_2}\right)\right\}_{\text{ret}} &= \frac{1 - v_2^2}{x_{12}^2}\\ &\quad - \int_0^1 \frac{dz}{1 - z^2}\left\{\mathfrak{R}'(z)x_{12}f_2\,\frac{\partial}{\partial v_2}\left[\frac{1 - v_2^2}{x_{12}^2}\right]\right\}.\end{aligned} \tag{30}$$

This yields the conjecture (21a) if we can show that

$$\begin{aligned}\int_0^1 \frac{dw}{1 - w^2}\,\mathfrak{R}'(w)\varphi &= \int_0^1 \frac{dw}{1 - w^2}\,\mathfrak{R}(w)\varphi\\ &\quad - \int_0^1 \frac{dz}{1 - z^2}\,\mathfrak{R}'(z)x_{12}f_2\,\frac{\partial}{\partial v_2}\int_0^1 \frac{du}{1 - u^2}\,\mathfrak{R}(u)\varphi\end{aligned} \tag{31}$$

for φ an arbitrary function of x_{12}, v_1, v_2; iterating (30) with the aid of (31) leads to a geometric series in the operator

$$\int_0^1 dz\,(1 - z^2)^{-1}\mathfrak{R}(z)x_{12}f_2\,\partial/\partial v_2,$$

which is summed by (21a). The formula (31) can be established by letting

$$\psi = (1 - w^2)^{-1}\mathfrak{R}'(w)x_{12}f_2\partial\varphi/\partial v_2$$

in (28) and integrating from 0 to 1 on w. The resulting formula involves a double integral $\int_0^1 dw \int_0^w dz$; interchanging the order of integration produces a double integral $\int_0^1 dz \int_z^1 dw$. The lower limit z on the dw integral can be made into a zero by going into a frame moving with velocity z. We therefore make the change of variables from w to u via the Lorentz velocity addition formula $w = (z + u)/(1 + uz)$. Using (29) and the definitions of $\mathfrak{R}$ and $\mathfrak{R}'$, establishes (31) and completes the proof of (21a). Equation (21b) then follows from integrating to the advanced point in the same way.

We must point out that the results (21a) and (21b) have been established formally, and not rigorously. The difficulty lies with the convergence of the geometric series in the operator

$$\int_0^1 dz\,(1 - z^2)^{-1}\mathfrak{R}(z)x_{12}f_2(\partial/\partial v_2);$$

a sufficient condition for this convergence [and hence the rigorous, rather than formal, establishment of (21a) and (21b)] is the convergence of the expansions (4) in inverse powers of x_{12} for E_1 and E_2.

We can summarize the content of the integrodifferential equations (21a) and (21b) by remarking that they exploit the invariance group (Lorentz group) to integrate back (forward) to the retarded (advanced) time once and for all, for all members

of the family of possible particle orbits. We now turn to the three-dimensional problem.

V. THE TWO-BODY PROBLEM IN THREE DIMENSIONS

In order to establish a notation and provide a convenient reference list of formulas, we write down the Lorentz point transformations, which relate kinematical quantities at the same world point. A caret over a vector denotes a unit vector: thus $\hat{\mathbf{z}}$ is a unit vector in the direction of the vector $\mathbf{z}$. I is the unit dyadic. We first define the dyadics

$$\mathsf{J}(\mathbf{z}) \equiv [\mathsf{I} - \hat{\mathbf{z}}\hat{\mathbf{z}}] + (1 - z^2)^{-\frac{1}{2}}\hat{\mathbf{z}}\hat{\mathbf{z}}, \tag{32}$$

$$\mathsf{K}(\mathbf{z}, \mathbf{v}) \equiv (1 - z^2)^{-1}(1 - \mathbf{z}\cdot\mathbf{v})^2 \times [\mathsf{J}(\mathbf{z}) - (1 - z^2)^{-\frac{1}{2}}\mathbf{z}\mathbf{v}]. \tag{33}$$

Then, if the frame S' is moving with a velocity $\mathbf{z}$ with respect to S, quantities at the same world point are related by the formulas

$$\begin{aligned} \mathbf{x}' &= \mathsf{J}(\mathbf{z})\cdot(\mathbf{x} - \mathbf{z}t), \\ t' &= (t - \mathbf{z}\cdot\mathbf{x})(1 - z^2)^{-\frac{1}{2}}, \\ \mathbf{v}' &= (1 - z^2)^{\frac{1}{2}}\mathsf{J}(\mathbf{z})\cdot(\mathbf{v} - \mathbf{z})/(1 - \mathbf{z}\cdot\mathbf{v}), \\ \mathbf{a}' &= \mathbf{a}\cdot\mathsf{K}^{-1}(\mathbf{z}, \mathbf{v}). \end{aligned} \tag{34}$$

The first step is to generalize (22) to three dimensions. Assume that $\mathbf{x}_1(t_0)$ and $\mathbf{x}_2(t_z)$ are simultaneous in S'. Then, if $[\mathbf{x}_1(t_0), t_0]$ are the coordinates in S of the world point P_0, and $[\mathbf{x}_2(t_z), t_z]$ are the coordinates in S of Q_z,

$$\begin{aligned} t'(P_0) &= [t_0 - \mathbf{z}\cdot\mathbf{x}_1(t_0)](1 - z^2)^{-\frac{1}{2}}, \\ t'(Q_z) &= [t_z - \mathbf{z}\cdot\mathbf{x}_2(t_z)](1 - z^2)^{-\frac{1}{2}}. \end{aligned}$$

But $t'(P_0) = t'(Q_z)$, hence

$$t_z - t_0 = -\mathbf{z}\cdot[\mathbf{x}_1(t_0) - \mathbf{x}_2(t_z)]. \tag{35}$$

We are now in a position to generalize the conditions (14). The use of a vector notation includes the rotational subgroup of the Lorentz group automatically; we thus need concern ourselves only with an arbitrary infinitesimal pure Lorentz transformation. Using the notation of Fig. 2 and letting $\mathbf{z} = \delta\boldsymbol{\beta}$, the use of (34) and (35) yields the generalization of (13):

$$\begin{aligned} \mathbf{x}_1'(P_0) &= \mathbf{x}_1(P_0) - t_0\,\delta\boldsymbol{\beta}, \\ \mathbf{x}_2'(Q_0') &= \mathbf{x}_2(Q_0) - \delta\boldsymbol{\beta} \cdot[\mathbf{x}_1(P_0) - \mathbf{x}_2(Q_0)]\mathbf{v}_2(Q_0) - t_0\,\delta\boldsymbol{\beta}, \\ \mathbf{v}_1'(P_0) &= \mathbf{v}_1(P_0) - \delta\boldsymbol{\beta}\cdot[\mathsf{I} - \mathbf{v}_1(P_0)\mathbf{v}_1(P_0)], \\ \mathbf{v}_2'(Q_0') &= \mathbf{v}_2(Q_0) - \delta\boldsymbol{\beta}\cdot[\mathbf{x}_1(P_0) - \mathbf{x}_2(Q_0)] \times \mathbf{a}_2(Q_0) - \delta\boldsymbol{\beta}\cdot[\mathsf{I} - \mathbf{v}_2(Q_0)\mathbf{v}_2(Q_0)], \\ \mathbf{a}_1'(P_0) &= \mathbf{a}_1(P_0) + \delta\boldsymbol{\beta} \cdot[2\mathbf{v}_1(P_0)\mathbf{a}_1(P_0) + \mathbf{a}_1(P_0)\mathbf{v}_1(P_0)]. \end{aligned} \tag{36}$$

We assume instantaneous action-at-a-distance equations of motion of the form

$$\begin{aligned} \mathbf{a}_1 &= \mathbf{f}_1(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2), \\ \mathbf{a}_2 &= \mathbf{f}_2(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2). \end{aligned} \tag{37}$$

We now write the Eqs. (37) in the frame S', insert (36), expand to first order in $\delta\boldsymbol{\beta}$, and demand agreement with the equations in S to arrive at the three-dimensional version of (14):

$$2\mathbf{v}_1\mathbf{f}_1 + \mathbf{f}_1\mathbf{v}_1 + \mathbf{L}\mathbf{f}_1 = -\mathbf{x}_{12}\mathbf{f}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1, \tag{38a}$$

$$2\mathbf{v}_2\mathbf{f}_2 + \mathbf{f}_2\mathbf{v}_2 + \tilde{\mathbf{L}}\mathbf{f}_2 = \mathbf{x}_{12}\mathbf{f}_1\cdot\frac{\partial}{\partial\mathbf{v}_1}\mathbf{f}_2, \tag{38b}$$

where the operators $\mathbf{L}$ and $\tilde{\mathbf{L}}$ are defined by

$$\mathbf{L} = -\mathbf{x}_{12}\mathbf{v}_2\cdot\partial/\partial\mathbf{x}_{12} + [\mathsf{I} - \mathbf{v}_1\mathbf{v}_1]\cdot\partial/\partial\mathbf{v}_1 + [\mathsf{I} - \mathbf{v}_2\mathbf{v}_2]\cdot\partial/\partial\mathbf{v}_2, \tag{39a}$$

$$\tilde{\mathbf{L}} = -\mathbf{x}_{12}\mathbf{v}_1\cdot\partial/\partial\mathbf{x}_{12} + [\mathsf{I} - \mathbf{v}_1\mathbf{v}_1]\cdot\partial/\partial\mathbf{v}_1 + [\mathsf{I} - \mathbf{v}_2\mathbf{v}_2]\cdot\partial/\partial\mathbf{v}_2. \tag{39b}$$

An attempt to generalize the method of Sec. IV (parametrizing an integral to the retarded point on the velocities of Lorentz-frames) is faced with the difficulty that the velocity of the limiting frame in which $\mathbf{x}_1(t_0)$ and $\mathbf{x}_2(t_{\text{ret}})$ are simultaneous, which is

$$\mathbf{z}_{\text{ret}} = [\mathbf{x}_1(t_0) - \mathbf{x}_2(t_{\text{ret}})]/|\mathbf{x}_1(t_0) - \mathbf{x}_2(t_{\text{ret}})|, \tag{40}$$

is not known *a priori* as it was in the one-dimensional case. We thus find it necessary to approximate successively to the upper limit as well as to the path of particle 2 in generalizing (21). Furthermore, we can no longer avail ourselves of the simplification achieved in one dimension by working with the field E_1 which transformed as a scalar under Lorentz transformations. Thus we work directly with the acceleration $\mathbf{f}_1$. A third problem absent in one dimension is the presence of radiation terms containing the acceleration of the second particle at the retarded point. We disregard the radiation terms and attack the first two problems first.

We begin with the correction formula

$$\varphi[\mathbf{x}_1(t_0) - \mathbf{x}_2(t_w) + (t_w - t_0)\mathbf{v}_2(t_w), \mathbf{v}_1(t_0), \mathbf{v}_2(t_w)]$$
$$= \varphi[\mathbf{x}_1(t_0) - \mathbf{x}_2(t_0), \mathbf{v}_1(t_0), \mathbf{v}_2(t_0)]$$
$$+ \int_{t_0}^{t_w} dt_z\, \mathbf{a}_2(t_z)\cdot\frac{\partial}{\partial \mathbf{v}_2(t_z)}$$
$$\times \varphi[\mathbf{x}_1(t_0) - \mathbf{x}_2(t_z) + (t_z - t_0)\mathbf{v}_2(t_z), \mathbf{v}_1(t_0), \mathbf{v}_2(t_z)] \tag{41}$$

and rewrite the integral as a line integral on $\mathbf{z}$. This line integral will be path independent as a consequence of the connection (35) between t_z and $\mathbf{z}$: only the component of $\mathbf{z}$ along $\mathbf{x}_1(t_0) - \mathbf{x}_2(t_z)$ matters. We adopt a short hand notation:

$$\mathbf{x}_z \equiv \mathbf{x}_1(t_0) - \mathbf{x}_2(t_z), \quad \mathbf{v}_1 \equiv \mathbf{v}_1(t_0), \quad \mathbf{v}_z \equiv \mathbf{v}_2(t_z),$$
$$\mathbf{x}_{ret} \equiv \mathbf{x}_1(t_0) - \mathbf{x}_2(t_{ret}), \quad \mathbf{v}_{ret} = \mathbf{v}_2(t_{ret}). \tag{42}$$

From (35), it follows that

$$dt_z = -(1 - \mathbf{z}\cdot\mathbf{v}_z)^{-1}\mathbf{x}_z\cdot d\mathbf{z}. \tag{43}$$

We define an operator $\mathcal{R}'(\mathbf{z})$, which is the generalization of (27):

$$\mathcal{R}'(\mathbf{z})\varphi(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$$
$$= \varphi[\mathbf{J}^{-1}(\mathbf{z})\cdot\mathbf{x}_z, (1 - z^2)^{\frac{1}{2}}\mathbf{J}(\mathbf{z})\cdot(\mathbf{v}_1 - \mathbf{z})/(1 - \mathbf{z}\cdot\mathbf{v}_1),$$
$$(1 - z^2)^{\frac{1}{2}}\mathbf{J}(\mathbf{z})\cdot(\mathbf{v}_z - \mathbf{z})/(1 - \mathbf{z}\cdot\mathbf{v}_z)] \tag{44}$$

for any function φ (which may have *any* transformation properties—$\mathcal{R}'$ operates only on the arguments). The straight-line approximation $\mathcal{R}$ to $\mathcal{R}'$ is defined by

$$\mathcal{R}(\mathbf{z})\varphi(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$$
$$= \varphi\{\mathbf{J}^{-1}(\mathbf{z})\cdot[\mathbf{x}_{12} + \mathbf{v}_2(\mathbf{z}\cdot\mathbf{x}_{12})/(1 - \mathbf{z}\cdot\mathbf{v}_2)],$$
$$(1 - z^2)^{\frac{1}{2}}\mathbf{J}(\mathbf{z})\cdot(\mathbf{v}_1 - \mathbf{z})/(1 - \mathbf{z}\cdot\mathbf{v}_1),$$
$$(1 - z^2)^{\frac{1}{2}}\mathbf{J}(\mathbf{z})\cdot(\mathbf{v}_2 - \mathbf{z})/(1 - \mathbf{z}\cdot\mathbf{v}_2)\}. \tag{45}$$

The assumed over-all Lorentz invariance of the theory permits writing $\mathbf{a}_2(t_z)$ as

$$\mathbf{a}_2(t_z) = [\mathcal{R}'(\mathbf{z})\mathbf{f}_2]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_2). \tag{46}$$

We now need the straight-line approximation to $\mathbf{x}_{ret}$. This can be obtained by making the straight-line approximation $\mathbf{x}_z \approx \mathbf{x}_{12} + \mathbf{v}_2(t_0 - t_z)$ in (35) and (40); the result of solving for $\mathbf{x}_{ret}$ is

$$\mathbf{x}_{ret} \approx \mathbf{x}_{12} + \mathbf{v}_2\{(\mathbf{x}_{12}\cdot\mathbf{v}_2)$$
$$+ [x_{12}^2 - (\mathbf{x}_{12}\times\mathbf{v}_2)^2]^{\frac{1}{2}}\}/(1 - v_2^2). \tag{47}$$

The correction formula (41) motivates the definition of an interpolating vector $\mathbf{S}(\mathbf{z})$ by making the replacements

$$\mathbf{x}_{12} \to \mathbf{x}_z + (t_z - t_0)\mathbf{v}_z = [\mathbf{1} - \mathbf{v}_z\mathbf{z}]\cdot\mathbf{x}_z,$$
$$v_2 \to \mathbf{v}_z \tag{48}$$

in (47); thus we define

$$\mathbf{S}(\mathbf{z}) \equiv \mathbf{x}_z + \mathbf{v}_z(1 - v_z^2)^{-1}\{(\mathbf{v}_z - \mathbf{z})\cdot\mathbf{x}_z$$
$$+ [(\mathbf{x}_z - \mathbf{v}_z\mathbf{z}\cdot\mathbf{x}_z)^2 - (\mathbf{x}_z\times\mathbf{v}_z)^2]^{\frac{1}{2}}\}. \tag{49}$$

$\mathbf{S}(\mathbf{z})$ has the property that $\mathbf{S}(\mathbf{z}_{ret}) = \mathbf{x}_{ret}$. $\mathbf{S}(0)$ is the straight-line approximation to $\mathbf{x}_{ret}$.

We now assume that the primitive expression for the contribution to the acceleration $\mathbf{a}_1$ of particle 1 from the retarded interaction has the form $\mathbf{F}_1(\mathbf{x}_{ret}, \mathbf{v}_1, \mathbf{v}_{ret})$. The straight-line approximation to $\mathbf{F}_1$, which we call $\mathbf{f}_1^{(0)}$, is

$$\mathbf{f}_1^{(0)}(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2) \equiv \mathbf{F}_1(\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2). \tag{50}$$

We assume that $\mathbf{F}_1$ is such that $\mathbf{f}_1^{(0)}$ satisfies (38a) with $\mathbf{f}_2$ set equal to zero [Eq. (B11) of Appendix B]. The result (B14) of Appendix B then implies that

$$[\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) = \mathbf{F}_1(\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2).$$

But $\mathbf{S}(\mathbf{z})$ was obtained from $\mathbf{S}(0)$ by the replacements (48); the same replacements in $\mathcal{R}(z)$ produce $\mathcal{R}'(z)$, since

$$[\mathbf{1} + (1 - \mathbf{z}\cdot\mathbf{v}_z)^{-1}\mathbf{v}_z\mathbf{z}]\cdot[\mathbf{1} - \mathbf{v}_z\mathbf{z}] = \mathbf{1}.$$

Hence

$$[\mathcal{R}'(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) = \mathbf{F}_1[\mathbf{S}(\mathbf{z}), \mathbf{v}_1, \mathbf{v}_z]. \tag{51}$$

We now use (43), (46), (50), (51), and the operator relation

$$\partial/\partial\mathbf{v}_z\mathcal{R}'(\mathbf{z})$$
$$= (1 - z^2)^{-\frac{1}{2}}(1 - \mathbf{z}\cdot\mathbf{v}_z)\mathbf{K}^{-1}(\mathbf{z}, \mathbf{v}_z)\cdot\mathcal{R}'(\mathbf{z})\,\partial/\partial\mathbf{v}_2 \tag{52}$$

in (41) with $\varphi = \mathbf{F}_1[\mathbf{S}(\mathbf{z}), \mathbf{v}_1, \mathbf{v}_z]$ and $t_w = t_{ret}$ to produce the result

$$\mathbf{F}_1(\mathbf{x}_{ret}, \mathbf{v}_1, \mathbf{v}_{ret}) = \mathbf{f}_1^{(0)}(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$$
$$- \int_0^{\mathbf{z}_{ret}} d\mathbf{z}\cdot(1 - z^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{z})$$
$$\cdot\left[\mathcal{R}'(\mathbf{z})\mathbf{f}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(0)}\right]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_z), \tag{53}$$

which is a three-dimensional analog of (30). We must next find a formula analagous to (31) which permits the iteration of (53) into an integrodifferential equation from which the time has disappeared.

We therefore consider an integral of the general form

$$\int_0^{z_{ret}} d\mathbf{w}\cdot(1-w^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{w})\cdot[\mathcal{R}'(\mathbf{w})\mathbf{x}_{12}\mathbf{G}]\cdot\mathbf{K}(\mathbf{w},\mathbf{v}_1), \quad (54)$$

where $\mathbf{G}$ is a function of $\mathbf{x}_{12}$, $\mathbf{v}_1$, $\mathbf{v}_2$ which transforms as a vector under the rotation group. In order to be able to approximate to the upper limit $\mathbf{z}_{ret}$, as well as to $\mathbf{x}_w$ and $\mathbf{v}_w$, we specify a particular path of integration—the path

$$\mathbf{w} = w\hat{\mathbf{x}}_w. \quad (55)$$

With this specification of path,

$$d\mathbf{w}\cdot\mathbf{J}(\mathbf{w})\cdot[\mathcal{R}'(\mathbf{w})\mathbf{x}_{12}] = |\mathbf{x}_w|\,dw \quad (56)$$

since $\mathbf{x}_w\cdot d\hat{\mathbf{x}}_w = 0$. We now use the scalar w as our dummy variable of integration.

We need next the straight-line approximation to $\mathbf{x}_w$ with the path (55). We use the straight-line approximation $\mathbf{x}_w \approx \mathbf{x}_{12} + \mathbf{v}_2(t_0 - t_w)$ in (35) and in (55) to obtain

$$\mathbf{x}_w \approx \mathbf{x}_{12} + \mathbf{v}_2\{w^2(\mathbf{v}_2\cdot\mathbf{x}_{12}) + w[x_{12}^2 - w^2(\mathbf{x}_{12}\times\mathbf{v}_2)^2]^{\frac{1}{2}}\}/(1-w^2v_2^2) \quad (57)$$

in straight-line approximation. The correction formula (41) again motivates the definition of an interpolating vector. We make the replacements (48) in (57) and define the interpolating vector

$$\mathbf{M}(w,z) \equiv \mathbf{x}_z + \mathbf{v}_z[1-w^2v_z^2]^{-1}\{(w^2\mathbf{v}_z - \mathbf{z})\cdot\mathbf{x}_z + w[(\mathbf{x}_z - \mathbf{v}_z\mathbf{z}\cdot\mathbf{x}_z)^2 - w^2(\mathbf{x}_z\times\mathbf{v}_z)^2]^{\frac{1}{2}}\}. \quad (58)$$

$\mathbf{M}$ has the property that $\mathbf{M}(w, w\hat{\mathbf{x}}_w) = \mathbf{x}_w$. $\mathbf{M}(w,0)$ is the straight-line approximation to $\mathbf{x}_w$ with the path specified by (55). Also, $\mathbf{M}(1,z) = \mathbf{S}(z)$. We define an interpolating operator $\mathcal{R}(w,z)$ by making the replacements (48) and the replacement $\mathbf{w} \to w\hat{\mathbf{M}}(w,z)$ in the definition of $\mathcal{R}(\mathbf{w})$:

$$\begin{aligned}\mathcal{R}(w,z)&\varphi(\mathbf{x}_{12},\mathbf{v}_1,\mathbf{v}_2)\\ &\equiv \varphi\{(1-w^2)^{\frac{1}{2}}\mathbf{M}(w,z),\,(1-w^2)^{\frac{1}{2}}\mathbf{J}[w\hat{\mathbf{M}}(w,z)]\\ &\quad\cdot[\mathbf{v}_1 - w\hat{\mathbf{M}}(w,z)]/[1-w\mathbf{v}_1\cdot\hat{\mathbf{M}}(w,z)],\\ &\quad(1-w^2)^{\frac{1}{2}}\mathbf{J}[w\hat{\mathbf{M}}(w,z)]\\ &\quad\cdot[\mathbf{v}_z - w\hat{\mathbf{M}}(w,z)]/[1-w\mathbf{v}_z\cdot\hat{\mathbf{M}}(w,z)]\}. \end{aligned} \quad (59)$$

It follows from definition (59) that

$$\mathcal{R}(w,0) = \mathcal{R}[w\hat{\mathbf{M}}(w,0)] \text{ and } \mathcal{R}(w,w\hat{\mathbf{x}}_w) = \mathcal{R}'(w\hat{\mathbf{x}}_w).$$

We now apply the correction formula (41) to (54) and use (43), (46), (58), and (59) to obtain

$$\begin{aligned}\int_0^1 dw\,(1-w^2)^{-\frac{1}{2}}\,&|\mathbf{x}_w|\,[\mathcal{R}'(w\hat{\mathbf{x}}_w)\mathbf{G}]\cdot\mathbf{K}(w\hat{\mathbf{x}}_w,\mathbf{v}_1)\\ &= \int_0^1 dw\,(1-w^2)^{-\frac{1}{2}}\,|\mathbf{M}(w,0)|\,\{\mathcal{R}[w,\hat{\mathbf{M}}(w,0]\mathbf{G}\}\cdot\mathbf{K}[w\hat{\mathbf{M}}(w,0),\mathbf{v}_1]\\ &\quad-\int_0^1 dw\,(1-w^2)^{-\frac{1}{2}}\int_0^{w\hat{\mathbf{x}}_w}\frac{d\mathbf{z}\cdot\mathbf{x}_z}{1-\mathbf{z}\cdot\mathbf{v}_z}[\mathcal{R}'(\mathbf{z})\mathbf{f}_2]\cdot\mathbf{K}(\mathbf{z},\mathbf{v}_z)\cdot\frac{\partial}{\partial\mathbf{v}_z}\\ &\quad\times\{|\mathbf{M}(w,z)|\,[\mathcal{R}(w,z)\mathbf{G}]\cdot\mathbf{K}[w\hat{\mathbf{M}}(w,z),\mathbf{v}_1]\}. \end{aligned} \quad (60)$$

We next specify a path for the z integration in the double integral in (60). We make again the choice (55): $\mathbf{z} = z\hat{\mathbf{x}}_z$. We then have an integral over the scalar z instead of a line integral; the double integral has the form $\int_0^1 dw\int_0^w dz$. If we interchange orders of integration, the double integral in (60) becomes

$$-\int_0^1\frac{|\mathbf{x}_z|\,dz}{1-z\hat{\mathbf{x}}_z\cdot\mathbf{v}_z}[\mathcal{R}'(z\hat{\mathbf{x}}_z)\mathbf{f}_2]\cdot\mathbf{K}(z\hat{\mathbf{x}}_z,\mathbf{v}_z)\cdot\frac{\partial}{\partial\mathbf{v}_z}\left\{\int_z^1 dw\,(1-w^2)^{-\frac{1}{2}}\,|\mathbf{M}(w,z\hat{\mathbf{x}}_z)|\,[\mathcal{R}(w,z\hat{\mathbf{x}}_z)\mathbf{G}]\cdot\mathbf{K}[w\hat{\mathbf{M}}(w,z\hat{\mathbf{x}}_z),\mathbf{v}_1]\right\}. \quad (61)$$

The substitution which disentangled things at this point in the one-dimensional case suggests what to do here: we go to a frame moving with velocity $z\hat{\mathbf{x}}_z$ for the evaluation of the integral on w in (61). Thus, we define a vector $\mathbf{u}$ via the Lorentz velocity addition formula

$$w\hat{\mathbf{M}}(w,z\hat{\mathbf{x}}_z) = \frac{(1-z^2)^{\frac{1}{2}}}{1+z\hat{\mathbf{x}}_z\cdot\mathbf{u}}\mathbf{J}(z\hat{\mathbf{x}}_z)\cdot(\mathbf{u}+z\hat{\mathbf{x}}_z). \quad (62)$$

The integral over w is now to be thought of as a line integral on $\mathbf{u}$ with the path specified by the relationship (62) between $\mathbf{u}$ and w. We adopt the convention that $\mathcal{R}'\mathbf{u} = \mathbf{u}$ [we could define a $\mathbf{u}'$ in terms of $\mathbf{x}_{12}$ and $\mathbf{v}_2$ which would be mapped into the $\mathbf{u}$ of Eq. (62) by $\mathcal{R}'(z\hat{\mathbf{x}}_z)$ and then allow $\mathcal{R}'$ to operate on this $\mathbf{u}'$, but to do so would introduce unnecessary complication]. With $\mathbf{u}$ defined by (62), it follows

from the definitions (44), (45), (59), and the formulas (B1), (B2), and (B3) that

$$\mathcal{R}(w, z\hat{\mathbf{x}}_z)\mathbf{v} = \mathbf{R}^{-1}[\mathbf{u}, z\hat{\mathbf{x}}_z, w\hat{\mathbf{M}}(w, z\hat{\mathbf{x}}_z)]\cdot\mathcal{R}'(z\hat{\mathbf{x}}_z)\mathcal{R}(\mathbf{u})\mathbf{v} \tag{63}$$

for $\mathbf{v}$ either $\mathbf{v}_1$ or $\mathbf{v}_2$. By direct computation it can be verified that

$$\mathcal{R}(w, z\hat{\mathbf{x}}_z)\mathbf{x}_{12} = \mathbf{R}^{-1}[\mathbf{u}, z\hat{\mathbf{x}}_z, w\hat{\mathbf{M}}(w, z\hat{\mathbf{x}}_z)]\cdot\mathcal{R}'(z\hat{\mathbf{x}}_z)\mathcal{R}(\mathbf{u})\mathbf{x}_{12} \tag{64}$$

and

$$dw\,(1 - w^2)^{-\frac{1}{2}}\,|\mathbf{M}(w, z\hat{\mathbf{x}}_z)| = d\mathbf{u}\cdot(1 - u^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{u})\cdot\mathcal{R}'(z\hat{\mathbf{x}}_z)\mathcal{R}(\mathbf{u})\mathbf{x}_{12}. \tag{65}$$

We use (63), (64), (B4), and the fact that $\mathbf{G}$ transforms as a vector under rotations to obtain

$$[\mathcal{R}(w, z\hat{\mathbf{x}}_z)\mathbf{G}]\cdot\mathbf{K}[w\hat{\mathbf{M}}(w, z\hat{\mathbf{x}}_z), \mathbf{v}_1] = \{\mathcal{R}'(z\hat{\mathbf{x}}_z)[\mathcal{R}(\mathbf{u})\mathbf{G}]\cdot\mathbf{K}(\mathbf{u}, \mathbf{v}_1)\}\cdot\mathbf{K}(z\hat{\mathbf{x}}_z, \mathbf{v}_1). \tag{66}$$

Furthermore, $w = z$ implies $\mathbf{u} = 0$, and $w = 1$ implies $\mathbf{u} = \mathcal{R}'(z\hat{\mathbf{x}}_z)\hat{\mathbf{S}}(0)$. We now insert (65) and (66) into (61) and then employ (52) to rewrite (60) as

$$\int_0^1 dw\,(1 - w^2)^{-\frac{1}{2}}\,|\mathbf{x}_w|\,[\mathcal{R}'(w\hat{\mathbf{x}}_w)\mathbf{G}]\cdot\mathbf{K}(w\hat{\mathbf{x}}_w, \mathbf{v}_1)$$
$$= \int_0^1 dw\,(1 - w^2)^{-\frac{1}{2}}\,|\mathbf{M}(w, 0)|\,\{\mathcal{R}[w\hat{\mathbf{M}}(w, 0)]\mathbf{G}\}\cdot\mathbf{K}[w\hat{\mathbf{M}}(w, 0), \mathbf{v}_1]$$
$$- \int_0^1 dz\,(1 - z^2)^{-\frac{1}{2}}\,|\mathbf{x}_z|\left\{\mathcal{R}'(z\hat{\mathbf{x}}_z)\mathbf{f}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}\int_0^{\hat{\mathbf{S}}(0)} d\mathbf{u}\cdot(1 - u^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{u})\cdot[\mathcal{R}(\mathbf{u})\mathbf{x}_{12}\mathbf{G}]\cdot\mathbf{K}(\mathbf{u}, \mathbf{v}_1)\right\}\cdot\mathbf{K}(z\hat{\mathbf{x}}_z, \mathbf{v}_1). \tag{67}$$

Equation (67) provides the needed generalization of (31). At this point, the path in the line integral over $\mathbf{u}$ is still specified by (62). However, alternate application of the results of Appendix C and Appendix D show that *this integral is actually independent of path at any stage of the iteration* of (53) by means of (67) if a corresponding iteration is carried on simultaneously to compute $\mathbf{f}_2$. Hence, the iterates of (53) are summed by the integrodifferential equation

$$\mathbf{f}_{1,r} = \mathbf{f}_1^{(0)} - \int_0^{\hat{\mathbf{S}}(0)} dz\cdot(1 - z^2)^{-\frac{1}{2}}\mathbf{J}(z)\cdot\left[\mathcal{R}(z)\mathbf{x}_{12}\mathbf{f}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_{1,r}\right]\cdot\mathbf{K}(z, \mathbf{v}_1). \tag{68}$$

$\mathbf{f}_{1,r}$ is the retarded contribution to the acceleration of particle 1; (68) provides the three-dimensional generalization of (21a). The advanced contribution $\mathbf{f}_{1,a}$ can be obtained by merely changing the upper limit on (68); thus

$$\mathbf{f}_{1,a} = \mathbf{f}_1^{(0)} - \int_0^{-\hat{\mathbf{S}}(0)} dz\cdot(1 - z^2)^{-\frac{1}{2}}\mathbf{J}(z)\cdot\left[\mathcal{R}(z)\mathbf{x}_{12}\mathbf{f}_2\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_{1,a}\right]\cdot\mathbf{K}(z, \mathbf{v}_1). \tag{69}$$

Equations for the retarded and advanced contributions $\mathbf{f}_{2,r}$ and $\mathbf{f}_{2,a}$ to $\mathbf{f}_2$ can be obtained from (68) and (69) by interchanging the subscripts 1 and 2 everywhere. The alternate use of the result of Appendices C and D, already used to establish path independence, also shows that the $\mathbf{f}_i$ obtained as any linear combination of the $\mathbf{f}_{1,r}$ and $\mathbf{f}_{1,a}$ of (68) and (69) satisfies the differential condition (38a); hence the principle of relativity is satisfied and the assumption (46) is justified.

The derivation of Eq. (68) required three assumptions:

(a) Convergence of the sum of the iterates of (53) to (68) [a sufficient condition for the convergence of this iteration is convergence of the expansions in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$ for $\mathbf{f}_1$ and $\mathbf{f}_2$],
(b) the primitive expression for the contribution to the acceleration a_1 from the retarded interaction has the form $\mathbf{F}_1(\mathbf{x}_{ret}, \mathbf{v}_1, \mathbf{v}_{ret})$, and
(c) $\mathbf{F}_1$ satisfies the differential condition

$$2\mathbf{v}_1\mathbf{F}_1[\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2] + \mathbf{F}_1[\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2]\mathbf{v}_1 + \mathbf{L}\mathbf{F}_1[\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2] = 0, \tag{70}$$

which was used to obtain (51). The condition (b) excludes the radiation terms in the $\mathbf{F}_1$ derived from the Lienard–Wiechert fields. We must therefore generalize assumption (b) to include the acceleration of particle 2 at the retarded point, and show that the resulting generalization of condition (c) is satisfied. We now take the primitive expression for the contribution to $\mathbf{a}_1$ from the retarded interaction to be of the form $\mathbf{A}_r(\mathbf{x}_{ret}, \mathbf{v}_1, \mathbf{v}_{ret}, \mathbf{a}_{ret})$, where $\mathbf{a}_{ret} \equiv \mathbf{a}_2(t_{ret})$. From (46) $\mathbf{a}_{ret} = [\mathcal{R}'(\mathbf{z}_{ret})\mathbf{f}_2]\cdot\mathbf{K}(\mathbf{z}_{ret}, \mathbf{v}_{ret})$. We obtain the needed generalization of (53)[*] by using (41) with $t_w = t_{ret}$, and

$$\varphi = \mathbf{A}_r\{\mathbf{S}(z), \mathbf{v}_1, \mathbf{v}_z, [\mathcal{R}(1, z)\mathbf{f}_2]\cdot\mathbf{K}[\hat{\mathbf{S}}(z), \mathbf{v}_z]\}.$$

We define

$$\mathbf{\Lambda}^{(0)}_{1,r}(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$$
$$\equiv \mathbf{\Lambda}_r\{\mathbf{S}(0), \mathbf{v}_1, \mathbf{v}_2, [\mathcal{R}(1, 0)\mathbf{f}_2]\cdot\mathbf{K}[\hat{\mathbf{S}}(0), \mathbf{v}_2]\}. \quad (71)$$

Since $\mathbf{\Lambda}^{(0)}_{1,r}$ is not known *a priori*, as was $\mathbf{f}^{(0)}_1$, we must approximate successively to it as we proceed to compute successive approximations to $\mathbf{f}_2$. We assume that

$$2\mathbf{v}_1\mathbf{\Lambda}^{(0)}_{1,r} + \mathbf{\Lambda}^{(0)}_{1,r}\mathbf{v}_1 + \mathbf{L}\mathbf{\Lambda}^{(0)}_{1,r} = 0 \quad (72)$$

holds with $\mathbf{f}_2$ replaced by any member of the sequence of approximations to $\mathbf{f}_2$ generated by expanding in inverse powers of $|\mathbf{x}_{12}|$. The derivation of the integrodifferential equation now goes through just as it did in the absence of the radiation terms; the only change is that $\mathbf{f}^{(0)}_1$ in (68) is replaced by $\mathbf{\Lambda}^{(0)}_{1,r}$. We must now show that (72), which replaces (70), is satisfied.

To establish (72), we begin with the primitive expression for the retarded contribution to $\mathbf{a}_1$ expressed in 4-vector form. We use the metric tensor $g^{00} = -1$, $g^{11} = g^{22} = g^{33} = +1$. The 4-velocities of particles 1 and 2 are $u_{i\mu} = (u_{i0}, \mathbf{u}_i)$ with $u_{i0} = (1 - v_i^2)^{-\frac{1}{2}}$, $\mathbf{u}_i = (1 - v_i^2)^{-\frac{1}{2}}\mathbf{v}_i$ for $i = 1, 2$. The 4-displacement to the retarded point is $R_\mu = (|\mathbf{x}_{\rm ret}|, \mathbf{x}_{\rm ret})$; the proper times of particles 1 and 2 are τ_1 and τ_2. The contribution of the retarded Lienard–Wiechert fields to the 4-acceleration $du_{1\mu}/d\tau_1$ is then

$$u_1^\nu F^{\rm ret}_{1\mu\nu} = (u_2^\alpha R_\alpha)^{-3}\{(1 + R_\beta\, du_2^\beta/d\tau_2)$$
$$\times [(u_1^\nu u_{2\nu})R_\mu - (u_1^\nu R_\nu)u_{2\mu}] - (u_2^\beta R_\beta)$$
$$\times [(u_1^\nu\, du_{2\nu}/d\tau_2)R_\mu - (u_1^\nu R_\nu)\, du_{2\mu}/d\tau_2]\}, \quad (73)$$

where all quantities referring to particle 2 are evaluated at the retarded time. The contribution (73) to $du_{1\mu}/d\tau_1$ has the form $\lambda_\mu(R_\alpha, u_{1\alpha}, u_{2\alpha}, du_{2\alpha}/d\tau_2)$ where λ_μ is a 4-vector function of the indicated arguments which satisfies the condition $u_1^\mu\lambda_\mu \equiv 0$ (required by $u_1^\mu\, du_{1\mu}/d\tau_1 = 0$). This condition implies that, if $\lambda_\mu = (\lambda_0, \boldsymbol{\lambda})$, then $\lambda_0 = \mathbf{v}_1\cdot\boldsymbol{\lambda}$. The 4-vectors on which λ_μ depends each have only three independent components (because of the restrictions $R^\alpha R_\alpha = 0$, $u_1^\alpha u_{1\alpha} = u_2^\alpha u_{2\alpha} = -1$, $u_2^\alpha\, du_{2\alpha}/d\tau_2 = 0$); hence a retarded contribution to the 4-acceleration $du_{1\mu}/d\tau_1$ of the form $\lambda_\mu(R_\alpha, u_{1\alpha}, u_{2\alpha}, du_{2\alpha}/d\tau_2)$ leads immediately to a retarded contribution to the 3-vector acceleration $\mathbf{a}_1$ of the form $\mathbf{\Lambda}_r(\mathbf{x}_{\rm ret}, \mathbf{v}_1, \mathbf{v}_{\rm ret}, \mathbf{a}_{\rm ret})$ via the relation

$$\mathbf{\Lambda}_r = (1 - v_1^2)[\mathbf{I} - \mathbf{v}_1\mathbf{v}_1]\cdot\boldsymbol{\lambda}. \quad (74)$$

If we write the dependence of $\boldsymbol{\lambda}$ on its arguments in the form

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}(R_0, \mathbf{R}; u_{10}, \mathbf{u}_1; u_{20}, \mathbf{u}_2; du_{20}/d\tau_2, d\mathbf{u}_2/d\tau_2),$$

then (71) implies that

$$\mathbf{\Lambda}^{(0)}_{1,r}(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2) = (1 - v_1^2)[\mathbf{I} - \mathbf{v}_1\mathbf{v}_1]$$
$$\cdot\boldsymbol{\lambda}[|\mathbf{S}(0)|, \mathbf{S}(0); u_{10}, \mathbf{u}_1; u_{20}, \mathbf{u}_2; \mu_0, \boldsymbol{\mu}], \quad (75)$$

where

$$\boldsymbol{\mu} = (1 - v_2^2)^{-1}[\mathbf{I} - \mathbf{v}_2\mathbf{v}_2]^{-1}\cdot[\mathcal{R}(\hat{\mathbf{S}}(0)\mathbf{f}_2]\cdot\mathbf{K}[\hat{\mathbf{S}}(0), \mathbf{v}_2]$$

and $\mu_0 = \mathbf{v}_2\cdot\boldsymbol{\mu}$. By direct computation,

$$\begin{aligned} \mathbf{L}\mathbf{S}(0) &= \mathbf{I}\,|\mathbf{S}(0)|, \\ \mathbf{L}\,|\mathbf{S}(0)| &= \mathbf{S}(0), \\ \mathbf{L}u_i &= \mathbf{I}u_{i0}, \\ \mathbf{L}u_{i0} &= \mathbf{u}_i, \end{aligned} \quad (76)$$

where $i = 1, 2$. By the use of (B6), (B8), (B16), the relation $\mathbf{L}\hat{\mathbf{S}}(0) = \mathbf{I} - \hat{\mathbf{S}}(0)\hat{\mathbf{S}}(0)$, and the definition of $\boldsymbol{\mu}$ and μ_0, we obtain

$$\mathbf{L}\boldsymbol{\mu} = \mathbf{I}\mu_0, \quad \mathbf{L}\mu_0 = \boldsymbol{\mu} \quad (77)$$

independent of the functional form of the $\mathbf{f}_2$ *on which* $\boldsymbol{\mu}$ *and* μ_0 *depend.* Thus, (77) holds whether we use the actual $\mathbf{f}_2$ or some member of the sequence of approximations to it. From (75) we find that

$$2\mathbf{v}_1\mathbf{\Lambda}^{(0)}_{1,r} + \mathbf{\Lambda}^{(0)}_{1,r}\mathbf{v}_1 + \mathbf{L}\mathbf{\Lambda}^{(0)}_{1,r}$$
$$= (1 - v_1^2)[(\boldsymbol{\lambda} - \mathbf{L}\lambda_0)\mathbf{v}_1 + \mathbf{L}\boldsymbol{\lambda} - \mathbf{I}\lambda_0]. \quad (78)$$

The fact that λ_μ transforms as a 4-vector function of 4-vector arguments under Lorentz transformation implies that

$$\delta w^{\mu\nu}\lambda_\nu = \delta w^{\alpha\nu}[R_\nu\,\partial/\partial R_\alpha + u_{1\nu}\,\partial/\partial u_{1\alpha}$$
$$+ u_{2\nu}\,\partial/\partial u_{2\alpha} + (du_{2\nu}/d\tau_2)\,\partial/\partial\,(du_{2\alpha}/d\tau_2)]\lambda_\mu. \quad (79)$$

For a pure Lorentz transformation $\delta w^{00} = \delta w^{ij} = 0$, $\delta w^{0i} = \delta w^{i0} = \delta\beta_i$, where i and j run from 1 to 3, and (79) becomes

$$\delta\boldsymbol{\beta}\cdot\boldsymbol{\lambda} = \delta\boldsymbol{\beta}\cdot\mathbf{L}'\lambda_0, \quad \delta\boldsymbol{\beta}\,\lambda_0 = \delta\boldsymbol{\beta}\cdot\mathbf{L}'\boldsymbol{\lambda}, \quad (80)$$

where

$$\mathbf{L}' \equiv \mathbf{S}(0)\,\partial/\partial\,|\mathbf{S}(0)| + |\mathbf{S}(0)|\,\partial/\partial\,\mathbf{S}(0)$$
$$+ \mathbf{u}_1\,\partial/\partial\,u_{10} + u_{10}\,\partial/\partial\mathbf{u}_1 + \mathbf{u}_2\,\partial/\partial u_{20} + u_{20}\,\partial/\partial\mathbf{u}_2$$
$$+ \boldsymbol{\mu}\,\partial/\partial\mu_0 + \mu_0\,\partial/\partial\boldsymbol{\mu}. \quad (81)$$

The results (76) and (77) show that $\mathbf{L} = \mathbf{L}'$ when the operand is $\boldsymbol{\lambda}$ or λ_0; hence (80) implies the vanishing of the right-hand side of (78) and the satisfaction of (72) *at every stage of the approximation to*

$\mathbf{f}_2$. The integrodifferential equation (69) for the advanced interaction contribution to $\mathbf{a}_1$ may be generalized to include radiation terms by making appropriate sign changes in the preceding argument.

As was noted in the Introduction, an instantaneous action-at-a-distance description requires only a decision to confine attention to particle world lines; it is thus, in principle, possible to include the effects of radiation reaction. The details of showing that the path independence and Lorentz covariance conditions continue to hold when a radiation contribution to the force is included are given in Appendix E. Thus, the electrodynamic two-body problem with half-advanced plus half-retarded interactions can be reduced to instantaneous action-at-a-distance form by solving (68) and (69) [with $\mathbf{f}_1^{(0)}$ replaced by $\mathbf{A}_{1,\mathrm{r}}^{(0)}$], where $\mathbf{f}_1 = \frac{1}{2}(\mathbf{f}_{1,\mathrm{a}} + \mathbf{f}_{1,\mathrm{r}})$ together with the corresponding equations for $\mathbf{f}_2$. The electrodynamic two-body problem with retarded interactions and a radiation reaction term can be reduced to instantaneous action-at-a-distance form by solving (68) [with $\mathbf{f}_1^{(0)}$ replaced by $\mathbf{A}_{1,\mathrm{r}}^{(0)}$], where $\mathbf{f}_1 = \mathbf{f}_{1,\mathrm{r}} + \mathbf{f}_1^{\mathrm{rr}}$ [$\mathbf{f}_1^{\mathrm{rr}}$ is given by ($E2$) of Appendix E] together with the corresponding equations for $\mathbf{f}_2$.

The results (68) and (69), as well as their generalization to include terms dependent on the acceleration at the retarded point (radiation terms) are not restricted to electrodynamics. Equations (68) and (69) can be used to re-express *any* Lorentz-invariant direct interaction along light cones in instantaneous action-at-a-distance form *if* the iteration process [expansion in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$ for electrodynamics] used to establish (68) and (69) converges. The results (68) and (69) may remain valid as formal re-expressions of the dynamics even if the iteration used to establish them does not converge, but this remains to be investigated. For the particular case of electrodynamics, the iteration may possibly be expected to converge for interparticle separations greater than something of the order of the classical charge radius. In any event, the integrodifferential equations, (68) and (69), should provide a powerful tool for the investigation of the character of the series in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$, which appears in the instantaneous action-at-a-distance reformulation of the electrodynamic two-body problem.

We now turn to a brief examination of the N-body problem.

VI. INSTANTANEOUS ACTION-AT-A-DISTANCE FOR N PARTICLES

The differential statement of the principle of relativity for a translationally invariant system of N particles is easily written down if one chooses to write the equation of motion of the ith particle in the form

$$\mathbf{a}_i = \mathbf{f}_i(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{i,i-1}, \mathbf{x}_{i,i+1}, \cdots, \mathbf{x}_{i,N}; \mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N). \tag{82}$$

The set of $N-1$ vectors $\{\mathbf{x}_{i1}, \cdots, \mathbf{x}_{i,i-1}, \mathbf{x}_{i,i+1}, \cdots, \mathbf{x}_{i,N}\}$ with $\mathbf{x}_{i,j} \equiv \mathbf{x}_i - \mathbf{x}_j$ is clearly sufficient for writing the most general translationally-invariant function of the N vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$. The argument which led to the conditions (30) can be immediately generalized to obtain the conditions for N particles, which are

$$2\mathbf{v}_i\mathbf{f}_i + \mathbf{f}_i\mathbf{v}_i - \sum_{\substack{j=1\\ j\neq i}}^{N} \mathbf{x}_{ij}\mathbf{v}_j\cdot\frac{\partial}{\partial\mathbf{x}_{ij}}\mathbf{f}_i + \sum_{i=1}^{N}[\mathbf{I} - \mathbf{v}_i\mathbf{v}_i]\cdot\frac{\partial}{\partial\mathbf{v}_i}\mathbf{f}_i + \sum_{\substack{j=1\\ j\neq i}}^{N}\mathbf{x}_{ij}\mathbf{f}_i\cdot\frac{\partial}{\partial\mathbf{v}_j}\mathbf{f}_i = 0. \tag{83}$$

If we choose to call the $\mathbf{f}_i$ of (82) forces, then it can be easily shown from (83) that $\mathbf{f}_i$ cannot be a sum of two-body forces in the form $\sum_{j\neq i}\mathbf{F}_{i,j}(\mathbf{x}_{ij}, \mathbf{v}_i, \mathbf{v}_j)$, where $\mathbf{F}_{i,j}$ is the interaction when only the ith and jth particles are present. This is essentially a consequence of the nonlinearity of (83), which invalidates superposition. Thus, *the principle of relativity demands the existence of many-body forces* in the N-particle instantaneous action-at-distance description. The same conclusion can be drawn by beginning to generalize (68) and (69); such considerations show that many-particle forces [which enter by virtue of the fact that $\mathbf{a}_2(t_z)$ in (41) now contains coordinates and velocities of all other particles] fall off as higher inverse powers of the distance than do the two-body forces.[12]

VII. SUMMARY AND DISCUSSION

The work of the preceding sections indicates that instantaneous action-at-a-distance is quite tenable in relativistic mechanics, and does not violate the principle of relativity. This conclusion has been drawn by deriving, and investigating the consequences of a set of differential conditions [Eq. (14) for two particles in one dimension; Eq. (83) for the general three-dimensional N-particle case] on the instantaneous equations of motion which guarantee world-line invariance and form invariance of

[12] The fact that expansion with order depression leads to many-body forces was recognized by H. Primakoff and T. Holstein, Phys. Rev. 55, 1218 (1939). I am indebted to Peter Havas for calling my attention to their work.

the equations of motion under Lorentz transformation. In the case of the perturbative reduction of the equations of motion of the two-body problem of electrodynamics to instantaneous interaction form, it is found, with the aid of dimensional analysis, that these differential conditions imply that not only all powers of c^{-1}, but also all powers of the coupling constant e^2, must appear in the instantaneous equations of motion if they are to be compatible with special relativity. The first three terms of a perturbation series in e^2 [Eqs. (4), (8), and (9)] for the instantaneous force for two charged particles moving in one dimension have been worked out by general, but somewhat pedestrian, methods in Sec. II. In Sec. IV, less general (restricted in their application to the reduction to instantaneous form of interactions originally given along light cones), but more powerful group-theoretic methods have been used to derive a pair of integrodifferential equations for the instantaneous forces [Eqs. (21)]. The equations of motion which are obtained by solving these integrodifferential equations also satisfy the partial differential conditions (14), thus guaranteeing compatibility with special relativity.

The methods have been extended to three dimensions in Sec. V to give integrodifferential equations for the instantaneous forces, which can be used to reduce any two-body problem involving interactions originally given along light cones to instantaneous interaction form. As in the one-dimensional case, these integrodifferential equations have been established by carrying an approximation-plus-correction procedure to infinite order to produce formally exact results. Convergence of this procedure has not been proven, so that the results are only formally, and not rigorously, established. In the case of electrodynamic interactions, convergence of the series in powers of e^2, which is actually a series in powers of the ratio of the classical charge radius e^2/mc^2 to the interparticle separation $|\mathbf{x}_{12}|$, is a sufficient condition for the rigorous establishment of the integrodifferential equations for the instantaneous forces. It has also been shown that self-interactions, such as the radiation-damping term given by Dirac, can be formally included.

Inasmuch as the whole area of instantaneous action-at-a-distance in relativistic mechanics is at present almost completely unexplored, the present work raises many questions. What are the necessary and sufficient conditions for the rigorous (as opposed to formal) validity of the integrodifferential equations for the instantaneous equations of motion? Does the series in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$ which occurs in the electrodynamic case have a finite radius of convergence? If so, what is the nature of the singularity controlling the convergence?

The conservation laws associated with the invariance of the theory under space and time translation, rotation, and pure Lorentz transformations are obscure when the equations of motion are written in the form (82) as systems of differential equations solved for the highest derivative. The usual route for the discussion of conservation laws in dynamical theories (i.e., via Hamilton's principle and Noether's theorem[3]) does not appear promising as a consequence of the zero-interaction theorem of Currie, Jordan, and Sudarshan.[13] This theorem, which shows that Dirac's relativistic Hamiltonian formulation is incompatible with inter-particle interaction, may possibly be circumvented by renouncing the demand that the Lorentz transformations be represented as canonical transformations. If such a renunciation allowed the physical coordinate to be canonical, there would be a Lagrangian, but the center-of-mass theorem could not be obtained via Noether's theorem. The existence and exhibition of conserved quantities is thus a nontrivial and presently unsolved problem in instantaneous action-at-a-distance dynamics.

Another set of questions revolves about the relationship between instantaneous action-at-a-distance and field theory. To what extent and under what conditions are instantaneous action-at-a-distance descriptions and more conventional field-theoretic descriptions of particle world lines in relativistic mechanics equivalent? The present work gives a construction for finding instantaneous action-at-a-distance equations of motion for those world lines which can be obtained by successive corrections to straight-line motion when the interaction is originally viewed as occurring along light cones. Does this family of world lines include all of those which are contained in the original equations if one imposes the further condition of asymptotic straight-line motion for widely separated particles? The method of Sec. II can be used to construct perturbative approximations to instantaneous action-at-a-distance equations of motion when the interaction is viewed as being carried by a finite mass field. Can integrodifferential equations for these instantaneous forces be found? One can also pose the question the other way: is instantaneous action-at-a-distance

[13] D. G. Currie, T. F. Jordon, and E. C. G. Sudarshan, Rev. Mod. Phys. 35, 350 (1963); D. G. Currie, J. Math. Phys. 4; 1470 (1963); J. T. Cannon and T. F. Jordon, *ibid* 5, 299 (1964). A zero-interaction theorem for N particles has been proven by H. Leutwyler, Nuovo Cimento 37, 556 (1965).

richer than field theory in the sense that there are instantaneous action-at-a-distance theories for which no field-theoretic alternative exists?

Dimensional considerations have shown that the series in powers of e^2 for the force of instantaneous action-at-a-distance electrodynamics is actually a series in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$. Such a series is reminiscent of the multipole expansions which occur in the description of the interactions of extended charge distributions; it suggests that it might be fruitful to view the charge, which was a point particle in the field-theoretic description, as an object with structure from the instantaneous action-at-a-distance vantage point. Such a view is supported by the fact that the principle of relativity demands many-body forces (Sec. VI); many-body forces in physics typically arise from the polarization of one structured body by another. That the complete elimination of the fields by passing to an instantaneous action-at-a-distance description should endow the particles with structure is not surprising; in the language of quantum field theory, we would attribute the structure to the dressing of the particles by a cloud of virtual photons. The present suggestion of structure is possibly some classical remnant of this dressing.

The fact that the series in e^2, which occurs in the present classical theory, is not consistent with the principle of relativity unless summed to all orders would seem to raise a serious question as to the meaning of the formal covariance, *at each stage* of the computation in powers of e^2, of conventional quantum electrodynamics (with or without renormalization). How can quantization restore covariance order by order in e^2? How is this restored covariance to be understood? This mystery is relieved in no visible way by the occurrence of the additional fundamental length $\hbar/mc$.

The whole question of quantization of the present theory is quite unclear. If a conventional Hamiltonian quantization is to be attempted, the zero-interaction theorem[13] must be circumvented by renouncing one (or possibly both) of the hypotheses of the theorem: (a) that the transformations of the inhomogeneous Lorentz group (Poincaré group) be canonical, and (b) that the physical coordinates be canonical coordinates. Inasmuch as a Hamiltonian formulation is not uniquely determined (to within a canonical transformation) by the classical equations of motion, this route appears to be plagued with ambiguity.

Nevertheless, any resolution of these difficulties should lead to a theory in which the present classical equations of motion can be viewed as quantum equations of motion in a Heisenberg representation with the position and velocity operators obeying (possibly quite complicated) commutation rules. The structure of the series for f_1 [in powers of $e^2/(mc^2|\mathbf{x}_{12}|)$] already indicates a source of divergence difficulties in such a theory: divergent integrals necessarily appear in a perturbation series in e^2 when the point $|\mathbf{x}_{12}| = 0$ is sampled in an intermediate state, even if the full f_1 obtained by summing the series in e^2 of the classical theory is not badly behaved at this point. These divergence difficulties would also appear to provide a hitherto unremarked source of divergence problems in conventional quantum electrodynamics. Further exploration of the questions associated with quantization will, however, have to be postponed until a better understanding of instantaneous action-at-a-distance at the classical level has been achieved.

Note added in proof: Since the present manuscript was submitted, it has been shown that the demand of asymptotic reduction to the usual Hamiltonian scheme for free particles uniquely determines the Hamiltonian scheme (to within a canonical transformation), thus removing this source of ambiguity in a conventional Hamiltonian quantization.[14] This demand also uniquely determines the conservation laws associated with the geometrical invariances of the inhomogeneous Lorentz group.

ACKNOWLEDGMENTS

The author is greatly obliged to Professor E. H. Kerner for encouragement and valuable discussions during the course of the present work, and to the University of Delaware Research Foundation for support during the summer of 1965 when the major portion of this work was accomplished.

APPENDIX A. COMPUTATION OF E_1 TO ORDER x_{12}^{-4}

For a computation of the series for $\dot{v}_1$ in Eq. (4) to order x_{12}^{-4}, the derivatives appearing in the series (6) and (7) are approximated with sufficient accuracy by

$$\frac{d^n v_2}{dt^n} \approx (-1)^n(1 - v_2^2)^{\frac{1}{2}}\{n!\,\tilde{e}_0(v_1 - v_2)^{n-1}x_{12}^{-n-1}$$
$$+ \tfrac{1}{2}(n+1)!\,\tilde{e}_1(v_1 - v_2)^{n-1}x_{12}^{-n-2}$$
$$- A_n x_{12}^2[\dot{\tilde{e}}_0 - 3v_2\tilde{e}_0/(1 - v_2^2)](v_1 - v_2)^{n+2}x_{12}^{-n-2}$$
$$- B_n\tilde{e}_0[x_{12}^2(\dot{v}_1 - \dot{v}_2)](v_1 - v_2)^{n-3}x_{12}^{-n-2}\}, \qquad \text{(A1)}$$

[14] R. N. Hill and E. H. Kerner, Phys. Rev. Letters 17, 1156 (1966); and another paper to be published.

where A_n and B_n, whose rules of formation are $A_n = (n-1)! + (n+1)A_{n-1}$ and $B_n = (n-2)(n-1)! + (n+1)B_{n-1}$ (with the initial conditions $A_1 = B_1 = 0$) have the values

$$A_n = \tfrac{1}{2}(n-1)n!,$$

$$B_n = (n+1)!\left[-\frac{5}{2} + \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right) + \frac{3}{n+1}\right]. \quad \text{(A2)}$$

From the zeroth approximation it is seen that

$$x_{12}^2[\dot{\mathcal{E}}_0 - 3v_2\dot{\mathcal{E}}_0/(1-v_2^2)] \approx (1-v_2^2)^{\frac{1}{2}}[3v_2\mathcal{E}_0/(1-v_2^2) - 2v_1\mathcal{E}_0/(1-v_1^2)],$$

$$[x_{12}^2(\dot{v}_1 - \dot{v}_2)] \approx (1-v_1^2)^{\frac{1}{2}}e_0 + (1-v_2^2)^{\frac{1}{2}}\mathcal{E}_0$$

to the accuracy needed in (A1). The series for $t - t_{\text{ret}}$ which appears in (6) can be summed in the approximation (A1) by using the formulas.

$$\sum_{n=1}^{\infty}\frac{A_n}{(n+1)!}\lambda^{n+1} = \frac{\lambda^2}{2(1-\lambda)} + \lambda + \ln(1-\lambda),$$

$$\sum_{n=1}^{\infty}\frac{B_n}{(n+1)!}\lambda^{n+1} = \frac{\lambda^2 - 6\lambda}{2(1-\lambda)} + \frac{2\lambda - 3}{1-\lambda}\ln(1-\lambda). \quad \text{(A3)}$$

The derivatives with respect to λ of the formulas (A3) are useful in summing the series for $(v_2)_{\text{ret}}$ in (7). A consistent calculation to order x_{12}^{-4} then yields the results (8) and (9).

APPENDIX B. COMPOSITION OF LORENTZ TRANSFORMATIONS

Inasmuch as the pure Lorentz transformations do not form a subgroup of the Lorentz group in three dimensions, the result of two successive pure Lorentz transformations can, in general, be written not as a pure Lorentz transformation but as a pure Lorentz transformation followed by a rotation. Thus, if we begin in a given Lorentz frame with a quantity $\mathbf{V}$ which transforms as a vector under the rotation group, the result of performing a pure Lorentz transformation to a second frame moving with velocity $\boldsymbol{\beta}$, which we write $\mathfrak{L}(\boldsymbol{\beta})\mathbf{V}$, and following it with a pure Lorentz transformation to a frame moving with a velocity $\boldsymbol{\alpha}$ with respect to the second frame can be written as

$$\mathfrak{L}(\boldsymbol{\alpha})[\mathfrak{L}(\boldsymbol{\beta})\mathbf{V}] = \mathsf{R}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})\cdot\mathfrak{L}(\boldsymbol{\gamma})\mathbf{V}, \quad \text{(B1)}$$

where

$$\boldsymbol{\gamma} = \frac{(1-\beta^2)^{\frac{1}{2}}}{1+\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}\mathsf{J}(\boldsymbol{\beta})\cdot(\boldsymbol{\alpha}+\boldsymbol{\beta}). \quad \text{(B2)}$$

The dyadic which effects the rotation is

$$\mathsf{R}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \left[\mathsf{J}(\boldsymbol{\alpha})\cdot\mathsf{J}(\boldsymbol{\beta}) + \frac{\boldsymbol{\alpha}\boldsymbol{\beta}}{(1-\alpha^2)^{\frac{1}{2}}(1-\beta^2)^{\frac{1}{2}}}\right]\cdot\mathsf{J}^{-1}(\boldsymbol{\gamma}). \quad \text{(B3)}$$

The correctness of (B2), which is just the velocity addition formula, can be seen by considering the velocity of the origin of the third frame, as seen in the first; (B3) can be verified by considering a particular representation of the Lorentz group, such as the (4-vector) representation formed by the coordinates $(\mathbf{x}, t)$ of a world point.

If we apply (B1) to an acceleration $\mathbf{a}$ in the first frame, and use $\tilde{\mathsf{R}} = \mathsf{R}^{-1}$ for a rotation, we obtain

$$\mathbf{a}\cdot\mathsf{K}^{-1}(\boldsymbol{\gamma}, \mathbf{v})\cdot\mathsf{R}^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \mathbf{a}\cdot\mathsf{K}^{-1}(\boldsymbol{\beta}, \mathbf{v})\cdot\mathsf{K}^{-1}(\boldsymbol{\alpha}, \mathfrak{L}(\boldsymbol{\beta})\mathbf{v})$$

from which the identity

$$\mathsf{R}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})\cdot\mathsf{K}(\boldsymbol{\gamma}, \mathbf{v}) = \mathsf{K}(\boldsymbol{\alpha}, \mathfrak{L}(\boldsymbol{\beta})\mathbf{v})\cdot\mathsf{K}(\boldsymbol{\beta}, \mathbf{v}) \quad \text{(B4)}$$

follows, since $\mathbf{a}$ is arbitrary.

Certain useful formulas can be derived by considering the composition of two Lorentz transformations, one of which is infinitesimal. Let $\varphi = \varphi(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$ be some function of the indicated arguments whose transformation properties is, for the moment, unspecified. If we make the replacements $\boldsymbol{\alpha} = \mathbf{z}$, $\boldsymbol{\beta} = \delta\boldsymbol{\beta}$ in (B1) and use the definition (45) of $\mathcal{R}$, we obtain

$$\mathcal{R}(\delta\boldsymbol{\beta})\mathcal{R}(\mathbf{z})\varphi = \mathsf{R}(\mathbf{z}, \delta\boldsymbol{\beta}, \boldsymbol{\gamma})\mathcal{R}(\boldsymbol{\gamma})\varphi. \quad \text{(B5)}$$

Equation (B2) implies $\boldsymbol{\gamma} = \mathbf{z} + [\mathsf{I} - \mathbf{zz}]\cdot\delta\boldsymbol{\beta}$ to first order in the infinitesimal $\delta\boldsymbol{\beta}$. Expanding (B5) to first order in $\delta\boldsymbol{\beta}$ on both sides yields

$$\delta\boldsymbol{\beta}\cdot\mathsf{L}\mathcal{R}(\mathbf{z})\varphi = -\delta\boldsymbol{\beta}\cdot[\mathsf{I} - \mathbf{zz}]\cdot\frac{\partial}{\partial\mathbf{z}}\mathcal{R}(\mathbf{z})\varphi + \mathfrak{M}, \quad \text{(B6)}$$

where, in general, $\mathfrak{M}$, which effects the rotation R, is

$$\mathfrak{M} = [1 + (1-z^2)^{\frac{1}{2}}]^{-1} \times \{[\mathcal{R}(\mathbf{z})\mathbf{x}_{12}]\cdot(\mathbf{z}\,\delta\boldsymbol{\beta} - \delta\boldsymbol{\beta}\,\mathbf{z})\cdot\partial/\partial\,[\mathcal{R}(\mathbf{z})\,\mathbf{x}_{12}] + [\mathcal{R}(\mathbf{z})\mathbf{v}_1]\cdot(\mathbf{z}\,\delta\boldsymbol{\beta} - \delta\boldsymbol{\beta}\,\mathbf{z})\cdot\partial/\partial\,[\mathcal{R}(\mathbf{z})\mathbf{v}_1] + [\mathcal{R}(\mathbf{z})\mathbf{v}_2]\cdot(\mathbf{z}\,\delta\boldsymbol{\beta} - \delta\boldsymbol{\beta}\,\mathbf{z})\cdot\partial/\partial\,[\mathcal{R}(\mathbf{z})\mathbf{v}_2]\}. \quad \text{(B7)}$$

If φ transforms as a scalar under the rotation group, $\mathfrak{M}$ vanishes. If $\varphi = \mathbf{V}$, where $\mathbf{V}$ transforms as a

vector under the rotation group, then

$$\mathfrak{M} = [1 + (1 - z^2)^{\frac{1}{2}}]^{-1}(\delta\beta\ \mathbf{z} - \mathbf{z}\ \delta\beta)\cdot[\mathcal{R}(\mathbf{z})\mathbf{V}]. \quad \text{(B8)}$$

If $\varphi = \mathbf{D}$, where $\mathbf{D}$ transforms as a dyadic under the rotation group, then

$$\mathfrak{M} = [1 + (1 - z^2)^{\frac{1}{2}}]^{-1}\{(\delta\beta\ \mathbf{z} - \mathbf{z}\ \delta\beta)\cdot(\mathcal{R}(\mathbf{z})\mathbf{D}) + [\mathcal{R}(\mathbf{z})\mathbf{D}]\cdot(\mathbf{z}\ \delta\beta - \delta\beta\ \mathbf{z})\}. \quad \text{(B9)}$$

If we perform the two transformations in the opposite order and let $\alpha = \delta\beta$, $\beta = \mathbf{z}$ in (B1), we derive, in the same manner, the result

$$\delta\beta\cdot\mathcal{R}(\mathbf{z})(\mathbf{L}\varphi) = -(1 - z^2)^{\frac{1}{2}}\ \delta\beta\cdot\mathbf{J}^{-1}(\mathbf{z})\cdot\frac{\partial}{\partial\mathbf{z}}\mathcal{R}(\mathbf{z})\varphi - \mathfrak{M}. \quad \text{(B10)}$$

Both (B6) and (B10) can be verified directly without making use of the group-theoretic argument.

If $\mathbf{f}_1^{(0)}(\mathbf{x}_{12}, \mathbf{v}_1, \mathbf{v}_2)$ is the approximation to the acceleration $\mathbf{a}_1$ of particle 1 reckoned on the assumption of straight-line motion for particle 2 ($\mathbf{f}_2 = 0$), then (38a) implies

$$\delta\beta\cdot\mathbf{L}\mathbf{f}_1^{(0)} + 2\ \delta\beta\cdot\mathbf{v}_1\mathbf{f}_1^{(0)} + \delta\beta\cdot\mathbf{f}_1^{(0)}\mathbf{v}_1 = 0. \quad \text{(B11)}$$

By operating on this with $\mathcal{R}(\mathbf{z})$ and multiplying the result on the right by $\mathbf{K}(\mathbf{z}, \mathbf{v}_1)$, we obtain

$$[\delta\beta\cdot\mathcal{R}(\mathbf{z})(\mathbf{L}\mathbf{f}_1^{(0)})]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) + [\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}][2\ \delta\beta\cdot\mathcal{R}(\mathbf{z})\mathbf{v}_1]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) + [\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\delta\beta\ \mathbf{v}_1\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) = 0. \quad \text{(B12)}$$

Since (B11) is the statement of the vanishing of the infinitesimal transformation on $\mathbf{f}_1\cdot\mathbf{K}$ in straight-line approximation, we should be able to show that $[\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1)$ is independent of $\mathbf{z}$ from (B12) by using (B10). From this we infer the existence of the identity

$$(1 - z^2)^{\frac{1}{2}}\ \delta\beta\cdot\mathbf{J}^{-1}(\mathbf{z})\cdot\frac{\partial}{\partial\mathbf{z}}\mathbf{K}(\mathbf{z}, \mathbf{v}) = [1 + (1 - z^2)^{\frac{1}{2}}]^{-1}(\mathbf{z}\ \delta\beta - \delta\beta\ \mathbf{z})\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}) - [2\ \delta\beta\cdot\mathcal{R}(\mathbf{z})\mathbf{v}]\mathbf{K}(\mathbf{z}, \mathbf{v}) - \delta\beta\ [\mathcal{R}(\mathbf{z})\mathbf{v}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}), \quad \text{(B13)}$$

which can be verified (with either $\mathbf{v} = \mathbf{v}_1$ or $\mathbf{v} = \mathbf{v}_2$) by direct computation. Hence, (B11) implies

$$[\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1) = \mathbf{f}_1^{(0)}. \quad \text{(B14)}$$

If we now place (B14) back into (B11), we obtain

$$(\delta\beta\cdot\mathbf{L})\{[\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1)\} + [\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1)\cdot[2(\delta\beta\cdot\mathbf{v}_1)\mathbf{I} + \delta\beta\ \mathbf{v}_1] = 0. \quad \text{(B15)}$$

We use (B6) and $\partial/\partial\mathbf{z}\{[\mathcal{R}(\mathbf{z})\mathbf{f}_1^{(0)}]\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1)\} = 0$ in (B15) and remark that $\mathbf{f}_1^{(0)}$ is arbitrary except for the restriction (B11) to derive the identity

$$\delta\beta\cdot\mathbf{L}\mathbf{K}(\mathbf{z}, \mathbf{v}) + \delta\beta\cdot[\mathbf{I} - \mathbf{z}\mathbf{z}]\cdot\frac{\partial}{\partial\mathbf{z}}\mathbf{K}(\mathbf{z}, \mathbf{v}) + 2(\mathbf{v}\cdot\delta\beta)\mathbf{K}(\mathbf{z}, \mathbf{v}) + \mathbf{K}(\mathbf{z}, \mathbf{v})\cdot\delta\beta\ \mathbf{v} + [(1 + (1 - z^2)^{\frac{1}{2}}]^{-1}(\mathbf{z}\ \delta\beta - \delta\beta\ \mathbf{z})\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}) = 0, \quad \text{(B16)}$$

which holds for either $\mathbf{v} = \mathbf{v}_1$ or $\mathbf{v} = \mathbf{v}_2$. Equation (B16), which is needed in Appendix D, can also be verified directly.

APPENDIX C. PATH INDEPENDENCE

In this Appendix, we show that, if

$$\mathbf{L}\mathbf{f}_1^{(n)} + 2\mathbf{v}_1\mathbf{f}_1^{(n)} + \mathbf{f}_1^{(n)}\mathbf{v}_1 = -\mathbf{x}_{12}\sum_{l=1}^{n-1}\mathbf{f}_2^{(l)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n-l)}, \quad \text{(C1)}$$

$$\tilde{\mathbf{L}}\mathbf{f}_2^{(m)} + 2\mathbf{v}_2\mathbf{f}_2^{(m)} + \mathbf{f}_2^{(m)}\mathbf{v}_2 = \mathbf{x}_{12}\sum_{l=1}^{m-1}\mathbf{f}_1^{(l)}\cdot\frac{\partial}{\partial\mathbf{v}_1}\mathbf{f}_2^{(m-l)} \quad \text{(C2)}$$

hold, then

$$\int_0^{\hat{\mathbf{s}}(0)} d\mathbf{z}\cdot\mathbf{T}^{(m,n)}(\mathbf{z}),$$

which appears at the $(m + n)$th stage of the iteration of (68), is independent of path, where

$$\mathbf{T}^{(m,n)}(\mathbf{z}) \equiv (1 - z^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{z})\cdot\left\{\mathcal{R}(\mathbf{z})\left[\mathbf{x}_{12}\mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)}\right]\right\}\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1). \quad \text{(C3)}$$

Path independence in the original "correction formula" is a consequence of the fact that, since $t_s - t_0 = -\mathbf{z}\cdot\mathbf{x}_s$ [Eq. (35)], only the component of $\mathbf{z}$ in the direction of $\mathbf{x}_s$ matters as far as moving along the world line of particle 2 is concerned. Thus, path independence can be expected to follow from considering infinitesimal Lorentz transformations in a direction perpendicular to $\mathbf{x}_{12}$. We therefore cross

$\mathbf{x}_{12}$ into (C1) and (C2) from the left (remembering that $\mathbf{x}_{12} \times \mathbf{L} = \mathbf{x}_{12} \times \tilde{\mathbf{L}}$) to obtain

$$\mathbf{x}_{12} \times [\mathbf{L}\mathbf{f}_1^{(n)} + 2\mathbf{v}_1\mathbf{f}_1^{(n)} + \mathbf{f}_1^{(n)}\mathbf{v}_1] = 0, \tag{C4}$$

$$\mathbf{x}_{12} \times [\mathbf{L}\mathbf{f}_2^{(m)} + 2\mathbf{v}_2\mathbf{f}_2^{(m)} + \mathbf{f}_2^{(m)}\mathbf{v}_2] = 0. \tag{C5}$$

We now procede to show that the path-independence condition $(\partial/\partial\mathbf{z}) \times \mathbf{T}^{(m,n)}(\mathbf{z}) = 0$ follows from (C4) and (C5). We use the identities

$$\partial/\partial\mathbf{z} \times ((1 - \mathbf{z}\cdot\mathbf{v}_2)^{-1}\{[\mathcal{R}(\mathbf{z})\mathbf{x}_{12})\cdot\mathbf{J}(\mathbf{z})]\}) = 0,$$

$$\partial/\partial\mathbf{z}[(1 - \mathbf{z}\cdot\mathbf{v}_2)(1 - z^2)^{-\frac{1}{2}}] = -(1 - z^2)^{-1}(1 - \mathbf{z}\cdot\mathbf{v}_2)\mathbf{J}(\mathbf{z})\cdot\mathcal{R}(\mathbf{z})\mathbf{v}_2$$

to obtain

$$\frac{\partial}{\partial\mathbf{z}} \times \mathbf{T}^{(m,n)}(\mathbf{z}) = \frac{[\mathcal{R}(\mathbf{z})\mathbf{x}_{12}]\cdot\mathbf{J}(\mathbf{z})}{1 - \mathbf{z}\cdot\mathbf{v}_2} \times \left(\left\{\frac{1 - \mathbf{z}\cdot\mathbf{v}_2}{1 - z^2}\mathbf{J}(\mathbf{z})\cdot[\mathcal{R}(\mathbf{z})\mathbf{v}_2] - \frac{\partial}{\partial\mathbf{z}}\right\} \times \left[\left(\mathcal{R}(\mathbf{z})\mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)}\right)\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1)\right]\right)\cdot \tag{C6}$$

For arbitrary vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, $\boldsymbol{\alpha}\cdot\mathbf{J}(\mathbf{z}) \times \boldsymbol{\beta} = (1 - z^2)^{-\frac{1}{2}}\boldsymbol{\alpha} \times \mathbf{J}^{-1}(\mathbf{z})\cdot\boldsymbol{\beta}$. We use this, together with (B10) and (B13), to reduce (C6) to the form

$$\frac{\partial}{\partial\mathbf{z}} \times \mathbf{T}^{(m,n)}(\mathbf{z}) = (1 - z^2)^{-\frac{1}{2}} \times \left\{\mathcal{R}(\mathbf{z})\mathbf{x}_{12} \times \left[\mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)} + \mathbf{L}\left(\mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)}\right) + \mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}(2\mathbf{v}_1\mathbf{f}_1^{(n)} + \mathbf{f}_1^{(n)}\mathbf{v}_1)\right]\right\}\cdot\mathbf{K}(\mathbf{z}, \mathbf{v}_1). \tag{C7}$$

If we now use the identity

$$\mathbf{x}_{12} \times \mathbf{L}\left(\mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)}\right) = \mathbf{x}_{12} \times [(\mathbf{L}\mathbf{f}_2^{(m)}) + \mathbf{v}_2\mathbf{f}_2^{(m)} + \mathbf{f}_2^{(m)}\mathbf{v}_2]\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n)} + \mathbf{f}_2^{(m)}\cdot\frac{\partial}{\partial\mathbf{v}_2}[\mathbf{x}_{12} \times \mathbf{L}\mathbf{f}_1^{(n)}] \tag{C8}$$

and then employ (C4) and (C5), we conclude that $(\partial/\partial\mathbf{z}) \times \mathbf{T}^{(m,n)}(\mathbf{z}) = 0$.

APPENDIX D. LORENTZ COVARIANCE

In this Appendix, we show that, if

$$\mathbf{f}_1^{(n)} = -\sum_{l=1}^{n-1}\int_0^{\hat{\mathbf{S}}(0)} d\mathbf{z}\cdot\mathbf{T}^{(l,n-l)}(\mathbf{z}) \tag{D1}$$

[where $\mathbf{T}$ is defined in (C3) and $n \neq 0$] holds, and if $(\partial/\partial\mathbf{z}) \times \mathbf{T}^{(l,n-l)}(\mathbf{z}) = 0$ so that the integral is path independent, then $\mathbf{f}_1^{(n)}$ satisfies

$$\mathbf{L}\mathbf{f}_1^{(n)} + 2\mathbf{v}_1\mathbf{f}_1^{(n)} + \mathbf{f}_1^{(n)}\mathbf{v}_1 = -\mathbf{x}_{12}\sum_{l=1}^{n-1}\mathbf{f}_2^{(l)}\cdot\frac{\partial}{\partial\mathbf{v}_2}\mathbf{f}_1^{(n-l)}. \tag{D2}$$

From (D1) it follows immediately that

$$\delta\boldsymbol{\beta}\cdot(\mathbf{L}\mathbf{f}_1^{(n)} + 2\mathbf{v}_1\mathbf{f}_1^{(n)} + \mathbf{f}_1^{(n)}\mathbf{v}_1) = -\sum_{l=1}^{n-1}\left\{\delta\boldsymbol{\beta}\cdot[\mathbf{L}\hat{\mathbf{S}}(0)]\cdot\mathbf{T}^{(l,n-l)}[\hat{\mathbf{S}}(0)] + \int_0^{\hat{\mathbf{S}}(0)} d\mathbf{z}\cdot[(\delta\boldsymbol{\beta}\cdot\mathbf{L})\mathbf{T}^{(l,n-l)}(\mathbf{z}) + 2(\mathbf{v}_1\cdot\delta\boldsymbol{\beta})\mathbf{T}^{(l,n-l)}(\mathbf{z}) + \mathbf{T}^{(l,n-l)}(\mathbf{z})\cdot\delta\boldsymbol{\beta}\mathbf{v}_1]\right\}. \tag{D3}$$

If we use (B6), (B9), (B16), and the identity

$$\delta\boldsymbol{\beta}\cdot[\mathbf{I} - \mathbf{z}\mathbf{z}]\cdot\partial/\partial\mathbf{z}\,[(1 - z^2)^{-\frac{1}{2}}\mathbf{J}(\mathbf{z})] = [1 + (1 - z^2)^{\frac{1}{2}}]^{-1}\mathbf{J}(\mathbf{z})\cdot(\mathbf{z}\,\delta\boldsymbol{\beta} - \delta\boldsymbol{\beta}\,\mathbf{z}) + [\mathbf{I}(\delta\boldsymbol{\beta}\cdot\mathbf{z}) + \delta\boldsymbol{\beta}\,\mathbf{z}]\cdot\mathbf{J}(\mathbf{z}), \tag{D4}$$

which can be verified by direct computation, the integral appearing in (D3) takes the form

$$\int_0^{\hat{\mathbf{S}}(0)} d\mathbf{z}\cdot\left\{-\delta\boldsymbol{\beta}\cdot[\mathbf{I} - \mathbf{z}\mathbf{z}]\cdot\frac{\partial}{\partial\mathbf{z}}\mathbf{T}^{(l,n-l)}(\mathbf{z}) + [\mathbf{I}(\mathbf{z}\cdot\delta\boldsymbol{\beta}) + \delta\boldsymbol{\beta}\,\mathbf{z}]\cdot\mathbf{T}^{(l,n-l)}(\mathbf{z})\right\}. \tag{D5}$$

For any dyadic function $\mathbf{D}(\mathbf{z})$,

$$\delta\boldsymbol{\beta}\cdot[\mathbf{I} - \mathbf{z}\mathbf{z}]\cdot\frac{\partial}{\partial\mathbf{z}}\mathbf{D} = \partial/\partial\mathbf{z}\,[\delta\boldsymbol{\beta}\cdot(\mathbf{I} - \mathbf{z}\mathbf{z})\cdot\mathbf{D}] + [\mathbf{I}(\mathbf{z}\cdot\delta\boldsymbol{\beta}) + \delta\boldsymbol{\beta}\,\mathbf{z}]\cdot\mathbf{D} - \delta\boldsymbol{\beta}\cdot[\mathbf{I} - \mathbf{z}\mathbf{z}] \times [\partial/\partial\mathbf{z} \times \mathbf{D}]. \tag{D6}$$

Using (D6) and the path-independence condition $(\partial/\partial\mathbf{z}) \times \mathbf{T}(\mathbf{z}) = 0$, (D5) becomes

$$-\int_0^{\hat{\mathbf{S}}(0)} d\mathbf{z}\cdot\frac{\partial}{\partial\mathbf{z}}[\delta\boldsymbol{\beta}\cdot(\mathbf{I} - \mathbf{z}\mathbf{z})\cdot\mathbf{T}^{(l,n-l)}(\mathbf{z})]. \tag{D7}$$

Performing the integration and using

$$\mathbf{L}\hat{\mathbf{S}}(0) = \mathbf{I} - \hat{\mathbf{S}}(0)\hat{\mathbf{S}}(0),$$

we obtain (D2).

APPENDIX E. RADIATION REACTION

The radiation reaction term given by Dirac takes the form

$$\Gamma_{1\mu} = \tfrac{2}{3}[d^2u_{1\mu}/d\tau_1^2 - (du_{1\lambda}/d\tau_1)(du_1^{\lambda}/d\tau_1)u_{1\mu}]. \tag{E1}$$

By rewriting this in 3-vector form, we obtain the radiation reaction contribution $\mathbf{f}_1^{rr}$ to $\mathbf{a}_1$:

$$\mathbf{f}_1^{rr} = \tfrac{2}{3}(1 - v_1^2)^{-\frac{1}{2}}[D\mathbf{f}_1 + 3(1 - v_1^2)^{-1}(\mathbf{v}_1 \cdot \mathbf{f}_1)\mathbf{f}_1], \quad \text{(E2)}$$

where

$$D \equiv (\mathbf{v}_1 - \mathbf{v}_2)\cdot\frac{\partial}{\partial \mathbf{x}_{12}} + \mathbf{f}_1 \cdot \frac{\partial}{\partial \mathbf{v}_1} + \mathbf{f}_2 \cdot \frac{\partial}{\partial \mathbf{v}_2} \quad \text{(E3)}$$

is the total time derivative operator. We define the operators $\boldsymbol{\Gamma}$ and $\tilde{\boldsymbol{\Gamma}}$ by

$$\begin{aligned} \boldsymbol{\Gamma}\mathbf{f} &\equiv 2\mathbf{v}_1\mathbf{f} + \mathbf{f}\mathbf{v}_1 + \mathbf{L}\mathbf{f} + \mathbf{x}_{12}\mathbf{f}_2 \cdot \frac{\partial}{\partial \mathbf{v}_2}\mathbf{f}, \\ \tilde{\boldsymbol{\Gamma}}\mathbf{f} &\equiv 2\mathbf{v}_2\mathbf{f} + \mathbf{f}\mathbf{v}_2 + \tilde{\mathbf{L}}\mathbf{f} - \mathbf{x}_{12}\mathbf{f}_1 \cdot \frac{\partial}{\partial \mathbf{v}_1}\mathbf{f}, \end{aligned} \quad \text{(E4)}$$

for $\mathbf{f}$ any function of $\mathbf{x}_{12}$, $\mathbf{v}_1$, and $\mathbf{v}_2$. Direct computation then shows that

$$\begin{aligned} \boldsymbol{\Gamma}\mathbf{f}_1^{rr} = \tfrac{2}{3}(1 - v_1^2)^{-\frac{1}{2}}\{ & D(\boldsymbol{\Gamma}\mathbf{f}_1) \\ & + (\tilde{\boldsymbol{\Gamma}}\mathbf{f}_2)\cdot\partial\mathbf{f}_1/\partial\mathbf{v}_2 + (\boldsymbol{\Gamma}\mathbf{f}_1)\cdot\partial\mathbf{f}_1/\partial\mathbf{v}_1 \\ & + (1 - v_1^2)^{-1}(\boldsymbol{\Gamma}\mathbf{f}_1)\cdot[\mathbf{I}(\mathbf{v}_1\cdot\mathbf{f}_1) + \mathbf{v}_1\mathbf{f}_1]\}. \end{aligned} \quad \text{(E5)}$$

An expansion of (E5) in inverse powers of $|\mathbf{x}_{12}|$ and application of the result of Appendix D, order by order, on the right-hand side, shows that each successive approximation to $\mathbf{f}_1^{rr}$ satisfies $\boldsymbol{\Gamma}\mathbf{f}_1^{rr} = 0$ to the order of the approximation at each stage of the approximation. Hence, the path-independence and Lorentz-covariance conditions continue to hold when $\mathbf{f}_1^{rr}$ is included in $\mathbf{f}_1$.

Functional Approach to Classical Non-Equilibrium Statistical Mechanics

IWAO HOSOKAWA
National Aerospace Laboratory, Jindaiji, Chofu, Tokyo
(Received 6 June 1966)

A new approach to classical non-equilibrium statistical mechanics is considered. The essential idea is to solve Bogoliubov's functional-differential equation for the generating functional embracing all distribution functions for the particle system with the aid of the technique of functional integration. It becomes possible with a minor modification of his formalism. A general solution is given. The solution for our state functional is reduced to the Hopf characteristic functional for the Vlasov self-consistent field when the interaction energy of particles is relatively small in each s-body Hamiltonian H_s.

1. INTRODUCTION

AT present there seems to be two general methods for approaching nonequilibrium statistical mechanics; one deals with the so-called BBGKY hierarchy[1] and the other was developed by Prigogine[2] and his school. The starting point for both is the Liouville equation for the N-body system and hence perfectly reliable from the mechanical point of view. In either method, however, no exact general solutions for distribution functions or correlations with any number of particles have been found without using an infinite-series expansion, e.g., in powers of an appropriate parameter like the density of particles. Such an expression, however, always includes the question of convergence, so that various statements based on this type of solution undergo considerable restriction of their applicability; they become already doubtful for too dense a gas, etc. We therefore still have a reason to seek for another general method, if any, which can yield *closed-form* general solutions, valid irrespectively of the magnitude of the parameters involved.

For this reason a *functional* approach to nonequilibrium statistical mechanics is sketched in this paper. The idea of using a functional to describe a statistical state is not new, since Bogoliubov[3] introduced it in 1946 to reform the BBGKY hierarchy into a single closed equation for the generating functional; his aim was to achieve the formalistic convenience in describing the state, but no more. But, if it is possible to directly integrate a *functional-differential* equation, it is evident that a new scope of development in statistical mechanics will emerge, since the general solution may be given in closed form for the generating functional, without treating the question of convergence. The present author shows here that this is possible with a necessary minor modification of the Bogoliubov *functional* formalism; a special type of complex-valued functional is introduced anew (Sec. 2). The modification, of course, holds the equivalence between our new formalism and the BBGKY hierarchy (for the limit $N \to \infty$). The general solution for our *state* functional is given in terms of a kind of *functional integration* (Sec. 3). As this mathematical tool may be somewhat novel, the Appendix is devoted to it.

Although our solution expressed in a functional integral is apparently sophisticated, we stress that the Monte Carlo approach makes its practical utility possible, since we can have the approximation scheme in which it is estimated by a multiple integral (cf. Appendix); the situation is similar to the case with the Feynmann integral. Finally, it is mentioned that the general solution may be approximated by the Hopf characteristic functional[4] for the turbulent one-body distribution function governed by the Vlasov kinetic equation if the potential energy is much less than the kinetic energy in each s-body Hamiltonian H_s ($s = 2, 3, \cdots$) (Sec. 4).

In this paper we treat the system of identical particles governed by classical mechanics. We assume (1) that the Liouville equation holds valid in the limit $N \to \infty$ with the condition that N/V remains a positive finite constant in such a limit (V is the volume which involves the total system), and (2) that more than one particle cannot occupy the same spatial location at the same time, or equivalently that a particle always has a hard repulsive core. At the same time it is natural to assume that the wall encircling V has a hard repulsive (potential) layer. Moreover, our N-body distribution function should be symmetric with the interchange between particles, as also assumed by

[1] Bogoliubov–Born–Green–Kirkwood–Yvon. See, for example, Uhlenbeck's special lectures in M. Kac, *Probability and Related Topics in Physical Sciences* (Interscience Publishers, Inc., New York, 1959).

[2] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).

[3] N. N. Bogoliubov, J. Phys. U.S.S.R. 10, 265 (1946); *Problems of a Dynamical Theory in Statistical Physics*, translated by E. Gora (Providence College, Providence, Rhode Island 1959).

[4] E. Hopf, J. Ratl. Mech. Anal. 1, 87 (1952).

Bogoliubov.[3] This premise makes it necessary for us to essentially treat an *ensemble* of the $6N$-dimensional phase points; even if we exactly specify a configuration of N bodies in the six-dimensional phase space at the initial time, we must start from not one but $N!$ points in the $6N$-dimensional phase space on the basis of assumption (2). That is to say, in this formalism we have, in principle, no chance of specifying only one point in the $6N$-dimensional phase space for any real physical state, although the situation is more amplified on the basis of indeterminism of the system observation or coarse-graining procedure. An interesting topological picture is given: the state functional introduced here, which corresponds to an ensemble of these phase points, is considered as a point in the new space consisting of a certain class of regular functionals; in this picture a possible steady (e.g., equilibrium) state must correspond to a fixed point for the automapping of this space, which represents the time development of the ensemble.

2. STATE FUNCTIONAL AND BASIC EQUATION

The essence of Bogoliubov's functional formalism for non-equilibrium statistical mechanics is that any (s-body *generic*) distribution function F_s should be embraced in one generating functional as a coefficient function in its *functional* Taylor series[5] (times a numerical factor) and that the functional should be governed by a single *functional-differential*[3,5] equation of the type

$$\partial\psi(y, t)/\partial t = \mathfrak{L}\psi(y, t), \tag{2.1}$$

where y is a function on the one-body phase space X, $y \in A$ (cf. Appendix), t the time variable, and $\mathfrak{L}$ a linear operator on functionals including functional differentiation; this operator is given in such a way that (2.1) should be equivalent to the BBGKY hierarchy. Now that we know the technique of a functional integration as defined in the Appendix, it is natural to expect the integrated form, i.e., general solution of (2.1), to be

$$\psi(y, t) = \int_A K(y, t/y', t') \times \psi(y',t')\, \delta y', \qquad t \geq t', \tag{2.2}$$

where K is the kernel functional which is absolutely integrable. [Absolute integrability of K is the essential property for the mapping on $\psi(y', t')$ defined by (2.2) to be *continuous*, as is seen from (2.9).] If so, in order for the integral in (2.2) to be mathematically meaningful, it is sufficient that

$$\max_{y\in A} |\psi(y', t')| < \infty. \tag{2.3}$$

If we consider our functional as an element of a norm space by introducing the norm

$$||\psi(y, t)|| = \max_{y\in A} |\psi(y, t)|, \tag{2.4}$$

(2.3) means that ψ should have a finite norm. Unfortunately Bogoliubov's original form of the functional cannot satisfy this condition, because his functional is expressed by the functional Taylor series with (nonvanishing) positive coefficients and increase endlessly as $y \to +\infty$. Moreover, for Bogoliubov's original functional to converge, the function space $\{y(x)\}$ should be limited in a complicated way,[3] over which we cannot make a meaningful functional integration of the type (2.2). Thus it is desirable for us to find a new form of the generating functional on A with a finite norm and still governed by the equation of the type (2.1).

Keeping in mind that we are going to treat the limiting state in which both the number of particles N and the spatial volume V containing them tend to infinity, holding the relation $N/V = \text{const}$, let us consider the complex-valued functional

$$\psi(y, t) = 1 + \sum_{s=1}^{\infty} \frac{i^s}{s!} \int_X \cdots \int_X F_s(x_1, \cdots, x_s, t) \times y(x_1) \cdots y(x_s)\, dx_1 \cdots dx_s. \tag{2.5}$$

Here F_s, as the s-body generic distribution function, is a real-valued symmetric function with respect to an interchange of $x_1, \cdots, x_s \in X$ and should obey

$$F_s(x_1, \cdots, x_s, t) = \lim_{V\to\infty} \int_X F_{s+1}(x_1, \cdots, x_s, x_{s+1}, t)\, dx_{s+1}/V, \qquad s \geq 1. \tag{2.6}$$

Apart from the factors i^s, this is in complete agreement with Bogoliubov's generating functional.[3] As ψ embraces all generic distribution functions and hence represents a state of our particle system exactly in the statistical sense, it may be called the state functional. Now, it is very possible that the state functional ψ in (2.5) has a finite form, since each coefficient in the series (2.5) is not always positive. (By definition $F_s \geq 0$.) For example, we consider the limit situation where all particles are independent; then

$$F_s(x_1, \cdots, x_s, t) = F_1(x_1, t) \cdots F_1(x_s, t) \tag{2.7}$$

so that

$$\psi(y, t) = \exp\left\{i \int_X F_1(x, t)y(x)\, dx\right\}, \tag{2.8}$$

the norm of which is unity. We may here have to

[5] V. Volterra, *Theory of Functionals* (Blackie & Son, London-Glasgow, 1930).

answer the problem of convergence of the series expression (2.5). We may assume that every realizable state of our particle system should have a finite norm. This is a reasonable assumption, because, first, we can see from (2.2) that if the initial functional $\psi(y', t')$ is given with a finite norm, at any time later the functional $\psi(y, t)$ continues to have a finite norm; that is,

$$||\psi(y, t)|| \leq \max_{y \in A} \int_A |K| \, \delta y' \, ||\psi(y', t')||. \tag{2.9}$$

Second, once our functional has a finite norm, the series in (2.5) should be bounded in absolute magnitude.

While all regular functionals (which can be everywhere Taylor-expanded) with a finite norm and satisfying (2.6) constitute a linear norm space, state functionals do not. A linear combination of state functionals $\alpha\psi_1 + \beta\psi_2$ can be another state functional only with $\alpha + \beta = 1$, since we have the condition

$$\psi(0, t) = 1, \tag{2.10}$$

as is seen from (2.5). If we denote the norm space by Ψ, we may say that a state functional is a point in Ψ which in particular satisfies (2.10). The condition (2.6) may be expressed by the following functional equation:

$$\psi = \lim_{V \to \infty} \int_X \frac{\delta\psi}{i \, \delta y(x)} \, dx/V, \tag{2.11}$$

which does not necessarily ensure (2.10).

Since (2.5) is a minor modification of Bogoliubov's functional, it is easy to establish the basic equation which governs ψ, i.e., the explicit form of (2.1); we have only to change the real argument in his theory[3] into the imaginary argument. Thus, we have

$$\begin{aligned}\frac{\partial\psi}{\partial t} &= i \int_X y(x) \left[H_1(x); \frac{\delta\psi}{i \, \delta y(x)} \right] dx \\ &- \frac{1}{2} \int_X \int_X \{y(x)y(x') - iny(x) - iny(x')\} \\ &\times \left[\phi(|q - q'|); \frac{\delta^2\psi}{i^2 \, \delta y(x) \, \delta y(x')} \right] dx \, dx',\end{aligned} \tag{2.12}$$

where $H_1(x)$ is the Hamiltonian for one body, $\phi(|q - q'|)$ the potential of interaction between two particles, $x = (q, p)$, q and p being the position and momentum vector, respectively, $n = N/V$, and $[\,;]$ denotes the Poisson bracket. To identify the equivalence of (2.12) and the BBGKY hierarchy, one may directly substitute (2.5) into (2.12).

Finally we show that only if the initial state functional obeys (2.11), the solution of (2.12) does so at any successive time. First, functional differentiation of (2.12) gives

$$\begin{aligned}&\frac{\partial}{\partial t} \frac{\delta\psi}{i \, \delta y(x)} \\ &= \left[H_1(x); \frac{\delta\psi}{i \, \delta y(x)} \right] + \int_X \{iy(x') + n\} \\ &\times \left[\phi(|q - q'|); \frac{\delta^2\psi}{i^2 \, \delta y(x) \, \delta y(x')} \right] dx' + \mathfrak{L} \frac{\delta\psi}{i \, \delta y(x)}.\end{aligned} \tag{2.13}$$

Here and hereafter $\mathfrak{L}$ represents the whole operation in the right-hand side of (2.12). Next, if we integrate (2.13) over the phase space X, we can see that the first two terms should vanish provided that any distribution function vanishes at the boundary points (including the infinity) of X; this condition is ensured by the assumption described in Sec. 1 (that the boundary of the container has a hard repulsive layer) and the fact that the mean kinetic energy of particles should be finite. In fact, as for the first term, we have

$$\begin{aligned}&\int_X \left[H_1(x); \frac{\delta\psi}{i \, \delta y(x)} \right] dx \\ &= \int_X dx \left(\frac{\partial H_1}{\partial q} \frac{\partial}{\partial p} - \frac{\partial H_1}{\partial p} \frac{\partial}{\partial q} \right) \\ &\times \sum_{s=1}^{\infty} \frac{i^{s-1}}{(s-1)!} \int_X \cdots \int_X F_s(x, x_2, \cdots x_s, t) \\ &\times y(x_2) \cdots y(x_s) \, dx_2 \cdots dx_s,\end{aligned} \tag{2.14}$$

which, on this condition, vanishes by the integration by parts with respect to $dx = dq \, dp$, as is easily checked. The same is true for the second term in (2.13). Thus, we establish

$$\frac{\partial}{\partial t} \int_X \frac{\delta\psi}{i \, \delta y(x)} \, dx/V = \mathfrak{L} \int_X \frac{\delta\psi}{i \, \delta y(x)} \, dx/V \tag{2.15}$$

in the limit $V \to \infty$. Comparing this with (2.1), we may have the integrated form based on the same kernel:

$$\begin{aligned}\lim_{V \to \infty} \int_X \frac{\delta\psi}{i \, \delta y(x)} \, dx/V &= \int_A K(y, t/y', t') \\ &\times \lim_{V \to \infty} \int_X \frac{\delta\psi}{i \, \delta y'(x)} \, dx/V \, \delta y'.\end{aligned} \tag{2.16}$$

If (2.11) holds at time t', this can be rewritten as

$$\begin{aligned}&\lim_{V \to \infty} \int_X \frac{\delta\psi}{i \, \delta y(x)} \, dx/V \\ &= \int_A K(y, t/y', t') \psi(y', t') \, \delta y' \\ &= \psi(y, t),\end{aligned} \tag{2.17}$$

which shows that the relation (2.11) is conserved

at any time t after t'. As a result, the associative condition (2.11) has only to be taken into account on the initial state functional.

Now, it is evident that the mapping (2.2) always maps the space Ψ into itself. Thus, [with the aid of (3.10)] we may understand that our state functional ψ moves in Ψ, drawing a line trajectory according to the dynamical rule (2.12). If we have the unique steady (e.g., equilibrium) state for the system considered, it is expected that all such trajectories in Ψ converge to the point corresponding to that state. As is known from (2.5), always

$$||\psi|| \geq 1, \tag{2.18}$$

so that any trajectory cannot penentrate into the unit (pseudo) sphere at the origin in Ψ.

3. GENERAL SOLUTION

We are in a position to solve our basic equation (2.12) [or (2.1) in abridged notation] with the conditions (2.10) and (2.11). A general solution for $\psi(y, t)$ is established by finding the explicit form of K in (2.2) and verifying that (2.2) satisfies (2.10); we already know (in Sec. 2) that the condition (2.11) is automatically satisfied by $\psi(y, t)$ if $\psi(y', t')$ is a state functional.

In he first place, it is important to examine at few properties of K. As is seen from (2.1) and (2.2), K should obey

$$\partial K(y, t/y', t')/\partial t = \mathfrak{L}K(y, t/y', t'). \tag{3.1}$$

Equation (2.2) requires

$$K(y, t'/y', t') = \delta[y - y'] \tag{3.2}$$

[cf. (A17)]. For the solution to be unique,

$$K(y, t/y', t') = \int_A K(y, t/y'', t'')K(y'', t''/y', t')\, \delta y'' \tag{3.3}$$

for $t' \leq t'' \leq t$. Next, immediately from (3.1) and (3.2), we obtain

$$\begin{aligned} K_{\Delta t}(y/y') &\equiv K(y, t' + \Delta t/y', t') \\ &= \delta[y - y'] + \Delta t\, \mathfrak{L}\, \delta[y - y'] + O(\Delta t^2), \end{aligned} \tag{3.4}$$

which may be called the infinitesimal kernel. Substitution of an appropriate expression of the delta functional into (3.4) leads to an explicit form of $K_{\Delta t}$, and hence K in general is constructed by the property (3.3). This scheme is nothing more than a *functional* version of the propagation kernel method for a usual differential equation and was recently used by Rosen[6] to solve the Hopf equation for turbulence.[4]

With (A17) and (2.12) substituted into (3.4), we have

$$\begin{aligned} K_{\Delta t}(y/y') &= \int_A \left\{1 + i\Delta t \int_X dx\, y(x)Qz(x) - \tfrac{1}{2}\Delta t \int_X \int_X dx\, dx'\, y(x)y(x') \right. \\ &\qquad \left. \times [\phi(|q - q'|); z(x)z(x')]\right\} \exp\left[i \int_X dx\, z(x)\{y(x) - y'(x)\}\right] \delta z + O(\Delta t^2) \\ &= \int_A \exp\left(i \int_X dx\, [z(x)\{y(x) - y'(x)\} + \Delta t y(x)Qz(x)] \right. \\ &\qquad \left. - \tfrac{1}{2}\Delta t \int_X \int_X dx\, dx'\, y(x)y(x')[\phi(|q - q'|); z(x)z(x')]\right) \delta z + O(\Delta t^2), \end{aligned} \tag{3.5}$$

where

$$Qz(x) \equiv [H_1(x); z(x)] + n \int_X dx'\, [\phi(|q - q'|); z(x)z(x')]. \tag{3.6}$$

The higher-order terms $O(\Delta t^2)$ may be calculated in an explicit manner but they are not important when Δt is sufficiently small; thus they are omitted in every equation below. Now, following the above-described scheme, an explicit form of $K(y, t/y_0, t_0)$ is given by the repeated use of (3.3) after dividing the time difference $t - t_0$ into infinitely small pieces; that is,

$$\begin{aligned} K(y, t/y_0, t_0) &= \lim_{\Delta t \to 0} \int_A \cdots \int_A \prod_{k=1}^{M} K_{\Delta t}(y_k/y_{k-1}) \prod_{k=1}^{M-1} \delta y_k \\ &= \lim_{\Delta t \to 0} \int_A \cdots \int_A \exp\left(i \int_X dx \sum_{k=1}^{M} [z_k(x)\{y_k(x) - y_{k-1}(x)\}\right. \end{aligned} \tag{3.7}$$

$$\left. + \Delta t y_k(x)Qz_k(x)] - \frac{\Delta t}{2} \int_X \int_X dx\, dx' \sum_{k=1}^{M} y_k(x)y_k(x')[\phi(|q - q'|); z_k(x)z_k(x')]\right) \prod_{k=1}^{M} \delta z_k \prod_{k=1}^{M-1} \delta y_k, \tag{3.8}$$

[6] G. Rosen, Phys. Fluids **3**, 519 (1960).

where $\Delta t = t_k - t_{k-1} = (t_M - t_0)/M$ and $t_M = t$, $y_M = y$.

To achieve the other requirement for the general solution we notice from (3.5) that

$$K_{\Delta t}(0/y') = \delta[y']. \tag{3.9}$$

With the aid of this and (3.7), our general solution (2.2) has the property that

$$\begin{aligned}\psi(0, t) &= \int_A K(0, t/y_0, t_0)\psi(y_0, t_0)\, \delta y_0 \\ &= \psi(0, t_0) = 1\end{aligned} \tag{3.10}$$

according to the initial condition.

Thus, (2.2) with (3.8) provides a general solution for the state functional, and F_s can be related to ψ by

$$F_s(x_1, \cdots, x_s, t) = \frac{\delta^s}{i^s\, \delta y(x_1) \cdots \delta y(x_s)} \psi(y, t)\Big|_{y=0} \tag{3.11}$$

according to (2.5). As described in the Appendix, a functional integral itself can be systematically computed. Our main concern may be in economy and accuracy in practically calculating (3.11) by a high-speed computer; with the Monte Carlo schemes this depends on many factors and is now under the author's investigation. The limit operation in (3.8), $\Delta t \to 0$, may be omitted on the approximate basis. Therefore, there is no reason that the solution we just obtained should remain of formalistic value only. In order to evade the expression in an infinitely repeated functional integral, we may introduce the space A' composed of functions with the X cross time argument,[6,7] a kind of the path-integral technique. The result is as follows:

$$\begin{aligned}\psi(y, t) = \int_{A'} \int_{A', \int^t \eta\, d\tau = y} &\times \exp\left(\int_{t_0}^{t} d\tau \left[i \int_X dx \left\{ \eta(x, \tau)\zeta(x, \tau) \right.\right.\right. \\ &+ \left.\left.\left. \int^{\tau} \eta(x, \sigma)\, d\sigma\, Q\zeta(x, \tau) \right\} \right. \right. \\ &- \frac{1}{2} \int_X \int_X dx\, dx' \int^{\tau} \eta(x, \sigma)\, d\sigma \int^{\tau} \eta(x', \sigma)\, d\sigma \\ &\left.\left. \times [\phi(|q - q'|); \zeta(x, \tau)\zeta(x', \tau)] \right] \right) \\ &\times \psi\left(\int^{t_0} \eta(x, \tau)\, d\tau, t_0\right) \delta\eta\, \delta\zeta,\end{aligned} \tag{3.12}$$

where $\eta, \zeta \in A'$, of which the time argument belongs to $[t_0, t] \equiv T$, A' as well as integration over it are defined by replacing X by XT in the Appendix, and the equation subscribed with the integral sign $\int_{A'}$ restricts the domain of integration with respect to η into the subset of A' which satisfies it. This expression is obtained from the previous one by the replacement

$$\zeta(x, t_k) = z_k(x), \tag{3.13}$$

$$\int^{t_k} \eta(x, \tau)\, d\tau = y_k(x). \tag{3.14}$$

It is remarkable that (3.12) has no such extra normalization factor as usually needed in the path-integral expression.[6]

The simplest type of initial state functionals which should satisfy (2.10) and (2.11) may be given by the type of (2.8) provided $\int_X F_1(x, t_0)\, dx/V = 1$. This appears to present only the case where all particles are independent, but after the initial time the correlation of particles does play a role. This is owing to the existence of the second term in the integral of the exponent in (3.8), as is known by comparison with (4.13)–(4.15). This term is considered to be in charge of the effect of short-range collision. In particular, as the state tends to approach the possible steady state, a particular initial condition is meaningless.

4. TURBULENT VLASOV FIELD APPROXIMATION

Let us consider the case where the interaction energy of particles is relatively small in each s-body Hamiltonian H_s ($s \geq 2$), so that ϕ can be factored by the small parameter ϵ. In this case, we can expand the factor with the quadratic term in y in (3.8) in powers of ϵ, and the first approximation is obtained by neglecting all but the first term, unity. That is,

$$\psi(y, t) = \lim_{\Delta t \to 0} \int_A \cdots \int_A \exp\left(i \int_X dx \sum_{k=1}^{M} [z_k(x)\{y_k(x) - y_{k-1}(x)\} + \Delta t\, y_k(x) Q z_k(x)] \right) \prod_{k=1}^{M} \delta z_k \prod_{k=1}^{M-1} \delta y_k\, \psi(y_0, t_0)\, \delta y_0$$

[7] I. Hosokawa, National Aerospace Laboratory (Tokyo) Report No. TR-103T (1966).

$$= \lim_{\Delta t \to 0} \int_A \cdots \int_A \exp\left(i \int_X dx \left[\sum_{k=1}^{M-1} y_k(x)\{z_k(x) + \Delta t\, Qz_k(x) - z_{k+1}(x)\} \right.\right.$$
$$\left.\left. + y(x)\{z_M(x) + \Delta t\, Qz_M(x)\} - z_1(x)y_0(x) \right] \right) \prod_{k=1}^{M-1} \delta y_k \prod_{k=1}^{M} \delta z_k\, \psi(y_0, t_0)\, \delta y_0$$

$$= \lim_{\Delta t \to 0} \int_A \cdots \int_A \prod_{k=1}^{M-1} \delta[z_{k+1} - (1 + \Delta t\, Q)z_k]$$
$$\times \exp\left[i \int_X dx\, \{y(1 + \Delta t\, Q)z_M - z_1 y_0\} \right] \prod_{k=1}^{M} \delta z_k\, \psi(y_0, t_0)\, \delta y_0$$

$$= \int_A \int_A \exp\left[i \int_X dx\, \{y(x)T^{t-t_0}z(x) - y_0(x)z(x)\} \right] \delta z\, \psi(y_0, t_0)\, \delta y_0, \tag{4.1}$$

where

$$T^{t-t_0}z = \lim_{\Delta t \to 0} (1 + \Delta t\, Q)^M z. \tag{4.2}$$

Obviously the mapping T^{t-t_0} gives the field $z(x)$ a time development governed by the equation

$$\partial z/\partial t = Qz. \tag{4.3}$$

On account of (3.6), this is rewritten as

$$\partial z/\partial t + (p/m)\, \partial z/\partial q - (\partial U/\partial q)\, \partial z/\partial p$$
$$= n \int_X dx' \frac{\partial \phi}{\partial q'} \frac{\partial z(x')}{\partial p'} z(x), \tag{4.4}$$

where

$$U = U_E(q) + \int_X n\phi(|q - q'|)z(x')\, dx'. \tag{4.5}$$

U_E being the potential of the external force. If

$$z(x') = 0 \quad \text{at} \quad |p'| = \infty, \tag{4.6}$$

the right-hand side of (4.4) vanishes, and then (4.4)–(4.6) constitute the Vlasov self-consistent field equation.[3]

Now, in comparison with the theory of Hopf,[4] (4.1) can formally be considered as the characteristic functional for the turbulent (stochastic) field $z(x)$. In fact, (4.1) is seen to be in complete agreement with his starting formula for the characteristic functional if we understand the Fourier transformation of functionals on the basis of (A18)–(A19).[5]

The other conditions necessary for a functional to be the characteristic functional are as follows:

$$\varphi(0, t) = 1, \quad \varphi^*(y, t) = \varphi(-y, t), \quad |\varphi(y, t)| \leq 1, \tag{4.7}$$

where * denotes the conjugate complex. They all are derived from the equation

$$\varphi(y, t) = \int_A \exp\left(i \int_X dx\, yz \right) \rho(z, t)\, \delta z, \tag{4.8}$$

in which $\rho \delta z$ should be a (differential) probability measure. The first two conditions in (4.7) are common to our state functionals [cf. (2.5)]. The other is interesting in that it requires

$$||\varphi(y, t)|| = 1, \tag{4.9}$$

which can coexist with (2.17). However, for the characteristic functional to be considered as a state functional, it should obey the further condition (2.11). In terms of (4.8), it can be rewritten as

$$\int_A \exp\left(i \int_X dx\, yz \right) \rho\, \delta z$$
$$= \lim_{V \to \infty} \int_X \int_A z(x) \exp\left(i \int_X dx\, yz \right) \rho\, \delta z\, dx/V,$$
$$\therefore \lim_{V \to \infty} \int_X z(x)\, dx/V = 1. \tag{4.10}$$

Thus, a characteristic functional φ for the turbulence of the Vlasov field associated with (4.10) can be a state functional within the present approximation. Then, $F_s(x_1, \cdots, x_s, t)$ is approximated by the correlation tensor

$$F_s(x_1, \cdots, x_s, t) \cong \langle z(x_1) \cdots z(x_s) \rangle_t$$
$$\equiv \int_A z(x_1) \cdots z(x_s) \rho(z, t)\, \delta z \tag{4.11}$$

according to (3.11). In order that $F_s \geq 0$, it is sufficient that

$$z(x) \geq 0, \tag{4.12}$$

but the usual Vlasov field obeys this condition well. It is noted that (4.10) and (4.12) considerably restrict the type of the measure factor ρ; we have

[5] I. Hosokawa, National Aerospace Laboratory (Toyko) Report No. TR-75 (1964).

to give exactly zero measure to the $z(x)$, which violates either condition.

Let us take (2.8) as the initial state functional in (4.1). Then,

$$\psi(y, t) = \int_A \exp\left\{i \int_X dx\, y(x) T^{t-t_0} z(x)\right\} \times \delta[z(x) - F_1(x, t_0)]\, \delta z = \exp\left\{i \int_X dx\, y(x) T^{t-t_0} F_1(x, t_0)\right\}. \tag{4.13}$$

Comparing this with (4.8), we obtain

$$\rho[z, t] = \delta[z(x) - T^{t-t_0} F_1(x, t_0)], \tag{4.14}$$

which shows that (4.13) corresponds to the (not turbulent but) *laminar* field exactly developed from $F_1(x, t_0)$ subject to the Vlasov kinetic equations (4.4)–(4.6). In this case, (4.11) becomes

$$F_s(x_1, \cdots, x_s, t) \cong T^{t-t_0} F_1(x_1, t_0) \cdots T^{t-t_0} F_1(x_s, t_0), \tag{4.15}$$

which is the same as (2.7); so that not only $F_1(x, t)$ but also any distribution function with $s \geq 2$ can be approximately calculated without dealing with any functional integration. From a practical point of view, this fact seems to be very advantageous. Indeed, it is a lucky result due to the twofold action of the weak interaction assumption and the initial condition just adopted. With another initial condition, we have, more or less, turbulence of the Vlasov field. But if the turbulence so obtained is weak, the laminar field approach would be a good approximation. This is analogous to the situation in fluid dynamics that the laminar Navier–Stokes equation can considerably well describe a real fluid flow, while a real fluid includes more or less turbulence. However, for a state with a strong turbulence in which (4.15) fails, the turbulent Vlasov field approximation is more general.

5. CONCLUSION

We modified Bogoliubov's formalism of non-equilibrium statistical mechanics to gain a more appropriate formalism, in which the state functional is defined and the basic equation governing it is solvable by means of functional integration. The state functional embraces all generic distribution functions as the coefficient functions in the functional Taylor expansion (except for the numerical factors). In the case where particles interact weakly, the state functional may be approximated by the characteristic functional for the turbulent Vlasov field. Our general solution for the state functional is given in terms of functional integrals, so that in order to make our solution fully useful for a general practical problem in statistical mechanics, we need the Monte Carlo approach to functional integrals by the use of a high-speed computer. The knowledge given in the Appendix is expected to be working for such a numerical research. Our formulation was given for the system of identical particles governed by classical mechanics. There is no difficulty in extending this to a multicomponent particle system.

Note added in proof: When we have m kinds of particles in the system, the state functional is defined as

$$\psi(y_1, \cdots, y_m, t) = 1 + i \sum_{a=1}^{m} \int_X F_a(x_1, t) y_a(x_1)\, dx_1 + \frac{i^2}{2!} \sum_{a,b=1}^{m} \int_X \int_X F_{ab}(x_1, x_2, t) \times y_a(x_1) y_b(x_2)\, dx_1\, dx_2 + \cdots, \tag{5.1}$$

where $F_{\underbrace{ab\ldots}_{k}}(x_1, \cdots, x_k, t)$ is the k-body *generic* distribution function for the particles specified by $a, b, \cdots$, which should be symmetric with respect to an interchange of $(a, x_1), (b, x_2), \cdots$. Introducing the one-body Hamiltonian H_a^1 for the a particle and the interaction potential $\phi_{ab}(q_1, q_2) = \phi_{ba}(q_2, q_1)$ between the a and b particles, we can establish the basic functional-differential equation for ψ as

$$\frac{\partial \psi}{\partial t} = i \sum_a \int_X y_a(x) \left[H_a^1(x); \frac{\delta \psi}{i\, \delta y_a(x)}\right] dx - \tfrac{1}{2} \sum_{a,b} \int_X \int_X \{y_a(x) y_b(x') - i n_b y_a(x) - i n_a y_b(x')\} \times \left[\phi_{ab}(q, q'); \frac{\delta^2 \psi}{i^2\, \delta y_a(x)\, \delta y_b(x')}\right] dx\, dx', \tag{5.2}$$

where n_a is the mean number density of the a particle. It is quite straight-forward to give a general solution of (5.2). Corresponding to (3.12), it may be formulated as

$$\psi = \int_{A'm} \int_{A'm, \int^t \eta_a d\tau = y_a} \exp\left(\int_{t_0}^{t} d\tau \left[i \int_X dx \times \sum_a \left\{\eta_a(x, \tau) \zeta_a(x, \tau) + \int^{\tau} \eta_a(x, \sigma)\, d\sigma\, Q_a \zeta_a(x, \tau)\right\} - \frac{1}{2} \int_X \int_X dx\, dx' \sum_{a,b} \int^{\tau} \eta_a(x, \sigma)\, d\sigma \int^{\tau} \eta_b(x', \sigma)\, d\sigma \times [\phi_{ab}(q, q'); \zeta_a(x, \tau) \zeta_b(x', \tau)]\right]\right) \times \psi\left(\int^{t_0} \eta_1\, d\tau, \cdots, \int^{t_0} \eta_m\, d\tau, t_0\right) \prod_a^m \delta\eta_a\, \delta\zeta_a, \tag{5.3}$$

with

$$Q_a z_a(x) = [H_a^1(x); z_a(x)] + \sum_b n_b \int_X dx' \, [\phi_{ab}(q, q'); z_a(x) z_b(x')]. \quad (5.4)$$

ACKNOWLEDGMENTS

The author is grateful to Professor T. Taniuti and other members of the Institute of Plasma Physics, Nagoya University, for valuable discussions on this work.

APPENDIX

We here try to fix the concept of our functional integration by a systematic method which can open the way to its numerical calculation. First, we introduce the set A of the functions defined by

$$y(x) = \sum_{i=1}^{\infty} a_i s_i(x), \quad (A1)$$

where $(a_1, a_2, \cdots, a_n, \cdots)$ indicates a point in the infinite-dimensional Euclidean space R^∞ and $\{s_i(x)\}$ makes a complete set of real orthonormal functions on the space X. If we define the inner product

$$(y_1, y_2) = \int_X y_1(x) y_2(x) \, dx = \sum_{i=1}^{\infty} a_{1i} a_{2i} \quad (A2)$$

and the norm

$$||y|| = (y, y) = \sum_{i=1}^{\infty} a_i^2, \quad (A3)$$

A is a linear norm space containing the Hilbert space. We call a mapping from A to the number space a functional.

Let us define an integration of the (complex-valued) functional $F(y)$ over the space A as

$$I_F^\mu = \int_{R^\infty} F\left[\sum_{i=1}^{\infty} a_i s_i(x)\right] \prod_{i=1}^{\infty} d\mu\,(a_i), \quad (A4)$$

where $\mu(a_i)$ is a certain measure of the space R subject to the condition that

$$\int_R d\mu\,(a_i) = 1. \quad (A5)$$

It is proved by Halmos[9] that the product $\prod_{i=1}^{\infty} d\mu(a_i)$ accompanied with (A5) makes itself a Lebesgue–Stieltjes measure of R^∞. Therefore if F is finite and measurable with the measure μ, it is (finitely) integrable with μ. There are typical examples of such a measure; i.e.,

$$\text{Gaussian measure: } d\mu(a_i) = e^{-\frac{1}{2}a_i^2} da_i/(2\pi)^{\frac{1}{2}}; \quad (A6)$$

$$\text{delta measure: } d\mu(a_i) = [(2\pi)^{\frac{1}{2}} \delta(a_i - a_{0i})] da_i/(2\pi)^{\frac{1}{2}}. \quad (A7)$$

If F is integrable, with the repeated use of Fubini's theorem factors $e^{-\frac{1}{2}a_i^2}$, $(2\pi)^{\frac{1}{2}}\delta(a_i - a_{0i})$, etc. may be absorbed in the integrand so as to make the new integrand F_1,

$$\begin{aligned} I_F^\mu = I(F_1) &= \int_{R^\infty} F\Big(\sum_{i=1}^{\infty} a_i s_i(x)\Big) f_n^\mu(a_1, a_2, \cdots, a_n) \prod_{i=1}^{n} [da_i/(2\pi)^{\frac{1}{2}}] \prod_{i=n+1}^{\infty} d\mu\,(a_i) \\ &= \lim_{n\to\infty} \int_{R^n} F_1\Big(\sum_{i=1}^{\infty} a_i s_i(x); a_1, a_2, \cdots, a_n, \cdots\Big) \prod_{i=1}^{n} [da_i/(2\pi)^{\frac{1}{2}}], \end{aligned} \quad (A8)$$

where f_∞^μ may be called a measure factor in the functional integral. Thus when F is integrable with some measure, there exists the finite limiting value of the n-dimensional-Euclidean volume integral of F_1 as $n \to \infty$, and if there exists such a value, let us define it as $I(F_1)$. This fact provides a promising way to numerically approach a functional integration with the help of a high-speed computer. A method of truncating an infinitely multiple integral to a finitely multiple one with a sufficiently exact result was given by Cameron,[10] who treats the Wiener integral. To make a technique like this for the present case is not difficult, with the use of the Monte Carlo method in mind.

The integral $I(F_1)$ defined by (A8), however, can be rewritten in a more useful form from the theoretical point of view. Let us consider $y(x) \in A$ as the limit of a step-function sequence,

$$y(x) = \lim_{\Delta x_k \to 0} \sum_k y_k c_k(x), \quad (A9)$$

where X was divided into many small hyperparallelepipeds each specified by a pair of vectors $(x_k, x_{k+1} = x_k + \Delta x_k)$ and $c_k(x)$ is the characteristic function such that it is 1 in the region $[x_k, x_{k+1})$ and otherwise vanishes. Comparison of (A1) and (A9) suggests the linear transformation of variables from $(a_1, a_2, \cdots, a_n, \cdots)$ to $(y_1, y_2, \cdots, y_n, \cdots)$ or its equivalent.

If we start from the n-dimensional approximation, we have the following n equations:

[9] P. R. Halmos, *Measure Theory* (D. Van Nostrand Co., Inc., Princeton, N. J., 1950), p. 154.

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

$$\sum_{i=1}^{n} a_i s_i(x_k') = \sum_{k=1}^{n} y_k c_k(x_k'), \quad x_k \le x_k' < x_{k+1}. \tag{A10}$$

Here we divided X into n cells. Multiplied by $s_i(x_k')\Delta x_k$ and summed up with respect to k, they lead to

$$a_i + \sum_{i=1}^{n} \epsilon_{ij} a_i = \sum_{k=1}^{n} y_k s_i(x_k')\, \Delta x_k, \tag{A11}$$

where $\epsilon_{ij} \to 0$ as $n \to \infty$ $(\Delta x_k \to 0)$, on account of the orthonormality of $\{s_i(x)\}$. Hence

$$\partial a_i/\partial y_k(\Delta x_k)^{\frac{1}{2}} = s_i(x_k')(\Delta x_k)^{\frac{1}{2}} + O(\epsilon) \equiv t_{ik}, \tag{A12}$$

which may become the elements of the transformation matrix under consideration. Now,

$$\begin{aligned} \sum_{k=1}^{n} t_{ik} t_{jk} &= \sum_{k=1}^{n} s_i(x_k') s_j(x_k')\, \Delta x_k + O(\epsilon) \\ &= \delta_{ij} + \epsilon_{ij} + O(\epsilon) \\ &= \delta_{ij} + O(\epsilon), \end{aligned} \tag{A13}$$

where δ_{ij} is the Kronecker delta. This shows that the transformation from $(a_1, a_2, \cdots, a_n)$ to $[y_1(\Delta x_1)^{\frac{1}{2}}, y_2(\Delta x_2)^{\frac{1}{2}}, \cdots, y_n(\Delta x_n)^{\frac{1}{2}}]$ becomes orthogonal or represents a particular rotation of the Euclidean space about the origin in the limit $n \to \infty$ $(\Delta x_k \to 0)$, when the equality of (A1) and (A9) is not approximate but exact.

It is easy to see that the Jacobian of this transformation is unity; $|(t_{ij})| = |(\sum_{k=1}^{n} t_{ik} t_{jk})|^{\frac{1}{2}} = 1$ from (A13). Thus (A8) may be re-expressed as

$$I(F_1) = \lim_{n\to\infty} \int_{R^n} F_1(y(x); a_1, a_2, \cdots, a_n, \cdots) \times \prod_{k=1}^{n} d[y_k(\Delta x_k/2\pi)^{\frac{1}{2}}]. \tag{A14}$$

Since a_i is given by $\int_X y(x)s_i(x)\, dx$, F_1 is perfectly a functional of $y(x)$ so that $a_1, a_2, \cdots, a_n, \cdots$ may be omitted in (A14). It is to be noted here that there are many different ways to take a sequence of partitions which divide the X space into small cells, but one of them (any one) should have been specified while considering the particular transformation (rotation) described above; in some case the simplest one may be to always divide X into equal cells so that $\Delta x_k \equiv \Delta x$. But the present result insists that *the functional integral of the type of* (A14) *is uniquely determined irrespective of the ways to take a partition sequence and equal to* (A8) *by any way.* In other words, a particular way to take a partition sequence corresponds to a particular rotation of the Euclidean space, to which the value of $I(F)$ should naturally be invariant. For simplicity, the notation

$$I(F_1) = \int_A F_1(y)\, \delta y \tag{A15}$$

is often used. Since to analyze $y(x)$ by another complete orthonormal set rather than $\{s_i(x)\}$ corresponds to a particular rotation of the space, $I(F_1)$ is invariant also to the way of taking the orthonormal set.

The measure factor f_∞^μ derived from (A6) and (A7) may be rewritten in the following forms:

$$\lim_{n\to\infty} \prod_{i=1}^{n} e^{-\frac{1}{2}a_i^2} = \exp\left[-\frac{1}{2}\int_X y(x)^2\, dx\right]; \tag{A16}$$

$$\begin{aligned} &\lim_{n\to\infty} \prod_{i=1}^{n} \{(2\pi)^{\frac{1}{2}}\, \delta(a_i - a_{0i})\} \\ &\quad = \lim_{n\to\infty} \prod_{k=1}^{n} \{(2\pi)^{\frac{1}{2}}\, \delta[y_k(\Delta x_k)^{\frac{1}{2}} - y_{0k}(\Delta x_k)^{\frac{1}{2}}]\} \\ &\quad \text{(by orthogonality)} \\ &\quad = \lim_{n\to\infty} \int_{R^n} \exp\left\{i \sum_{k=1}^{n} (y_k - y_{0k}) z_k\, \Delta x_k\right\} \\ &\qquad\qquad \times \prod_{k=1}^{n} d[z_k(\Delta x_k/2\pi)^{\frac{1}{2}}] \\ &\quad = \int_A \exp\left[i \int_X \{y(x) - y_0(x)\} z(x)\, dx\right] \delta z \\ &\quad \equiv \delta[y - y_0], \end{aligned} \tag{A17}$$

where, of course,

$$y_0(x) = \sum_{i=1}^{\infty} a_{0i} s_i(x) = \lim_{n\to\infty} \sum_{k=1}^{n} y_{0k} c_k(x).$$

$\delta[\ ]$ may be called a delta functional, which plays a role analogous to a delta function for functionals; a delta functional is symmetric with the argument just like a delta function. When F_1 has the measure factor (A16), (A14) reproduces the functional integral with the Gaussian measure established by Friedrichs.[11]

In many cases of analyzing functionals, it is useful to introduce the *functional* Fourier transformation. Let us see

$$\psi(y) = \int_A \chi(u) \exp\left[i \int_X y(x)u(x)\, dx\right] \delta u. \tag{A18}$$

If $\int_A |\chi(u)|\, \delta u < \infty$, $\psi(y)$ is meaningful and on account of (A17)

$$\begin{aligned} \int_A \psi(y) \exp\left[-i \int_X y(x)v(x)\, dx\right] \delta y &= \int_A \chi(u)\, \delta[u - v]\, \delta u \\ &= \chi(v). \end{aligned} \tag{A19}$$

[11] K. O. Friedrichs and H. N. Shapiro, lecture notes, Courant Institute of Mathematical Sciences, New York University (1957).

Series Representation Method for Obtaining the Emission Coefficient from the Integrated Intensity Distribution for Asymmetrical Light Sources*

C. D. MALDONADO
Northrop Space Laboratories, Hawthorne, California
(Received 9 February 1966)

A procedure for inverting the spectroscopic integral equation which relates the emission coefficient to the integrated intensity distribution for an optically thin and asymmetrical light source is presented. In this procedure the emission coefficient is expanded in terms of a complete set of orthogonal polynomials which are "invariant in form" to a rotation of axes and use is made of the spectroscopic integral equation to determine the unknown expansion coefficients in terms of known information for the integrated intensity distribution. In order to check its validity, the resultant series representation was summed exactly for a hypothetical example corresponding to an asymmetrical source whose emission coefficient possesses elliptical symmetry and diminishes with position in a Gaussian manner. Also the same hypothetical example was used to test the accuracy of the numerical procedure which was developed for summing the series representation in practical situations where the known information on the integrated intensity distribution is presented in the form of experimental data.

I. INTRODUCTION

KNOWLEDGE of the spatial distribution of the emission coefficient for a plasma in local thermodynamic equilibrium (LTE) is important because the temperature and particle density profiles are related to this quantity. For an optically thin plasma, this information is measured indirectly in the form of an integrated intensity distribution which is obtained by integrating the emission coefficient along the line of sight in the y direction for a sequence of detector positions along the x axis; i.e.,

$$g(\xi, x) = \int_{-\infty}^{+\infty} dy\, f(x', y'), \qquad (1.1)$$

where $g(\xi, x)$ is the measured integrated intensity distribution, $f(x', y')$ is the emission coefficient, and ξ denotes the angle of rotation of the source coordinate system (x', y') relative to the fixed laboratory coordinate system (x, y) as shown in Fig. 1. Up to the present time the problem of inverting this integral equation for the special case of circularly symmetric light sources has been treated quite thoroughly in the literature,[1-7] while Freeman and Katz[8] have presented an approximate method for the case of light sources possessing slight asymmetry. Therefore, it is the purpose of this paper to present a mathematical procedure for inverting Eq. (1.1), which is sufficiently general, so as to account for light sources of arbitrary symmetry and spatial extent as well.

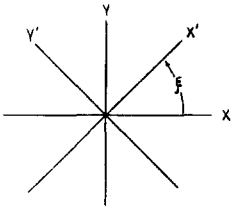

FIG. 1. Diagram of source coordinate system (x', y') rotated by an angle ξ relative to the fixed laboratory coordinate system (x, y).

The mathematical procedure is a straightforward one in which the emission coefficient is expanded in terms of a complete orthonormal set of functions and use is made of Eq. (1.1) to evaluate the unknown expansion coefficients in terms of the integrated intensity distribution $g(\xi, x)$. Obviously, there are an infinity of complete orthonormal sets which one can use for this purpose. However, the choice of any particular set must be done with care in order to avoid dealing with a cumbersome set of algebraic equations for the evaluation of the unknown coefficients. The important contribution of this paper is that we were able to circumvent this difficulty by expanding in terms of a set of complete orthogonal functions[9] which allowed us to obtain a generalized Fourier expression for evaluating the unknown expansion coefficients from known information on the integrated intensity distribution, $g(\xi, x)$. The details of this procedure and the resultant series representation for the emission coefficient are given in Sec. II.

* This work was supported by the Northrop Independent Research Funds.

[1] M. P. Freeman and S. Katz, J. Opt. Soc. Am. 50, 826 (1960).
[2] K. Bockasten, J. Opt. Soc. Am. 51, 943 (1961).
[3] W. L. Barr, J. Opt. Soc. Am. 52, 885 (1962).
[4] O. H. Nestor and H. N. Olsen, S.I.A.M. Rev. 2, 200 (1963).
[5] D. R. Paquette and W. L. Wiese, Appl. Opt. 3, 291 (1964).
[6] R. G. Buser and J. J. Kainz, J. Opt. Soc. Am. 55, 12 (1965).
[7] W. G. Braun, Rev. Sci. Instr. 36, 802 (1965).
[8] M. P. Freeman and S. Katz, J. Opt. Soc. Am. 53, 1172 (1963).
[9] C. D. Maldonado, J. Math. Phys. 6, 1935 (1965).

The validity of this series representation is checked in Sec. III by means of a hypothetical example for an asymmetrical source whose emission coefficient possesses elliptical symmetry and diminishes with position in a Gaussian manner. Finally, a numerical procedure for summing the series representation for the emission coefficient is presented in Sec. IV and applied to the hypothetical example of Sec. III.

II. SERIES REPRESENTATION

For most laboratory plasma light sources it is reasonable to assume that the emission coefficient is a squared integrable function on the entire two-dimensional plane and hence may be expanded in terms of a complete set of functions. As stated previously, there are an infinity of complete sets which span the space of squared integrable functions on the entire plane that could be used for this purpose; however, some care must be given to the choice of any particular set in order to avoid dealing with a cumbersome system of equations for evaluating the unknown expansion coefficients. A set of functions which allowed us to bypass this difficulty was constructed[9] and in terms of the elements of this set the emission coefficient can be expanded as

$$f(x', y') = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \epsilon_m [C^m_{m+2k}(\alpha) U^m_{m+2k}(\alpha x', \alpha y') + C^{-m}_{m+2k}(\alpha) U^{-m}_{m+2k}(\alpha x', \alpha y')] \exp[-\alpha^2(x'^2 + y'^2)], \tag{2.1}$$

where the symbol ϵ_m is equal to $\frac{1}{2}$ for $m = 0$ and 1 for $m = 1, 2, 3, \cdots$; α is a scale factor; $C^{\pm m}_{m+2k}(\alpha)$ are the unknown expansion coefficients; and $U^{\pm m}_{m+2k}(\alpha x', \alpha y')$ are polynomials which are "invariant in form" to a rotation of axes[9] and are orthonormal with respect to the Gaussian weight function $\exp[-\alpha^2(x'^2 + y'^2)]$, i.e.,

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'\, dy'\, U^m_{m+2k}(\alpha x', \alpha y') \tilde{U}^r_{r+2\beta}(\alpha x', \alpha y') \times \exp[-\alpha^2(x'^2 + y'^2)] = \delta^{r+2\beta}_{m+2k} \tag{2.2}$$

with $\sim$ denoting complex conjugate. The orthogonal polynomials $U^{\pm m}_{m+2k}(\alpha x', \alpha y')$ are given explicitly by[9]

$$U^{\pm m}_{m+2k}(\alpha x', \alpha y') = (-1)^k(\alpha/\pi^{\frac{1}{2}})[k!/(m+k)!]^{\frac{1}{2}} \times [\alpha^2(x'^2 + y'^2)]^{m/2} \exp(\pm im\varphi) L^m_k[\alpha^2(x'^2 + y'^2)], \tag{2.3}$$

where in this expression

$$L^m_k[\alpha^2(x'^2 + y'^2)] = \sum_{s=0}^{k} (-1)^s[(m+k)!/(k-s)!\,(m+s)!\,s!] \times [\alpha^2(x'^2 + y'^2)]^s \tag{2.4}$$

are the associated Laguerre polynomials[10] of argument $\alpha^2(x'^2 + y'^2)$, and $\varphi = \tan^{-1}(y'/x')$.

Now the orthonormal condition for $U^{\pm m}_{m+2k}(\alpha x', \alpha y')$ given by Eq. (2.2) allows us to obtain from Eq. (2.1) the following generalized Fourier expression:

$$C^{\pm m}_{m+2k}(\alpha) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'\, dy'\, f(x', y') \times \tilde{U}^{\pm m}_{m+2k}(\alpha x', \alpha y') \tag{2.5}$$

for the unknown expansion coefficients provided the emission coefficient is a known quantity. However, in actual practice the emission coefficient is an unknown quantity and the only known information is a series of spectrums for the integrated intensity distribution corresponding to different angles of rotation of the source relative to the fixed laboratory coordinate system. For this situation use must be made of Eq. (1.1) in order to infer information about the unknown expansion coefficients.

To carry out this alternate procedure for obtaining the unknown expansion coefficients, Eq. (2.1) is first introduced into Eq. (1.1), then subsequent use is made of the integral given by Eq. (A13) of Appendix A to write, for the integrated intensity distribution, the following series representation:

$$g(\xi, x) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \epsilon_m \left[\frac{1}{k!\,(m+k)!}\right]^{\frac{1}{2}} \left[\frac{1}{2^{(m+2k)}}\right] \times [C^m_{m+2k}(\alpha) \exp(im\xi) + C^{-m}_{m+2k}(\alpha) \exp(-im\xi)] \times H_{m+2k}(\alpha x) \exp(-\alpha^2 x^2), \tag{2.6}$$

where

$$H_{m+2k}(\alpha x) = \sum_{t=0}^{[\frac{1}{2}(m+2k)]} (-1)^t \left\{\frac{(m+2k)!}{t!\,[m+2(k-t)]!}\right\} (2\alpha x)^{m+2(k-t)} \tag{2.7}$$

are Hermite polynomials[10] of order $m + 2k$ with the upper limit $[\frac{1}{2}(m+2k)]$ being equal to $\frac{1}{2}(m+2k)$ or $\frac{1}{2}(m+2k-1)$ according as m is even or odd. Now, the orthogonality conditions for the Hermite polynomials and harmonic functions given, respectively, by

[10] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, in *Higher Transcendental Functions* (McGraw-Hill Book Company, Inc., New York, 1963), Vols. I and II.

$$(\alpha/\pi^{\frac{1}{2}})\left[\frac{1}{(m+2k)!\,(\nu+2\beta)!\,2^{m+2k}2^{\nu+2\beta}}\right]^{\frac{1}{2}} \times \int_{-\infty}^{+\infty} dx\, H_{m+2k}(\alpha x)H_{\nu+2\beta}(\alpha x)\exp(-\alpha^2x^2) = \delta_{m+2k}^{\nu+2\beta} \tag{2.8}$$

and

$$\left(\frac{1}{2\pi}\right)\int_{-\pi}^{\pi} d\xi \exp(\pm im\xi)\exp(\mp i\nu\xi) = \delta_m^{\nu}, \tag{2.9}$$

allow us to obtain from Eq. (2.6)

$$C_{m+2k}^{\pm m}(\alpha) = \left(\frac{\alpha}{2\pi^{\frac{1}{2}}}\right)\left[\frac{(k!\,(m+k)!)^{\frac{1}{2}}}{(m+2k)!}\right] \times \int_{-\pi}^{\pi} d\xi \exp(\mp im\xi)\int_{-\infty}^{+\infty} dx\, g(\xi, x)H_{m+2k}(\alpha x) \tag{2.10}$$

as the generalized Fourier expression for the unknown expansion coefficients. In Appendix B we show that Eqs. (2.5) and (2.10) for the unknown expansion coefficients, despite their apparently dissimilar appearance, do yield equivalent results for the hypothetical example of Sec. III.

Finally on combining Eqs. (2.1) and (2.10) we arrive at the important result of this paper, that is, the series representation for the emission coefficient of a completely asymmetrical source,

$$f(x', y') = \left(\frac{\alpha}{\pi^{\frac{1}{2}}}\right)\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\epsilon_m\left\{\frac{[k!\,(m+k)!]^{\frac{1}{2}}}{(m+2k)!}\right\} \times \mathrm{Re}\left\{\left[\int_{-\pi}^{\pi} d\xi \exp(-im\xi) \times \int_{-\infty}^{+\infty} dx\, g(\xi, x)H_{m+2k}(\alpha x)\right]U_{m+2k}^{m}(\alpha x', \alpha y')\right\} \times \exp[-\alpha^2(x'^2+y'^2)], \tag{2.11}$$

where Re { } denotes the real part of { }. In writing Eq. (2.11) use has been made of the properties $\tilde{U}_{m+2k}^{\pm m}(\alpha x', \alpha y') = U_{m+2k}^{\mp m}(\alpha x', \alpha y')$ and $\tilde{C}_{m+2k}^{\pm m}(\alpha) = C_{m+2k}^{\mp m}(\alpha)$, where the former follows from Eq. (2.3) and the latter from Eq. (2.10) since $g(\xi, x)$ and $H_{m+2k}(\alpha x)$ are real quantities.

III. HYPOTHETICAL EXAMPLE

A mathematical procedure for summing the series representation of Eq. (2.11) corresponding to an asymmetrical source whose emission coefficient possesses elliptical symmetry and diminishes with position in a Gaussian manner as

$$f(x', y') = B\exp[-(\lambda^2x'^2 + \beta^2y'^2)], \tag{3.1}$$

with the amplitude B and parameters λ and β being fixed quantities, are presented here. The ultimate goal of this procedure is to check the validity of the series representation for the emission coefficient by showing that the closed form expression of Eq. (3.1) is recovered when Eq. (2.11) is summed exactly.

The first step in this procedure is to obtain the integrated intensity distribution corresponding to this assumed form for the emission coefficient as given by Eq. (3.1). This one can readily do by first transforming Eq. (3.1) to the unprimed coordinate system by means of the transformation, $x' = x\cos\xi + y\sin\xi$ and $y' = -x\sin\xi + y\cos\xi$, then introducing the resultant expression for $f(x', y')$ into Eq. (1.1) and using a tabulated result[11] to evaluate the resultant integral. This leads to a spectrum for $g(\xi, x)$ of the form

$$g(\xi, x) = B[\pi/(\lambda^2\sin^2\xi + \beta^2\cos^2\xi)]^{\frac{1}{2}} \times \exp[-\beta^2\lambda^2x^2/(\lambda^2\sin^2\xi + \beta^2\cos^2\xi)], \tag{3.2}$$

which is a Gaussian whose amplitude and spread is a function of the angle of rotation ξ.

Next, Eqs. (2.3) and (3.2) are incorporated into Eq. (2.11) and use was made of the property that $g(\xi, x)$ is an even function with respect to ξ, that is $g(\xi, x) = g(-\xi, x)$, to rewrite Eq. (2.11) as

$$f(x', y') = \left(\frac{\alpha^2}{\pi^{\frac{1}{2}}\lambda}\right)B\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\epsilon_m(-1)^k\left[\frac{k!}{(m+2k)!}\right]\left[\int_{-\pi}^{\pi} d\xi\,[\cos(m\xi)/(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)^{\frac{1}{2}}] \times \int_{-\infty}^{+\infty} dx \exp[-\beta^2x^2/(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)]H_{m+2k}(\alpha x)\right] \times \cos(m\varphi)[\alpha^2(x'^2+y'^2)]^{\frac{1}{2}m}L_k^m[\alpha^2(x'^2+y'^2)]\exp[-\alpha^2(x'^2+y'^2)]. \tag{3.3}$$

The second integral in the above expression vanishes for odd values of m; a result which follows from the fact that the spectrum for $g(\xi, x)$ is also an even function of x while $H_{m+2k}(\alpha x)$ is an odd function of x for odd values of m. For the even values of m the explicit evaluation of the double integral in Eq.

[11] W. Gröbner and N. Hofreiter, *Intergraltafel erster teil Unbestimmte Integrale* (Springer-Verlag, Vienna, 1961).

(3.3) with α set equal to β follows directly from Eqs. (B9) and (B13) of Appendix B and this yields

$$\int_{-\pi}^{\pi} d\xi \left[\frac{\cos(2p\xi)}{(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)^{\frac{1}{2}}}\right] \int_{-\infty}^{+\infty} dx \exp\left[-\frac{\beta^2 x^2}{(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)}\right] H_{2(p+k)}(\beta x)$$
$$= \left(\frac{2\pi^{\frac{1}{2}}}{\beta}\right)\left[\frac{(2p+2k)!\,(2p+2k)!}{k!\,(p+k)!\,(2p+k)!}\right]\left[\left(\frac{\beta}{2\lambda}\right)^2 - \left(\frac{1}{2}\right)^2\right]^{p+k}, \quad (3.4)$$

with m equal to $2p$ for $p = 0, 1, 2, 3, \cdots$. Now, if we set α equal to β, replace m by $2p$, and introduce Eq. (3.4) for the double integral in Eq. (3.3), we obtain

$$f(x', y') = 2B\left(\frac{\beta}{\lambda}\right)\sum_{p=0}^{\infty}\epsilon_p\left(\frac{\beta}{2\lambda}\right)^{2p}[(\lambda^2 - \beta^2)(x'^2 + y'^2)]^p \cos(2p\varphi)$$
$$\times \sum_{k=0}^{\infty}\left[\frac{(2p+2k)!}{(2p+k)!\,(p+k)!\,2^{2k}}\right]\left[1 - \left(\frac{\beta}{\lambda}\right)^2\right]^k L_k^{2p}[\beta^2(x'^2 + y'^2)] \exp[-\beta^2(x'^2 + y'^2)] \quad (3.5)$$

as the desired series representation for the emission coefficient which we wish to sum exactly.

There is no unique way to sum the series representation of Eq. (3.5) and the one we have chosen for this purpose makes use of the expansion[10]

$$M\left[a, c; \frac{zt}{(t-1)}\right] = (1-t)^a\left[\frac{\Gamma(c)}{\Gamma(a)}\right] \times \sum_{k=0}^{\infty}\left[\frac{\Gamma(a+k)}{\Gamma(c+k)}\right] t^k L_k^{c-1}(z) \quad (3.6)$$

for the confluent hypergeometric function, which is valid for $c > 0$, $|t| < 1$, and $z > 0$ to write the infinite series over the k index in Eq. (3.5) more compactly as

$$\sum_{k=0}^{\infty}\left[\frac{(2p+2k)!}{(2p+k)!\,(p+k)!\,2^{2k}}\right]\left[1 - \left(\frac{\beta}{\lambda}\right)^2\right]^k L_k^{2p}[\beta^2(x'^2 + y'^2)]$$
$$= \left[\frac{(\lambda/\beta)^{2p+1}}{p!}\right] M[p + \tfrac{1}{2}, 2p + 1; (\beta^2 - \lambda^2)(x'^2 + y'^2)], \quad (3.7)$$

where in the latter expression $(\beta/\lambda)^2$ is restricted to the range $0 < (\beta/\lambda)^2 < 2$ and $\beta^2(x'^2 + y'2) > 0$. For the set of parameters $a = p + \frac{1}{2}$ and $c = 2p + 1$, the confluent hypergeometric function of Eq. (3.7) is related to the modified Bessel's function of the first kind through the relationship[10]

$$M[p + \tfrac{1}{2}, 2p + 1; (\beta^2 - \lambda^2)(x'^2 + y'^2)]$$
$$= [p!\,2^{2p}/(\beta^2 - \lambda^2)^p(x'^2 + y'^2)^p]$$
$$\times I_p[\tfrac{1}{2}(\beta^2 - \lambda^2)(x'^2 + y'^2)]$$
$$\times \exp[\tfrac{1}{2}(\beta^2 - \lambda^2)(x'^2 + y'^2)]. \quad (3.8)$$

Use of Eqs. (3.7) and (3.8) to replace the infinite series over the k index in Eq. (3.5) yields the result

$$f(x', y') = 2B\sum_{p=0}^{\infty}\epsilon_p \cos(2p\varphi)$$
$$\times I_p[\tfrac{1}{2}(\beta^2 - \lambda^2)(x'^2 + y'^2)]$$
$$\times \exp[-\tfrac{1}{2}(\beta^2 + \lambda^2)(x'^2 + y'^2)], \quad (3.9)$$

which we can rewrite as

$$f(x', y') = B \exp(-\{\beta^2[1 - \cos(2\varphi)]$$
$$+ \lambda^2[1 + \cos(2\varphi)]\}[\tfrac{1}{2}(x'^2 + y'^2)]) \quad (3.10)$$

when use is made of the generating function[10]

$$2\sum_{p=0}^{\infty}\epsilon_p \cos(p\psi) I_p(z) = \exp(z\cos\psi) \quad (3.11)$$

with ψ set equal to 2φ and $z = \frac{1}{2}(\beta^2 - \lambda^2)(x'^2 + y'^2)$. Finally, one can readily verify by means of the identities $2\cos^2\varphi = 1 + \cos(2\varphi)$ and $2\sin^2\varphi = 1 - \cos(2\varphi)$ and the transformation from polar to Cartesian coordinates given by $x'^2 = (x'^2 + y'^2)\cos^2\varphi$ and $y'^2 = (x'^2 + y'^2)\sin^2\varphi$ that Eq. (3.10) is equivalent to the closed form expression of Eq. (3.1).

IV. NUMERICAL PROCEDURE

For laboratory light sources encountered in plasma research the integrated intensity distribution is never known explicitly but is an experimentally measured quantity. In order for the method of this paper to be of practical value in situations of this type the series representation of Eq. (2.11) must be summed numerically. Therefore, it is the purpose of this section to present such a numerical procedure and to check the accuracy of this procedure by applying it to the hypothetical example of the previous section.

The procedure we wish to consider here, like all

other numerical procedures, is based on the assumption that the series representation of Eq. (2.11) may be approximated to any degree of accuracy by a finite double sum in the following manner:

$$f(x', y') \approx \left(\frac{\alpha}{\pi^{\frac{1}{2}}}\right) \sum_{k=0}^{K} \sum_{m=0}^{M} \epsilon_m \left[\frac{(k!\,(m+k)!)^{\frac{1}{2}}}{(m+2k)!}\right] \times \mathrm{Re}\left\{\left[\int_{-\pi}^{\pi} d\xi \exp(-im\xi) \int_{-\infty}^{+\infty} dx\, g(\xi, x) H_{m+2k}(\alpha x)\right] U^m_{m+2k}(\alpha x', \alpha y')\right\} \exp -[\alpha^2(x'^2 + y'^2)] \tag{4.1}$$

or if use is made of Eq. (2.3) to replace $U^m_{m+2k}(\alpha x', \alpha y')$ as

$$f(x', y') \approx \left(\frac{\alpha}{\pi}\right)^2 \sum_{k=0}^{K} \sum_{m=0}^{M} \epsilon_m (-1)^k \left[\frac{k!}{(m+2k)!}\right] [B^m_{m+2k}(\alpha) \cos(m\varphi) + D^m_{m+2k}(\alpha) \sin(m\varphi)] \times [\alpha^2(x'^2 + y'^2)]^{\frac{1}{2}m} L^m_k[\alpha^2(x'^2 + y'^2)] \exp[-\alpha^2(x'^2 + y'^2)], \tag{4.2}$$

where the coefficients $B^m_{m+2k}(\alpha)$ and $D^m_{m+2k}(\alpha)$ are given, respectively, by

$$B^m_{m+2k}(\alpha) = \int_{-\pi}^{\pi} d\xi \cos(m\xi) \int_{-\infty}^{+\infty} dx\, g(\xi, x) H_{m+2k}(\alpha x) \tag{4.3}$$

for $m = 0, 1, 2, \cdots$, and

$$D^m_{m+2k}(\alpha) = \int_{-\pi}^{\pi} d\xi \sin(m\xi) \int_{-\infty}^{+\infty} dx\, g(\xi, x) H_{m+2k}(\alpha x) \tag{4.4}$$

for $m = 1, 2, 3, \cdots$. Since $g(\xi, x)$ is not an explicitly known quantity, the double integrals of Eqs. (4.3) and (4.4) must be evaluated numerically; and for this purpose we have chosen a crude numerical procedure but one which has yielded excellent results.

The double integrals were evaluated numerically by noting that for every spectrum $g(\xi, x)$ which is encountered experimentally, there is always a circular domain of radius R outside of which $g(\xi, x)$ can be set equal to zero. This procedure for truncating $g(\xi, x)$ introduces a negligible error and permits us to approximate Eqs. (4.3) and (4.4) by finite integrals as follows:

$$B^m_{m+2k}(\alpha) \approx \int_{-\pi}^{\pi} d\xi \cos(m\xi) \int_{-R}^{R} dx\, g(\xi, x) H_{m+2k}(\alpha x) \tag{4.5}$$

and

$$D^m_{m+2k}(\alpha) \approx \int_{-\pi}^{\pi} d\xi \sin(m\xi) \int_{-R}^{R} dx\, g(\xi, x) H_{m+2k}(\alpha x). \tag{4.6}$$

The integrals in the latter two equations were evaluated numerically by partitioning the circular domain of radius R into segmented radial sectors, then approximating the surface for $g(\xi, x)$ by constant step values over these sectors. A typical step approximation to the surface for $g(\xi, x)$ at the point $(\xi = \xi_i + \Delta\xi_i, x = x_i + \Delta x_i$, where $2\Delta\xi_i = \xi_{i+1} - \xi_i$ and $2\Delta x_i = x_{i+1} - x_i)$ is illustrated in Fig. 2.

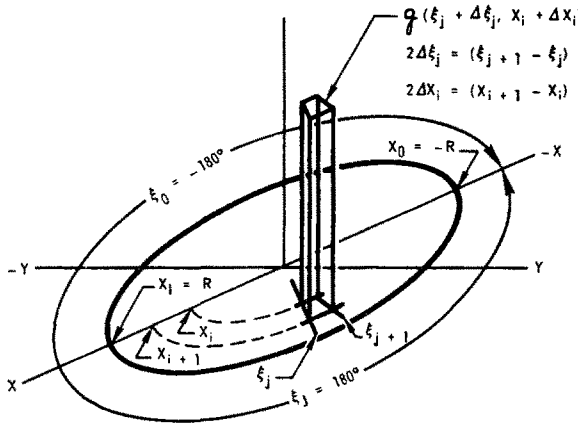

FIG. 2. Qualitative sketch of a typical step approximation to the surface of $g(\xi, x)$ at the point $(\xi = \xi_i + \Delta\xi_i, x = x_i + \Delta x_i)$ together with symbols and nomenclature that are pertinent to the numerical procedure for evaluating the double integrals of Eqs. (4.3) and (4.4).

In terms of this numerical approximation for the surface of $g(\xi, x)$ Eqs. (4.5) and (4.6) may be expressed by finite double sums which in accordance with the symbols and nomenclature of Fig. 2 we can write as

$$B^m_{m+2k}(\alpha) \approx \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} g(\xi_i + \Delta\xi_j, x_i + \Delta x_i) \times \int_{\xi_i}^{\xi_{i+1}} d\xi \cos(m\xi) \int_{x_i}^{x_{i+1}} dx\, H_{m+2k}(\alpha x) \tag{4.7}$$

and

$$D^m_{m+2k}(\alpha) \approx \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} g(\xi_i + \Delta\xi_j, x_i + \Delta x_i) \times \int_{\xi_i}^{\xi_{i+1}} d\xi \sin(m\xi) \int_{x_i}^{x_{i+1}} dx\, H_{m+2k}(\alpha x), \tag{4.8}$$

where $g(\xi_i + \Delta\xi_i, x_i + \Delta x_i)$ is the value of $g(\xi, x)$ evaluated at the point $(\xi = \xi_i + \Delta\xi j, x = x_i + \Delta x_i)$,

$\xi_0 = -\pi$, $x_0 = -R$, $\xi_J = \pi$, and $x_I = R$. The integrals with respect to ξ are elementary and well known while those with regard to x one can easily evaluate with the aid of the derivative formula for the Hermite polynomials.[10] When expressed in terms of these evaluated integrals Eqs. (4.7) and (4.8) become

$$B^m_{m+2k}(\alpha) \approx \left[\frac{1}{2\alpha m(m+2k+1)}\right] \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} g(\xi_j + \Delta\xi_j, x_i + \Delta x_i) \times [\sin(m\xi_{j+1}) - \sin(m\xi_j)][H_{m+2k+1}(\alpha x_{i+1}) - H_{m+2k+1}(\alpha x_i)] \quad (4.9)$$

and

$$D^m_{m+2k}(\alpha) \approx -\left[\frac{1}{2\alpha m(m+2k+1)}\right] \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} g(\xi_j + \Delta\xi_j, x_i + \Delta\xi_i) \times [\cos(m\xi_{j+1}) - \cos(m\xi_j)][H_{m+2k+1}(\alpha x_{i+1}) - H_{m+2k+1}(\alpha x_i)], \quad (4.10)$$

respectively.

Numerical Results

The accuracy of the above numerical procedure was checked by applying it to the hypothetical example of the previous section. For this application the computed coefficients for $D^m_{m+2k}(\alpha)$ based on Eq. (4.10) were zero for all allowed values of m ($m = 1, 2, 3, \cdots$); it can be shown that these results are consistent with those predicted theoretically from either Eq. (4.4) or (4.6) when use is made of the symmetry property that $g(\xi, x)$ is an even function with respect to ξ. Also, for the same reason as was given previously with regard to the second integral of Eq. (3.3), the theoretical values for $B^m_{m+2k}(\alpha)$ based on Eq. (4.3) vanish for all odd values of m and these results were shown to be in complete accord with those computed from Eq. (4.9). For the even values of m the finite double sum of Eq. (4.2) with $D^m_{m+2k}(\alpha)$ set equal to zero, m replaced by $2p$ with $p = 0, 1, 2, 3, \cdots, P$, and α set equal to β; that is,

$$f(x', y') \approx \left(\frac{\beta}{\pi}\right)^2 \sum_{k=0}^{K} \sum_{p=0}^{P} \epsilon_p(-1)^k \left[\frac{k!}{(2p+2k)!}\right] B^{2p}_{2(p+k)}(\beta) \cos(2p\varphi) \times [\beta^2(x'^2 + y'^2)]^p L^{2p}_k[\beta^2(x'^2 + y'^2)] \exp[-\beta^2(x'^2 + y'^2)] \quad (4.11)$$

was evaluated strictly in a numerical manner by making use of the approximate numerical expression

$$B^{2p}_{2(p+k)}(\beta) \approx \left[\frac{1}{4\beta p(2p+2k+1)}\right] \sum_{i=0}^{I-1} \sum_{j=0}^{J-1} g(\xi_j + \Delta\xi_j, x_i + \Delta x_i) \times [\sin(2p\xi_{j+1}) - \sin(2p\xi_j)][H_{2(p+k)+1}(\beta x_{i+1}) - H_{2(p+k)+1}(\beta x_i)], \quad (4.12)$$

which is Eq. (4.9) with m replaced by $2p$ and α set equal to β, to compute the resultant coefficients $B^{2p}_{2(p+k)}(\beta)$. In computing these coefficients the input data $g(\xi_j + \Delta\xi_j, x_i + \Delta x_i)$ were obtained from the spectrum of Eq. (3.2) and limited to the domain of the unit circle by setting the parameters λ and β equal to the values of 4.0 and 5.33, respectively. For simplicity in carrying out the numerical computations the domain of the unit circle was partitioned into segmented radial sections in a uniform manner such that $2\Delta x_i$ and $2\Delta\xi_j$ are fixed quantities for all values of i and j. Also actual numerical computations based on Eq. (4.11) revealed that a partitioning of the unit circle no finer than $I = 160$ zones of width $2\Delta x_i = 0.0125$ along the x axis and $J = 72$ zones of width $2\Delta\xi_j = 5°$ along the direction of angular position ξ, was necessary to ensure sufficient accuracy.

The results of these numerical computations for the spatial distribution of the emission coefficient based on Eq. (4.11) with K and P set equal to 4 and 5, respectively, are tabulated in Table I along with the theoretical values of Eq. (3.1). These results are plotted as a function of radial position in Fig. 3(a) for $\varphi = \tan^{-1}(y'/x') = 0°$, 45°, and 90°; and for completeness the spectrums for $g(\xi, x)$ from Eq. (3.2) corresponding to $\xi = 0°$, 45°, and 90°, respectively, are illustrated in Fig. 3(b).

For further details with regard to experimental accuracy and number of $g(\xi, x)$ values required,

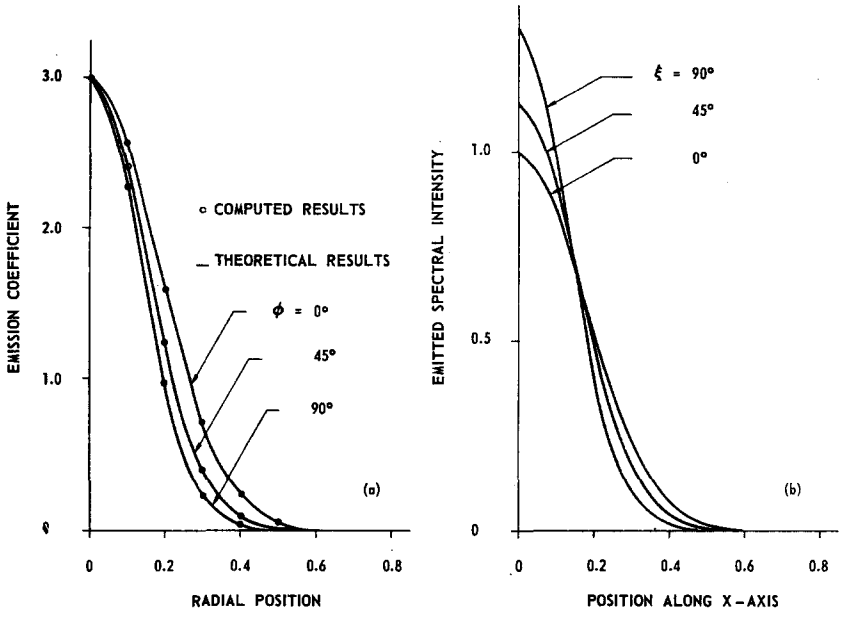

FIG. 3. For comparison, the computed and theoretical results for the spatial distribution of the emission coefficient for the hypothetical example of Sec. III are shown in (a) for $\varphi = 0°$, 45°, and 90°. The integrated intensity distribution curves corresponding to $\xi = 0°$, 45°, and 90°, respectively, are also shown in (b).

computer programs, criteria for choosing K, P, I, J, α, etc., reference is made to a published report[12] on this subject.

V. FINAL REMARKS

Recently[13] the method of this paper was applied to the degenerate case of a circularly symmetric plasma light source corresponding to the positive column of a free-burning argon arc. Despite the success which was achieved in that particular application, the initial impetus for the work presented in this paper as stated previously, was motivated by the need of a procedure for inverting Eq. (1.1) which would yield information on the spatial distribution of the emission coefficient from known experimental data on the integrated intensity distribution for completely asymmetrical light sources. We feel that the series representation method of this paper does accomplish this objective and in fact, it is applied[14] to an asymmetrical plasma light source which is obtained experimentally through the interaction of a free-burning argon arc with a cross magnetic field. The spatial distribution of the emission coefficient corresponding to the atomic spectral line is presented in Ref. 14, which is, as far as we know, the first publication of such information in the open literature.

TABLE I. Comparison of the computed numerical and theoretical results for the spatial distribution of the emission coefficient for the hypothetical example of Sec. III.[a]

r	$\varphi=0°$	15°	30°	45°	60°	75°	90°
0	3.008[b]						
	3.007[c]						
0.1	2.562	2.539	2.480	2.407	2.337	2.284	2.264
	2.562	2.541	2.484	2.408	2.335	2.282	2.263
0.2	1.589	1.535	1.399	1.236	1.093	0.998	0.966
	1.586	1.534	1.401	1.237	1.093	0.998	0.965
0.3	0.713	0.661	0.539	0.407	0.308	0.252	0.233
	0.712	0.661	0.539	0.408	0.308	0.251	0.233
0.4	0.232	0.203	0.142	0.086	0.053	0.037	0.032
	0.232	0.204	0.142	0.086	0.052	0.036	0.032
0.5	0.055	0.045	0.025	0.012	0.005	0.003	0.002
	0.055	0.045	0.025	0.012	0.005	0.003	0.002
0.6	0.010	0.007	0.003	0.001	0.000	0.000	0.000
	0.009	0.007	0.003	0.001	0.000	0.000	0.000
0.7	0.001	0.001	0.000	0.000	0.000	0.000	0.000
	0.001	0.001	0.000	0.000	0.000	0.000	0.000
0.8	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000

[a] $r = (x'^2 + y'^2)^{\frac{1}{2}}$; $\varphi = \tan^{-1}(y'/x')$. $\alpha = \beta = 5.33$; $\lambda = 4.0$; $K = 40$; $P = 5$; $I = 160$; $J = 72$.
[b] Numerical results based on Eq. (4.11).
[c] Theoretical results based on Eq. (3.1).

ACKNOWLEDGMENT

The author is indebted to Professor A. Baños, Jr., for several helpful discussions.

APPENDIX A

The essential steps which are necessary in order to evaluate the integral

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = \int_{-\infty}^{+\infty} dy\, U^{\pm m}_{m+2k}(\alpha x', \alpha y') \exp(-\alpha^2 y^2) \qquad \text{(A1)}$$

[12] H. N. Olsen, C. D. Maldonado, G. D. Duckworth, and A. P. Caron, "Investigation of the Interaction of an External Magnetic Field with an Electric Arc," Final Report under Contract AF33(615)-1105, ARL 66-0016 (December 1965).
[13] C. D. Maldonado, A. P. Caron, and H. N. Olsen, J. Opt. Soc. Am. 55, 1247 (1965).
[14] C. D. Maldonado and H. N. Olsen, J. Opt. Soc. Am. 56, 1305 (1966).

are presented in this Appendix. It follows from this integral that the first and obvious step is to express $U^{\pm m}_{m+2k}(\alpha x', \alpha y')$ in terms of the laboratory coordinate system (x, y). This is readily done by making use of the "invariant in form" property of the orthogonal polynomials to a rotation of axes given by[9] $U^{\pm m}_{m+2k}(\alpha x', \alpha y') = \exp(\pm im\xi) U^{\pm m}_{m+2k}(\alpha x, \alpha y)$. The introduction of this "invariant in form" property for $U^{\pm m}_{m+2k}(\alpha x', \alpha y')$ into Eq. (A1) yields

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = \exp(\pm im\xi) \times \int_{-\infty}^{+\infty} dy\, U^{\pm m}_{m+2k}(\alpha x, \alpha y) \exp(-\alpha^2 y^2), \quad \text{(A2)}$$

which we can write as

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = (-1)^k(\alpha/\pi^{\frac{1}{2}}) \exp(\pm im\xi)\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} \sum_{s=0}^{m} (\pm i)^s \left[\frac{m!}{s!\,(m-s)!}\right](\alpha x)^{m-s} \times \int_{-\infty}^{+\infty} dy\, (\alpha y)^s L^m_k[\alpha^2(x^2+y^2)] \exp(-\alpha^2 y^2) \quad \text{(A3)}$$

when $U^{\pm m}_{m+2k}(\alpha x, \alpha y)$ is replaced by the finite series

$$U^{\pm m}_{m+2k}(\alpha x, \alpha y) = (-1)^k(\alpha/\pi^{\frac{1}{2}})\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} L^m_k[\alpha^2(x^2+y^2)] \times \sum_{s=0}^{m} (\pm i)^s \left[\frac{m!}{s!\,(m-s)!}\right](\alpha x)^{m-s}(\alpha y)^s. \quad \text{(A4)}$$

Equation (A4) is Eq. (2.3) with $[\alpha^2(x'^2 + y'^2)]^{\frac{1}{2}m} \exp(\pm im\varphi)$ replaced by the binomial expansion for $[\alpha(x' \pm iy')]^m$ and (x', y') by (x, y). The integral in Eq. (A3) is nonzero only for even values of s and for these even values of s we use the finite sum[10]

$$L^m_k[\alpha^2(x^2+y^2)] = \sum_{r=0}^{k} L^p_{k-r}(\alpha^2x^2) L^q_r(\alpha^2y^2), \quad \text{(A5)}$$

which is valid for $m = p + q + 1$, to write Eq. (A3) as

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = (-1)^k\left(\frac{\alpha}{\pi^{\frac{1}{2}}}\right) \exp(\pm im\xi)\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} \times \sum_{n=0}^{[\frac{1}{2}m]} (-1)^n \left[\frac{m!}{(2n)!\,(m-2n)!}\right](\alpha x)^{m-2n} \times \sum_{r=0}^{k} L^p_{k-r}(\alpha^2x^2) \int_{-\infty}^{+\infty} dy\, (\alpha y)^{2n} L^q_r(\alpha^2y^2) \exp(-\alpha^2y^2), \quad \text{(A6)}$$

where in the latter expression s has been set equal to $2n$ for $n = 0, 1, 2, 3, \cdots$, and the upper limit $[\frac{1}{2}m]$ on the first sum is equal to $\frac{1}{2}m$ or $\frac{1}{2}(m-1)$ according as m is even or odd.

To evaluate the integral in Eq. (A6) which is equivalent to

$$\int_{-\infty}^{+\infty} dy\, (\alpha y)^{2n} L^q_r(\alpha^2y^2) \exp(-\alpha^2y^2) = \left(\frac{1}{\alpha}\right) \int_0^{\infty} dt\, t^{n-\frac{1}{2}} L^q_r(t) \exp(-t) \quad \text{(A7)}$$

use is made of the tabulated result[15] for the Laplace transform of $t^{n-\frac{1}{2}}L^q_r(t)$ to write

$$\int_0^{\infty} dt\, t^{n-\frac{1}{2}} L^q_r(t) \exp(-t) = \frac{\pi^{\frac{1}{2}}(2n)!\,\Gamma(q+r-n+\frac{1}{2})}{n!\,2^{2n}\Gamma(r+1)\Gamma(q-n+\frac{1}{2})}. \quad \text{(A8)}$$

Combining Eqs. (A7) and (A8) to eliminate the integral in Eq. (A6) and the subsequent use of the finite sum[10]

$$L^{m-n-\frac{1}{2}}_k(\alpha^2x^2) = \sum_{r=0}^{k} \left[\frac{\Gamma(q+r-n+\frac{1}{2})}{\Gamma(r+1)\Gamma(q-n+\frac{1}{2})}\right] L^p_{k-r}(\alpha^2x^2) \quad \text{(A9)}$$

yields

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = (-1)^k \exp(\pm im\xi)\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} \times \sum_{n=0}^{[\frac{1}{2}m]} (-1)^n \left[\frac{m!}{n!\,(m-2n)!\,2^{2n}}\right] \times (\alpha x)^{m-2n} L^{m-n-\frac{1}{2}}_k(\alpha^2x^2). \quad \text{(A10)}$$

Finally with the aid of the following differential expression:

$$L^{m-n-\frac{1}{2}}_k(\alpha^2x^2) = (-1)^k[(\alpha x)^{-(m-2n)}/k!\,2^{2k}] \times \exp(\alpha^2x^2)[d^{2k}/d(\alpha x)^{2k}][(\alpha x)^{m-2n} \exp(-\alpha^2x^2)], \quad \text{(A11)}$$

Eq. (A10) may be expressed as

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = \exp(\pm im\xi)[k!\,(m+k)!]^{-\frac{1}{2}} 2^{-(m+2k)} \times \exp(\alpha^2x^2) \frac{d^{2k}}{d(\alpha x)^{2k}} \left\{\exp(-\alpha^2x^2) \sum_{n=0}^{[\frac{1}{2}m]} (-1)^n \left[\frac{m!}{n!\,(m-2n)!}\right](2\alpha x)^{m-2n}\right\}, \quad \text{(A12)}$$

[15] A. Erdelyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi, *Tables of Integral Transforms* (McGraw-Hill Book Company, Inc., New York, 1963), Vol. I.

or more compactly by

$$I^{\pm m}_{m+2k}(\xi, \alpha x) = \exp(\pm im\xi)[k!\,(m+k)!]^{-\frac{1}{2}} \times 2^{-(m+2k)} H_{m+2k}(\alpha x), \quad \text{(A13)}$$

since

$$H_m(\alpha x) = \sum_{n=0}^{[\frac{1}{2}m]} (-1)^n \left[\frac{m!}{n!(m-2n)!}\right](2\alpha x)^{m-2n} \quad \text{(A14)}$$

and

$$H_{m+2k}(\alpha x) = \exp(\alpha^2 x^2)[d^{2k}/d(\alpha x)^{2k}] \times [\exp(-\alpha^2 x^2) H_m(\alpha x)]. \quad \text{(A15)}$$

APPENDIX B

In this Appendix we use the hypothetical example of Sec. III to show that the Fourier expressions of Eqs. (2.5) and (2.10) for the unknown expansion coefficients yield equivalent results. With no loss in generality the scale factor α is set equal to β in order to simplify the calculations. For the time being we restrict our attention to Eq. (2.5) with α set equal to β. This yields for the unknown expansion coefficients the result

$$C^{\pm m}_{m+2k}(\beta) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dx'\,dy'\, f(x', y') \times \tilde{U}^{\pm m}_{m+2k}(\beta x', \beta y'), \quad \text{(B1)}$$

which we can express in terms of Eq. (3.1) and the complex conjugate of Eq. (A4) with α set equal to β and (x, y) replaced by (x', y') as

$$C^{\pm m}_{m+2k}(\beta, \lambda) = (-1)^k(\beta/\pi^{\frac{1}{2}})\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} B \times \sum_{s=0}^{m} (\pm i)^s \left[\frac{m!}{s!\,(m-s)!}\right] \times \int_{-\infty}^{+\infty} dx'\,(\beta x')^{m-s} \exp(-\lambda^2 x'^2) \times \int_{-\infty}^{+\infty} dy'\,(\beta y')^s L^m_k[\beta^2(x'^2 + y'^2)] \exp(-\beta^2 y'^2). \quad \text{(B2)}$$

Replacing $L^m_k[\beta^2(x'^2 + y'^2)]$ in the double integral of Eq. (B2) by the finite series representation[10]

$$L^m_k[\beta^2(x'^2 + y'^2)] = \sum_{n=0}^{k} (-1)^n \left[\frac{(m+k)!}{(k-n)!\,(m+n)!\,n!}\right] \times [\beta^2(x'^2 + y'^2)]^n \quad \text{(B3)}$$

and using the binomial theorem to expand $[\beta^2(x'^2 + y'^2)]^n$ we obtain

$$C^{\pm m}_{m+2k}(\beta, \lambda) = (-1)^k\left(\frac{\beta}{\pi^{\frac{1}{2}}}\right)\left[\frac{k!}{(m+k)!}\right]^{\frac{1}{2}} B \sum_{s=0}^{m} (\mp i)^s [m!/s!\,(m-s)!] \sum_{n=0}^{k} (-1)^n \left[\frac{(m+k)!}{(k-n)!\,(m+n)!\,n!}\right] \times \sum_{q=0}^{n}\left[\frac{n!}{q!\,(n-q)!}\right] \int_{-\infty}^{+\infty} dx'\,(\beta x')^{2(n-q)+m-s} \exp(-\lambda^2 x'^2) \int_{-\infty}^{+\infty} dy'\,(\beta y')^{2q+s} \exp(-\beta^2 y'^2). \quad \text{(B4)}$$

From the integrals in this last expression it follows that the coefficients $C^{\pm m}_{m+2k}(\beta, \lambda)$ are nonzero only for even values of m and s. And for these even values of m and s which we denote by $m = 2p$ and $s = 2r$ for $p, r = 0, 1, 2, 3, \cdots$, Eq. (B4) becomes

$$C^{\pm 2p}_{2(p+k)}(\beta, \lambda) = (-1)^k\left(\frac{\pi^{\frac{1}{2}}}{\lambda}\right)\left[\frac{k!}{(k+2p)!}\right]^{\frac{1}{2}}\left(\frac{\beta}{2\lambda}\right)^{2p} B \sum_{r=0}^{p} (-1)^r \left\{\frac{(2p)!}{(2r)!\,[2(p-r)]!}\right\}\left(\frac{\lambda}{\beta}\right)^{2r} \times \sum_{n=0}^{k} (-1)^n \left[\frac{(k+2p)!}{(k-n)!\,(n+2p)!\,n!}\right]\left(\frac{\beta}{2\lambda}\right)^{2n} \sum_{q=0}^{n} \left\{\frac{n!\,[2(q+r)]!\,[2(n-q+p-r)]!}{q!\,(n-q)!\,(q+r)!\,(n-q+p-r)!}\right\}\left(\frac{\lambda}{\beta}\right)^{2q}, \quad \text{(B5)}$$

where use has been made of the tabulated results[15]

$$\int_{-\infty}^{+\infty} dx'\,(\beta x')^{2(n-q+p-r)} \exp(-\lambda^2 x'^2) = \left(\frac{\pi^{\frac{1}{2}}}{\lambda}\right)\left\{\frac{[2(n-q+p-r)]!}{(n-q+p-r)!}\right\}\left(\frac{\beta}{\lambda}\right)^{2(n-q+p-r)} \quad \text{(B6)}$$

and

$$\int_{-\infty}^{+\infty} dy'\,(\beta y')^{2(q+r)} \exp(-\beta^2 y'^2) = \left(\frac{\pi^{\frac{1}{2}}}{\beta}\right)\{[2(q+r)]!/(q+r)!\}. \quad \text{(B7)}$$

Calculations based on Eq. (B5) were carried out for k, $p = 0, 1, 2$, and 3, and the obtained results for the normalized coefficients $\lambda C^{\pm 2p}_{2(p+k)}(\beta, \lambda)/\pi^{\frac{1}{2}}B$ are tabulated in Table II.

To check whether Eq. (2.10), when applied to the hypothetical example of Sec. III, yields results

equivalent to those of Eq. (B5), we now concern ourselves with the evaluation of the integral

$$C^{\pm m}_{m+2k}(\beta) = \left(\frac{\beta}{2\pi^{\frac{1}{2}}}\right)\left\{\frac{[k!\,(m+k)!]^{\frac{1}{2}}}{(m+2k)!}\right\}\int_{-\pi}^{\pi} d\xi\,\cos(m\xi)\int_{-\infty}^{+\infty} dx\; g(\xi, x)H_{m+2k}(\beta x), \tag{B8}$$

which is Eq. (2.10) for a spectrum $g(\xi, x)$ that is an even function of ξ and α set equal to β. When expressed in terms of the spectrum for $g(\xi, x)$ as given by Eq. (3.2), Eq. (B8) becomes

$$C^{\pm m}_{m+2k}(\beta, \lambda) = \left(\frac{\beta}{2\pi\lambda}\right)\left\{\frac{[k!\,(m+k)!]^{\frac{1}{2}}}{(m+2k)!}\right\}B\int_{-\pi}^{\pi} d\xi\,[\cos(m\xi)/(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)^{\frac{1}{2}}]$$
$$\times\int_{-\infty}^{+\infty} dx\,\exp[-\beta^2x^2/(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)]H_{m+2k}(\beta x) \tag{B9}$$

and for the same reason as was given previously with regard to Eq. (3.3), this expression is nonzero only for even values of m. And for these even values of m which we denote by $m=2p$ for $p=0, 1, 2, 3, \cdots$, the second integral in Eq. (B9) is the Gauss transform of $H_{2(p+k)}(\beta x)$ given by the tabulated result[10]

$$\int_{-\infty}^{+\infty} dx\,\exp\left[-\frac{\beta^2x^2}{(\sin^2\xi + (\beta/\lambda)^2\cos^2\xi)}\right]H_{2(p+k)}(\beta x)$$
$$= \left(\frac{\pi^{\frac{1}{2}}}{\beta}\right)\left\{\frac{[2(p+k)]!}{(p+k)!}\right\}\left[\left(\frac{\beta}{\lambda}\right)^2 - 1\right]^{p+k}$$
$$\times\left[\sin^2\xi + \left(\frac{\beta}{\lambda}\right)^2\cos^2\xi\right]^{\frac{1}{2}}\cos^{2(p+k)}(\xi). \tag{B10}$$

Setting $m = 2p$ in Eq. (B9) and subsequently using Eq. (B10) to eliminate the second integral yield for $C^{\pm 2p}_{2(p+k)}(\beta, \lambda)$ the expression

$$C^{\pm 2p}_{2(p+k)}(\beta, \lambda)$$
$$= \left(\frac{1}{2\pi^{\frac{1}{2}}\lambda}\right)\left\{\frac{[k!\,(k+2p)!]^{\frac{1}{2}}}{(p+k)!}\right\}\left[\left(\frac{\beta}{\lambda}\right)^2 - 1\right]^{p+k}$$
$$\times B\int_{-\pi}^{\pi} d\xi\,\cos(2p\xi)\,\cos^{2(p+k)}(\xi), \tag{B11}$$

which we can rewrite in terms of the tabulated result[15]

$$\int_{-\pi}^{\pi} d\xi\,\cos(2p\xi)\,\cos^{2(p+k)}(\xi) = \left\{\frac{2\pi[2(p+k)]!}{k!\,(k+2p)!\,2^{2(p+k)}}\right\} \tag{B12}$$

as

$$C^{\pm 2p}_{2(p+k)}(\beta, \lambda)$$
$$= \left(\frac{\pi^{\frac{1}{2}}}{\lambda}\right)\left\{\frac{[2(p+k)]!}{[k!\,(k+2p)!]^{\frac{1}{2}}(p+k)!}\right\}$$
$$\times\left[\left(\frac{\beta}{2\lambda}\right)^2 - \left(\frac{1}{2}\right)^2\right]^{p+k}B. \tag{B13}$$

Finally, we can readily verify by direct computations that the results obtained from Eq. (B13) for the normalized coefficients $\lambda C^{\pm 2p}_{2(p+k)}(\beta, \lambda)/\pi^{\frac{1}{2}}B$ are equivalent and in complete accord with the results tabulated in Table II based on Eq. (B5).

TABLE II. Normalized coefficients $\lambda C^{\pm 2p}_{2(p+k)}(\beta\lambda)/\pi^{\frac{1}{2}}B$ for $k, p \leq 3$.

	k			
p	0	1	2	3
0	1	$2\chi^a$	$6\chi^2$	$20\chi^3$
1	$(2)^{\frac{1}{2}}\chi$	$2(6)^{\frac{1}{2}}\chi^2$	$10(3)^{\frac{1}{2}}\chi^3$	$28(5)^{\frac{1}{2}}\chi^4$
2	$(6)^{\frac{1}{2}}\chi^2$	$2(30)^{\frac{1}{2}}\chi^3$	$14(10)^{\frac{1}{2}}\chi^4$	$12(210)^{\frac{1}{2}}\chi^5$
3	$2(5)^{\frac{1}{2}}\chi^3$	$4(35)^{\frac{1}{2}}\chi^4$	$18(35)^{\frac{1}{2}}\chi^5$	$22(210)^{\frac{1}{2}}\chi^6$

[a] $\chi = [(\beta/2\lambda)^2 - (1/2)^2]$.

Relaxation to Equilibrium of a Dilute Electron Plasma

J. B. McLeod
Fellow of Wadham College, Oxford University, Oxford, England
AND
R. S. B. Ong*
U.K.A.E.A. Culham Laboratory, Abingdon, Berkshire, England
(Received 21 April 1966)

The relaxation of the velocity distribution to equilibrium in an electron plasma in which the dominant collisions are described by the Fokker–Planck operator is studied. It is shown mathematically that the linearized collision operator possesses a continuous spectrum of eigenvalues which extends over the entire real interval from zero to infinity. Consequently the decay to equilibrium is not uniformly exponential. For large values of the time the decay to Maxwellian is shown to be of the order of the inverse power of the time variable. Moreover the rate of decay depends also on the initial perturbation.

1. INTRODUCTION

ONE of the fundamental problems in the kinetic theory of gases is the study of the approach to equilibrium. Given an arbitrary initial velocity distribution we know from the H theorem that the collisions among the particles in the gas will in the course of time bring it to an equilibrium distribution. However, the H theorem fails to disclose how rapidly the final equilibrium distribution is attained. A knowledge of this relaxation time is nevertheless essential for the kinetic theoretical derivation of the complete set of hydrodynamic equations describing the fluidlike behavior of the gas.

In this article we are in particular interested in the relaxation problem of a dilute, fully ionized gas consisting of electrons imbedded in a stationary, uniformly smeared-out, neutralizing ion background, in which the dominant collisions are described by a Fokker–Planck collision operator. For simplicity only electron–electron collisions are taken into account. While this problem has attracted the attention of several authors, most of the work was either done numerically[1] or by assuming *a priori* that the linearized Fokker–Planck collision operator may be expanded in a series of orthogonal polynomials.[2] The latter method is equivalent to the *a priori* assumption that the Fokker–Planck operator possesses a discrete spectrum of eigenvalues bounded away from zero. However, it is well known that in the case of neutral gases the eigenvalue spectrum of the Boltzmann collision operator for soft potentials has a continuous part extending from a finite eigenvalue all the way to zero. On the basis of this it is reasonable to suspect that the eigenvalue spectrum of the Fokker–Planck collision operator for a Coulomb potential might have a similar property. Hence it is justified that one makes a careful study of the mathematical nature of the eigenvalue spectrum of the linearized Fokker–Planck operator. It as not until very recently that attempts were made in this direction. Su and also Lewis gave nonrigorous indications of the existence of a continuous eigenvalue spectrum.[3] In this article we rigorously determine the extent of the continuous spectrum and also show how this influences the decay time of the perturbed velocity distribution function of the particles in the gas.

The argument of this paper is partly physical (or at least formal mathematically) and partly rigorous mathematically. In Sec. 2, which is the former, we deduce the linearized form of the Fokker–Planck equation which we intend to use. It turns out that this can be written as

$$(L_1 + L_2)g = -\partial g/\partial t, \tag{1.1}$$

where $g = g(c, t)$, c is the velocity variable $(0 < c < \infty)$, and t is the time. In (1.1), L_1 is a second-order differential operator and L_2 an integral operator, both involving only the variable c.

At this point the mathematics becomes rigorous. We start by studying the spectrum associated with the operator $L_1 + L_2$. It turns out that L_2 is completely continuous, so that the continuous spectrum

* On leave from The University of Michigan, Department of Aerospace Engineering. This work is partially supported by the U. S. Air Force Office of Scientific Research under Grant AF-AFOSR-825-65.

[1] W. M. Macdonald, M. N. Rosenbluth, and W. Chuck, Phys. Rev. **107**, 350 (1957); A. Dolinsky, Phys. Fluids **8**, 438 (1965).

[2] F. J. Jankulak, L. G. de Sobrino, and Y. K. S. Tam, Can. J. Phys. **42**, 1743 (1964).

[3] C. H. Su, Seventh Conference on Ionization Phenomena in Gases, Belgrade (1965); J. Lewis (private communication).

is the continuous spectrum determined by L_1 alone, and we are able to show that this consists of exactly the positive real axis. This work is done in Sec. 3, although the more complicated mathematics is removed to the Appendix.

We then want to use this in Sec. 4 to examine the decay behavior of solutions of (1.1), and at this point we make two further assumptions. Both these assumptions are physically extremely reasonable, but they are mathematically assumptions which should ideally be proved from Eq. (1.1), and we hope to return to their proof in a later paper. The first assumption is that in discussing the decay behavior of solutions of (1.1) we can in fact ignore L_2 so that what we are really discussing is the decay behavior of solutions of

$$L_1 g = -\partial g/\partial t. \tag{1.2}$$

The second assumption is that L_1 has no negative eigenvalues. Not only is this physically reasonable, but numerical work on the computer also suggests it most strongly.

With these assumptions we show in Sec. 4 that the solution of (1.1) [or, more accurately of (1.2)] satisfying any reasonable initial conditions decays at least as fast as $O(t^{-1})$, and although the rate of decay depends on the initial conditions, we can show that at least "in general" the decay is not as fast as exponential.[4]

2. FORMULATION OF THE RELAXATION PROBLEM

Let us consider a spatially homogeneous electron gas in a stationary, neutralizing, uniformly smeared-out ion background such that the dominant collision process may be described by the Fokker–Planck collision operator with a cutoff. For simplicity only electron–electron interactions are taken into account. In the absence of external forces the basic equation governing the relaxation problem is

$$\frac{\partial}{\partial t}\hat{f}(\mathbf{v}, t) = \frac{2\pi n e^4 \ln \Lambda}{m^2}\int_{-\infty}^{\infty} d\mathbf{v}_1 \times \frac{\partial}{\partial \mathbf{v}}\cdot\left[\frac{g^2\mathsf{I} - \mathbf{gg}}{g^3}\cdot\left(\frac{\partial}{\partial \mathbf{v}} - \frac{\partial}{\partial \mathbf{v}_1}\right)\hat{f}(\mathbf{v}, t)\hat{f}(\mathbf{v}_1, t)\right], \tag{2.1}$$

where $\Lambda \equiv \lambda_D/r_L$, $\lambda_D = (KT/4\pi ne^2)^{\frac{1}{2}}$ is the Debye length, $r_L = e^2/KT$ is the mean distance of closest approach in a Coulomb encounter, $g = \mathbf{v} - \mathbf{v}_1$ is the relative velocity, I is the unit dyadic. The other symbols used in (2.1) are those commonly employed in plasma kinetic theory and need no further explanation.

It may be verified by direct substitution that the Maxwell–Boltzmann distribution

$$\hat{f}_0(\mathbf{v}) = n(m/2\pi KT)^{\frac{3}{2}}\exp(-mv^2/2KT), \tag{2.2}$$

where the number density n and the temperature T are constants, causes the right-hand side of Eq. (2.1) to vanish. This shows that $\hat{f}_0(\mathbf{v})$ is a solution of the stationary problem. Moreover, from the H theorem it is a well-known fact that eventually the gas will relax to the distribution $\hat{f}_0(\mathbf{v})$ and that this is the unique equilibrium distribution.

Let us now suppose that the gas is slightly disturbed from its initial equilibrium state. We wish to know how rapidly $\hat{f}(\mathbf{v}, t)$ relaxes to equilibrium. Accepting the fact that the Maxwell–Boltzmann distribution will be reached eventually, we put in Eq. (2.1)

$$\hat{f}(\mathbf{v}, t) = \hat{f}_0(\mathbf{v})[1 + \hat{h}(\mathbf{v}, t)] \tag{2.3}$$

and neglect quadratic terms in $\hat{h}(\mathbf{v}, t)$. We then obtain a linearized Fokker–Planck equation. Since the particle interactions are assumed to be governed by a central force law, it may be shown by direct calculation that the angular dependence of the Fokker–Planck operator in (2.1) is just the spherical harmonic operator. Thus by writing the kinetic equation (2.1) in spherical polar coordinates and letting

$$\hat{h}(\mathbf{v}, t) = \sum_{l,m}\hat{g}_{l,m}(v, t)Y_{l,m}(\theta, \phi), \tag{2.4}$$

the equation for the angular dependence $Y_{l,m}(\theta, \phi)$ of the perturbation distribution function $\hat{h}(\mathbf{v}, t)$ may be uncoupled from the equation governing the radial distribution $\hat{g}_{l,m}(\mathbf{v}, t)$. Futhermore it is convenient to measure the time in units of the "relaxation time" and the velocity in units of the thermal velocity by means of the transformations

$$\tau = (3/2\sqrt{2\pi})(\omega_p \ln \Lambda/\Lambda)t, \quad \mathbf{c} = (m/2KT)^{\frac{1}{2}}\mathbf{v}, \tag{2.5}$$

where $\omega_p = (4\pi ne^2/m)^{\frac{1}{2}}$ is the electron plasma frequency.

The linearized Fokker–Planck equation for the perturbation in the radial distribution function then

[4] The authors are indebted to the Referee for drawing the attention that a similar problem arises in the theory of neutron thermalization. For references see, e.g., R. E. Marshak, Rev. Mod. Phys. 19, 185 (1947); B. Davison and I. B. Sykes, *Neutron Transport Theory* (Clarendon Press, Oxford, England, 1957); J. Lehner and G. M. Wing, Commun. Pure Appl. Math. 8, 217 (1955); D. S. Selengut, in *Neutron Physics: Proceedings*, M. L. Yeater, Ed. (Academic Press Inc., New York, 1962).

assumes the form

$$\frac{\partial}{\partial\tau}\vartheta_{l,m}(c,\tau) = \left[\frac{\Phi(c)}{c^3} - \frac{e^{-c^2}}{c^2}\right]\frac{\partial^2\vartheta_{l,m}}{\partial c^2} + \left[\frac{e^{-c^2}}{c^3}(1+4c^2) - \frac{\Phi(c)}{c^4}(1+2c^2)\right]\frac{\partial\vartheta_{l,m}}{\partial c} + \left\{4e^{-c^2} - \frac{l(l+1)}{2c^4}\left[\frac{\Phi(c)}{c}(2c^2-1) + e^{-c^2}\right]\right\}\vartheta_{l,m} + 4\int_0^\infty c_1^2 e^{-c_1^2}K_l(c,c_1)\vartheta_{l,m}(c_1,\tau)\,dc_1,$$

where

$$\Phi(c) \equiv \int_0^c e^{-x^2}\,dx,$$

$$K_l(c,c_1) = \begin{cases} \dfrac{2}{2l+1}\left\{\dfrac{c_1^l}{c^{l+1}}\left[\dfrac{(l+1)(l+2)}{2l+3}c_1^2 - 1\right] - \dfrac{l(l-1)c_1^l}{(2l-1)c^{l-1}}\right\} & \text{for } c_1 < c, \\ \dfrac{2}{2l+1}\left\{\dfrac{c^l}{c_1^{l+1}}\left[\dfrac{(l+1)(l+2)}{2l+3}c^2 - 1\right] - \dfrac{l(l-1)c^l}{(2l-1)c^{l-1}}\right\} & \text{for } c_1 > c. \end{cases}$$

[The independence of (2.6) from m shows that the subscript m in $g_{l,m}$ can be—and is—dropped.] This equation has been obtained also by Su. By direct substitution it may be verified that for $l = 0$ ($m = 0$), the functions $\vartheta_0 = 1$ and $\vartheta_0 = c^2$ make the right-hand side of Eq. (2.6) vanish. This corresponds to the conservation of mass and energy, respectively. Similarly, for $l = 1$ ($m = 0, \pm 1$), we have $g_1 = c$ as a solution to the linearized Fokker–Planck equation in three dimensions, which corresponds to the conservation of momentum.

It is now convenient to introduce the transformation

$$g_l(c,\tau) = ce^{-c^2/2}\vartheta_l(c,\tau). \tag{2.7}$$

Equation (2.6) then reduces to

$$-(\partial/\partial\tau)g_l = (L_1 + L_2)g_l, \tag{2.8}$$

where

$$L_{1,2} \equiv -(\partial/\partial c)[p(c)(\partial/\partial c)] + q(c),$$

$$L_2 \equiv -4\int_0^\infty cc_1 e^{-\frac{1}{2}(c^2+c_1^2)}K_l(c,c_1)\,dc_1,$$

$$p(c) \equiv (1/c^3)(\Phi[c] - ce^{-c^2}),$$

$$q(c) \equiv (3/c^5)(\tfrac{2}{3}c^3e^{-c^2} + ce^{-c^2} - \Phi[c]) + (1/c^2)q^*(c),$$

$$q^*(c) \equiv c(\Phi[c] - 7ce^{-c^2}) + \tfrac{1}{2}l(l+1)\{[\Phi(c)/c^3](2c^2-1) + (e^{-c^2}/c^2)\},$$

$$\Phi(c) \equiv \int_0^c e^{-y^2}\,dy.$$

3. ASSOCIATED EIGENVALUE PROBLEM

In order to study the behavior of the distribution function we first consider the spectral behavior of the operator $L_1 + L_2$ in the space $L^2(0, \infty)$. This square integrability condition imposed on the distribution function is in a sense connected with the existence of the entropy as was shown by Waldmann.[5] We first note that L_2 is a symmetric Hilbert–Schmidt operator, since it is an integral operator with a kernel $H(c, c_1)$, say, which is symmetric and satisfies

$$\int_0^\infty\int_0^\infty H^2(c,c_1)\,dc\,dc_1 < \infty.$$

Hence L_2 is certainly compact and by a well-known theorem due to Weyl (p. 367 of Ref. 6) the set of limit points of the spectrum of $L_1 + L_2$ (provided L_1 is self-adjoint) will be the same as that for L_1 alone. What we are going to do in the present section is to show that the spectrum of L_1 is continuous over $(0, \infty)$ and is otherwise at most discrete, and then that the same must hold also for $L_1 + L_2$.

To make L_1 essentially self-adjoint, we have to introduce a boundary condition at $c = 0$, at least when $l = 0$, and in view of (2.7), we take the boundary condition to be $g_l(0, \tau) = 0$. We can then verify that L_1 is essentially self-adjoint, or, what amounts to the same thing, that we have the limit-point case at both ends of the c interval $[0, \infty)$.

That the problem is limit point at $c = \infty$ means that there is only one linearly independent solution of $L_1 y = \lambda y$ which is $L^2(A, \infty)$, A being a positive constant and λ a nonreal number. This implies that an eigenvalue problem is properly defined without the need for any further boundary conditions at $c = \infty$. Similarly for $c = 0$ (cf. Ref. 7, Chap. II).

To verify this at $c = 0$, we note that near $c = 0$ we have

$$p(c) = \tfrac{2}{3} + O(c^2),$$

$$q(c) = \tfrac{2}{3}l(l+1)/c^2 + O(1).$$

Hence if $l = 0$, there is no singularity of the equation

[5] L. Waldmann, in *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1958), Vol. XII.

[6] F. Riesz and B. Sz-Nagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955).

[7] E. C. Titchmarsh, *Eigenfunction Expansions* (Clarendon Press, Oxford, England, 1962), Vol. 1.

at $c = 0$, and there is thus one (and apart from an unimportant multiplicative constant) only one solution which satisfies the boundary condition $g_l(0) = 0$. Hence the problem is limit point.

If $l \neq 0$, it is once again standard work to see that there are two linearly independent solutions which for small c are of order c^{l+1} and c^{-l}. Only multiples of the former can vanish at $c = 0$, or indeed ever belong to the Hilbert space $L^2(0, \infty)$. Hence if $l \neq 0$, the problem is limit point at $c = 0$ even without the boundary condition.

For the behavior for large values of c, and also for the purpose of discussing the nature of the spectrum, it is best to simplify the differential operator by applying the Liouville transformation:

$$U_l = p^{\frac{1}{2}} g_l, \qquad x = \int_0^c \{p(\xi)\}^{-\frac{1}{2}}\, d\xi.$$

The equation $L_1 g_l = -\lambda g_l$ then takes the form

$$(d^2u_l/dx^2) + \{\lambda - Q(x)\}u_l = 0 \qquad (0 < x < \infty), \tag{3.1}$$

where

$$Q(x) = \frac{3}{c^5}(\Phi[c] - ce^{-c^3}) - \frac{2}{c^2}e^{-c^3} - e^{-c^2} - \frac{c}{16}\frac{\{2e^{-c^3} - (3/c^3)(\Phi[c] - ce^{-c^3})\}^2}{\Phi(c) - ce^{-c^3}} + q(c).$$

Both x and c become large together. In fact, for large c, $x \sim c^{5/2}$ and so we see that $Q(x) \to 0$ as $x \to \infty$. This at once implies that the eigenvalue problem associated with (3.1) is limit point at $x = \infty$, and furthermore that the spectrum is discrete for $\lambda < 0$ and continuous for $\lambda \geq 0$ (cf. Sec. 5.5 of Ref. 7).

However, for the purpose of the next section, we need to know more than just that the spectrum is continuous for $\lambda \geq 0$. We need to show also that the "spectral density" (an intuitive idea that is defined explicitly in the Appendix) is bounded for small positive value of λ, and this we prove in the Appendix using the fact that not only does $Q(x) \to 0$, but it does so very smoothly.

4. DECAY TO EQUILIBRIUM

The existence of the continuous spectrum extending to the origin influences the relaxation process a great deal. We cannot expect a uniform exponential decay to equilibrium. Of particular importance are the $\lambda \approx 0$ modes, since they dominate the behavior of the relaxation process for large values of time. For these values of λ near zero it is plausible to assume that the relaxation process is dominated by the continuous spectrum of the differential operator L_1. To estimate the decay time corresponding to these very small values of λ, we therefore study the behavior as $\tau \to \infty$ of the solution $u_l(x, \tau)$ of the partial differential equation

$$(\partial/\partial\tau)u_l(x, \tau) = [(\partial^2/\partial x^2) - Q(x)]u_l(x, \tau) \qquad (x, \tau \geq 0), \tag{4.1}$$

which satisfies the boundary condition

$$u_l(x, 0) = w(x), \tag{4.2}$$

where $w(x)$ is the initial perturbation of the radial distribution function. We suppose, as is physically perfectly reasonable, that $w(x)$ satisfies the following conditions:

(i) $w(x)$ is twice differentiable and $L(0, \infty)$, while both $w(x)$ and $Q(x)w(x) - w''(x)$ are $L^2(0, \infty)$.

(ii) $w(0) = 0$.

(iii) For every $\alpha > 0$, as $x \to \infty$

$$w(x) = o(e^{\alpha x}); \qquad w'(x) = o(e^{\alpha x}).$$

We then have the following result.

Theorem: If $Q(x)$ satisfies the conditions (i)–(iv) in the Appendix, and if $w(x)$ satisfies the conditions (i)–(iii) above, and if the eigenvalue problem associated with the equation

$$(d^2y/dx^2) + [\lambda - Q(x)]y = 0 \tag{4.3}$$

and the boundary condition $y(0) = 0$ (the boundary condition being required only in the case $l = 0$) has no negative eigenvalues, then we can construct a solution of (4.1) and (4.2) which has the property that, for fixed x and as $\tau \to \infty$, the solution decays at least as fast as τ^{-1}.

Before giving the proof of the theorem, we remember that the assumption that (4.3) has no negative eigenvalues has not been proved for the particular $Q(x)$ of (3.2), although it is physically reasonable. We remark also that we say nothing about the uniqueness of the solution. There is no reason to suppose that there is only one solution satisfying (4.1) and (4.2), although the solution would be unique if we imposed further conditions. Since, however, the solution we do obtain is the physically relevant one, we leave the question of uniqueness for the moment.

Proof: We give this for the case $l = 0$, since this is the case that is discussed in detail in the

Appendix. There is no difficulty in extending the arguments to $l \neq 0$.

We first show the existence of a solution. By the spectral theorem given in Chap. III of Ref. 7 we can express $w(x)$ in terms of eigenfunctions of (4.3) by

$$w(x) = \frac{1}{\pi}\int_0^\infty \phi(x, \lambda)\, dF(\lambda), \tag{4.4}$$

the integral is taken only over positive values of λ because we are assuming no negative spectrum. Here

$$F(\lambda) = \int_0^\infty \chi(y, \lambda)w(y)\, dy,$$

$$\chi(y, \lambda) = \int_0^\lambda \phi(y, \mu)\, dk(\mu),$$

where the functions ϕ and k are defined in the Appendix. From the continuity properties of the spectral density given there, we can write

$$\chi(y, \lambda) = -\int_0^\lambda \phi(y, \mu)\ \mathrm{Im}\ m(\mu)\, d\mu$$

and so

$$F(\lambda) = -\int_0^\lambda \mathrm{Im}\ m(\mu)\left[\int_0^\infty \phi(y, \mu)w(y)\, dy\right] d\mu,$$

$$F'(\lambda) = -\mathrm{Im}\ m(\lambda)\left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right],$$

and (4.4) appears as

$$w(x) = -\frac{1}{\pi}\int_0^\infty \phi(x, \lambda) \times \left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right] \mathrm{Im}\ m(\lambda)\, d\lambda. \tag{4.5}$$

Now consider the function

$$u_l(x, \tau) = -\frac{1}{\pi}\int_0^\infty e^{-\lambda\tau}\phi(x, \lambda) \times \left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right] \mathrm{Im}\ m(\lambda)\, d\lambda. \tag{4.6}$$

In view of the convergence of (4.5) we know that (4.6) must be absolutely convergent if $\tau > 0$, and from the absolute convergence of the integral obtained by differentiating with respect to τ under the integral sign we obtain

$$\frac{\partial u_l}{\partial \tau} = -\frac{1}{\pi}\int_0^\infty e^{-\lambda\tau}[-\lambda\phi(x, \lambda)] \times \left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right] \mathrm{Im}\ m(\lambda)\, d\lambda$$

$$= -\frac{1}{\pi}\int_0^\infty e^{-\lambda\tau}[\phi'' - Q(x)\phi] \times \left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right] \mathrm{Im}\ m(\lambda)\, d\lambda$$

$$= -Q(x)u_l + \partial^2 u_l/\partial x^2. \tag{4.7}$$

[To justify the last step, we can derive from the Appendix, although it is not specifically stated there, that, for large λ, $|\mathrm{Im}\ m(\lambda)| \leq \lambda^{\frac{1}{2}}$, while

$$\int_0^\infty \phi(y, \lambda)w(y)\, dy$$

is uniformly bounded for all λ. The necessary differentiations under the integral are then easily justified.]

We have therefore found a solution, and next show that this solution vanishes at least as fast as τ^{-1}, x being fixed. For by the remarks justifying (4.7), and the fact that $m(\lambda)$ is continuous for $\lambda \geq 0$, we have

$$|u_l(x, \tau)| \leq c\left[\int_0^1 e^{-\lambda\tau}\, d\lambda + \int_1^\infty \lambda^{\frac{1}{2}}e^{-\lambda\tau}\, d\lambda\right],$$

where c is some constant, and this implies that, for large τ,

$$u_l(x, \tau) = O(\tau^{-1}).$$

The decay of $u_l(x, \tau)$ may, of course, be faster than τ^{-1}. If, for example, $\mathrm{Im}\ m(\lambda) = O(\lambda)$ as $\lambda \to 0$ then we could conclude that $u_l(x, \tau) = O(\tau^{-2})$. But in order that $u_l(x, \tau)$ should vanish exponentially, so that $u_l(x, \tau) = O(e^{-\sigma\tau})$ for some $\sigma > 0$, we should have in (4.6) that

$$\phi(x, \lambda)\left[\int_0^\infty \phi(y, \lambda)w(y)\, dy\right] \mathrm{Im}\ m(\lambda)$$

vanish identically for $0 \leq \lambda \leq \sigma$, and for "general" $w(y)$, this is possible only if $\mathrm{Im}\ m(\lambda) = 0$ for $0 \leq \lambda < \sigma$, which implies that the spectrum is empty for the interval $0 \leq \lambda < \sigma$. However, we know this to be untrue. This concludes the proof of the theorem.

ACKNOWLEDGMENTS

One of the authors (R.S.B.O.) would like to thank the members of the Theory Division for their hospitality during his stay at the U.K.A.E.A. Culham Laboratory. He is also greatly indebted to Professor G. E. Uhlenbeck of the Rockefeller Institute, New York City, for valuable discussions concerning this problem.

APPENDIX

Our object is to consider the spectral density associated with the operator

$$Q(x) - d^2/dx^2 \tag{A1}$$

in the Hilbert space $L^2(0, \infty)$. We impose the following conditions on $Q(x)$, and it is trivial to verify that these conditions are in fact satisfied by the specific function $Q(x)$ given by (3.2).

(i) $Q(x)$ is continuous for $x > 0$ and three times continuously differentiable if x is sufficiently large, say $x \geq A$.
(ii) $Q(x) \geq x^{-2+\epsilon}$ for some $\epsilon > 0$ if $x \geq A$, and $Q(x) \to 0$ as $x \to \infty$.
(iii) For $x \geq A$,

$$\frac{Q'(x)}{Q(x)} \sim -\frac{1}{x}, \quad \frac{Q''(x)}{Q'(x)} = O\left(\frac{1}{x}\right), \quad \frac{Q'''(x)}{Q'(x)} = O\left(\frac{1}{x^2}\right).$$

(iv) For small x,

$$Q(x) = [l(l+1)/x^2] + Q^*(x),$$

where $Q^*(x)$ is continuous for $x \geq 0$. (It is possible to allow $Q^*(x)$ to have singularities of a limited kind at $x = 0$, but this is unnecessary for the present paper.)

It has already been pointed out in Sec. 3 that the eigenvalue problem associated with (A1) is limit point at both ends of the interval $[0, \infty)$ if $l \neq 0$, so that the spectral problem is well defined. If $l = 0$, we have to impose the additional condition that the situation vanishes at $x = 0$. The analysis is slightly different in the two cases $l = 0$, $l \neq 0$. We follow in detail the case $l = 0$, indicating where necessary the slight changes required to deal with the case when $l \neq 0$.

Titchmarsh's analysis of this sort of problem assures us that, if $\operatorname{Im} \lambda \neq 0$, then the equation

$$d^2y/dx^2 + [\lambda - Q(x)]y = 0 \tag{A2}$$

has just one linearly independent solution which is $L^2(0, \infty)$.[7] If we define $\vartheta(x, \lambda)$, $\phi(x, \lambda)$ to be the solutions of (A2) which satisfy

$$\phi(0, \lambda) = 0, \qquad \phi'(0, \lambda) = -1,$$
$$\vartheta(0, \lambda) = 1, \qquad \vartheta'(0, \lambda) = 0,$$

and if we define $m(\lambda)$ by saying that the L^2 solution must take the form

$$\psi(x, \lambda) = \vartheta(x, \lambda) + m(\lambda)\phi(x, \lambda),$$

then Titchmarsh's form of the spectral theorem is made to depend upon $m(\lambda)$, which we now investigate.

{If $l \neq 0$, we have to set up two eigenvalue problems associated with (A2), one for $[1, \infty)$ with the boundary condition $y(1) = 0$; and the other for the interval $[0, 1]$, with the boundary condition $y(1) = 0$. We then have two functions $m_1(\lambda)$, $m_2(\lambda)$ defined in a similar way to $m(\lambda)$ above; $m_1(\lambda)$ for the problem over $[0, 1]$ and $m_2(\lambda)$ for the problem $[1, \infty)$. Since the problem over $[0, 1]$ certainly has a discrete spectrum, $m_1(\lambda)$ is a meromorphic function of λ, all its poles lying on the real axis and being otherwise real on the real axis. And $m_2(\lambda)$ will behave very much like $m(\lambda)$ in the case $l = 0$. Titchmarsh gives the spectral theorem in terms of $m_1(\lambda)$, $m_2(\lambda)$, and using this form in place of that involving $m(\lambda)$ alone, we are able to conclude, as we do for $m(\lambda)$, that the spectral density is bounded.}

We know that $m(\lambda)$ exists and is analytic for $\operatorname{Im} \lambda > 0$. It is our object to show that under the condition we have, the function can be extended by continuity to the whole first quadrant $\operatorname{Re} \lambda \geq 0$, $\operatorname{Im} \lambda \geq 0$.

Let $\lambda = \mu + i\nu$; $\mu \geq 0$, $\nu > 0$. To find $m(\lambda)$ we look for the $L^2(0, \infty)$ solution of (A2), and to do this, we make the change of variable

$$\xi(x) = \int_0^x [\lambda - Q(t)]^{\frac{1}{2}}\, dt, \qquad \eta = [\lambda - Q(x)]^{\frac{1}{4}} y.$$

{We can choose the sign of $[\lambda - Q(t)]^{\frac{1}{2}}$ to be such that $0 < \arg [\lambda - Q(t)]^{\frac{1}{2}} < \frac{1}{2}\pi$.} Then Eq. (A2) reduces to

$$\frac{d^2\eta}{d\xi^2} + \eta = -\left\{\frac{1}{4}\frac{Q''(x)}{[\lambda - Q(x)]^2} + \frac{5}{16}\frac{Q'^2(x)}{[\lambda - Q(x)]^3}\right\}\eta, \tag{A3}$$

and, formally at least, this is satisfied by a solution of the integral equation

$$\eta(x) = e^{i\xi(x)} + \int_\xi^\infty \sin(\xi - \tau) \times \left\{\frac{1}{4}\frac{Q''(t)}{\{\lambda - Q(t)\}^2} + \frac{5}{16}\frac{Q'^2(t)}{[\lambda - Q(t)]^3}\right\}\eta(\tau)\, d\tau, \tag{A4}$$

where $\tau = \xi(t)$. This integral equation can be solved by iteration following the method of Titchmarsh (Paragraph 22.26 of Ref. 8), provided that we solve for $x \geq B$, say, and that

$$\int_B^\infty \left[\frac{|Q''(t)|}{|\lambda - Q(t)|^{3/2}} + \frac{Q'^2(t)}{|\lambda - Q(t)|^{5/2}}\right] dt \tag{A5}$$

exists. This is so if B is sufficiently large, depending on λ, since then $|\lambda - Q(t)| \geq \frac{1}{2}|\lambda|$ for $t \geq B$, and

[8] Reference 7, Vol. 2 (1958).

the integral converges by comparison with $\int_B^\infty dt/t^2$, using conditions (ii) and (iii) on $Q(x)$. Further, if we denote by S_δ the region in the closed first quadrant for which $|\lambda| \geq \delta$, δ being some fixed positive number, then one value of B will do uniformly for all λ in S_δ, and so the iterative process converges uniformly in λ and the limit function is continuous in λ. (Note that though we started thinking of $\operatorname{Im} \lambda = \nu > 0$, the above argument does allow us to include $\nu = 0$.) Finally, for λ in S_δ and $x \geq B$, the iterative process shows that the solution satisfies

$$\eta(x) = e^{i\xi(x)}\left[1 + O\left(\int_B^\infty dt/t^2\right)\right], \qquad \text{(A6)}$$

which, with $y = [\lambda - Q(x)]^{-\frac{1}{4}}\eta$, leads to an $L^2(0, \infty)$ solution of (A2) if $\operatorname{Im} \lambda > 0$, since $e^{i\xi(x)}$ is exponentially small as $x \to \infty$.

We now have a solution $y(x, \lambda)$ of (A2) which is continuous in λ for any $x \geq B$, λ being in S_δ, and which, if $\operatorname{Im} \lambda > 0$, is the $L^2(0, \infty)$ solution. Hence, for $\operatorname{Im} \lambda > 0$, this must be a constant multiple of $\vartheta(x, \lambda) + m(\lambda)\phi(x, \lambda)$ and so we have, using the boundary condition at $x = 0$, that

$$m(\lambda) = -y'(0, \lambda)/y(0, \lambda).$$

Since $y(B, \lambda)$, and similarly $y'(B, \lambda)$, are continuous functions of λ in S_δ, and since there are no singularities of the differential equation for x in $[0, B]$, it follows by standard theorems on differential equations that $y(0, \lambda)$ and $y'(0, \lambda)$ are also continuous functions of λ in the same region, and therefore so must $m(\lambda)$, provided that $y(0, \lambda) \neq 0$.

In fact, $y(0, \lambda) \neq 0$ if $\operatorname{Im} \lambda > 0$, for this would imply that λ is an eigenvalue and the problem is a self-adjoint one which can have no complex eigenvalues. Nor can $y(0, \lambda) = 0$ if λ is real, for then $y(x, \lambda)$ would be a constant multiple of $\phi(x, \lambda)$, which is real if λ is real, and $y(x, \lambda)$ is certainly not a multiple of a real-valued function, as we see by looking at the asymptotic behavior given by (A6). Hence $m(\lambda)$, originally defined for $\operatorname{Im} \lambda > 0$, can be extended by continuity to λ in S_δ.

Now the expansion of an arbitrary function is given as a Stieltjes integral with respect to $k(\lambda)$ where, for real λ,

$$k(\lambda) = \lim_{\epsilon \to 0} \int_L^\lambda -\operatorname{Im} m(u + i\epsilon)\, du.$$

Since we are interested only in $dk(\lambda)$, the precise value of L is unimportant, but it is convenient to suppose it positive. Then, in view of the continuity properties of $m(\lambda)$ just found, if $\lambda > 0$, we have

$$k(\lambda) = \int_L^\lambda -\operatorname{Im} m(u)\, du,$$

$$k'(\lambda) = -\operatorname{Im} m(\lambda).$$

This implies that $-\operatorname{Im} m(\lambda)$, which we call the spectral density, will appear in the expansion of an arbitrary function, so that it becomes important to estimate the value of this as $\lambda \to 0$.

To do this, we have to look at the $L^2(0, \infty)$ solution of (A2) more carefully. Let $p(\lambda)$ be the solution [guaranteed unique by conditions (ii) and (iii)] of $\lambda - Q(x) = 0$. Define

$$\zeta(x) = \int_{p(x)}^x [\lambda - Q(t)]^{\frac{1}{2}}\, dt, \qquad \eta = [\lambda - Q(x)]^{\frac{1}{4}}h.$$

(Until further notice λ is small and positive.) Then as before (A2) transforms to

$$(d^2\eta/d\zeta^2) + [1 + (5/36\zeta^2)]\eta = g(x)\eta,$$

where

$$g(x) = \frac{5}{36\zeta^2} - \frac{Q''(x)}{4[\lambda - Q(x)]^2} - \frac{5}{16}\frac{Q'^2(x)}{[\lambda - Q(x)]^3}$$

and, formally at least, this is satisfied by a solution of the integral equation

$$\eta(x) = (\tfrac{1}{2}\pi\zeta)^{\frac{1}{2}}H_{\frac{1}{3}}^{(1)}(\zeta) + i\frac{\pi}{2}\int_x^\infty [H_{\frac{1}{3}}^{(1)}(\zeta)J_{\frac{1}{3}}(\vartheta) - J_{\frac{1}{3}}(\zeta)H_{\frac{1}{3}}^{(1)}(\vartheta)] \times \zeta^{\frac{1}{2}}\vartheta^{\frac{1}{2}}g(t)\eta(t)[\lambda - Q(t)]^{\frac{1}{2}}\, dt, \qquad \text{(A7)}$$

where $\vartheta = \zeta(t)$ (cf. Paragraph 22.27 of Ref. 8).

This integral equation can be solved by iteration, as in Ref. 9, Paragraph 4. In Ref. 9, λ is small and negative and the function $U(t)$, corresponding to $Q(t)$ here, is negative, but this does not alter the proof that, in the present notation,

$$\int_{P(k\lambda)}^\infty |g(t)|\, |\lambda - Q(t)|^{\frac{1}{2}}\, dt = O\{\lambda^{-\frac{1}{4}}[p(\lambda)]^{-\frac{1}{4}}\},$$

where k is some fixed constant exceeding unity; and $\lambda^{-\frac{1}{4}}[p(\lambda)]^{-\frac{1}{4}} \to 0$ as $\lambda \to 0$ since $Q(x) \geq x^{-2+\epsilon}$, and so $p(\lambda) \geq \lambda^{-1}/(2 - \epsilon)$. Also,

$$\int_A^{p(k\lambda)} |g(t)|\, |\lambda - Q(t)|^{\frac{1}{2}}\, dt \leq C\int_A^{p(k\lambda)} \left[\frac{|Q''(t)|}{Q^{3/2}(t)} + \frac{Q'^2(t)}{Q^{5/2}(t)}\right] dt \leq C\int_A^\infty t^{-1-\frac{1}{4}\epsilon}\, dt,$$

[9] J. B. McLeod, Proc. London Math. Soc. 11, 139 (1961).

by conditions (ii) and (iii),

$$\leq CA^{-\frac{1}{2}\epsilon} \tag{A8}$$

for positive constants C not necessarily the same at each appearance.

The iteration process of Ref. 9 now assures us that the solution of (A7) satisfies, uniformly for $x \geq A$ and $\lambda > 0$,

$$\eta(x) = (\tfrac{1}{2}\pi\zeta)^{\frac{1}{2}}H_{\frac{1}{2}}^{(1)}(\zeta) + O(A^{-\frac{1}{2}\epsilon}e^{-\mathrm{Im}\zeta}) \tag{A9}$$

with an expression for $\eta'(x)$ which is obtained by formal differentiation. Further, for any fixed $\lambda > 0$, and $x \to \infty$, $\eta(x)$ satisfies

$$\eta(x) = (\tfrac{1}{2}\pi\zeta)^{\frac{1}{2}}H_{\frac{1}{2}}^{(1)}(\zeta) + o(1),$$

which from the asymptotic expansion for Bessel functions, assures us that $[\lambda - Q(x)]^{-\frac{1}{4}}\eta(x)$ is a constant multiple of $\lim_{\nu\to 0} y(x, \lambda + i\nu)$, where $y(x, \lambda)$ is the $L^2(0, \infty)$ solution previously discussed. Hence

$$\begin{aligned}\lim_{\nu\to 0} m(\lambda + i\nu) &= \lim_{\nu\to 0}\left[-\frac{y'(0, \lambda + i\nu)}{y(0, \lambda + i\nu)}\right] \\ &= -\frac{(d/dx)\{[\lambda - Q(x)]^{-\frac{1}{4}}\eta(x, \lambda)\}}{[\lambda - Q(x)]^{-\frac{1}{4}}\eta(x, \lambda)}\bigg|_{x=0}. \end{aligned} \tag{A10}$$

But $\eta'(A, \lambda)/\eta(A, \lambda)$ is a continuous function of λ for $\lambda > 0$, from the uniform nature of (A9), and we can see also quite easily that $\eta'(A, \lambda)/\eta(A, \lambda)$ tends to a limit as $\lambda \to 0$. Hence we can extend the definition of $\eta'(A, \lambda)/\eta(A, \lambda)$ to include $\lambda = 0$ by continuity, and deduce, as we did for $y'(0, \lambda)/y(0, \lambda)$ previously that the last member in (A10) will also be continuous for $\lambda \geq 0$, provided that the denominator does not vanish at $\lambda = 0$. This last possibility can be excluded as it was when we discussed $y'(0, \lambda)/y(0, \lambda)$.

We have thus established the continuity and boundedness of $m(\lambda)$ for $\lambda \geq 0$, and so, for $\lambda > 0$,

$$k'(\lambda) = -\mathrm{Im}\, m(\lambda)$$

and $-\mathrm{Im}\, m(\lambda) \to a$ as $\lambda \to 0$, for some constant a. The only remaining question is whether there is a jump in $k(\lambda)$ at $\lambda = 0$. This possibility can be excluded by returning to (A4). If $\lambda \leq 0$, there is no difficulty in carrying through the original iteration process for $x \geq A$, since we cannot then have $\lambda - Q(x) = 0$. The relevant integral (A5) involved in the iteration process converges even when $\lambda = 0$ as in (A8). Hence we can solve (A3) for $\lambda = 0$ by solving (A4) with $e^{i\xi(x)}$ replaced by

$$Ce^{i\xi(x)} + De^{-i\xi(x)},$$

C, D being arbitrary constants, and this shows that the solutions are oscillatory and hence no solution is $L^2(0, \infty)$, which would not be the case if $k(\lambda)$ had a jump at $\lambda = 0$ (cf. Ref. 7).

Theory of the Linear Fokker-Planck Collision Operator*

C. H. Su†
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received 26 May 1966)

The linearized Fokker–Planck collision integral with coulomb interaction is expanded in terms of surface spherical harmonics. The radial part of the distribution function is shown to be governed by a set of decoupled differential-integral equations. The differential operators are shown to be self-adjoint while the integral operators are symmetric and completely continuous. The spectrum of the eigenvalues contains, on top of discrete points, a continuous part ranging from zero to minus infinity. The discrete part of the spectrum is obtained by requiring that the corresponding eigenvectors have a definite asymptotic behavior at large velocity. A variation principle is constructed for the computation of the discrete spectrum.

I. INTRODUCTION

SINCE the discovery of the eigenspectrum of the linearized Boltzmann collision operator by Wang–Chang and Uhlenbeck,[1] and independently by Waldmann[2] for the intermolecular force equal to Kr^{-5} (Maxwell molecules), there is considerable interest in studying the spectrum of the same operator with other intermolecular forces. Recently, Grad[3] studied the spectrum for general intermolecular forces. He found, with the assumption of a cutoff in potential, that the spectrum for general intermolecular forces consists of two parts: a discrete spectrum and a continuous spectrum. The latter is bounded away from zero for hard potentials (with the force exponent >5), but approaches zero for soft potentials (with the force exponent <5). In both cases the potentials are cut off at large intermolecular distance to avoid the divergence of distant collisions. The result of soft potentials is of special interest. It indicates that the decay of an arbitrary initial disturbance is not necessary exponential. Since the case of soft potentials includes gases with the coulomb interaction, it is of interest to investigate the spectrum of the Fokker–Planck collision integral directly. It is known that, as long as $\ln \Lambda$ (Λ is the ratio of Debye length to minimum impact parameter) is much greater than unity, the unshielded Fokker–Planck collision integral gives an accurate description of the effect of collisions between particles interacting with coulomb potential. Since the Fokker–Planck equation is structurally simpler than the Boltzmann equation, we feel that much light can be shed on the nature of the eigenspectrum and hence the relaxation phenomena in ionized gases by a careful analysis of the Fokker–Planck collision operator.

We consider a spatially homogeneous electron plasma in a uniformly neutralizing positive background. Extension to include discrete ions can be easily made by taking advantage of the small mass ratio between electrons and ions.

In Sec. II, we make all possible simplifications of the Fokker–Planck collision integral, which includes the linearization around the thermal equilibrium and the use of surface spherical harmonics. We then obtain a set of second-order differential-integral equations for the radial functions of the distribution function. In Sec. III, we show that the obtained differential operator is self-adjoint and the integral operators are symmetric and completely continuous, i.e., the kernels are square integrable. The continuous part of the spectrum of the Fokker–Planck collision operator is then displayed by the study of the continuous spectrum of the second-order differential operator (Strum–Liouville operator). This is possible because the integral operators in the final equations are completely continuous. It is found that the eigenvalues are continuous and extends throughout the half-axis from 0 to $-\infty$. The corresponding eigenfunctions oscillate very fast at large velocity and they are not vectors in a Hilbert space. The latter is a basic feature of the eigenvectors associated with continuous eigenvalues. They are not normalizable, but must be included in order to make the set of the eigenfunctions complete. In Sec. IV, we study the nature of the discrete eigenvalues by converting the dif-

* The first part of this paper was presented at the Seventh International Conference on Phenomena in Ionized Gases, Belgrade, Yugoslavia (1965).

† Present Address: Plasma Physics Laboratory, Princeton University, Princeton, New Jersey.

[1] C. S. Wang-Chang and G. E. Uhlenbeck, "On the Propagation of Sound in Monatomic Gases," Engineering Research Institute Report, University of Michigan (1952).

[2] L. Waldmann, *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1958), Vol. 12.

[3] H. Grad, *Third Symposium on Rarefied Gas* (Academic Press Inc., New York, 1963).

ferential-integral equation into fourth-order differential equations. By studying the behavior of the solutions near the singular points of these fourth-order equations, it is found that, if we require the eigenfunctions to go to zero sufficiently fast as the velocity tends to infinity, the eigenvalues become discrete. A variational principle is constructed and used to compute these discrete eigenvalues and eigenfunctions by choosing the trial functions which tend to zero as

$$\exp(-\tfrac{1}{2}c^2)$$

in the limit of $c \to \infty$ (where c is the normalized velocity as defined below).

II. LINEARIZED FOKKER-PLANCK OPERATION

We start our discussion by writing down the Fokker–Planck equation in the following form:

$$\frac{\partial f}{\partial t} = (2\pi)^3 \frac{n\pi}{m^2} \int d\mathbf{v}_2 \frac{\partial}{\partial \mathbf{v}_1} \cdot \left\{ \left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) f(1) f(2) \cdot \int d\mathbf{K} \, |\phi(K)|^2 \, \mathbf{K}\mathbf{K} \delta[\mathbf{K}\cdot(\mathbf{v}_1 - \mathbf{v}_2)] \right\}. \quad (1)$$

Here, we have assumed the spatial homogeneity. The positive ions are assumed to play the role as neutralization background only. The dynamics of the ions are completely neglected. The collision integral involving identical particles are much more difficult than the collision between ions and electrons, because, in the latter case, there is a great simplification due to the smallness of the mass ratio.

Equation (1) can be obtained from the plasma kinetic equation by putting the dielectric function equal to unity (no screening) in the Lenard–Balescu equation.[4,5] It can also be derived from the Liouville's equation by making the usual weak coupling approximation.[6,7] The presence of the δ function in Eq. (1) indicates that the collisions are weak, i.e., the exchange of momentum during a collision is small and in the direction perpendicular to the velocity of approach of the colliding particles. In the cases we are interested in, where

$$\phi(K) = [4\pi e^2/(2\pi)^3]K^{-2} \quad \text{(coulomb potential)},$$

the K integration in Eq. (1) diverges both at $K = 0$ and ∞. Following Spitzer,[8] we cut the integration off at $K = 1/\lambda_D$ (inverse Debye length) and

$K = 3kT/2e^2$ (inverse minimum impact parameter).

We then obtain the more familiar form of the Fokker–Planck collision integral, i.e.,

$$\frac{\partial f}{\partial t} = \frac{2\pi e^4 n \ln \Lambda}{m^2} \int d\mathbf{v}_2 \frac{\partial}{\partial \mathbf{v}_1} \cdot \left[\frac{v_{12}^2 \mathbf{I} - \mathbf{v}_{12}\mathbf{v}_{12}}{v_{12}^3} \cdot \left(\frac{\partial}{\partial \mathbf{v}_1} - \frac{\partial}{\partial \mathbf{v}_2} \right) f(1) f(2) \right], \quad (2)$$

where

$$\Lambda = 3\lambda_D kT/2e^2.$$

This form of the Fokker–Planck collision integral can also be obtained by expanding Boltzmann integral in term of the smallness of the momentum exchange during collisions. It was done first by Landau[9] and sometimes referred to as Landau's collision integral.

We now linearize around the Maxwellian distribution f_M, i.e., we let

$$f(\mathbf{v}) = f_M(v)[1 + \phi(\mathbf{v})] \quad (3a)$$

and

$$f_M(v) = n(m/2\pi kT)^{\frac{3}{2}} \exp(-mv^2/2kT) \quad (3b)$$

in Eq. (2). By making the following change of notations,

$$\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{v}_{12} \to \mathbf{g}, \qquad \mathbf{v}_1 \to \mathbf{v}, \qquad \mathbf{v}_2 \to \mathbf{v}_1,$$

we obtain the linearized Fokker–Planck equation as follows:

$$\frac{m^2 f_M}{2\pi e^4 \ln \Lambda} \frac{\partial \phi}{\partial t} = J(\phi) \equiv \int d\mathbf{v}_1 \frac{\partial}{\partial \mathbf{v}} \cdot \left[f_M(v) f_M(v_1) \times \frac{g^2 \mathbf{I} - \mathbf{g}\mathbf{g}}{g^3} \cdot \left(\frac{\partial \phi}{\partial \mathbf{v}} - \frac{\partial \phi}{\partial \mathbf{v}_1} \right) \right]. \quad (4)$$

We choose the nondimensional velocity as

$$c = (m/2kT)^{\frac{1}{2}}\mathbf{v}.$$

With a certain amount of manipulation we obtain

$$\partial\phi/\partial t = [3/2(2\pi)^{\frac{1}{2}}](\omega_p \ln \Lambda/\Lambda) J_1(\phi), \quad (5)$$

where

$$J_1(\phi) \equiv \left\{ \left[\frac{\operatorname{erf}(c)}{2c^3}(2c^2 - 1) + \frac{\exp(-c^2)}{2c^2} \right] \mathbf{I} - \left[\frac{\operatorname{erf}(c)}{2c^3}(2c^2 - 3) + \frac{3\exp(-c^2)}{2c^2} \right] \frac{\mathbf{cc}}{c^2} \right\} : \left[\frac{1}{2} \frac{\partial^2 \phi}{\partial \mathbf{c}\, \partial \mathbf{c}} - \mathbf{c} \frac{\partial \phi}{\partial \mathbf{c}} \right] + \left(\frac{\exp(-c^2)}{c^2} - \frac{\operatorname{erf}(c)}{c^3} \right) \mathbf{c} \cdot \frac{\partial \phi}{\partial \mathbf{c}} + 2\exp(-c^2)\phi(\mathbf{c}) + \frac{1}{\pi} \int d\mathbf{c}_1 \exp(-c_1^2) \frac{(c^2 - 1)g^2 - (\mathbf{c}\cdot\mathbf{g})}{g^3} \phi(\mathbf{c}_1) \quad (6)$$

[4] A. Lenard, Ann. Phys. (N. Y.) **10**, 391 (1960).
[5] R. Balescu, Phys. Fluids **3**, 52 (1960).
[6] E. A. Frieman, J. Math. Phys. **4**, 410 (1963).
[7] C. H. Su, J. Math. Phys. **5**, 1273 (1964).
[8] L. Spitzer, *Physics of Fully Ionized Gases* (Interscience Publishers, Inc., New York, 1956).
[9] L. Landau, Physik Z. Sowjetunion **10**, 154 (1936).

and

$$\text{erf}\,(c) = \int_0^c \exp(-x^2)\,dx. \tag{7}$$

The characteristic time in Eq. (5) is given by

$$\tau = [\tfrac{2}{3}(2\pi)^{\frac{1}{2}}](\Lambda/\omega_p \ln \Lambda) \quad \text{with} \quad \omega_p^2 = 4\pi ne^2/m. \tag{8}$$

This time should be of the order of the mean free time. It is related to the "deflection time t_D" defined by Spitzer[8] as

$$\tau = [2.8(2\pi)^{\frac{1}{2}}/9]t_D.$$

Except for the zero eigenvalue, we expect that the first discrete eigenvalue of the operator defined in Eq. (6) will have the same order of magnitude as $1/\tau$. Therefore, we normalize the time by τ, i.e., $t = \tau t_1$, then Eq. (5) reduces to

$$\partial\phi/\partial t_1 = J_1(\phi). \tag{9}$$

We then express the vector $\mathbf{c}$ in a spherical polar coordinate system, i.e., $\mathbf{c}(c, \theta, \chi)$. The collision integral $J_1(\phi)$ in this coordinate takes the following form:

$$\begin{aligned} J_1(\phi) &= \frac{1}{2}\left[\frac{\text{erf}\,(c)}{c^3} - \frac{\exp(-c^2)}{c^2}\right]\frac{\partial^2\phi}{\partial c^2} \\ &+ \frac{1}{2}\left[\frac{\exp(-c^2)}{c^3}(1+4c^2) - \frac{\text{erf}\,(c)}{c^4}(1+2c^2)\right]\frac{\partial\phi}{\partial c} \\ &+ \frac{1}{2}\left[\frac{\text{erf}\,(c)}{2c^3}(2c^2-1) + \frac{\exp(-c^2)}{2c^2}\right] \\ &\times \left[\frac{1}{c^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi}{\partial\theta}\right) + \frac{1}{c^2\sin^2\theta}\frac{\partial^2\phi}{\partial\chi^2}\right] \\ &+ 2\exp(-c^2)\phi(\mathbf{c}) + \frac{1}{\pi}\int c_1^2 \sin\theta'\,dc_1\,d\theta'\,d\chi' \\ &\times \exp(-c_1^2)\frac{(c^2-1)g^2 - (\mathbf{c}\cdot\mathbf{g})}{g^3}\phi(\mathbf{c}_1), \end{aligned} \tag{10}$$

where (θ, χ) and (θ_1, χ_1) are angles used to specify the direction of $\mathbf{c}$ and $\mathbf{c}_1$, respectively, in a given spherical-polar coordinate system, while (θ', χ') are polar angles of $\mathbf{c}_1$ in the spherical coordinate system with $\mathbf{c}$ as the polar axis.

We now expand $\phi(\mathbf{c}, t_1)$ in terms of the surface spherical harmonics, i.e.,

$$\phi(\mathbf{c}, t_1) = \sum_l \psi_l(c, t) Y_{lm}(\theta, \chi), \tag{11}$$

where, by definition, Y_{lm} satisfies the following equation:

$$\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial Y_{lm}}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 Y_{lm}}{\partial\chi^2} = -l(l+1)Y_{lm}. \tag{12}$$

The integral term in Eq. (10) can be simplified by the use of the addition theorem of the spherical harmonics, i.e.,

$$\int_0^{2\pi} d\chi'\, Y_{lm}(\theta_1, \chi_1) = 2\pi P_l(\cos\theta') Y_{lm}(\theta, \chi), \tag{13}$$

where P_l is the Legendre's polynomials.

After separating out the angular dependence of $\phi(\mathbf{c}, t_1)$, we obtain a set of equations governing the radial function $\psi_l(c, t_1)$ as follows:

$$\begin{aligned} \frac{\partial\psi_l}{\partial t_1} &= \frac{1}{2}\left[\frac{\text{erf}\,(c)}{c^3} - \frac{\exp(-c^2)}{c^2}\right]\frac{\partial^2\psi_l}{\partial c^2} \\ &+ \frac{1}{2}\left[\frac{\exp(-c^2)}{c^3}(1+4c^2) - \frac{\text{erf}\,(c)}{c^4}(1+2c^2)\right]\frac{\partial\psi_l}{\partial c} \\ &+ 2\exp(-c^2)\psi_l - \frac{l(l+1)}{4c^4} \\ &\times \left[\frac{\text{erf}\,(c)}{c}(2c^2-1) + \exp(-c^2)\right]\psi_l \\ &+ 2\int_0^\infty c_1^2 \exp(-c_1^2) K_l(c, c_1)\psi_l(c_1)\,dc_1, \end{aligned} \tag{14}$$

where

$$\begin{aligned} K_l(c, c_1) &= \int_0^\pi \sin\theta'\,d\theta'\,\frac{(c^2-1)g^2 - (\mathbf{c}\cdot\mathbf{g})^2}{g^3}P_l(\cos\theta') \\ &= \int_{-1}^1 d\mu\, P_l(\mu)\frac{c^2c_1^2(1-\mu^2) - g^2}{g^3} \end{aligned} \tag{15}$$

with $\mu = \cos\theta'$ and $g = (c^2 + c_1^2 - 2cc_1\mu)^{\frac{1}{2}}$.

The integral in Eq. (15) can be carried out explicitly by making use of the properties of Legendre's polynomials. We then have

$$\begin{aligned} K_l(c, c_1) = \frac{2}{2l+1}\Bigg\{&\frac{c_1^l}{c^{l+1}}\left[\frac{(l+1)(l+2)}{2l+3}c_1^2 - 1\right] \\ &- \frac{l(l-1)}{2l-1}\frac{c_1^l}{c^{l-1}}\Bigg\}, \quad \text{for} \quad c_1 < c, \end{aligned} \tag{16}$$

interchange c and c_1 for $c_1 > c$. We can make Eq. (14) more symmetric by letting

$$\psi_l = \exp(c^2/2)\hat{\psi}_l, \tag{17}$$

we then have

$$\partial\psi_l/\partial t_1 = L\hat{\psi}_l = (2c^2)^{-1}(L_1 + L_2)\hat{\psi}_l, \qquad l = 0, 1, \cdots, \tag{18}$$

where

$$L_1 = \frac{\partial}{\partial c}\left\{\left[\frac{\text{erf}\,(c)}{c} - \exp(-c^2)\right]\frac{\partial}{\partial c}\right\} + q(c), \tag{19}$$

$$L_2 = 4\int_0^\infty dc_1\, c^2 c_1^2 \exp\left(-\frac{c^2+c_1^2}{2}\right) K_l(c, c_1), \qquad (20)$$

$$q(c) = 7c^2\exp(-c^2) - c\,\mathrm{erf}\,(c) - \frac{l(l+1)}{2}\left[\frac{\mathrm{erf}\,(c)}{c^3}(2c^2-1) + \frac{\exp(-c^2)}{c^2}\right]. \qquad (21)$$

Equation (18) is the decoupled differential-integral equation for the radial functions $\hat{\psi}_l$ we have set out to find. It differs structurally from the linearized Boltzmann collision integral by the presence of the differential operator L_1. In the latter case, it is known that the eigenfunctions corresponding to the continuous eigenvalues are delta functions. For the Fokker–Planck collision operator [Eq. (18)], we show in the next section that there is a continuous spectrum for the operator L and the corresponding eigenfunctions are oscillatory waves at large c. In this respect, we see the sensitivity of the spectrum to the perturbation, i.e., even though the Fokker–Planck collision operator is considered to be an approximation to the Boltzmann collision integral for weak collisions, the natures of the eigenfunctions in both cases are quite different!

III. PROPERTIES OF THE OPERATOR L—CONTINUOUS SPECTRUM

Before we discuss the properties of L, we show the functions $\hat{\psi}_l$, which satisfy Eq. (18), are vectors of a Hilbert space. This comes about because of our desire to define a physical quantity entropy, or the Boltzmann's H function. It is readily shown that the relevant H for $\phi(\mathbf{c})$ is

$$H = \int d\mathbf{c}\, F_M \phi^2(\mathbf{c}) = \sum_l \int Y_{lm}^2(\theta, \chi)\sin\theta\, d\theta\, d\chi \int_0^\infty \hat{\psi}_l^2 c^2\, dc. \qquad (22)$$

From this, it is obvious that, in order to have H defined, we must choose the functions $\hat{\psi}_l$ from a Hilbert space with the norm defined by the inner product

$$\int_0^\infty \hat{\psi}_l^2 c^2\, dc < \infty. \qquad (23)$$

We now study the properties of the operators L_1 and L_2. Since $K(c, c_1)$ is symmetry with respect to the interchange of c and c_1, the integral operator L_2 is self-adjoint. We have for L_1

$$\int_0^\infty (\hat{\chi} L_1\hat{\psi} - \hat{\psi} L_1\hat{\chi})\, dc = \left[\frac{\mathrm{erf}\,(c)}{c} - \exp(-c^2)\right]\left[\left(\hat{\chi}\frac{\partial\psi}{\partial c} - \hat{\psi}\frac{\partial\chi}{\partial c}\right)\right]_{c=0}^{\infty} = 0,$$

since

$$\lim_{c\to\infty}\left[\frac{\mathrm{erf}\,(c)}{c} - \exp(-c^2)\right] = \lim_{c\to 0}\left[\frac{\mathrm{erf}\,(c)}{c} - \exp(-c^2)\right] = 0. \qquad (24)$$

Therefore, L_1 is a self-adjoint and thus $(L_1 + L_2)$ is self-adjoint. One concludes that the eigenfunctions corresponding to different eigenvalues must be orthogonal to one another with respect to the weighting function c^2. The negative definiteness of the operator L is difficult to show directly. However, from Eq. (1) or Eq. (2), it is easily shown that

$$\frac{dH}{dt} = \frac{d}{dt}\int d\mathbf{c}\, f(\mathbf{c})\ln f(\mathbf{c}) < 0. \qquad (25)$$

From this and Eq. (22), we see that the operator L is indeed negative definite, and thus all the eigenvalues of L cannot be positive.

Furthermore, it is an easy matter to show that the kernel of the integral operator L_2 is square integrable, i.e.,

$$\left|\int_0^\infty\int_0^\infty c^2 c_1^2\exp(-c^2 - c_1^2)K_l^2(c, c_1)\, dc\, dc_1\right| < \infty, \qquad (26)$$

thus the operator L_2 is a completely continuous operator.[10] Therefore, our operator in Eq. (18) is a sum of a self-adjoint operator L_1 and a completely continuous operator L_2. Now, the addition of a completely continuous operator L_2 to a self-adjoint operator L_1 does not alter the limit points (which include the continuous eigenvalue) of the latter.[11] Hence, in order to study the continuous eigenvalues of $L_1 + L_2$, it is only necessary to investigate the continuous eigenvalue of L_1. This amounts to the study of the eigenvalue problems of the following Strum–Liouville equation;

$$\frac{d}{dc}\left\{\left[\frac{\mathrm{erf}\,(c)}{c} - \exp(-c^2)\right]\frac{d\hat{\psi}_l}{dc}\right\} + (q - 2c^2\lambda)\hat{\psi}_l = 0. \qquad (27)$$

It is quite difficult to obtain the exact solution of Eq. (27) because of its complicated coefficients. However, in order to investigate the nature of its eigenvalue, we need only to study the behavior of its solution around its singular points. Now, the

[10] L. A. Liusternik and V. J. Soholev, *Elements of Functional Analysis* (Frederick Ungar Publishing Company, New York, 1961), p. 131.

[11] F. Riesz and B. SzNagy, *Functional Analysis* (Frederick Ungar Publishing Company, New York, 1955), p. 367.

linear differential equation (27) has a regular singularity point at $x = 0$ and an irregular singularity point at $c \to \infty$. The behavior of its solutions around these two points can be readily obtained.

At $c \to 0$ Eq. (27) reduces to

$$\hat{\psi}_l'' + \frac{2}{c}\hat{\psi}_l' + \left[9 - \frac{3}{2}\lambda - \frac{l(l+1)}{c^2}\right]\hat{\psi}_l = 0. \quad (28)$$

The two independent solutions of Eq. (28) are

$$\hat{\psi}_l = Ac^l + (B/c^{l+1}), \quad (29)$$

where A and B are arbitrary constants.

At $c \to \infty$, Eq. (27) becomes

$$\hat{\psi}_l - \frac{1}{c}\hat{\psi}_l + (a^2c^3 - c^2)\hat{\psi}_l = 0, \quad (30)$$

where $a^2 = -4\lambda/\pi^{\frac{1}{2}} > 0$. After introducing the following transformation:

$$c = x^{2/5}, \qquad \hat{\psi}_l = x^{1/10}y \quad (31)$$

into Eq. (30), we obtain

$$y + (4/25)(a^2 - x^{-2/5})y = 0. \quad (32)$$

The solutions of this equation at large x are oscillatory, i.e.,

$$Y = A \sin \tfrac{2}{5}ax + B \cos \tfrac{2}{5}ax, \quad (33)$$

where A and B are arbitrary constants. Transforming back to the old variables, we have

$$\hat{\psi}_l \simeq (c)^{-\frac{1}{4}} \exp\left[\pm\tfrac{2}{5}aic^{5/2}\right]. \quad (34)$$

Equation (32) can be looked upon as the equation for the scattering of a particle with an energy $(4/25)a^2$ by a potential which goes to $(4/25)x^{-2/5}$ asymptotically. It is apparent that the incoming particle can assume any momentum or energy. This implies that the eigenvalue $a^2 = -4\lambda/\pi^{\frac{1}{2}} > 0$ is continuous and λ can range from zero to minus infinity. For $a^2 < 0$, the eigenvalues are discrete. Since only the continuous eigenvalues of Eq. (27) are relevant to that of Eq. (18), we disregard the discrete part of the spectrum of Eq. (27).

We see that, asymptotically, the eigenfunction corresponding to the continuous eigenvalues in our problem is very similar to the eigenfunction of a free particle in quantum-mechanical problems. They are not the vectors of the Hilbert space. However, by using the continuous normalization one can express any vector in the Hilbert space as the sum over the discrete eigenfunctions plus the integral over the continuous eigenfunctions.

In the linearized Boltzmann equation, the collision integral is a pure integral operator. The continuous spectrum there comes solely from an algebraic operator $\nu(c)$. It is seen that the corresponding eigenfunctions are Dirac's delta functions. They are also not square integrable, but quite different from the continuous eigenfunctions for the linearized Fokker–Planck operator which are circular functions asymptotically. This demonstrates the sensitivity of the eigenspectrum caused by perturbation of the operator. Even though the Fokker–Planck operator is an approximation of the Boltzmann collision integral, the natures of the spectra (at least their continuous spectra) in the two cases are dramatically different.

IV. DISCRETE EIGENVALUES OF THE OPERATOR L

In the above section, we found that the operator L has the continuous eigenvalues ranging from zero to minus infinity. In the present section, we investigate the nature of its discrete eigenspectrum and thus establish a scheme which enables us to construct the discrete part of the spectrum of the operator L.

For the sake of simplicity, we focus our attention on the case where $l = 0$. Extension to $l = 1, 2 \cdots$ can be readily made. For $l = 0$, the eigenvalue problem of Eq. (18), i.e.,

$$(L_1 + L_2)\hat{\psi}_0 = 2c^2\lambda\hat{\psi}_0, \quad (35)$$

can be transformed into a pure fourth-order differential equation by proper differentiations of Eq. (35). The resulting equation with the omission of the subscript 0 at $\hat{\psi}_0$ is as follows:

$$\begin{aligned}
&\frac{1}{c}\left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2)\right]\hat{\psi}^{IV} + \left\{6\exp(-c^2)\right. \\
&+ \left(2 - \frac{5}{c^2} - \frac{4}{2c^2-1}\right)\left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2)\right]\Big\}\hat{\psi}''' \\
&+ \left\{\left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2)\right]\left(\frac{12}{c^3} - \frac{7}{c} + \frac{10}{c}\frac{1}{2c^2-1}\right)\right. \\
&- \left(\frac{8}{c} - 2c\right)\exp(-c^2) - \frac{16c\exp(-c^2)}{2c^2-1} - 2\lambda c\Big\}\hat{\psi}'' \\
&+ \left\{\left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2)\right]\right. \\
&\times \left(-\frac{12}{c^4} + \frac{8}{c^2} - \frac{9}{c^2}\frac{1}{2c^2-1} + 1 - 2c^2\right) \\
&+ \left(2 - 14c^2 + \frac{8}{c^2}\right)\exp(-c^2) + \frac{4\lambda}{2c^2-1} - 4\lambda c^2\Big\}\hat{\psi}' \\
&+ \left\{\left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2)\right]\right.
\end{aligned}$$

$$\times\left(\frac{1}{c}+c-c^3+\frac{1}{c}\frac{1}{2c^2-1}\right)$$

$$+\left(2c-10c^3-\frac{8c}{2c^2-1}\right)\exp(-c^2)$$

$$+\frac{12\lambda c}{2c^2-1}-2\lambda c(1+c^2)\Big\}\hat{\psi}=0. \tag{36}$$

We first study the solutions of Eq. (36) around its singular points which are at $c = 0, 1/\sqrt{2}, \infty$.

At $c \to 0$, the singularity is a regular one and the equation around $c = 0$ has the following form:

$$\hat{\psi}^{IV}+\frac{4}{c}\hat{\psi}'''+\left(\frac{74}{5}-\frac{3\lambda}{2}\right)\hat{\psi}''$$

$$+\left(\frac{76}{5}-\frac{3\lambda}{2}\right)\frac{1}{c}\hat{\psi}'+\left(15-\frac{21\lambda}{2}\right)\hat{\psi}=0. \tag{37}$$

It is readily shown that there are three analytic solutions and the fourth solution is of the form of $1/c$ (bad solution).

At $c \to 1/\sqrt{2}$ Eq. (36) reduces to

$$\left(c-\frac{1}{\sqrt{2}}\right)\hat{\psi}^{IV}-\hat{\psi}'''$$

$$+F_1(c)\hat{\psi}''+F_2(c)\hat{\psi}'+F_3(c)\hat{\psi}=0, \tag{38}$$

where F_1, F_2, and F_3 are analytic functions of c around $c = 1/\sqrt{2}$. It can be shown that Eq. (38) has three analytic solutions and one bad solution of the form of $\ln(c - 1/\sqrt{2})$.

At $c \to \infty$, the singularity is irregular, we expect the solutions to have exponential forms. The equation for large c is

$$\hat{\psi}^{IV}+2c\hat{\psi}'''+a^2c^3\hat{\psi}''+2a^2c^4\hat{\psi}'+a^2c^5\hat{\psi}=0, \tag{39}$$

and its four independent solutions are

$$\hat{\psi}\sim(c)^{-\frac{1}{2}}\exp\left[\pm\frac{2ai}{5}c^{5/2}\right],\ \exp[-\tfrac{1}{2}c^2\pm c]. \tag{40}$$

We see that the first two oscillatory solutions are exactly the same as the one we have found for the differential operator. If we admit these oscillatory waves as the possible eigenfunctions we have continuous eigenspectrum which is the same as the one given by Eq. (27). However, if we restrict ourselves to functions which tend to zero at least as fast as $\exp(-\frac{1}{2}c^2)$ for large c, the first two solutions in Eq. (40) have to be ruled out as bad solutions. Now, for a fourth-order differential equation, we have three degrees of freedom (four arbitrary integration constants minus one scale factor) to satisfy various restrictions on the solution of the equation imposed by the regularity of the solutions throughout the whole range of integration. In our present problem, there are four restrictions [for functions to decay as $\exp(-\frac{1}{2}c^2)$ as $c \to \infty$]: one at $c = 0$, one at $c = 1/\sqrt{2}$, and two at ∞. In order to satisfy all these restrictions, we must make use of the parameter λ in Eq. (36). Therefore, the eigenvalue λ in Eq. (36) cannot take arbitrary values; they must be discrete. This also proves that the eigenfunctions which correspond to the discrete eigenvalues must go to zero as fast as $\exp(-\frac{1}{2}c^2)$ in the limit of $c \to \infty$.

The asymptotic property of the discrete eigenfunction discussed in the previous paragraph enables us to construct the discrete part of the spectrum of L. First, we construct a variational principle for the operator L by defining the following two functionals:

$$I = (\chi, L\chi), \tag{41}$$

$$R = (\chi, \chi). \tag{42}$$

The inner products in (41) and (42) are defined with the weighting function c^2, i.e.,

$$(\chi, \chi) = \int_0^\infty dc\, c^2\chi^2(c). \tag{43}$$

It can be shown (see Appendix) that the first nonzero eigenvalue of the operator L is the maximum value of the following functionals

$$\lambda(\chi) = -\tfrac{1}{2}[I(\chi)/R(\chi)]. \tag{44}$$

In choosing trial functions to maximize $\lambda(\chi)$, we adopt the Ritz method by expression trial functions as linear combination of Ritz manifold, i.e.,

$$\chi(c) = \sum_{i=1}^{n} a_n q_n(c), \tag{45}$$

where q_n's are chosen to be orthogonal polynomials with the convergent factor $\exp(-\frac{1}{2}c^2)$. For our present problem, we take the polynomials to be the Laguerre's polynomials (or equivalently Sonnine polynomials). Furthermore, to make the computation most economic, we make the vectors $q_i(c)$ orthonormal, i.e.,

$$\int_0^\infty dc\, c^2 q_i(c)q_j(c) = \delta_{ij}. \tag{46}$$

It can be shown that the original eigenvalue problem of the operator L is then equivalent (up to the order of approximation, n) to the eigenvalue problem of the following determinant:

$$|A_{ij} + 2\lambda\,\delta_{ij}| = 0, \tag{47}$$

where

$$A_{ij} = \int_0^\infty dc \left\{ \left[\frac{\operatorname{erf}(c)}{c} - \exp(-c^2) \right] \times \frac{dg_i}{dc}\frac{dg_j}{dc} - q(c) g_i(c) g_j(c) \right\} - 4 \int_0^\infty dc \int_0^c dc_1 \, \hat{K}_0(c, c_1)[g_i(c)g_j(c_1) + g_i(c_1)g_j(c)] \tag{48}$$

and

$$\hat{K}_0(c, c_1) = c_1^2 c^2 \exp\left(-\frac{c^2 + c_1^2}{2} \right) K_0(c, c_1). \tag{49}$$

By choosing $q_i(c)$ as the properly normalized Laguerre's polynomials, one can get better and better approximation to the eigenvalues of the operator L. A numerical computation of the eigenvalues and their corresponding eigenfunctions is under way. The results will be reported in a future publication.

V. DISCUSSION

We have shown that the spectrum of the Fokker–Planck collision integral consists of two parts: (1) discrete spectrum with the eigenvectors decay asymptotically as $e^{-\frac{1}{2}c^2}$ as $c \to \infty$; (2) continuous spectrum with the eigenvectors oscillating very rapidly for large c and consequently are not vectors in Hilbert space. Since the collision operator is self-adjoint, the eigenfunctions are complete. However, for most physical disturbances, it is reasonable to assume that they have only a finite extent in the velocity space and are cut off as $e^{-\frac{1}{2}c^2}$ for large c, because of the finiteness of the energy. One can then investigate this problem by expanding the perturbed function in terms of discrete eigenspectrum alone. This in fact is intimately connected to the possibility of computing transport coefficients of the Fokker–Planck equation by the variational method. In our forthcoming paper, we shall compute extensively the discrete eigenspectrum and apply the results to study the collisional effect on plasma waves.

ACKNOWLEDGMENTS

This work was supported by the Advanced Research Projects Agency (Ballistic Missile Defense Office) and technically administered by the Fluid Dynamics Branch of the U. S. Office of Naval Research under Contract Nonr-1841(93).

APPENDIX. VARIATIONAL PRINCIPLE FOR THE DISCRETE EIGENVALUE OF L

Suppose $\chi_0(c)$ maximizes the functional

$$\lambda[\chi(c)] = -\tfrac{1}{2}\{I[\chi(c)]/R[\chi(c)]\}, \tag{A1}$$

we let $\chi(c) = \chi_0(c) + \epsilon f(c)$ be any nearby function. Then, by definition, we have

$$\frac{\delta}{\delta\epsilon}\left(-\frac{1}{2}\frac{I}{R} \right) = -\frac{1}{2R}\left[\frac{\delta I}{\delta\epsilon} + 2\lambda \frac{\delta R}{\delta\epsilon} \right]_{\epsilon=0} = 0. \tag{A2}$$

Now,

$$\left.\frac{\delta I}{\delta\epsilon}\right|_{\epsilon=0} = -2\int_0^\infty dc \, \frac{\delta\chi(c)}{\delta\epsilon}\left\{ \frac{d}{dc}\left[\left(\frac{\operatorname{erf} c}{c} + e^{-c^2} \right)\frac{d\chi_0}{dc} \right] + q(c)\chi_0(c) + 4\int_0^\infty \hat{K}_1\chi_0(c_1)\, dc_1 \right\}$$

and

$$\left.\frac{\delta R}{\delta\epsilon}\right|_{\epsilon=0} = 2\int_0^\infty dc \, c^2 \frac{\delta\chi}{\delta\epsilon}\chi_0(c).$$

Substitute these into Eq. (A2) and we obtain

$$\frac{d}{dc}\left[\left(\frac{\operatorname{erf} c}{c} - e^{-c^2} \right)\frac{d\chi_0}{dc} \right] + (q - 2\lambda c^2)\chi_0 + 4\int_0^\infty \hat{K}_1\chi_0(c_1)\, dc_1 = 0. \tag{A3}$$

This shows that $\chi_0(c)$, which maximizes the functional $\lambda[\chi(c)]$, is a solution of our eigenvalue problem, and the corresponding $\lambda[\chi_0(c)]$ is the eigenvalue.

Relations between Field-Plus Source and Fokker-Type Action Principles

JAMES L. ANDERSON*
Stevens Institute of Technology, Hoboken, New Jersey
AND
SAMUEL SCHIMINOVICH†
City College of the City University of New York, New York
(Received 4 March 1966)

The problem of constructing a Fokker-type action that will yield equations of motion for the sources of an original field-plus-source system is considered. It is shown that such an action can be constructed for a large class of systems by substituting for the fields appearing in the original field-plus-source action a solution of the field equations in terms of the source variables provided that certain conditions are satisfied. In general, these conditions will be satisfied only by half-advanced, half-retarded-type solutions of the field equations and then only for a field-plus-source Lagrangian that differs from the usual one by a complete divergence. The problem of constructing a Fokker action for a system that possesses a gauge-type covariance is complicated by the existence of differential identities that are satisfied by the left-hand sides of the field equations and that arise as a consequence of the gauge covariance. These difficulties can be overcome by the introduction of gauge conditions on the field variables. The problem of finding solutions to the field equations suitably modified with the help of the gauge conditions that satisfy both the gauge conditions and the conditions for the construction of the Fokker action is discussed. As an application of the method, we have obtained the Fokker action of electrodynamics using a number of different gauge conditions, arriving thereby at several different expressions for this action. We have also considered the construction of a Fokker action when the field equations are solved by an approximation method. In such cases we have shown that to determine the Fokker action up to the nth order of the approximation it is only necessary to solve the field equations up to order $\frac{1}{2}(n+2)$ or $\frac{1}{2}(n+1)$ for n even or odd, respectively. The method is then applied to obtain the Darwin action of electrodynamics.

I. INTRODUCTION

IN this paper we study the problem of constructing a Fokker action principle for a system of interacting particles by using an extension of the substitution method first introduced by Plebanski and Bazanski.[1] The results of this study are thus a contribution to the overall problem of the relationship between the two chief ways of describing the interactions of material particles: action-at-a-distance or field-mediated interactions.

While the action-at-a-distance concept preceded the field concept historically and was adequate for the description of static interactions, the field concept appeared to be essential for a description of those interaction processes that involve the emission and transmission of energy. Nevertheless, over the years there has been a continuing effort to eliminate the field, either directly or indirectly, as an essential element in the description of nature. In the case of electrodynamics this approach has been most thoroughly discussed by Feynman and Wheeler,[2] who based their work on an old idea of Schwarzschild and Fokker.[3] They were able to show in this case that the action-at-a-distance description contained all of the physics originally contained in Maxwell's equations when one could neglect retardation effects. These latter effects were then obtained as a consequence of statistical mechanics in a completely absorbing universe. While not its primary goal, a consequence of Dirac's work on the classical theory of radiating electrons[4] was a set of equations of motion for a number of interacting electrons from which the field elements could be easily eliminated.

In the field of general relativity a large body of literature has grown up which concerns itself with the problem of deriving the equations of motion for material bodies from the field equations of Einstein. Einstein and Grommer[5] were the first to show that these equations of motion follow directly from the field equations and did not have to be postulated separately from them. The problem of actually obtaining these equations by an iteration scheme was first solved by Einstein, Infeld, and

* Work supported in part by the National Science Foundation.
† Work partially supported by Yeshiva University, New York, New York.
[1] J. Plebanski and S. Bazanski, Acta Phys. Polon. **18**, 307(1959).
[2] R. P. Feynman and J. A. Wheeler, Rev. Mod. Phys. **21**, 425(1949).
[3] K. Schwarzschild, Göttinger Nachr. **128**, 132 (1903); A. D. Fokker, Z. Physik **58**, 386 (1929); Physica **9**, 33 (1929).
[4] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938).
[5] A. Einstein and J. Grommer, Sitzber. Preuss. Akad. Wiss. 1, 2 (1927).

Hoffmann[6] and since that time they have been rederived and studied extensively by many workers, notably Infeld and his school in Poland. Again, in all of these works, the essential idea is to obtain a set of equations of motion for material bodies in which the gravitational field does not appear directly.

The usual procedure employed to eliminate the fields from the description of particle interactions is to start with the field-plus-source equations of motion. One looks first for a solution of the field equations in terms of the particle variables which appear in the source terms of these equations. One then uses this solution to eliminate the field variables that appear in the force terms of the particle equations of motion, obtaining thereby what are called the Fokker equations of motion for the sources of the field. In electrodynamics, such a procedure leads to the action-at-a-distance equations of motion for charged particles.

In attempting to obtain the Fokker equations of motion directly from the field-plus-source equations of motion two types of problems arise. First, the Fokker equations are not unique due to the fact that one must impose boundary conditions in order to obtain solutions of the field equations. Thus, in the case of electrodynamics, one has available for substitution into the particle equations of motion either the retarded or the advanced solutions or some linear combination thereof, each leading to different Fokker equations of motion. In particular, in these equations, one can or cannot obtain a radiative damping term, depending upon the solution chosen. The second problem is not one of principle as is the first, but arises as a consequence of the enormous amount of labor involved in obtaining the Fokker equations when the field equations must be solved by an approximation procedure. As a consequence we have been led to examine the question of when it is possible to construct an action principle from the original field-plus-source action that yields directly a set of Fokker equations of motion.

The main result of our work is that it is indeed possible to construct a Fokker action for a large class of systems that overcomes, at least in part, the above-mentioned problems. While such actions have been given in the past, they were usually presented as a new postulate[3] or derived in a heuristic manner[7] from the action for the field-plus-source equations of motion. In this paper we derive the conditions that allow one to construct a Fokker action in a rigorous manner from the field-plus-source action. In particular we show that such a construction is possible only if one uses half-advanced, half-retarded solutions of the field equations. Thus, of all possible Fokker equations of motion, we show that only one can be derived from an action principle and that this action is unique.

In this paper, we give particular attention to the important case when the original field-plus-source equations of motion possess a gauge-type covariance property as in electrodynamics and general relativity. Here the construction of a Fokker action is complicated by the existence of the so-called Bianchi identities which are satisfied by the field equations and which arise as a consequence of the gauge covariance of the theory. Because of these identities one cannot solve the field equations for arbitrary motions of the sources but only for a subset of such motions, and in general one does not know what this subset is until one has obtained a solution of the field equations. The way out of the difficulty indicated here is to impose gauge conditions (coordinate conditions in general relativity) on the field variables and to use these conditions to obtain a new set of field equations that do not satisfy any identities. One can now look for solutions of these modified equations that both satisfy the gauge conditions and at the same time lead to a Fokker action. In the case of electrodynamics such solutions are shown to exist and lead to various equivalent expressions for the Fokker action, depending on the form of the gauge condition employed.

The last part of this paper is devoted to a discussion of the problems involved in constructing a Fokker action when the field equations are to be solved by some approximation procedure. It is here that the procedural advantage of obtaining Fokker equations of motion via an action principle becomes apparent. If one attempts to find these equations by substituting directly into the equations of motion of the sources for the fields, it is necessary to solve the field equations up to the nth order to obtain Fokker equations of motion that are good to this order. However, we show that it is possible to obtain the corresponding Fokker action, and hence the Fokker equations of motion to this order, by solving the field equations only up to order $\frac{1}{2}(n+2)$ or $\frac{1}{2}(n+1)$ for n even or odd, respectively. As an application we apply the method to electrodynamics to obtain the Darwin Lagrangian. Elsewhere we will show that for the gravitational case it leads to particle equations of motion that agree, up to the

[6] A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. **39**, 65 (1938).

[7] L. Infeld, Rev. Mod. Phys. **29**, 398 (1957).

order considered, with the Fokker equations of motion obtained by Bertotti and Plebanski[8] and by Kerr.[9]

There are a number of important questions connected with the construction of Fokker actions that we do not discuss here. Foremost among these are the various self-energy problems and the problem of renormalization. Also, there is the question of the convergence of the method used here when it is necessary to solve the field equations by an approximation procedure.

II. CONSTRUCTION OF FOKKER ACTIONS

Let us consider a field-source system described by the field variables φ_A and source coordinates q_i. We assume that the equations of motion for these quantities follow from a variational principle in the usual manner and that the action has the form[10]

$$S = S_F(\varphi) + S_M(q) + S_I(\varphi, q),$$
$$\equiv \int_\Omega d^4x \, \{\mathfrak{L}_F(\varphi) + \mathfrak{L}_M(q) + \mathfrak{L}_I(\varphi, q)\}. \qquad (2.1)$$

The equations of motion for the field take the form

$$[\delta\mathfrak{L}_F(\varphi)/\delta\varphi_A] + [\delta\mathfrak{L}_I(\varphi, q)/\delta\varphi_A] = 0, \qquad (2.2)$$

and those of the sources take the form

$$[\delta\mathfrak{L}_M(q)/\delta q] + [\delta\mathfrak{L}_I(\varphi, q)/\delta q_i] = 0. \qquad (2.3)$$

They are obtained from the requirement that the action be stationary when subjected to a set of allowed variations, namely variations that vanish at the boundaries of the domain Ω.

To obtain a set of Fokker equations of motion for the sources alone, one would first solve Eq. (2.2) for the φ_A in terms of the q_i after imposing boundary conditions. When we substitute the solution $\varphi_A = \varphi_A[q]$ into Eqs. (2.3), we obtain

$$[\delta\mathfrak{L}_M(q)/\delta q_i] + [\hat{\delta}\mathfrak{L}_I(\varphi[q], q)/\hat{\delta}q_i] = 0, \qquad (2.4)$$

where $\hat{\delta}/\hat{\delta}(q_i)$ in the second term indicates that we are to take the variational derivative with respect on'y to those q's that appeared explicitly in $\mathfrak{L}_I$ before the substitution for the φ's in terms of the q's.

The question now arises, do the Fokker equations of motion (2.4) follow from an action principle, called a Fokker action principle, and if so what is this action?

One may believe that substitution of our solution of $\varphi_A(q)$ directly into the action (2.1) would yield this action.[1] In this section, we analyze the conditions under which this substitution method does indeed give correct results or, as we often say, "a good Fokker action." By the latter, we understand a Fokker action that is made stationary by trajectories of the q's as determined by Eqs. (2.4). This analysis may also be considered as a study of the conditions under which the field φ is a completely redundant element in the description of our system.

Stationary variation of the substituted action $S^F(q)$ implies

$$\delta S^F = \int_\Omega d^4x \left\{\frac{\delta S_F}{\delta\varphi} + \frac{\delta S_I}{\delta\varphi}\right\} \delta^*\varphi + \int_\Sigma d^3S_\mu P^\mu_\varphi \, \delta^*\varphi$$
$$+ \int_\Omega d^4x \left\{\frac{\hat{\delta} S_I}{\hat{\delta} q} + \frac{\delta S_M}{\delta q}\right\} \delta q + \int_\Sigma d^3S_\mu P^\mu_q \, \delta q = 0 \qquad (2.5)$$

or

$$\delta S^F = \left[\left(\frac{\delta S_F}{\delta\varphi} + \frac{\delta S_I}{\delta\varphi}\right)\cdot \delta^*\varphi\right]$$
$$+ \left[\left(\frac{\delta S_M}{\delta q} + \frac{\hat{\delta} S_I}{\hat{\delta} q}\right)\cdot \delta q\right] = 0, \qquad (2.6)$$

where we have used the symbolic compact notation $[(\delta S/\delta q)\cdot\delta q]$. It stands for the sum of all the volume and surface integrals that we obtain when the variation of S with respect to q is expressed linearly in terms of only δq so that no derivatives of δq occur. Thus

$$[(\delta S/\delta q)\cdot\delta q] \equiv S(q + \delta q) - S(q). \qquad (2.7)$$

We have also introduced the notation δ^*q, which stands for an induced variation of φ when the q's on which it functionally depends are varied. That is,

$$\delta^*\varphi \equiv \int d^4x' \, \frac{\delta\varphi[q(x)]}{\delta q(x')} \, \delta q(x'). \qquad (2.8)$$

If the first bracket in (2.6) is equal to zero, the Fokker equations of motion (2.4) follow because δq is the set of allowed variations for the q's in the original action (2.1), namely the set of arbitrary variations that vanish on the boundaries. Conversely, if Eq. (2.4) is satisfied and follows from a variation of S^F, the first bracket has to vanish. Thus we have

$$[(\delta S_F/\delta\varphi + \delta S_I/\delta\varphi)\cdot\delta^*\varphi] = 0 \qquad (2.9)$$

as a necessary and sufficient condition that Eqs.

---

[8] B. Bertotti and J. Plebanski, Ann. Phys. (N. Y.) 11, 169 (1960).
[9] R. P. Kerr, Nuovo Cimento 13, 469, 492 (1959).
[10] In general relativity the term S_M is missing in the action. The presence or absence of such a term, however, is not crucial to the discussion that follows.

(2.4) follow from a variational principle whose action S^F is obtained by the substitution method. Let us note here that condition (2.9) can, with the help of Eqs. (2.5) and (2.2) also be written as

$$\int_\Omega d^3S_\mu P_\varphi^\mu \, \delta^*\varphi = 0. \tag{2.10}$$

Careful consideration of condition (2.9) for each particular case shows that the substitution method gives a good Fokker action for a wide class of field-plus-source actions but then only for very specific solutions of the field equations. In particular, condition (2.9) is satisfied for a solution $\varphi_A(q)$ of the field equations whenever the $\delta^*\varphi$ are a subset of the allowed variations for the φ in the original action (2.1). This is seldom the case, however, because the $\delta^*\varphi$ do not vanish, in general, on the boundaries and so we find that condition (2.9) imposes restrictions on the solutions $\varphi_A(q)$ that will give rise to a good Fokker action. For example, it is enough that the derivatives of the field solution tend to zero as $1/r^2$ when r goes to infinity and that the motion of the sources be bounded in order to have all surface integrals in (2.10) go to zero in the limit, thus assuring the satisfaction of the condition. It is with this criteria that the usual approaches to the substitution method stop.[1] By doing this they leave out of consideration important cases like fast motion electrodynamics, for which in fact the original Fokker action was formulated.

We may further discuss the following important example in which we do not assume any asymptotic behavior for the field solutions. Let the original action S be of the form

$$S = \int_\Omega d^4x \, \{\varphi \mathfrak{D}\varphi - \varphi^n j(q) + \mathfrak{L}_M(q)\}, \tag{2.11}$$

where $\mathfrak{D}$ is a self-adjoint differential operator. It then follows that φ satisfies

$$\mathfrak{D}\varphi = \tfrac{1}{2}n\varphi^{n-1} j \tag{2.12}$$

while $\delta^*\varphi$ in its turn satisfies

$$\mathfrak{D}\,\delta^*\varphi = \tfrac{1}{2}n(n-1)\varphi^{n-2}\,\delta^*\varphi\cdot j + \tfrac{1}{2}n\varphi^{n-1}\,\delta^* j. \tag{2.13}$$

With the help of these two relations, condition (2.9) for the action (2.11) becomes

$$\int_\Omega d^4x \, \varphi^{n-1}[\varphi\,\delta^* j + (n-2)\,\delta^*\varphi\cdot j] = 0. \tag{2.14}$$

This shows that for the particular case $n = 2$ no solution gives a good Fokker action since $\delta^* j$ is arbitrary within the region Ω. This was to be expected because for any solution when $n = 2$ we have $S^F = S_M$. For the usual case $n = 1$ Eq. (2.14) requires symmetric solutions. Indeed, for $n = 1$, any solution of the field equations may be written, with the help of a Green's function $D(x, x')$ associated with the operator $\mathfrak{D}$, as

$$\varphi(x) = \int_R d^4x' \, D(x, x') j(x') + \varphi^0(x), \tag{2.15}$$

where φ^0 is some solution of the homogeneous field equations and the domain of integration R is all of space–time. Substitution of (2.15) into (2.14) leads then, provided the domain of integration Ω coincides with R, to

$$\varphi^0 = 0 \quad \text{and} \quad D(x, x') = D(x', x) \tag{2.16}$$

as sufficient conditions for the satisfaction of the Eq. (2.14). The question of whether they are also necessary depends on the functional dependence of j on the basic variables of the theory. It can be seen that the answer is affirmative for all cases of interest considered below.

It is instructive to consider the following example, where the two previous cases $n = 1$ and 2 appear mixed together:

$$S = \int_\Omega d^4x \, \{\varphi \mathfrak{D}\varphi + \varphi\Psi^2 + \Psi\mathfrak{D}\Psi\}. \tag{2.17}$$

Here, both φ and Ψ appear to have, at first, the same field status. However, we can eliminate only φ in favor of Ψ to get the good Fokker action

$$S^F = \int_R d^4x \, \{\Psi\mathfrak{D}\Psi - \frac{1}{8}\int_R d^4x'\Psi^2(x) \times [D(x, x') + D(x', x)]\Psi^2(x')\}. \tag{2.18}$$

From the physical point of view, one would then tend to consider Ψ as the "particle" or "matter" field, whereas φ mediates the interaction of Ψ with itself. The discussion of the general nonlinear cases is difficult except in the framework of an approximation procedure which achieves linearization at each of its stages.

In our example (2.11) we have made a particular choice for the Lagrangian of the φ field which differs in fact from the usual one by a complete divergence. The latter is immaterial for the equations of motion of the field but it contributes to the condition (2.9). Hence, for the construction of a good Fokker action it *does* matter from which particular expression of the action we choose to start. The question then arises as to whether or not the Fokker action obtained for a given field-plus-source theory is unique.

To discuss the question of uniqueness we restrict ourselves to linear theories, since, even in the nonlinear case, one solves a set of linear equations at

each stage of an approximation procedure. Given a Lagrangian in the form (2.11) together with our symmetric solutions, we ask whether condition (2.9) can be satisfied again after adding to the Lagrangian a suitable complete divergence $\partial_\mu Q^\mu(\varphi; \varphi_{,\nu})$ and to the symmetric solution, a solution χ of the homogeneous field equations that nevertheless depends on the j. Because Eq. (2.14) is satisfied by the symmetric solution, it follows that condition (2.9) is satisfied provided that

$$\frac{1}{2}\int_\Omega d^4x\ (\delta^*\chi\cdot j - \delta^* j\cdot\chi) = \int_\Omega d^4x\ \delta^*(\partial_\mu Q^\mu). \quad (2.19)$$

The following possibilities may now be discussed.

(a) if $\partial_\mu Q^\mu$ is such that, upon substitution,

$$\int_\Omega d^4x\ \delta^*(\partial_\mu Q^\mu) = \int_\Sigma d^3S_\mu\ \delta^* Q^\mu = 0, \quad (2.20)$$

we recover by means of Eq. (2.19) the condition that χ should again be purely symmetric. Every complete divergence that does not vanish upon substitution gives rise to an expression that differs from the good Fokker action by a complete divergence. Such a term, however, is immaterial for an action principle since its variation is always zero.

(b) We drop the assumption (2.20) in order to ask ourselves whether, by adding a proper $\partial_\mu Q^\mu$ to the Lagrangian (2.11), solutions other than the symmetric ones could give us a good Fokker action. Those nonsymmetric solutions differ from a symmetric solution by a solution χ of the homogeneous field equation. It can be written very generally as

$$\chi(x) = \int_R d^4x'\ K(x, x')j(x') + \chi_0(x), \quad (2.21)$$

where $K(x, x')$ is not symmetric and $\chi_0(x)$ is functionally independent of j. We assume $\chi_0 = 0$. The complete divergence $\partial_\mu Q^\mu$ is a function of the ϕ's and derivatives and can be expanded as

$$\partial_\mu Q^\mu = \sum_{m=1}^{m=n}\int d^4x \int d^4x' \cdots$$
$$\times \int d^4x^m\ Q^{(m)}(x, x'x'', \cdots, x^m)j(x')j(x'')\cdots j(x^m). \quad (2.22)$$

Its contribution to the δ^* variation of S^F can be written as

$$\delta^*\int_\Omega \partial_\mu Q^\mu\, d^4x = \sum_{m=1}^{m=n}\int d^4x' \int d^4x'' \cdots$$
$$\times \int d^4x^m\ \bar{Q}^{(m)}(x', x'', \cdots, x^{(m)})$$
$$\times\ \delta^* j(x')j(x'')\cdots j(x^m), \quad (2.23)$$

where the important point is that the $\bar{Q}^{(m)}$ are symmetric in their arguments. From Eq. (2.19) and the arbitrariness of the $\delta^* j(x)$ we have

$$\int dx''\ [K(x', x'') - K(x'', x')]j_\mu(x'')$$
$$= \sum_{m=1}^{m=n}\int d^4x'' \int d^4x''' \cdots$$
$$\times \int d^4x^m\ \bar{Q}^{(m)}(x'x'', \cdots, x^m)j(x'')j(x''')\cdots j(x^m). \quad (2.24)$$

Multiplying both sides by $j(x')$ and integrating, we get

$$\sum_{m=1}^{m=n}\int d^4x' \int d^4x'' \int d^4x''' \cdots$$
$$\times \int d^4x^m\ \bar{Q}^m(x', x'', x''' \cdots, x^m)$$
$$\times\ j(x')j(x'')j(x''')\cdots j(x^m) = 0. \quad (2.25)$$

Because this relation holds for arbitrary $j(x)$ we conclude that the $\bar{Q}^{(m)}$ have to vanish and hence $\delta^*\int \partial_\mu Q^\mu\, d^4x$ vanishes, thus contradicting our starting assumption. This tells us that there is no way one can add an antisymmetric part to our solution and still obtain a good Fokker action.

(c) The only other non-uniqueness that might arise could occur if the field equations admitted more than one symmetric solution. We can rule out this case if we require that field equations must yield a unique solution for a given set of boundary conditions.

III. THEORIES WITH GAUGES

A special problem arises when one applies the substitution method to a system of equations that admit a gauge-type covariance group. Because of this gauge covariance, the field equations for the φ variables satisfy a set of differential identities equal in number to the number of arbitrary functions needed to define an element of the group.[11] If now the φ field is coupled to sources these identities lead to integrability conditions to be satisfied by the sources. Thus, in electrodynamics they imply the conservation of charge, while in general relativity they form the basis for the Einstein, Infeld, and Hoffmann approximation method.[6] As a consequence, it is not possible to solve the field equations directly to obtain φ as a function of the source

[11] See, for example, A. Trautman, in *Gravitation, an Introduction to Current Research*, L. Witten, Ed., (John Wiley & Sons, Inc., New York, 1962), Chap. 5.

motions; only a restricted subclass of such motions allows such a solution. But since it is just the motion of the sources that we wish to determine by means of our substitution method we seem to have arrived at an impasse.

To investigate this situation in more detail let us return to the equations of motion (2.2) and (2.3). If we introduce the abbreviation

$$\mathfrak{L}^A(\varphi) \equiv \delta\mathfrak{L}_F/\delta\varphi_A, \tag{3.1}$$

then, because of the gauge covariance the $\mathfrak{L}^A$ satisfy a number of "Bianchi" identities:

$$\mathfrak{B}_\alpha(\mathfrak{L}^A) \equiv 0. \tag{3.2}$$

[In electrodynamics, $\mathfrak{L}^A \Rightarrow F^{\mu\nu}{}_{,\nu}$ and $\mathfrak{B}_\alpha(\mathfrak{L}^A) \Rightarrow F^{\mu\nu}{}_{,\mu\nu} \equiv 0$.] When the ϕ field is coupled to a source, the field equations take the form (2.2), which we write as

$$\mathfrak{L}^A(\varphi) = j^A(\varphi, q), \tag{3.3}$$

where the "current" j^A is given by

$$j^A(\varphi, q) \equiv \delta\mathfrak{L}_I(\varphi, q)/\delta\varphi_A. \tag{3.4}$$

Then, because of the Bianchi identities (3.2), the sources must satisfy

$$\mathfrak{B}_\alpha(j^A) = 0. \tag{3.5}$$

Only for source motions that satisfy Eqs. (3.5) will it be possible to solve Eqs. (3.3) for φ as a function of the q. These equations therefore constitute a set of integrability conditions on solutions of the field equations.

In order to solve the field equations (3.3) without having to take account of the integrability conditions, one imposes gauge conditions (coordinate conditions in general relativity) on the φ. These are a set of local, noncovariant relations

$$f_\alpha(\varphi) = 0 \tag{3.6}$$

equal in number to the number of Bianchi identities that can always be satisfied by an appropriate choice of gauge. With their help one then constructs a modified set of field equations

$$\tilde{\mathfrak{L}}^A(\varphi) = j^A(\varphi, q). \tag{3.7}$$

Now, the modified $\tilde{\mathfrak{L}}^A(\varphi)$ no longer satisfy Bianchi identities and so it is possible, at least in principle, to solve these modified field equations for arbitrary source motions. Of course, not every solution of the modified equations is a solution of the original field equations (3.3). One must retain only those solutions $\varphi(q)$ that simultaneously satisfy the gauge conditions (3.6) for motions of the sources that are allowed by the conditions (3.5). In the simple case of electrodynamics this condition is easily met as we shall see. Furthermore, the solutions of the modified equations (3.7) that lead to a good Fokker action are just a subset of such solutions. In the general case the situation is more complicated and the desired solutions must be obtained by an approximation procedure.[12]

Once a solution of the modified equations that satisfies the above-stated requirement is obtained, one can proceed as before and substitute it into the equations of motion (2.3) for the sources to obtain Fokker equations of motion for the sources and to try to construct a Fokker action for them by the method of Sec. II. In order to apply these methods in the present case, we must, of course, first find a modified action that leads to the modified field equations (3.7). The question of whether or not such an action exists must, at present time, be examined in each individual case. One might imagine that the resulting Fokker action principle is complicated by the existence of the integrability conditions (3.5) acting as constraints on allowed variations of the q's. Fortunately such is not the case. Trautman[11] has shown that any solution of the Fokker equations of motion automatically satisfies the integrability conditions. Thus, the only additional complication for the construction of a Fokker action principle in the case of a gauge-covariant theory is the construction of solutions of the modified field equations that satisfy the gauge conditions when the integrability conditions are satisfied. In the next section we illustrate the procedure for the case of the electromagnetic field. The general case of constructing a Fokker action for a gauge-covariant theory will be discussed in a future paper.

IV. FOKKER ACTION FOR THE ELECTROMAGNETIC INTERACTIONS

As an example of the application of the substitution method to a covariant theory, we discuss the case of electrodynamics. In this theory the interactions between sources are mediated by the electromagnetic potential A_μ and described by the action[13]

$$S = -\frac{1}{4}\int F^{\mu\nu}F_{\mu\nu}\,d^4x - \int A_\mu j^\mu\,d^4x + S_M, \tag{4.1}$$

[12] For a review of these methods in general relativity see J. M. Goldberg in Ref. 11. Chap. 3.

[13] In the following, Greek indices take on the values 0, 1, 2, 3. We employ the Einstein summation convention and the comma notation for differentation: $\phi, \mu \equiv \partial\phi/\partial x^\mu$. Indices are raised and lowered with the Minkowski metric $\eta_{\mu\nu}$ = diag $(1, -1, -1, -1)$.

where

$$F_{\mu\nu} \equiv A_{\mu,\nu} - A_{\nu,\mu}. \tag{4.2}$$

The corresponding field equations are

$$F^{\mu\nu}{}_{,\nu} = j^{\mu}. \tag{4.3}$$

The Bianchi identity $F^{\mu\nu}{}_{,\nu\mu} \equiv 0$ leads in this case to the single integrability condition

$$j^{\mu}{}_{,\mu} = 0. \tag{4.4}$$

Since the fields A_μ do not appear explicitly in this condition as they do in the general case, the problem of constructing a Fokker action is greatly simplified.

In order to solve the field equations (4.3) we must impose a gauge condition on the A_μ. If we require that the Lorentz gauge condition

$$A^{\mu}{}_{,\mu} = 0 \tag{4.5}$$

be satisfied, we may consider instead of Eqs. (4.3) the set of modified field equations

$$\Box \tilde{A}^{\mu} = j^{\mu}, \tag{4.6}$$

where we denote by $\tilde{A}_\mu$ the solutions to these equations. As a trial solution let us consider the symmetric solution

$$\tilde{A}^{\mu}(x) = \int d^4x'\, D^S(x - x') j^{\mu}(x'), \tag{4.7}$$

which is valid for an arbitrary $j^\mu(x)$. Having obtained such a solution we must now check to see if the gauge condition (4.5) is satisfied when the current is conserved, i.e., when the integrability condition (4.4) is satisfied. Because we have already discarded solutions to the homogeneous wave equations it is sufficient to have a spatially bounded charge distribution. Indeed,

$$\tilde{A}^{\mu}{}_{,\mu}(x) = \int_R d^4x'\, D^S(x - x') j^{\mu}{}_{,\mu}(x') + \int_S d^3S_\mu\, D^S(x - x') j^{\mu}(x') = 0 \tag{4.8}$$

for any finite x, because the contributions of the surface integral are from the points at infinity on the light cone at x, and there j^μ vanishes because it is spatially bounded. In order to apply the substitution method, we must find a modified action of the form (2.11) from which Eqs. (4.6) follow. This action is

$$\tilde{S} = \int_R d^4x\, \{\tfrac{1}{2}\tilde{A}_\mu \tilde{A}^{\mu}{}_{,\nu}{}^{,\nu} - \tilde{A}_\mu j^{\mu}\} + S_M. \tag{4.9}$$

The substitution of the symmetric solution (4.7) into (4.9) gives the well-known expression

$$S_F = -\frac{1}{2} \iint_R d^4x\, d^4x'\, D^S(x - x') j^{\mu}(x) j_{\mu}(x') + S_M, \tag{4.10}$$

which is a good Fokker action for (4.6). Because the satisfaction of condition (4.4) is an automatic consequence of the Fokker equations of motion, (4.10) also is a good Fokker action for (4.3). Thus, in the case of a Dirac field coupled minimally to the electromagnetic field the continuity equation for the current

$$j^{\mu} = ie\bar{\Psi}\gamma^{\mu}\Psi \tag{4.11}$$

is satisfied identically for any A_μ when Ψ satisfies the Dirac equations of motion. For this reason, the Ψ determined by the Fokker equations of motion

$$\gamma^{\mu}\, \partial_{\mu}\Psi - m\Psi = e\gamma^{\mu}\Psi \int d^4x'\, D(x - x')\bar{\Psi}(x')\gamma_{\mu}\Psi(x') \tag{4.12}$$

satisfy the condition (4.4).

It is interesting to apply the substitution method to solutions of the Maxwell equations that satisfy gauge conditions other than the Lorentz condition. In this way one obtains expressions equivalent to (4.10) that may be useful for applications. Gauge conditions may imply differential constraints, algebraic constraints, or a mixture of both. Because the Lorentz gauge condition is an example of the first type of condition, it is instructive to consider the imposition of a pure algebraic constraint like $A_1 = 0$. This leads to a set of modified field equations

$$\tilde{A}^{s,\mu}{}_{\mu} - \tilde{A}^{t,s}{}_{t} = j^{s}, \tag{4.13a}$$

$$\tilde{A}^{t,1}{}_{t} = -j^{1}, \qquad s, t \neq 1. \tag{4.13b}$$

It turns out to be impossible to find a modified action that leads to Eqs. (4.13). For this reason, we consider the equivalent problem in which

$$\tilde{A}^{s\mu}{}_{,\mu} - \tilde{A}^{t}{}_{,s}{}_{t} = j^{s}, \qquad \tilde{A}^{1} = \text{const} = 0, \qquad s, t \neq 1 \tag{4.14}$$

are taken to be the set of modified field equations, and

$$-\tilde{A}^{t,1}{}_{t} = j^{1} \tag{4.15}$$

is taken to be the gauge condition. For these equations, a modified action exists, given by

$$S = \int d^4x\, \{\tfrac{1}{2}\tilde{A}_s(\tilde{A}^{s,\mu}{}_{\mu} - \tilde{A}^{t,s}{}_{t}) - \tilde{A}_s j^{s}\} + S_M. \tag{4.16}$$

We have to proceed likewise in the case of the more useful Coulomb gauge, $A^{r}{}_{,r} = 0$; $r = 1, 2, 3$.

We solve the equivalent problem given by the set of modified field equations

$$-\tilde{A}^{0,r}{}_{,r} = j^0 \equiv \rho, \qquad \tilde{A}^{r,\mu}{}_{,\mu} = j^{rT}, \qquad r = 1, 2, 3, \tag{4.17}$$

with the gauge condition

$$\tilde{A}^{0}{}_{,0}^{,r} = j^{rL}. \tag{4.18}$$

We have introduced j^{rL}, the longitudinal part of the current and j^{rT}, the transverse part of the current defined by [14]

$$j_r^T(x, t) \equiv j_r - \partial_r \int d^3x' \frac{\partial'_s j_s(\mathbf{x}', t)}{-4\pi |\mathbf{x} - \mathbf{x}'|} \equiv j_r - j_r^L, \tag{4.19}$$

which satisfies $j^T_{r,r} = 0$. We notice that, once again under our assumptions on the current distribution, θ defined by

$$j_r^L \equiv \partial_r \theta = \partial_r \partial_s \int d^3x' \frac{j_s(\mathbf{x}', t)}{-4\pi |\mathbf{x} - \mathbf{x}'|} \tag{4.20}$$

behaves like R^{-2} when x_μ tends to infinity.

Taking now the symmetric solution of (4.17)

$$\tilde{A}_0(x, t) = \int d^3x' \frac{\rho(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|},$$

$$\begin{aligned}\tilde{A}_r(x) &= \int d^4x\, D^s(x, x') j_r^T(x') \\ &= \int d^4x'\, D^s(x, x') \Big[j_r(x) \\ &\quad - \partial'_r \int d^3x'' \frac{\partial''_s j_s(\mathbf{x}'', t)}{-4\pi |\mathbf{x}' - \mathbf{x}''|} \Big],\end{aligned} \tag{4.21}$$

we verify that (4.18) holds only when $j_{\mu,\mu} = 0$.

The modified Lagrangian for Eqs. (4.17) is

$$\tilde{S} = -\int d^4x' \{ \tfrac{1}{2} \tilde{A}_r \tilde{A}_{r,\mu\mu} + \tfrac{1}{2} \tilde{A}_0 \tilde{A}_{0,rr} - \tilde{A}_r j_r{}^T + \tilde{A}_0 \rho \} + S_M. \tag{4.22}$$

From it we can obtain the Fokker action in this gauge provided $j^\mu(x)$ satisfies the charge-continuity equation $j^\mu{}_{,\mu} = 0$. Such a current is given, for example, by a point charge by

$$j^\mu(x) = \sum_a e_a \int d\tau_a\, \delta^4(x - z_a) \frac{dz_a^\mu}{d\tau_a}, \tag{4.23}$$

where $z_a^\mu(\tau_a)$ are the coordinates of the ath charge and τ_a is its proper time. The charge e_a is constant along the ath trajectory. If, in the substitution method, we use this current we can be assured of the satisfaction of the Coulomb gauge for all variations.

Substitution of (4.21) into (4.22) gives for S_F

$$S_F = \frac{1}{2} \int d^4x \left\{ \int d^4x'\, j_r^T(x) D^S(x, x') j_r^T(x') - \int d^3x' \frac{\rho(\mathbf{x}, t)\rho(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} \right\} + S_M. \tag{4.24}$$

One can easily establish the equivalence of expressions (4.6) and (4.24) by direct computation, keeping in mind the assumption $j^\mu{}_{,\mu} = 0$.

The form (4.24) is particularly useful when dealing with problems in which a slow-motion approximation is justified. For example, to get the action to order v^2/c^2, which is Darwin's action,[15] we notice that j_r^T is of order v/c and so it is enough to put for $D^s(x, x')$ the zero-order term $\delta(t - t')/4\pi |\mathbf{x} - \mathbf{x}'|$ of its expansion in powers of v/c. We then get[16]

$${}^{0+2}S_F = \frac{1}{2} \int d^4x \int d^3x' \times \left\{ \frac{j_r^T(\mathbf{x}, t) j_r^T(\mathbf{x}', t) - \rho(\mathbf{x}, t)\rho(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} \right\} + {}^{0+2}S_M, \tag{4.25}$$

which is also equal to

$$\frac{1}{2} \int d^4x \int d^3x' \left\{ \frac{j_r(\mathbf{x}, t) j_r(\mathbf{x}', t) - j_r^L(\mathbf{x}, t) j_r{}^L(\mathbf{x}', t) - \rho(\mathbf{x}, t)\rho(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} \right\} + {}^{0+2}S_M \tag{4.26}$$

because

$$\int d^4x \int d^3x' \frac{j_r{}^L(\mathbf{x}, t) j_r{}^T(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} = 0 \tag{4.27}$$

by virtue of the definitions and asymptotic properties of j_r^L and j_r^T.

We may now write

$$\int d^4x \int d^3x' \frac{j_r^L(\mathbf{x}, t) j_r^L(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|} = \int d^4x\, \theta(\mathbf{x}, t) \theta(\mathbf{x}, t), \tag{4.28}$$

[14] In what follows we do not distinguish between upper and lower Latin indices and $\mathbf{x}$ stands for the components x_1, x_2 and x_3 of the four-vector x_μ.

[15] C. G. Darwin, Phil. Mag. **39**, 537 (1920).

[16] In the following, a left-hand superscript on a symbol indicates the nth-order term in an expansion of that quantity. The sum of a number of superscripts indicates the sum of the respective terms of the expansion, e.g., ${}^{1+2}S = {}^1S + {}^2S$.

which, on account of the relation

$$\frac{1}{16\pi^2}\int d^3x \int d^3x' \int d^3x''\, \partial_r \frac{1}{|\mathbf{x}-\mathbf{x}'|}\, \partial_s \frac{1}{|\mathbf{x}-\mathbf{x}''|}\, j_r(\mathbf{x}', t) j_s(\mathbf{x}'', t)$$
$$= \frac{1}{8\pi}\int d^3x' \int d^3x''\, \partial_r\, \partial_s\, |\mathbf{x}'-\mathbf{x}''|\, j_r(\mathbf{x}', t) j_s(\mathbf{x}'', t), \qquad (4.29)$$

leads us to give, as another form of (4.25)

$$^{0+2}S_F = \int d^4x \int d^3x'$$
$$\times \left\{ \frac{j_r(\mathbf{x}, t) j_r(\mathbf{x}' t) - \rho(\mathbf{x}, t)\rho(\mathbf{x}', t)}{4\pi\,|\mathbf{x}-\mathbf{x}'|} \right. \qquad (4.30)$$
$$\left. + \frac{1}{8\pi}\,\partial_r\,\partial'_s\,|\mathbf{x}-\mathbf{x}'|\, j_r(\mathbf{x}, t) j_s(\mathbf{x}', t) \right\} + {}^{0+2}S_M.$$

For charge pole singularities, that is, for currents of the form (4.23), this becomes

$$^{0+2}S_F = \int dt \left\{ \frac{1}{2} \sum_a \sum_b \left[\frac{e^{(a)}e^{(b)}}{2R_{ab}c^2} \right.\right.$$
$$\left.\left. \times\, [v_r{}^{(a)} v_r{}^{(b)} + (v_r{}^{(a)} n_r)(v_s{}^{(b)} n_s)] - \frac{e^{(a)}e^{(b)}}{R_{ab}} \right] \right\} + S_M, \qquad (4.31)$$

where n is a unit vector from charge a to charge b.

V. THEORIES WITH AN APPROXIMATION PROCEDURE

We now consider an interesting feature of the Fokker action approach to the Fokker equations of motion that results in considerable savings of computational effort whenever it is necessary to solve the field equations by an approximation procedure. Suppose that we want to construct the Fokker equations of motion only up to a given order of accuracy in an expansion

$$\phi = {}^0\phi + {}^1\phi + {}^2\phi + \cdots \qquad (5.1)$$

of the field that is being eliminated. If one takes into account this expansion, our Lagrangian in its turn decomposes into a sum of terms of increasing order:

$$\mathfrak{L} = {}^0\mathfrak{L} + {}^1\mathfrak{L} + {}^2\mathfrak{L} + \cdots, \qquad (5.2)$$

and the same happens with S. Given an ${}^n\mathfrak{L}$ we notice that the ${}^m\phi$ with $n > m > \frac{1}{2}n$ enter in it only linearly. If we split ${}^n\mathfrak{L}$ into a part ${}^n\mathfrak{L}^{(m)H}$ linearly homogeneous in ${}^m\phi$ and a part ${}^n\mathfrak{L}^I$ independent of the ${}^m\phi$, then ${}^nS$ can be written as

$${}^nS = \int d^4x \sum_m \left({}^m\phi \frac{\delta^n \mathfrak{L}^{(m)H}}{\delta\, {}^m\phi} \right) + {}^n\mathfrak{L}^I + {}^n\Sigma, \qquad (5.3)$$

where the ${}^n\Sigma$ are surface integrals that contain the ${}^m\phi$ which resulted from the partial integrations performed to eliminate the derivatives of the ${}^m\phi$. Now we have the interesting fact that the $\delta^n\mathfrak{L}^{(m)H}/\delta\, {}^m\phi$ are the field equations for the ${}^m\phi$,[17] and so the first term in (5.3) vanishes for any solution (5.1). Thus, there is no need to know these higher-order solutions for the construction of the Fokker action provided we can ignore the surface integrals ${}^n\Sigma$.

We prove that these integrals can indeed be ignored provided we choose our exact Lagrangian so that, to lowest order, it is equal to

$$^{0+2}S = \int d^4x\, [{}^1\phi \mathfrak{D}\, {}^1\phi - {}^1\phi\, {}^1\zeta_M(q)] + {}^0S_M(q). \qquad (5.4)$$

In addition we must require that ${}^k\phi$ and its derivatives approach zero at infinity at least as fast as r^{-1}. In this case we should then expect contributions to ${}^n\Sigma$ only from terms of ${}^nS$ that are quadratic in ${}^k\phi$, i.e., from the terms

$$\int d^4x \sum_{k=0}^{n} {}^k\phi \mathfrak{D}^{n-k}\phi \qquad (5.5)$$

when we perform the integrations by parts necessary to bring (5.5) to the form of (5.3). These surface integrals can be written as a volume integral equal to[18]

$$\int d^4x \sum_{k=0}^{k\leq \frac{1}{2}n} ({}^{n-k}\phi \mathfrak{D}^k \phi - {}^k\phi \mathfrak{D}^{n-k}\phi). \qquad (5.6)$$

This volume integral vanishes for symmetric solutions ${}^k\phi$. So we arrive at the conclusion that the surface integrals ${}^n\Sigma$ of Eq. (5.3) can indeed be disregarded whenever we use symmetric solutions in the substitution method.

One still has to prove that under these conditions a good Fokker action is obtained to all orders of the approximation procedure. The nth-order contribution to the Fokker equations of motion is, by definition,

$$\delta^n \mathfrak{L}(\phi[q])/\delta q, \qquad (5.7)$$

[17] E. Newman and P. G. Bergmann, Rev. Mod. Phys. 29, 443(1957).

[18] We assume throughout that $\mathfrak{D}$ is a self-adjoint differential operator for functions that vanish at infinity.

whereas the δ^* variation of ${}^nS^F$ leads to

$$\int d^4x \left[\frac{\delta^n \mathcal{L}(\phi[q])}{\delta q} + \sum_{k=1}^{n} \left\{ \frac{\partial^n \mathcal{L}(\phi[q]}{\partial^k \phi} \frac{\delta^* {}^k\phi}{\delta q} + \frac{\partial^n \mathcal{L}(\phi[q])}{\partial^k \phi_{,\mu}} \left(\frac{\delta^* {}^k\phi}{\delta q} \right)_{,\mu} + \frac{\partial^n \mathcal{L}(\phi[q])}{\partial^k \phi_{,\mu\nu}} \left(\frac{\delta^* {}^k\phi}{\delta q} \right)_{,\mu\nu} \right\} \right] \delta q \tag{5.8}$$

or

$$\int d^4x \left[\frac{\delta^n \mathcal{L}(\phi[q])}{\delta q} + \sum_{k=1}^{n} \frac{\delta^n \mathcal{L}(\phi[q])}{\partial^k \phi} \frac{\delta^* {}^k\phi}{\delta q} \right] \delta q + {}^n\Sigma'. \tag{5.9}$$

The summation in the volume integral vanishes upon substitution because the $\delta^n \mathcal{L}/\delta^k \phi$ are the equations of motion for the ${}^{(n-k)}\phi$. The surface integrals ${}^n\Sigma'$ that do not vanish on account of the asymptotic behavior of the ϕ turn out to be equal to the volume integral

$${}^n\Sigma' = \int d^4x \sum_{k=1}^{n} \left({}^k\phi \mathfrak{D} \frac{\delta^* {}^{(n-k)}\phi}{\delta q} - \frac{\delta^* {}^{(n-k)}\phi}{\delta q} \mathfrak{D}^n \phi \right) \delta q, \tag{5.10}$$

which also vanishes for symmetric solutions. So S^F, up to any order, is a good Fokker action because the variation (5.8) of each ${}^nS^F$ leads to the corresponding contribution (5.7) to the Fokker equation of motion.

As an example of the above discussion let us consider the case where solutions of Maxwell's equations are expanded in a slow-motion approximation, that is, in powers of v/c. The field variables are expanded according to[19]

$$\begin{aligned} A_0 &= {}^0A_0 + {}^2A_0 + {}^4A_0 + \cdots, \\ A_r &= {}^1A_r + {}^3A_r + {}^5A_r + \cdots, \end{aligned} \tag{5.11}$$

whereas ρ is a quantity of zero order and the j_r are first order, and each time derivation raises by one the order of a quantity.

If we likewise expand the modified action for the Lorentz gauge

$$S = \int_\Omega d^4x \, (\tfrac{1}{2} A_\mu A^{\mu,\nu}{}_{,\nu} - A^\mu j_\mu) + S_M, \tag{5.12}$$

we find

$$\begin{aligned} {}^0S &= -\int d^4x \, [\tfrac{1}{2}({}^0A_0 {}^0A_{0,rr}) + {}^0A_0\rho] + {}^0\Sigma + {}^0S_M, \\ {}^2S &= \int d^4x \, [-{}^2A_0 {}^0A_{0,rr} - {}^2A_0\rho + \tfrac{1}{2}({}^1A_r {}^1A_{r,ss}) \\ &\quad + \tfrac{1}{2}({}^0A_0 {}^0A_{0,00}) + {}^1A_r j_r] + {}^2\Sigma + {}^2S_M, \qquad (5.13) \\ {}^4S &= \int d^4x \, [-{}^4A_0 {}^0A_{0,rr} \\ &\quad - \tfrac{1}{2}({}^2A_0 {}^2A_{0,rr}) - {}^4A_0\rho + {}^3A_r {}^1A_{r,ss} \\ &\quad + {}^2A_0 {}^0A_{0,00} + {}^3A_r j_r - \tfrac{1}{2}({}^1A_r {}^1A_{r,00})] + {}^4\Sigma + {}^4S_M. \end{aligned}$$

It is easy to check that the surface integrals vanish upon substitution of symmetric solutions and that only a knowledge of ${}^0A_0$, ${}^2A_0$, and ${}^1A_r$ is required to get the Fokker action to fourth order. We have indeed

$$\begin{aligned} {}^{0+2+4}S^F &= \int d^4x \, [-\tfrac{1}{2}({}^0A_0 {}^0A_{0,rr}) - {}^0A_0\rho \\ &\quad + \tfrac{1}{2}({}^1A_r {}^1A_{r,ss}) + \tfrac{1}{2}({}^0A_0 {}^0A_{0,00}) + {}^1A_r j_r - \tfrac{1}{2}({}^2A_0 {}^2A_{0,rr}) \\ &\quad + {}^2A_0 {}^0A_{0,00} - \tfrac{1}{2}({}^1A_r {}^1A_{r,00})] + {}^{0+2+4}S_M \end{aligned} \tag{5.14}$$

with

$$\begin{aligned} {}^0A_0 &= \int d^3x' \frac{\rho(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|}, \\ {}^1A_r &= \int d^3x' \frac{j_r(\mathbf{x}', t)}{4\pi |\mathbf{x} - \mathbf{x}'|}, \\ {}^2A_0 &= \int d^3x' \frac{{}^0A_{0,00}}{-4\pi |\mathbf{x} - \mathbf{x}'|}. \end{aligned}$$

Notice that this expression up to second order coincides with the form (4.26) of Darwin's expression, which was derived in the previous section from an expansion of the Fokker action in the Coulomb gauge. On the other hand, we have derived the same expression in this section by using in the substitution method approximate solutions for the Lorentz gauge condition. Indeed, Darwin's formula is determined only by the approximation procedure in a way completely independent of the choice of gauge.

[19] S. Bazanski, in *Recent Developments in General Relativity* (Pergamon Press, Inc., New York, 1962), p. 137.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8, NUMBER 2 FEBRUARY 1967

Maximal Analytic Extension of the Kerr Metric*

ROBERT H. BOYER†
Department of Applied Mathematics, University of Liverpool, Liverpool, England
AND
RICHARD W. LINDQUIST
Scott Laboratory of Physics, Wesleyan University, Middletown, Connecticut
(Received 19 July 1966)

Kruskal's transformation of the Schwarzschild metric is generalized to apply to the stationary, axially symmetric vacuum solution of Kerr, and is used to construct a maximal analytic extension of the latter. In the low angular momentum case, $a^2 < m^2$, this extension consists of an infinite sequence Einstein–Rosen bridges joined in time by successive pairs of horizons. The number of distinct asymptotically flat sheets in the extended space can be reduced to four by making suitable identifications. Several properties of the Kerr metric, including the behavior of geodesics lying in the equatorial plane, are examined in some detail. Completeness is demonstrated explicitly for a special class of geodesics, and inferred for all those that do not strike the ring singularity.

I. INTRODUCTION

A FAMILIAR feature of exact solutions to the field equations of general relativity is the presence of singularities. Although long recognized in such special metrics as those of Schwarzschild and Reissner–Nordström, and in the spherically symmetric Friedmann models, these singularities have been dismissed as nonphysical, due perhaps to the high degree of symmetry assumed for the solutions.[1] Recent work by Penrose and Hawking[2] suggests that this traditional view may need to be revised; it is quite possible that gravitational collapse leads inevitably to a singular state if trapped surfaces are once formed[3]—provided that one stubbornly persists in applying the classical field equations to regions of arbitrarily high curvature.[4] Thus there is some point in studying the complete geometry of exact solutions, even though they are idealizations of physically reasonable gravitating systems, in order to learn something about the type of behavior to be expected in more realistic models. We here lay the foundations for such a study of stationary axially symmetric models of rotating bodies, by analyzing in some detail the empty space metric of Kerr.[5]

Of course one must be careful to distinguish true singularities, to which the Penrose–Hawking theorems refer, from "pseudosingularities" that reflect merely a poor choice of coordinates. The latter can always be removed by the familiar device of covering the manifold with a family of coordinate patches. It has long been recognized that the singularity at $r = 2m$ in the standard form of the Schwarzschild metric is of this type; transformations to coordinate frames which remove this apparent singularity have been given by several authors.[6]

Closely related to the problem of singularities is that of completeness.[7] Solutions of the field equa-

* Work supported in part by the Aerospace Research Laboratories, Office of Aerospace Research, and by the Office of Scientific Research, U. S. Air Force.

† Deceased.

[1] See, e.g., E. M. Lifshitz and I. M. Khalatnikov, Zh. Eksperim. i Teor. Fiz. **39**, 149, 800 (1960) [English transl.: Soviet Phys.—JETP **12**, 108, 558 (1961)].

[2] R. Penrose, Phys. Rev. Letters **14**, 57 (1965); S. W. Hawking, *ibid.* **15**, 689 (1965). See also S. W. Hawking and G. F. R. Ellis, Phys. Letters **17**, 246 (1965).

[3] The collapse of a spherically symmetric mass has been studied in detail by M. M. May and R. H. White, Phys. Rev. **141**, 1232 (1966). They find that unless the collapse is halted before any fraction of the total mass has fallen within its own Schwarzschild radius, a singularity invariably ensues.

[4] Wheeler has repeatedly emphasized that, from a deeper point of view, these singularities must be nonphysical, since quantum effects will necessarily alter the complexion of the problem completely in regions of high curvature. See B. K. Harrison, K. S. Thorne, M. Wakano, and J. A. Wheeler, *Gravitation Theory and Gravitational Collapse* (University of Chicago Press, Chicago, Ill., 1965).

[5] R. P. Kerr, Phys. Rev. Letters **11**, 237 (1963).

[6] A. S. Eddington, Nature **113**, 192 (1924); this transformation was rediscovered by D. Finkelstein, Phys. Rev. **110**, 965 (1958). Other forms of the metric which are nonsingular at the Schwarzschild radius were exhibited by G. Lemaître, Ann. Soc. Sci. Bruxelles **53A**, 51 (1933), and by J. L. Synge, Proc. Roy. Irish Acad. **A53**, 83 (1950). C. Fronsdal, Phys. Rev. **116**, 778 (1959); and J. Plebański, Bull. Acad. Polon. Sci. **10**, 373 (1962), independently constructed the maximal analytic extension of the Schwarzschild manifold by imbedding it in a six-dimensional flat space. A transformation to a coordinate frame which displays this maximal extension was first given by M. D. Kruskal, Phys. Rev. **119**, 1742 (1960).

[7] By "completeness" we mean affine completeness: the property that all geodesics can be continued to arbitrarily large values of their affine path parameters. A standard theorem of Riemannian geometry states that geodesic completeness is equivalent to completeness as usually defined—namely, that all Cauchy sequences converge—if the metric is positive definite; however, this theorem breaks down on manifolds with an indefinite metric. For a careful discussion of different possible and inequivalent definitions of completeness for manifolds with an indefinite metric, see W. Kundt, Z. Physik **172**, 488 (1963); C. W. Misner, J. Math. Phys. **4**, 924 (1963); and M. Fierz and R. Jost, Helv. Phys. Acta **38**, 137 (1965).

tions are usually presented locally, in terms of a coordinate system adapted to the symmetries of a given problem. Thus the manifold on which the metric is to be defined is left unspecified at the start, and determined only afterwards by imposing various global and topological conditions. Ideally, one would like the resulting manifold to be geodesically complete and free of singularities, but the Penrose–Hawking theorems show that in many cases these two aims are incompatible. Of course, one can always eliminate singularities from a given metric by redefining the manifold to exclude the singular points; this is not a very satisfactory solution, however, if physical test particles run off the edge of the space–time so obtained in a finite proper time. It seems preferable, we believe, to require geodesic completeness, insofar as possible, even though this leads in general to singularities. Thus if $\mathfrak{M}$ and $\mathfrak{M}'$ are analytic manifolds, with $\mathfrak{M}' \supset \mathfrak{M}$, we call $\mathfrak{M}'$ a *maximal analytic extension* of $\mathfrak{M}$ if every geodesic of $\mathfrak{M}'$ is either complete or terminates at a singular point. Such an extension need not exist,[8] and even if it does it will not be unique, because of the freedom available to identify points in $\mathfrak{M}'$ in a variety of ways without disturbing analyticity.

Analytic Extensions of the Schwarzschild and Reissner–Nordström Metrics

Eddington[6] long ago pointed out that the transformation

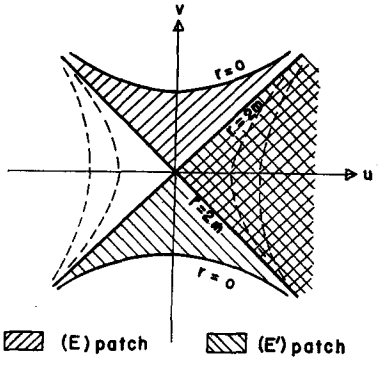

FIG. 1. A Kruskal diagram for the Schwarzschild metric, showing a portion of the (u, v) coordinate plane. In this plane null curves appear as straight lines of slope ± 1; the null lines $u \pm v = 0$ define the horizons. The curves $r = \text{const}$ are hyperbolas having these horizons as asymptotes, with the $r = 0$ hyperbola a singular surface. The shaded regions show the portion of the (u, v) plane in which the Eddington coordinates (E) and (E') are regular; their overlap is the region of regularity $u > 0$, $u^2 > v^2$ $(r > 2m, -\infty < \bar{t} < \infty)$ for standard Schwarzschild coordinates. The fourth unshaded region, $u < 0$ and $u^2 > v^2$, is also asymptotically flat, and represents the second sheet of an Einstein–Rosen bridge.

$$r = \bar{r}, \qquad \theta = \bar{\theta}, \qquad \varphi = \bar{\varphi}, \tag{1.1}$$
$$dt = d\bar{t} + 2m\, d\bar{r}/(\bar{r} - 2m)$$

(where, to agree with our later notation, we use bars to denote the original coordinates of Schwarzschild) leads to a form of the Schwarzschild metric free of singularities at $\bar{r} = 2m$:

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2) - dt^2 + (2m/r)(dr + dt)^2. \tag{1.2}$$

This metric, interpreted on the manifold $r > 0$, $0 \le \theta \le \pi$, $0 \le \varphi < 2\pi$, $-\infty < t < \infty$, is in fact an analytic extension of the original; it is complete for $t \to +\infty$—excepting those geodesics which strike the true singularity at $r = 0$—but not for $t \to -\infty$. One notes that this form has the structure of a flat space metric plus the square of a null vector; this feature serves as the point of departure in the Kerr–Schild[9] theory, to which we refer briefly in Sec. II. The alternative transformation

$$r' = \bar{r}, \qquad \theta' = \bar{\theta}, \qquad \varphi' = \bar{\varphi}, \qquad dt' = d\bar{t} - 2m d\bar{r}/(\bar{r} - 2m) \tag{1.3}$$

leads to an expression for ds^2 similar to (1.2), but with dt replaced by $-dt'$, which is complete for $t' \to -\infty$. One thus obtains two coordinate patches, shown in Fig. 1, whose domain of overlap is the region external to the Schwarzschild pseudosingularity. This is still not the maximal analytic extension, however. The latter has been given in a particularly simple form by Kruskal,[6] and is summarized by the transformation equations

$$u \pm v = |(r/2m) - 1|^{\frac{1}{2}} \exp[(r \pm \bar{t})/4m]. \tag{1.4}$$

The corresponding manifold, illustrated by the now familiar Kruskal diagram (Fig. 1), consists of two asymptotically flat universes whose spacelike sections are joined together on a 2-sphere of minimum area—the "throat" of an Einstein–Rosen bridge.[10] The area of this throat changes with time, reaching a maximum value of $4\pi(2m)^2$ at $u = 0$, $v = 0$, and collapsing to zero both in the future and the past, at those points ($u = 0$, $v = \pm 1$, for instance) for which $r = 0$.

A similar analysis for the Reissner–Nordström solution (the metric outside a single mass m with charge q) has been carried through by Graves and

[8] A simple counter-example has been constructed by C. W. Misner, University of Maryland Tech. Rept. No. 529 (1965), p. 14.

[9] R. P. Kerr and A. Schild, Am. Math. Soc. Symposium, New York, 1964; also Galileo Quatercentenary, Florence, 1964.

[10] A. Einstein and N. Rosen, Phys. Rev. 48, 73 (1935).

Brill.[11] This metric has two pseudosingularities at

$$r = r_{\pm} \equiv m \pm (m^2 - q^2)^{\frac{1}{2}} \qquad (1.5)$$

and a true singularity at $r = 0$. Graves and Brill displayed transformations of the Kruskal type which would eliminate either pseudosingularity, but not both. Nevertheless, their results are sufficient to portray the full manifold, for by piecing together successive coordinate patches one can extend the metric first across one singularity and then across the other. It turns out that the resulting manifold is still not complete, however, and one must continue to add similar patches indefinitely. The picture which emerges is thus altogether different from the Schwarzschild case. One can view the manifold as an Einstein–Rosen bridge, whose throat oscillates between a minimum and a maximum area, given by $4\pi r_-^2$ and $4\pi r_+^2$. After each cycle the bridge is found to be attached to a new pair of (asymptotically flat) sheets, isometric to the original pair but nevertheless topologically distinct from them (but see the end of Sec. III).

Horizons

The pseudosingularities $r = r_{\pm}$, like $r = 2m$ in the Schwarzschild metric, have an important geometric and physical significance: They determine the location of horizons. In stationary space–times, these may provisionally[12] be defined as stationary null hypersurfaces: Whenever, for some function $f(r, \theta, \varphi)$, the equation $(\text{grad } f)^2 = 0$ is satisfied on $f = \text{const}$, then this hypersurface is called a horizon. It acts locally as a one-way membrane of infinite red shift (see note added in proof).

Within the horizons of the Schwarzschild and Reissner–Nordström metrics, when such occur, there are trapped surfaces in the sense of Penrose.[2] Moreover, both metrics contain real singularities, located at $r = 0$ in the coordinates normally used. Timelike paths which cross the horizons of the latter, however, need not strike the singularity, but can be continued onto another sheet and thence out to (another) spatial infinity. The surface $r = 0$ is simply a barrier that prevents continuation of the metric to negative values of r. As we see presently, on the Kerr manifold $r = 0$ defines a similar barrier, but one of lower dimension, and the extension to negative values of r is meaningful and necessary. Trapped surfaces exist here, too; however, the Penrose–Hawking theorems do not apply to the Reissner–Nordström and Kerr manifolds, for these do not admit the required Cauchy initial hypersurface.

When $q^2 > m^2$, Eq. (1.5) breaks down; there are then no horizons and much of the previous discussion ceases to apply. In fact, the original coordinate patch covers the maximal extension.

Kerr Metric

We have reviewed the above results at length since they bear directly on the problem we wish to consider—that of obtaining a corresponding maximal analytic continuation of the Kerr metric. This solution, it may be recalled, describes a possible exterior field outside a rotating body; it is the only known example of a stationary vacuum metric with gravitational mass and rotation that is asymptotically flat. Like the Schwarzschild metric it is algebraically special (i.e., of type D in the Petrov–Pirani classification[13]); thus it contains two geodesic shear-free null congruences.[14] It admits two Killing vectors, associated with time translations and rotations about an axis of symmetry, and it contains two parameters m and a which can be identified with the total mass and angular momentum per unit mass of the source. Although no one has yet succeeded in displaying explicitly an interior metric which fits smoothly onto Kerr's exterior solution, there do not appear to be any difficulties in principle in integrating the combined equations of hydrodynamics and gravitation for the interior case, provided that the shape of the body is chosen appropriately.[15] For purposes of this paper, however, we neglect the presence of any such source, and ask rather for the maximal extension of the empty space metric.

Kerr's solution contains two event horizons, which

[11] J. C. Graves and D. R. Brill, Phys. Rev. **120**, 1507 (1960). The case $m^2 = q^2$ has been studied by B. Carter, Phys. Letters **21**, 423 (1966).

[12] B. Carter (private communication) has stated a theorem which suggests a more stringent definition of a horizon. Briefly: "Let a space–time admit a group of isometries with p-dimensional integral surfaces, and let the tangent p vector be $(4\text{-}p)$-surface orthogonal. Then the locus of nullity of the Killing p vector is itself a null hypersurface." This surface—the "Killing horizon"—appears in the Schwarzschild and Reissner–Nordström metrics with $p = 1$ or 3 and in the Kerr metric (as we see later) with $p = 2$. For a different definition of an "event horizon," see W. Rindler, Monthly Notices Roy. Astron. Soc. **116**, 662 (1956).

[13] A. Z. Petrov, Sci. Notices, Kazan State Univ. **114**, 55 (1954); F. A. E. Pirani, Phys, Rev. **105**, 1089 (1957).

[14] R. Debever, Compt. Rend. **249**, 1324 (1959); R. Penrose, Ann. Phys. (N. Y.) **10**, 171 (1960); R. K. Sachs, Proc. Roy. Soc. (London) **A264**, 309 (1961).

[15] R. H. Boyer, Proc. Cambridge Phil. Soc. **61**, 527 (1965). A. G. Doroshkevich, Ya. B. Zel'dovich and I. D. Novikov, Zh. Eksperim. i Teor. Fiz. **49**, 170 (1965) [English transl.: Soviet Phys.—JETP **22**, 122 (1966)] claim that the interior solution must also display some type of vortex or convective motion, because the Kerr metric has other off-digaonal components besides $g_{\varphi t}$. However, their argument is based upon a false premise; see for example Eq. (2.13).

in an appropriate coordinate system are located at

$$r = r_{\pm} \equiv m \pm (m^2 - a^2)^{\frac{1}{2}} \tag{1.6}$$

plus a true singularity again formally defined by $r = 0$. The correspondence with the Reissner–Nordström result (1.5) is rather striking. Thus one might expect the maximal extension of the Kerr manifold to be topologically very similar to the Graves–Brill construction. A main object of our work is to justify this expectation, by displaying a transformation analogous to that of Kruskal. We give this in detail in Sec. III, after first exhibiting several other useful coordinate frames in Sec. II. Kruskal's method, as generalized by Graves and Brill, cannot be applied directly—except to the two-dimensional subspace containing the symmetry axis and the time coordinate[16]—because the metric depends on the polar angle θ in a complicated way. But when we combine a transformation analogous to (1.4) with an appropriate change in the azimuthal angle φ, whose effect is to straighten out the null congruences in the neighborhood of the event horizon being considered, we find, happily, that the resulting metric is regular across the horizon at all values of θ.

The proof that our extension is maximal requires a demonstration that geodesics which do not strike a true singularity can be continued to infinite length. We show this in Sec. IV by studying the geodesic equations. Some particular features of the geodesics themselves, which seemed to us to be interesting and curious, are described briefly in Sec. V.

II. PROPERTIES OF THE KERR METRIC

Kerr–Schild Theory

Kerr and Schild[9] have studied solutions of the vacuum field equations for which the metric has the form[17]

$$g_{\alpha\beta} = \eta_{\alpha\beta} + 2Hk_\alpha k_\beta. \tag{2.1a}$$

Here $\eta_{\alpha\beta}$ is the metric of Minkowski space, k_α a null vector field, H a scalar field. It does not matter whether k_α is defined to be null with respect to the flat background metric $\eta_{\alpha\beta}$ or the full metric $g_{\alpha\beta}$, since

$$g^{\alpha\beta}k_\beta = \eta^{\alpha\beta}k_\beta \equiv k^\alpha$$

and therefore $g^{\alpha\beta}k_\alpha k_\beta = 0$ implies $\eta^{\alpha\beta}k_\alpha k_\beta = 0$ and conversely. In fact,

$$g^{\alpha\beta} = \eta^{\alpha\beta} - 2Hk^\alpha k^\beta. \tag{2.1b}$$

[16] B. Carter, Phys. Rev. 141, 1242 (1966), has independently worked out the analytic extension of this subspace.

[17] *Notational conventions*: Greek letters range and sum from 1 to 4; $g_{\alpha\beta}$ has signature $(+++-)$.

Kerr and Schild showed that the vacuum field equations require the null congruence to be *geodesic* (with respect to either $g_{\alpha\beta}$ or $\eta_{\alpha\beta}$). By choosing H suitably, k_α can be so normalized that

$$k^\beta \nabla_\beta k^\alpha = 0, \tag{2.2a}$$

∇_β being the covariant derivative based upon $g_{\alpha\beta}$. We suppose this to be the case, and define an affine parameter μ by $k^\alpha = dx^\alpha/d\mu$. Then, as Kerr and Schild noted, one finds

$$R_{\alpha\beta\gamma\delta}k^\beta k^\delta = -(d^2H/d\mu^2)k_\alpha k_\gamma. \tag{2.2b}$$

It follows that k^α is a multiple Debever–Penrose[14] vector, and consequently, by the Goldberg–Sachs[18] theorem, that the null congruence defined by k^α is *shear-free* as well.

In their analysis, Kerr and Schild give rules for constructing the general empty-space metric of the form (2.1). We do not quote these here, but merely remark that with the exception of the Kerr metric the representation (2.1) is unique, so that the metric is of type II in the Petrov–Pirani classification.

Kerr Metric in Explicit Form

Let us consider the exceptional case, in which the line element can be represented in the form (2.1) in two distinct ways:

$$g_{\alpha\beta} = \eta_{\alpha\beta} + 2Hk_\alpha k_\beta, \tag{2.3a}$$

$$\eta'_{\alpha\beta} + 2H'l_\alpha l_\beta. \tag{2.3b}$$

Both k^α and l^α are then principal null vectors (i.e., double Debever–Penrose vectors), which implies that the metric is of type D (type I degenerate). This case is of particular interest, for it describes the stationary vacuum solution with rotation first obtained in a different way by Kerr.[5] He gives the explicit presentation

$$ds^2 = dx^2 + dy^2 + dz^2 - dt^2 + \frac{2mr^3}{(r^4 + a^2z^2)} \times \left[\frac{r(x\,dx + y\,dy) + a(x\,dy - y\,dx)}{r^2 + a^2} + \frac{z\,dz}{r} + dt\right]^2, \tag{2.4}$$

r being defined by

$$[(x^2 + y^2)/(r^2 + a^2)] + z^2/r^2 = 1, \tag{2.5}$$

which corresponds to one of the two forms (2.3). We denote the coordinate system (x, y, z, t) as the (M) frame, and identify it with the form (2.3a) in which

[18] J. N. Goldberg and R. K. Sachs, Acta Phys. Polon. 22, 13 (1962).

the **k** congruence is displayed. The (altogether different) coordinate system (x', y', z', t') adapted to the **l** congruence we call the (M') frame. The transformation equations relating these two systems can be worked out from formulas given later in this section.

The form (2.4) of the Kerr metric is inconvenient for many applications, since r is a complicated function of x, y, z. Kerr has given alternative and more suitable form, in which r is one of the coordinates; he applies to Eq. (2.4) the transformation

$$\begin{aligned} x &= (r^2 + a^2)^{\frac{1}{2}} \sin\theta \cos[\varphi - \tan^{-1}(a/r)], \\ y &= (r^2 + a^2)^{\frac{1}{2}} \sin\theta \sin[\varphi - \tan^{-1}(a/r)], \\ z &= r\cos\theta, \end{aligned} \tag{2.6}$$

and obtains

$$\begin{aligned} ds^2 &= dr^2 + 2a\sin^2\theta\, dr\, d\varphi + (r^2 + a^2)\sin^2\theta\, d\varphi^2 \\ &+ \Sigma\, d\theta^2 - dt^2 + (2mr/\Sigma)(dr + a\sin^2\theta\, d\varphi + dt)^2. \end{aligned} \tag{2.7}$$

Here

$$\Sigma(r, \theta) \equiv r^2 + a^2\cos^2\theta. \tag{2.8}$$

This is a generalization of the Eddington[6] form of the Schwarzschild metric [cf. Eq. (1.2)]; accordingly, we refer to the coordinate system (r, θ, φ, t) as the (E) frame. Note that in the asymptotic region, $r \to \infty$, Eq. (2.7) reduces to Eq. (1.2), which justifies the interpretation of the parameter m as the total mass of the source. (We therefore assume $m > 0$.) By comparing higher-order terms with the weak-field metric for a rotating body as given by Landau–Lifshitz,[19] or by calculating the corrections to the perihelion precession formula and comparing with the Lense–Thirring result,[20] one can show that the parameter a is just the angular momentum per unit mass of the source. We see later that the Kerr solution radically changes its character for $|a| > m$.

(r, θ, φ) Coordinate Surfaces

To explore the properties of the principal null congruences in greater detail, we find it convenient to study the relation between the (E) and (M) coordinates, by viewing the surfaces $r, \theta, \varphi =$ const in a Euclidean 3-space whose Cartesian coordinates are (x, y, z). From Eq. (2.5) it is clear that the surfaces $r =$ const are confocal ellipsoids, while from

$$\frac{x^2 + y^2}{a^2\sin^2\theta} - \frac{z^2}{a^2\cos^2\theta} = 1 \tag{2.9}$$

it follows that the $\theta =$ const surfaces are hyperboloids of one sheet, confocal to the ellipsoids. Actually, since z has the same sign as $\cos\theta$, the surface $\theta =$ const is only a half-hyperboloid, truncated at its waist, lying in the half-space $z \gtrless 0$ according as $\theta \lesseqgtr \frac{1}{2}\pi$. Note that at $r = 0$ the ellipsoid degenerates to a disk, $x^2 + y^2 = a^2\sin^2\theta$, $z = 0$. The boundary of this disk [where $r = 0$, $\theta = \frac{1}{2}\pi$ and therefore $\Sigma(r, \theta) = 0$] is of particular importance, since it is precisely this set of points at which the metric (2.7) becomes singular.

The surfaces $\varphi =$ const have the appearance of bent planes, which are approximately vertical for large r but flatten out and become horizontal at the edge of the disk (Fig. 2). Letting φ be fixed but arbitrary, set

$$\xi = x\cos\varphi + y\sin\varphi, \qquad \eta = -x\sin\varphi + y\cos\varphi.$$

Equation (2.6) then yields

$$\xi = r\sin\theta, \qquad \eta = -a\sin\theta, \qquad z = r\cos\theta,$$

or equivalently

$$(a/\eta)^2 - (z/\xi)^2 = 1. \tag{2.10}$$

This defines a ruled quartic surface, whose generators are given by $\eta = -a\sin\theta$, $z = \xi\cot\theta$, with θ held constant on a given generator. (Hence each line is also a generator of the corresponding hyperboloid $\theta =$ const.) Since $\xi = x$, $\eta = y$ for $\varphi = 0$, we have in fact constructed the surface $\varphi = 0$. And because of the way (ξ, η) are related to (x, y), it is obvious that all the other $\varphi =$ const surfaces are

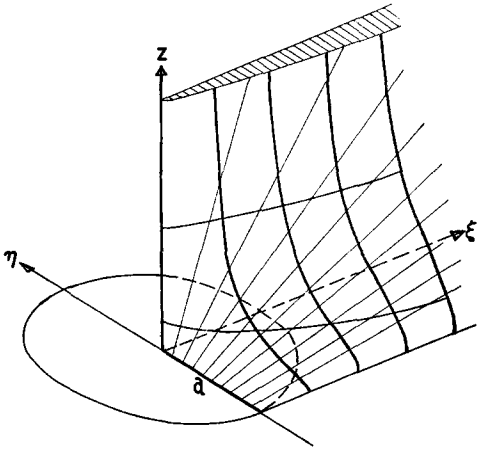

FIG. 2. The ruled quartic surface $\varphi =$ const, shown imbedded in a Euclidean 3-space with Cartesian coordinates (ξ, η, z).

[19] L. D. Landau and E. M. Lifshitz, *Classical Theory of Fields* (Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962), 2nd ed., p. 359.

[20] R. H. Boyer and T. G. Price, Proc. Cambridge Phil. Soc. **61**, 531 (1965).

merely replicas of this one, and can be obtained by rotating it about the z axis.

Throughout the foregoing we have tacitly assumed that r is positive. However, the disk $r = 0$, which has been viewed as a membrane of discontinuity, may just as well be considered a two-sided aperture to a second sheet on which r is negative. Such a continuation to negative r values is permissible because the metric (2.7) remains regular at $r = 0$ (provided $\theta \neq \frac{1}{2}\pi$). We hasten to remark that this second sheet is not to be confused with the other side of an Einstein–Rosen bridge,[10] which one encounters in the familiar Kruskal procedure. We construct an analog of the latter in Sec. III; it has properties quite different from the $r < 0$ extension contemplated here. For one thing, the behavior of the metric as $r \to -\infty$, with $m > 0$ by definition and therefore m/r negative, describes the sort of geometry one would expect far from a particle of *negative* mass. On the other hand, the two sides of an Einstein–Rosen bridge are isometric, so both describe the field outside a positive-mass body.

In the Kerr–Schild theory the most general metric is determined by an analytic function of one complex variable. For the special case of the Kerr solution, this function has a branch point at the ring singularity $r = 0$, $\theta = \frac{1}{2}\pi$. Thus one can properly view the continuation of (2.7) to negative r values as an analytic continuation onto the second Riemann surface of this function. By passing twice in the same direction through the ring one returns to the original Riemann sheet, and thus to a manifold isometric to the original one, which may for simplicity be identified with it.

"Schwarzschild-Like" Coordinates

Making the transformation

$$\begin{aligned} \bar{r} &= r, \qquad \bar{\theta} = \theta, \\ d\bar{\varphi} &= d\varphi + a\, dr/\Delta(r), \\ d\bar{t} &= dt - 2mr\, dr/\Delta(r), \end{aligned} \tag{2.11}$$

where

$$\Delta(r) \equiv r^2 - 2mr + a^2, \tag{2.12}$$

we find that

$$\begin{aligned} ds^2 = \Sigma(dr^2/\Delta + d\theta^2) + (r^2 + a^2)\sin^2\theta\, d\bar{\varphi}^2 - d\bar{t}^2 \\ + (2mr/\Sigma)(a\sin^2\theta\, d\bar{\varphi} + d\bar{t})^2. \end{aligned} \tag{2.13}$$

This form has only one off-diagonal component, $g_{\bar{\varphi}\bar{t}}$, and is thus invariant under the transformation $\bar{\varphi} \to -\bar{\varphi}$, $\bar{t} \to -\bar{t}$. We refer to $(r, \theta, \bar{\varphi}, \bar{t})$ as (S) coordinates, since they reduce to the standard Schwarzschild coordinates when $a = 0$.

The metric (2.13) has, in addition to the true singularity at $\Sigma = 0$, a pair of pseudosingularities at the real zeros of $\Delta(r)$. The latter are located at

$$r_\pm = m \pm (m^2 - a^2)^{\frac{1}{2}} \tag{1.6}$$

and thus exist only when $a^2 \leq m^2$. The close similarity of this equation with the one arising in the Reissner–Nordström problem has already been remarked, as has the identification of the surfaces $r = r_\pm$ with stationary null surfaces, or horizons. Let $f = 0$ be any null hypersurface containing the Killing vectors $\partial/\partial\varphi$ and $\partial/\partial t$; then

$$f = f(r, \theta), \quad (\text{grad } f)^2 = 0,$$

and from the contravariant form of the metric [given in Eq. (2.15)] we deduce

$$\Delta(r)(\partial f/\partial r)^2 + (\partial f/\partial\theta)^2 = 0. \tag{2.14}$$

The only solutions of this equation periodic in θ are the ellipsoids $r = r_\pm$. The cases $a = m$ and $a = 0$ (Schwarzschild) are seen to be exceptional: In the former case the two horizons coalesce, while in the latter case $r_- = 0$, which is not a null surface at all but an essential singularity.

Because of these additional spurious singularities the (S) coordinates are an inappropriate tool to use in studying the analytic properties of the vacuum metric; evidently in any such investigation the (E) coordinates are far superior.[21] but the (S) coordinates have one important advantage which we exploit in later sections, namely, they treat both of the principal null congruences on an equal footing.

Principal Null Congruences

The inverse (contravariant) form of Eq. (2.7) is

$$\begin{aligned} (\text{grad})^2 = \Sigma^{-1}\Big[(r^2 + a^2)\Big(\frac{\partial}{\partial r}\Big)^2 - 2a\Big(\frac{\partial}{\partial r}\Big)\Big(\frac{\partial}{\partial\varphi}\Big) \\ + \frac{1}{\sin^2\theta}\Big(\frac{\partial}{\partial\varphi}\Big)^2 + \Big(\frac{\partial}{\partial\theta}\Big)^2\Big] \\ - \Big(\frac{\partial}{\partial t}\Big)^2 - \Big(\frac{2mr}{\Sigma}\Big)\Big(\frac{\partial}{\partial r} - \frac{\partial}{\partial t}\Big)^2, \end{aligned} \tag{2.15}$$

which is again of the form: flat-space metric plus the square of a null vector [cf. Eq. (2.1b)]. Thus the contravariant components of the null congruence $\mathbf{k}$ can be read off immediately:

$$\mathbf{k} \equiv (k^r, k^\theta, k^\varphi, k^t) = (-1, 0, 0, 1). \tag{2.16a}$$

[21] This feature of the original Eddington coordinates was first clearly recognized by D. Finkelstein.[6]

With this normalization Eq. (2.2a) is satisfied, and consequently t (or r) is an affine parameter along the congruence. Having chosen $k^t > 0$, we can say that $\mathbf{k}$ is future-pointing, and since $k^r < 0$, we conclude that the $\mathbf{k}$ congruence is *ingoing*.

Next recall that $\mathbf{k}$ must appear as a *linear* congruence when expressed in (M) coordinates. Since $k^\theta = k^\varphi = 0$, such lines must lie in the surfaces $\theta, \varphi = \text{const}$, and are therefore the common generators of these surfaces. As t increases, the rays proceed inwards toward the disk $r = 0$, cross it, then emerge onto the $r < 0$ sheet keeping the same θ and φ values. The rays lying in the equatorial plane $\theta = \frac{1}{2}\pi$ are exceptions. These are tangent to the ring singularity at $\Sigma = 0$, meeting it after a finite lapse of affine parameter r.[22]

It is not easy to detect the remaining principal null vector $\mathbf{l}$ which is hidden in (2.7) or (2.15). Equation (2.13) offers a clue; it is invariant under the transformation $d\bar{r} \to -d\bar{r}$ which interchanges ingoing and outgoing rays. In system (S), therefore, $\mathbf{k}$ and $\mathbf{l}$ differ only in the sign of their radial components. Applying the transformation (2.11) to $\mathbf{k}$ gives us the following:

$$\mathbf{k} = (k^{\bar{r}}, k^{\bar{\theta}}, k^{\bar{\varphi}}, k^{\bar{t}}) = [-1, 0, -a/\Delta, (r^2 + a^2)/\Delta] \qquad (2.16b)$$

and consequently

$$\mathbf{l} = (l^{\bar{r}}, l^{\bar{\theta}}, l^{\bar{\varphi}}, l^{\bar{t}}) = [+1, 0, -a/\Delta, (r^2 + a^2)/\Delta]. \qquad (2.17b)$$

It is then a simple matter to determine the components of $\mathbf{l}$ in the (E) frame:

$$\mathbf{l} = (l^r, l^\theta, l^\varphi, l^t) = [+1, 0, -2a/\Delta, (r^2 + a^2 + 2mr)/\Delta]. \qquad (2.17a)$$

Note that $\bar{r} = r$ is an affine parameter along *both* ray systems, but that t is an affine parameter only for the ingoing congruence.

Equation (2.17a) fails at the horizons $r = r_\pm$, but if we rewrite it as

$$\mathbf{l} = N(\Delta, 0, -2a, r^2 + a^2 + 2mr)$$

with N an unspecified normalization factor, we can take the limit $r \to r_\pm$ without difficulty. The result,

$$\mathbf{l}_\pm = N_\pm(0, 0, -2a, 4mr_\pm), \qquad (2.18)$$

displays the key role of $\mathbf{l}_\pm$ as the *null generators* of the corresponding horizons (see Fig. 3). The affine parameter for this special case is not obvious—it evidently cannot be r, and it turns out that t does not work either—but we postpone this question until Sec. IV [see in particular Eq. (4.11)].

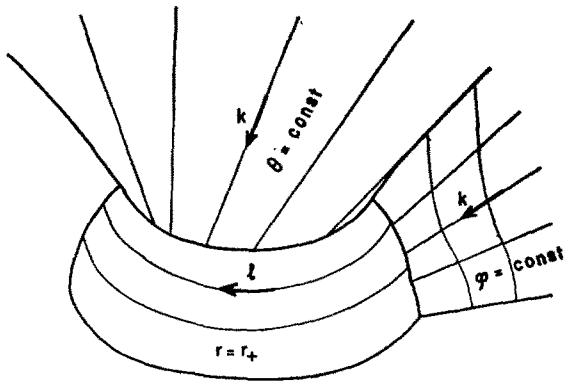

FIG. 3. The horizon $r = r_+$, viewed from the (M) frame with (x, y, z) as Cartesian coordinates. The principal null congruence $\mathbf{l}$ lying in the horizon is shown, as well as portions of the $\mathbf{k}$ congruence lying in the coordinate surface $\theta = \pi/4$ and $\varphi = \text{const}$.

This result appears at first sight to violate the symmetry between the two principal null congruences. However, from Eq. (2.3b), there must exist another coordinate frame—(E') say—adapted to the $\mathbf{l}$ congruence, in which the roles of $\mathbf{k}$ and $\mathbf{l}$ are interchanged. This frame is not hard to find. Using Eq. (2.11) as a guide, we make the obvious transformation

$$\bar{r} = r', \qquad \bar{\theta} = \theta', \qquad d\bar{\varphi} = d\varphi' - a\,dr'/\Delta, \qquad d\bar{t} = dt' + 2mr'\,dr'/\Delta \qquad (2.19)$$

and obtain a line element just like (2.7), except that the sign of dr (or equivalently, of $d\varphi$ and dt) is reversed:

$$ds^2 = dr^2 - 2a\sin^2\theta\,dr\,d\varphi' + (r^2 + a^2)\sin^2\theta\,d\varphi'^2 + \Sigma\,d\theta^2 - dt'^2 + (2mr/\Sigma)(dr - a\sin^2\theta\,d\varphi' - dt')^2. \qquad (2.20)$$

A further transformation, from (E') to the associated Minkowski (M') frame (x', y', z', t'), proceeds essentially via Eqs. (2.6), except that $\varphi - \tan^{-1}(a/r)$ is replaced by $\varphi' + \tan^{-1}(a/r)$. One finds

$$\mathbf{k} = (k^{r'}, k^{\theta'}, k^{\varphi'}, k^{t'}) = [-1, 0, -2a/\Delta, (r^2 + a^2 + 2mr)/\Delta], \qquad (2.16c)$$

$$\mathbf{l} = (l^{r'}, l^{\theta'}, l^{\varphi'}, l^{t'}) = (+1, 0, 0, +1), \qquad (2.17c)$$

which is just the reverse of (2.16a), (2.17a), as

[22] This shows, incidentally, that the ring singularity is real, for it is easily seen that $d^2H/d\mu^2$, which is an algebraic invariant of the Riemann tensor [see Eq. (2.2a)], becomes infinite there.

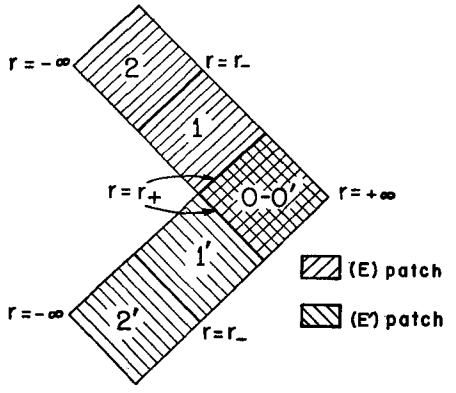

FIG. 4. The pair of (E) and (E') coordinate patches for the Kerr metric, drawn in a fashion roughly analogous to Fig. 1. Even-numbered regions, which are asymptotically flat, are separated by an odd-numbered interior region whose boundaries are the two horizons $r_\pm$. Although the drawing is only meant to be symbolic, the compression of the asymptotic regions is suggestive of a Penrose conformal transformation [R. Penrose, *Relativity, Groups and Topology*, B. DeWitt and C. DeWitt, Eds. (Gordon and Breach Science Publishers, New York, 1964)].

is to be expected. Note that in this frame it is the **k** congruence which provides the generators for the horizons.

This result is not paradoxical, for the two coordinate frames are not equivalent. Combining Eqs. (2.11) and (2.19), one gets the direct transformation equations $(E) \rightarrow (E')$. For $a^2 > m^2$ this transformation is one-to-one over the entire range of r values, but for $a^2 \leq m^2$, which is the interesting case, the (E) and (E') coordinate patches only partially overlap. It is a straightforward matter to show that

$$t' = t - \frac{2m}{(m^2 - a^2)^{\frac{1}{2}}}\left[r_+ \ln\left|\frac{r - r_+}{2m}\right| - r_- \ln\left|\frac{r - r_-}{2m}\right|\right] \qquad (2.21)$$

and

$$\varphi' = \varphi + \frac{a}{(m^2 - a^2)^{\frac{1}{2}}} \ln\left|\frac{r - r_+}{r - r_-}\right|,$$

from which it is clear that either (t, φ) or (t', φ') must diverge at $r_\pm$. We thus obtain three inequivalent analytic extensions of the original (E) frame, according as we choose r to lie in the different intervals $-\infty < r < r_-$, $r_- < r < r_+$, $r_+ < r < \infty$. Putting these several pieces together to make a smooth manifold is the job of the next section.

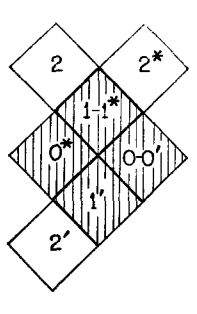

FIG. 5. The diagram of Fig. 4, enlarged by the addition of another (E') patch labeled $\{0^*, 1^*, 2^*\}$. The shaded portion defines the Kruskal or (K) patch.

III. ANALYTIC EXTENSIONS

"Kruskal-Like" Coordinates

Let us mark the three domains $r > r_+$, $r_- < r < r_+$, $r < r_-$ within the (E) patch as regions $\{0, 1, 2\}$, respectively. Transforming from (E) to (E') via Eq. (2.21) we obtain three similar domains, $\{0', 1', 2'\}$. We assume that the transformation is one-to-one (and therefore analytic) between $\{0\}$ and $\{0'\}$; it then follows that $\{1\}$ and $\{1'\}$ are inequivalent—one has a boundary crossed by the **k** congruence, the other by the **l** congruence—and likewise for $\{2\}$ and $\{2'\}$. Putting the two patches together, we obtain the enlarged domain shown schematically in Fig. 4.

Now take another (E') patch, whose sections are labeled $\{0^*, 1^*, 2^*\}$, and tie it into the middle region of the original (E) patch, by requiring that the domains $\{1\}$ and $\{1^*\}$ coincide. The manifold has now grown to the size of Fig. 5.

We have drawn the two regions $\{0^*\}$ and $\{1'\}$ as contiguous, but this is unfair, for there is no reason in the foregoing to suppose that they are in any way related. We now wish to demonstrate that the picture is in fact a reasonable one, by displaying a single coordinate system in which the four regions $\{0, 1, 0^*, 1'\}$ are linked together as shown. Such a linkage looks very similar to a Kruskal diagram (Fig. 1). Our object, therefore, is to find a transformation analogous to (1.4), such that both segments of the horizon $r = r_+$ appear regular when viewed from these new coordinates.

The task is not a trivial one. For one thing, the principal null congruences of the Schwarzschild metric are straightened out in Kruskal coordinates to the lines $u \pm v = \text{const}$, $\theta, \tilde{\varphi} = \text{const}$. One might hope, by combining a transformation of the azimuthal angle with a suitable generalization of Eq. (1.4), to achieve a similar simplification here. But the corresponding vectors **k** and **l** of the Kerr metric are not in general 2-surface-forming, so that such a search is pointless. We settle for a frame in which the principal null rays merely lie in the surfaces $u \pm v = \text{const}$. This solution, while less elegant than Kruskal's, is adequate for our purposes, and is perhaps as close as one can come to a generalization of it.

Since we want to treat both congruences impar-

tially, we start from the (S) frame. From the expressions for $\mathbf{k}$ and $\mathbf{l}$ in these coordinates [Eqs. (2.16b), (2.17b)], it may be seen that as $r \to r_+$ both become asymptotic to a family of helices winding around the horizon, given by

$$dr : d\theta : d\tilde{\varphi} : d\tilde{l} = 0 : 0 : -a : 2mr_+ .$$

We seek a transformation of the azimuthal coordinate which will straighten out the helices. Many obvious possibilities suggest themselves, but one encounters later difficulties unless one picks the new coordinate to be a linear combination of $\tilde{\varphi}$ and $\tilde{l}$ with *constant* coefficients, plus any function of r and θ, and this restricts the choice to

$$w = \tilde{\varphi} + (a/2mr_+)\tilde{l} + \psi(r, \theta). \tag{3.1}$$

For simplicity we take $\psi(r, \theta) = 0$. It is quite possible that other choices might lead to a more tractable form for the metric; we have not investigated this in any detail.

Through Eq. (3.1) we have untwisted the null curves in the limit $r \to r_+$ (at the expense, incidentally, of twisting them in the neighborhood of spatial infinity, which may be disagreeable but is not serious—Kruskal-type coordinates are not at all well adapted to the asymptotically flat portions of the manifold anyway.) The metric computed from (2.13) after this single change of variables is still singular at r_+, of course, but now we know how to deal with it. We make a further transformation $(r, t) \to (u, v)$, such that the integral curves of the two principal congruences lie in the hypersurfaces $u \pm v = \text{const}$, although not in $w = \text{const}$. According to Eqs. (2.16b), (2.17b), these integral curves satisfy

$$dr/\Delta(r) = \mp d\tilde{l}/(r^2 + a^2),$$

or

$$\left(\frac{r - r_+}{2m}\right)\left(\frac{r - r_-}{2m}\right)^{-\nu} \exp\left(\frac{r \pm \tilde{l}}{\sigma_+}\right) = \text{const} = F(u \pm v), \tag{3.2}$$

with

$$\sigma_\pm \equiv mr_\pm(m^2 - a^2)^{-\frac{1}{2}} \tag{3.3}$$

and

$$\nu \equiv r_-/r_+ .$$

From the work of Kruskal and Graves–Brill we are led to take[23] $F(x) = x^2$, and consequently

$$u \pm v = \left(\frac{r - r_+}{2m}\right)^{\frac{1}{2}}\left(\frac{r - r_-}{2m}\right)^{-\nu/2} \exp\left(\frac{r \pm \tilde{l}}{2\sigma_+}\right). \tag{3.4}$$

[23] Had we wanted the transformation to yield a metric regular across r_-, the appropriate choice would have been $F(x) = x^{-2\nu}$.

Although this is only defined for $r > r_+$, $\tilde{l}$ finite (i.e., for $u > |v|$), the inverse transformation, given implicitly by

$$\Psi(r) \equiv \left(\frac{r - r_+}{2m}\right)\left(\frac{r - r_-}{2m}\right)^{-\nu} e^{r/\sigma_+} = u^2 - v^2 \tag{3.5}$$

and

$$\tilde{l} = \sigma_+ \tanh^{-1}\left[2uv/(u^2 + v^2)\right]$$

is well defined over the cut plane $u^2 \neq v^2$. In particular, $\Psi(r)$ increases monotonically from $-\infty$ to 0 to $+\infty$ as r runs from r_- to r_+ to $+\infty$, and is an analytic function of r over this interval, so that $r(u, v)$ is an analytic function over the u, v plane, while $\tilde{l}(u, v)$ diverges at $u = \pm v$ and is analytic elsewhere. As the coordinates (u, v, w, θ)—which we henceforth call the (K) frame—vary over their respective ranges

$$-\infty < u < \infty, \qquad -\infty < v < \infty,$$
$$0 \leq w < 2\pi, \qquad 0 \leq \theta \leq \pi$$

they cover the four regions $\{0, 1, 1'$ and $0^*\}$ of Fig. 5; the boundaries between adjacent regions are the pair of lines $u = \pm v$ on which $r = r_+$. Note that region $\{0^*\}$, in which $u < |v|$ (and therefore $r > r_+$), forms the second sheet of an Einstein–Rosen bridge.

In the following it is convenient to know the direct transformation from (E) to (K) coordinates. This is given by

$$\begin{aligned} u + v &= e^{(r+t)/2\sigma_+}, \\ u - v &= \Psi(r)e^{-(r+t)/2\sigma_+}, \\ w &= \varphi + \left(\frac{a}{2mr_+}\right)t - \frac{a}{r_+}\ln\left(\frac{r - r_-}{2m}\right). \end{aligned} \tag{3.6}$$

The transformation from (E') to (K) is obtained similarly, with t replaced by $-t'$. Equation (3.6) evidently defines an analytic homeomorphism of the domain $r > r_-$, $-\infty < t < \infty$ onto the half-plane $u + v > 0$, i.e., onto the region $\{0, 1\}$.

On the other hand, we could have mapped the domain $-\infty < r < r_+$, $-\infty < t < \infty$ homeomorphically into a (different) Kruskal patch by regularizing about the other event horizon, $r = r_-$.[23] The latter transformation, $(E) \to (K')$ say, takes the form

$$\begin{aligned} u' + v' &= -e^{-(r+t)/2\sigma_-}, \\ u' - v' &= -\left(\frac{r_- - r}{2m}\right)\left(\frac{r_+ - r}{2m}\right)^{-1/\nu} e^{-(r-t)/2\sigma_-}, \\ w' &= \varphi + \left(\frac{a}{2mr_-}\right)t - \frac{a}{r_-}\ln\left(\frac{r_+ - r}{2m}\right). \end{aligned} \tag{3.7}$$

We postpone a discussion of the full import of this result until later in the section, in order to settle a fundamental point.

Analyticity of the Metric in (K) Coordinates

There remains the very important but rather tedious job of working out the components of the Kerr metric in (K) coordinates, and demonstrating that they are indeed analytic functions of u and v throughout the (u, v) plane. We spare you the details, and merely assert that after much labor one can cast the metric in the following form:

$$\begin{aligned} ds^2 &= \Sigma\, d\theta^2 + 4\sigma_+^2 f\Sigma(r^2 + a^2)^{-2}(du^2 - dv^2) \\ &\quad + \Sigma^{-1}\sin^2\theta[(r^2 + a^2)\, dw + a(m^2 - a^2)^{-\frac{1}{2}} f \\ &\quad \times (r + r_+)(r - r_-)^{-1}(v\, du - u\, dv)]^2 \\ &\quad + (\Sigma\Delta)^{-1}\{[2\sigma_+ f\Sigma(r^2 + a^2)^{-1}(v\, du - u\, dv)]^2 \\ &\quad - [2\sigma_+ f\Sigma_+(r_+^2 + a^2)^{-1}(v\, du - u\, dv) \\ &\quad - a\sin^2\theta\Delta\, dw]^2\}. \end{aligned} \tag{3.8}$$

Here we have introduced

$$\Sigma_\pm = \Sigma(r_\pm, \theta) \tag{3.9a}$$

and

$$\begin{aligned} f(r) &= \Delta(r)/\Psi(r) \\ &= 4m^2[(r - r_-)/2m]^{2m/r_+} e^{-r/\sigma_+}. \end{aligned} \tag{3.9b}$$

Since Σ and f are both analytic and nonzero throughout the whole (u, v) plane, it is clear that only the term in (3.8) with $\Delta(r)$ in the denominator can lead to difficulty. Closer inspection shows, however, that the numerator of this term has a simple zero at $r = r_+$ also, so the quotient remains analytic.

Had we chosen to regularize across r_- instead, using Eq. (3.7) or its equivalent, we would have obtained much the same result, except for the inevitable branch point at $r(u, v) = 0$ and $\theta = \frac{1}{2}\pi$.

The line element (3.8) is much too cumbersome to be of any direct use. Fortunately, once we have established its analyticity we need make no further use of it. This is to some extent regrettable, for it would have been desirable to have a coordinate system in which, like Kruskal's, the null curves took an especially simple form (see Fig. 6). Cohen and Brill,[24] in their study of metrics for slowly rotating bodies, have applied the Kruskal procedure to a line element differing from the Schwarzschild form by terms of first order in the angular velocity. We recover their results if we drop from Eq. (3.8) all terms of order a^2; this leaves

$$\begin{aligned} ds^2 &\simeq r^2(d\theta^2 + \sin^2\theta\, dw^2) \\ &\quad + (32m^3/r)e^{-r/2m}(du^2 - dv^2) \\ &\quad + (4a/r)\sin^2\theta e^{-r/2m} \\ &\quad \times (r^2 + 2mr + 4m^2)\, dw\, (v\, du - u\, dv). \end{aligned} \tag{3.10}$$

To this order, the curves $u \pm v = \text{const}$, w, $\theta = \text{const}$ are null, just as in the Kruskal form of the Schwarzschild metric.

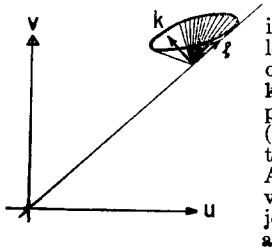

FIG. 6. The null cone at a point in the horizon $r = r_+$. The cone is lopsided and spills out of the angle defined by the principal null vectors k and l; in addition, the shaded portion of the cone behind the (u, v) plane is smaller than the portion jutting out in front of it. Although the shape of the cone varies from point to point, the projections of k and l onto the plane always have slope ± 1.

Construction of a Maximal Analytic Manifold

What we have done to extend the domain $\{0, 1\}$ we can also do to the domain $\{1, 2\}$. We use the alternative transformation (3.7) to produce a new (K') patch, consisting of the three regions $\{1, 2, 2^*\}$ already introduced plus a new domain $\{3\}$, which, like $\{1\}$, is bounded by two pairs of horizons (see Fig. 7). Here also the asymptotically flat portions, $\{2\}$ and $\{2^*\}$, form two sheets of an Einstein–Rosen bridge. While isometric to each other, these are not isometric to $\{0\}$ and $\{0^*\}$—they have, it may be recalled, the sort of geometry one would associate with a negative mass source.

A similar extension can be applied to the domain $\{1', 2'\}$. In view of the picture which is rapidly emerging, we find it preferable to change our notation slightly, and to write $\{-1, -2\}$ instead of $\{1', 2'\}$. Thus the new Kruskal patch encompasses the regions $\{-1, -2, -2^*, -3\}$ of Fig. 7.

This procedure can clearly be continued indefinitely in both directions, and generates an infinite chain of overlapping (E) and (E') patches. The result can be viewed even more simply as a chain of overlapping "Kruskal" patches, $\{K_{2n}\}$, isometric to (K) or (K') according as n is even or odd. We can, in fact, eliminate any reference whatever to the auxiliary (E) and (E') frames, by presenting the transformation equations that connect (K) and (K'); these take the surprisingly simple form

$$\begin{aligned} (u + v)^{r_+}(-u' - v')^{r_-} &= 1, \\ (-u + v)^{r_+}(u' - v')^{r_-} &= 1, \end{aligned} \tag{3.11a}$$

[24] D. R. Brill and J. M. Cohen, Phys. Rev. 143, 1011 (1966).

and

$$w' - w = (2a/r_-) \ln [(u + v)/(-u + v)]$$
$$= -(2a/r_+) \ln [(u' - v')/(-u' - v')]. \quad (3.11b)$$

These transformations are one-to-one and analytic in the domains $\cdots \{1\}, \{3\}, \cdots$ of overlap, and have the following simple properties: They preserve straight lines of slope ± 1 as well as the straight lines $v/u = \text{const}$ and the hyperbolas $v^2 - u^2 = \text{const}$, while mapping the north quadrant of (K) onto the south quadrant of (K').

The manifold we have constructed has many curious properties. It has infinitely many disjoint and asymptotically flat sheets, for one thing, and therefore does not admit a Cauchy surface. In fact, one can get from any such even-numbered sheet to any other—with the sole exception of the companion sheet on the other side of the "bridge"—by following a suitable timelike or lightlike curve. The timelike curve $u = 0 (u' = 0)$, w, $\theta = \text{const}$ is particularly noteworthy, for it defines the location of the "throat" of an Einstein–Rosen bridge. The throat itself (i.e., the 2-surface $u = 0$, $v = \text{const}$) has an area which pulsates with time, in close analogy with the Reissner–Nordström case.[12] Since the throat is deformed, due to the effects of rotation, its area is no longer given by the simple formula $4\pi r^2$ obtained for the Schwarzschild and Reissner–Nordström metrics, but by the more complicated expression

$$\text{area} = 2\pi\{(r^2 + a^2) + (g/a)(-\Delta)^{-\frac{1}{2}} \sin^{-1} [a(-\Delta/g)^{\frac{1}{2}}]\} \quad (3.12)$$

with $g(r) = (r^2 + a^2)^2 - a^2\Delta$. At the horizons $r = r_\pm$, where the extrema occur, this simplifies to

$$\text{area}_\pm = 8\pi m r_\pm. \quad (3.12)$$

The manifold sketched in Fig. 7 is not the only possible maximal extension, because of the freedom still remaining to make topological identifications. The simplest manifold of this more general type arises if we identify $\{K_{2n}\}$ with $\{K_{2n\pm4}\}$. It is covered by only two Kruskal patches, $\{K_0\}$ and $\{K_2\}$ say, with regions $\{-1\}$ and $\{3\}$ identified, so it consists of four asymptotically flat sheets (two with positive mass, two negative) glued together with two interior regions.[25] This manifold, like the others obtained in the same way, is violently acausal—a properly aimed signal will emerge, after crossing four horizons, in the past light cone of the source which emitted it—and this may be felt to be a bit unrealistic.[26]

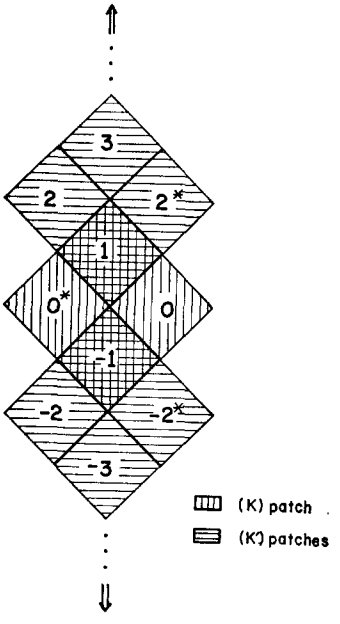

FIG. 7. A chain of Kruskal patches, with those of (K) and (K') type alternating and partially overlapping. The complete figure extends indefinitely in both directions, to form a maximal analytic extension of the Kerr manifold.

The foregoing construction presupposes, of course, that the metric admits two distinct horizons. For $a^2 > m^2$ the horizons disappear and the transformation to Kruskal coordinates loses all meaning; all that remains are two asymptotically flat spaces joined at the disk $r = 0$. The exceptional case $|a| = m$ deserves special comment. Carter has studied this in detail[16]; although his work was confined to the symmetry axis $(\theta = 0, \pi)$, it is clear that his conclusions apply with equal force to the full metric: One builds up a ladder of alternating (E) and (E') patches (with or without identifications) in the simple sequence $\{\cdots -2, 0, 2, \cdots\}$. The odd-numbered sheets disappear; there is nothing to correspond to the second Kruskal sheets $\{\cdots -2^*, 0^*, 2^*, \cdots\}$; and in fact there is no need at all to introduce (K)-type coordinates to cover the manifold in this case.

IV. GEODESIC COMPLETENESS

Geodesic Equations, First Integrals

For the study of geodesics the original (E) coordinates are particularly convenient. Let μ be an affine path parameter, normalized to give proper time along timelike geodesics, and use a dot to

[25] This particular choice was made by Graves and Brill.[12] It is clear from their work, however, that the identification is unnecessary; one can also describe the Reissner–Nordström metric by an infinite chain of Kruskal patches.

[26] B. Carter has pointed out that for small negative values of r and values of θ near the equator $g_{\varphi\varphi}$ goes negative. Hence as long as one adopts a topology in which φ is treated as an angular coordinate there will necessarily exist closed timelike curves in this region.

denote differentiation with respect to μ. The geodesic equations can be extracted in the usual way from the variational principle $\delta\int L\, d\mu = 0$, with a Lagrange function given by

$$\begin{aligned} 2L = {} & \dot{r}^2 + 2a\sin^2\theta\dot{r}\dot{\varphi} \\ & + (r^2 + a^2)\sin^2\theta\dot{\varphi}^2 + \Sigma\dot{\theta}^2 - \dot{t}^2 \\ & + (2mr/\Sigma)(\dot{r} + a\sin^2\theta\dot{\varphi} + \dot{t})^2 \end{aligned} \tag{4.1}$$

as in Eq. (2.7).

We are chiefly interested not in the equations of motion themselves, but rather in their first integrals. Since φ and t are cyclic, we obtain two integrals immediately:

$$\begin{aligned} p_\varphi &\equiv \partial L/\partial\dot{\varphi} = l, \\ p_t &\equiv \partial L/\partial\dot{t} = -\gamma, \end{aligned} \tag{4.2a}$$

with l and γ representing the angular momentum/mass and energy/mass of a test particle moving along the given geodesic. (On null geodesics μ is defined only up to a linear transformation, and consequently only the ratio l/γ is meaningful. We remove the arbitrariness by taking $\gamma = 1$ in this case.) A third integral is given by L itself:

$$2L = -\epsilon, \tag{4.2b}$$

the indicator ϵ being $+1$, 0, -1 for timelike, null, and spacelike geodesics, respectively.

The remaining two conjugate momenta are readily found to be

$$\begin{aligned} p_r &= (\Sigma\dot{r} + al + 2mr\gamma)/\Delta, \\ p_\theta &= \Sigma\dot{\theta}, \end{aligned} \tag{4.3}$$

which shows, incidentally, that we can expect difficulties with the equations of motion at the ring singularity ($\Sigma = 0$) and at the horizons ($\Delta = 0$). It is clear from the form of L that $p_\theta = 0$ is consistent with the equations of motion if $\theta = 0$ or π (axial case) or if $\theta = \frac{1}{2}\pi$ (equatorial case); in these special cases the first integrals (4.2) yield a complete solution by quadratures, which we consider in greater detail in Sec. V.

From Eq. (4.2a) we can express $\dot{\varphi}$ and $\dot{t}$ in terms of $\dot{r}$ and the constants of integration; the resulting formulas are

$$\begin{aligned} \dot{\varphi} &= \Sigma^{-1}[(l/\sin^2\theta) - ap_r], \\ \dot{t} &= \Sigma^{-1}[\gamma(\Sigma + 2mr) + 2mrp_r], \end{aligned} \tag{4.4}$$

with p_r given by (4.3).

We note also the following form of Eq. (4.2b), which is important in later arguments:

$$\begin{aligned} \Sigma^2\dot{r}^2 - {} & (al + 2mr\gamma)^2 \\ = {} & \Delta[-\epsilon\Sigma + (\Sigma + 2mr)\gamma^2 - p_\theta^2 - (l^2/\sin^2\theta)]. \end{aligned} \tag{4.5}$$

Problem of Completeness

The Kerr metric evidently cannot be imbedded isometrically in a complete analytic manifold, since there exist geodesics—the principal null rays lying in the equatorial plane, for example—which strike the ring singularity at finite values of their affine parameters. If one excludes all such geodesics, it is reasonable to ask whether the remainder can be continued to arbitrarily large values of μ. We argue below that this is so, provided that the manifold is chosen to be the analytic extension described previously. In this sense, therefore, the extension can be regarded as maximal.

Since it is impossible to solve the geodesic equations exactly, except in a few very special cases, we base the argument on the first integrals (4.4) and (4.5). Starting with any set of initial values, one extends the solution either until a singular point is reached, or until one or more coordinates diverge. There is clearly no problem if r diverges—since the metric is asymptotically flat and therefore complete for $r \to \pm\infty$—but only if $\dot{\varphi}$, $\dot{t}$, or $\dot{\theta}$ diverge at finite values of r. As one sees from the above equations, this happens at $\Sigma = 0$, $\sin\theta = 0$ or $\Delta = 0$. If the first possibility occurs nothing can be done about it; the second is obviously a consequence of the spheroidal-type coordinates here employed, and can be eliminated by transforming to those of the (M) type. Thus it is only the apparent divergence of $\dot{\varphi}$ and $\dot{t}$ at the horizons that needs to be examined carefully. [It follows from Eq. (4.5) that $\dot{r}$ remains finite at $r_\pm$; the same is true of $\dot{\theta}$.] In fact, $\dot{\varphi}$ and $\dot{t}$ diverge if and only if p_r diverges, and it is easy to see that this happens if and only if $\dot{r}_\pm$ has the same sign as $al + 2mr_\pm\gamma$, for Eqs. (4.3) and (4.5) imply

$$p_r = \frac{-\epsilon\Sigma + (\Sigma + 2mr)\gamma^2 - p_\theta^2 - (l^2/\sin^2\theta)}{\Sigma\dot{r} - (al + 2mr\gamma)}, \tag{4.6}$$

which yields a finite limit whenever $\dot{r}_\pm$ and $(al + 2mr_\pm\gamma)$ have opposite sign—that is, whenever the geodesic is ingoing. [Some confusion over the meaning of the term "ingoing" can arise here, unless one is careful. On the positive r sheet the hypersurfaces $t = \text{const}$ are everywhere spacelike: $(\text{grad}\, t)^2 = -1 - 2mr/\Sigma$, which is certainly negative for $r > 0$. Hence an ingoing path is properly defined as one

for which $dr/dt < 0$. For some values of l and γ Eqs. (4.3), (4.4) suggest that $dr/d\mu$ is positive on an ingoing path, but this simply means that the affine parameter has been chosen to increase into the past.]

We therefore conclude that all ingoing paths can be continued across the horizons. Some subsequently strike the ring singularity, others continue on to $r = -\infty$ (and are thus complete), a third class reaches a turning point and start back. On the return trip, however, both $\dot{t}$ and $\dot{\varphi}$ diverge as the horizon is approached. This is not surprising; the particle has merely run off the (E) coordinate patch. To continue to follow its motion transform to an (E') system, or equivalently to an appropriate set of (K) coordinates. It is clear that a result similar to Eq. (4.6) must also apply to the outgoing trajectory in this case; hence the particle re-emerges into an asymptotically flat "universe," but of course on a sheet different from the first. If the effective total energy $\Gamma \equiv \gamma^2 - \epsilon$ is positive, the particle escapes to infinity; if negative, it reaches another turning point and starts back toward $r = r_+$ again, in which case we transform to another set of (E) coordinates and repeat the cycle. Clearly μ can be made as large as we please by piecing together sufficiently many such cycles, which proves completeness for this case.

Of course there are many geodesics which do not fall into any of the above categories, such as the ones which oscillate to and fro between a maximum and minimum radius outside r_+, and those which spiral in towards (or out from) an unstable circular orbit. However, these lead in general to no difficulties with completeness, since $\dot{t}$ remains finite as t increases to infinity.

There remains one further class to be considered, namely, the geodesics for which $\dot{r} = 0$ at $r_\pm$, or equivalently, for which

$$al = 2mr_\pm\gamma. \tag{4.7}$$

This class includes spacelike geodesics tangent to the horizons (these present no problems), geodesics of all three types which approach the horizon asymptotically as $t \to \pm\infty$, and finally, and most importantly, the principal null geodesics that are the generators of the horizons. We study the latter in detail below.

Completeness within the Horizons

To obtain the solution for the principal null ray 1 lying in $r_\pm$ set $\Delta = 0$, $\epsilon = 0$ and also

$$\dot{\theta} = 0, \qquad \gamma = l = 0. \tag{4.8}$$

The first integrals (4.5), (4.6) then yield the common solution

$$\dot{\varphi} = -(a/2mr_\pm)\dot{t} \tag{4.9}$$

in agreement with Eq. (2.18). To find $t(\mu)$ it is necessary to solve one of the geodesic equations, and the simplest to use is $\dot{p}_r = \partial L/\partial r$. A straightforward computation gives

$$\ddot{t} = \mp(2\sigma_\pm)^{-1}\dot{t}^2,$$

with $\sigma_\pm$ defined by Eq. (3.3). Hence

$$t(\mu) = \pm 2\sigma_\pm \ln(\mu - \mu_0) \tag{4.10a}$$

or

$$\mu - \mu_0 = e^{\pm t/2\sigma_\pm}. \tag{4.10b}$$

We emphasize that this result is exact.

On the event horizon $r = r_+$, $\mu \to +\infty$ as $t \to +\infty$, while $\mu \to \mu_0$ as $t \to -\infty$. This is just what one would have expected, since the (E) coordinates are known to be incomplete at r_+ for $t \to -\infty$. A full picture of the horizon $r = r_+$ is provided by viewing it in (K) coordinates; then the path equations become

$$u(\mu) = v(\mu) = \tfrac{1}{2}(\mu - \mu_0)e^{r_+/2\sigma_+}, \qquad w(\mu) = \text{const} \tag{4.11}$$

and are evidently complete. A similar result applies at $r = r_-$, but with the time directions reversed; this too is consistent with the complete picture of this horizon when viewed from (K') coordinates (as in Fig. 7).

Qualitatively similar results emerge when one analyzes the geodesics that approach $r_\pm$ asymptotically (i.e., as $t \to \pm\infty$). All timelike curves of this type reach a vertex such as $u = v = 0$ in a finite proper time and can be extended without difficulty; similar remarks apply to the null curves. There are others, necessarily spacelike, which approach r_+ from region $\{1\}$ as $t \to +\infty$ (or r_- as $t \to -\infty$), and these are in fact complete.

V. Equatorial Geodesics

The first integrals obtained in the previous section yield a complete description when $\theta = 0$ or π or when $\theta = \frac{1}{2}\pi$. The former case, which corresponds to motion along the symmetry axis, has been investigated by Carter[16]; we confine our attention, therefore, to the latter. Boyer and Price[20] have shown that the orbit equation for equatorial geodesics leads to a precession of the pericenter in agreement, through third order, with the approximate

calculations of Lense–Thirring. In the present section we concentrate on the qualitative features of geodesics in the strong field regions. Our analysis is patterned after the study of orbits in the Schwarzschild metric first carried out by Darwin[27] and later extended by Mielnik and Plebański.[28] Of course, all these studies are largely academic exercises, since in realistic situations (except possibly the late phases of gravitational collapse[3]) the geometry in these regions, and hence the geodesics themselves, differ considerably from the empty-space results, due to the nonzero stress-energy tensor there. Nevertheless, the study of the Kerr metric, even as an ideal case, has, we believe, some real value, for it helps to clarify the role which angular momentum plays in general relativistic models. In particular, because the character of the orbits in the interior regions changes markedly as soon as the central body is given some angular momentum, it seems worthwhile to point out those features of Darwin's analysis which are unique to the Schwarzschild problem, and those which persist in the $a \neq 0$ case as well.

Null Geodesics

It is simplest, and most instructive, to begin with the null rays. We normalize the affine parameter along the rays by taking $\gamma = 1$. Then the energy integral (4.5) reduces to

$$\dot{r}^2 = 1 + (a^2 - l^2)r^{-2} + 2m(a + l)^2 r^{-3}. \qquad (5.1)$$

For convenience set $\beta = 2m$ and introduce the dimensionless variables

$$\rho = r/\beta, \qquad \lambda = l/\beta, \qquad \alpha = a/\beta. \qquad (5.2)$$

From (5.1) it is clear that turning points occur at the zeros of the cubic polynomial

$$\psi(\rho) = \rho^3 + (\alpha^2 - \lambda^2)\rho + (\alpha + \lambda)^2. \qquad (5.3)$$

The location of these zeros is thus fundamental to a qualitative understanding of the null trajectories.

Applying the rule of signs, we see that $\psi(\rho) = 0$ has always one real negative root: a ray sent in from $r = -\infty$ is thus repelled and ultimately deflected back to $-\infty$. (The outgoing principal null ray, with $\lambda = -\alpha$, is an exceptional case; it alone strikes the $r = 0$ singularity from this direction.) On the positive sheet there are consequently either zero or two turning points, depending on the relative magnitudes of α and λ. If $\lambda^2 < \alpha^2$ there are clearly no real positive roots, so collapse to the singularity inevitably occurs. Conversely, if λ^2 is sufficiently large, incoming rays reach a pericenter and return to infinity, while outgoing rays from $r = 0$ reach an apocenter and fall back in. Thus one expects that there should exist two critical impact parameters, $\lambda_1 > 0$ and $\lambda_2 < 0$ say, such that light signals spiral in to $r = 0$ if $\lambda_2 < \lambda < \lambda_1$, and "bounce" back out to infinity otherwise. At these critical values of λ, $\psi(\rho)$ has a double zero, which defines the corresponding critical radii ρ_1, ρ_2. A light ray at such a radius, and with the correct value for λ, will travel around in a circular orbit indefinitely, but such an orbit is, of course, unstable.

These predictions are borne out in the Schwarzschild case by Darwin's analysis. There, it may be recalled, the critical radii are both located at $\rho = \frac{3}{2}$ (or $r = 3m$), and the critical impact parameters are $\lambda = \pm 3(\frac{3}{2})^{\frac{1}{2}}$. Most importantly, there do not exist any light rays whose pericenters lie inside $r = 3m$. If one imagines the parameter α (or a) being increased gradually from zero, one expects $\rho_1(\alpha)$ and $\rho_2(\alpha)$ to depart smoothly from the Darwin value, and this is precisely what happens. In fact, one can give fairly simple closed expressions for the critical radii. These must be double zeros of $\psi(\rho)$ and thus must satisfy the condition

$$\rho(\rho - \tfrac{3}{2})^2 - 2\alpha = 0.$$

Solving this by standard methods, one finds[29]

$$\begin{aligned} \rho_1 &= \tfrac{3}{2} + \alpha \sec(\tfrac{1}{3}\cos^{-1} 2\alpha), \\ \rho_2 &= \tfrac{3}{2} + \alpha \sec(\tfrac{1}{3}\cos^{-1} 2\alpha + \tfrac{2}{3}\pi), \end{aligned} \qquad (5.4a)$$

provided that $\alpha \leq \frac{1}{2}$; there is only one critical radius, at

$$\rho_1 = \tfrac{3}{2} + \alpha \operatorname{sech}(\tfrac{1}{3}\cosh^{-1} 2\alpha), \qquad (5.4b)$$

if $\alpha > \frac{1}{2}$. (Note that ρ_1 is always larger than ρ_2. This is to be expected, since the centrifugal barrier is stronger if λ is positive.)

However, when $\alpha \neq 0$ a completely novel feature emerges: for a small range of impact parameters, $\lambda_3 \leq \lambda < -\alpha$, pericenters exist *inside the inner horizons*, and, in fact, for all values of ρ between 0 and ρ_-. This remarkable property permits one to transmit information from one positive sheet to another by bouncing a light signal off a centrifugal barrier inside ρ_-.

[27] C. Darwin, Proc. Roy. Soc. (London) **A249**, 180 (1959); **A263**, 39 (1961).

[28] B. Mielnik and J. Plebański, Acta Phys. Polon. **21**, 239 (1962).

[29] A third zero, $\rho_3 = \frac{3}{2} + \alpha \sec(\frac{1}{3}\cos^{-1} 2\alpha - \frac{2}{3}\pi)$, has a somewhat different interpretation: it describes the maximum possible apocenter within the inner horizon.

This result is illustrated by Fig. 8, which shows the portion of the λ–ρ plane in which light trajectories are possible. The boundary curve separating allowed and forbidden regions is obtained from Eq. (5.3) with $\psi(\rho) = 0$; it is given in explicit form as

$$\lambda = (1 - \rho)^{-1}\{-\alpha \pm \rho[(\rho - \rho_+)(\rho - \rho_-)]^{\frac{1}{2}}\}. \quad (5.5)$$

Not surprisingly, the shape of this curve is qualitatively different in the two cases $\alpha < \frac{1}{2}$ (where horizons exist) and $\alpha > \frac{1}{2}$ (where they do not). In the latter case pericenters occur in the region $\rho > \rho_1$ for λ positive [with ρ_1 given by Eq. (5.4b)], and over the entire region $\rho > 0$ for λ negative.

Timelike Geodesics

A very similar analysis can be carried through for the timelike geodesics; it is complicated, however, by the presence of an additional energy parameter, $p_t = -\gamma$ or equivalently $\Gamma = \gamma^2 - 1$, which governs the type of motion that results. In the Schwarzschild case, it may be recalled, the minimum possible pericenter for a particle trajectory changes with the effective total energy Γ in the fashion shown by Fig. 9. There are no pericenters for $\rho < \frac{3}{2}$. To reach a point between $\frac{3}{2} < \rho < 2$ and be deflected back out again the particle must come from infinity with

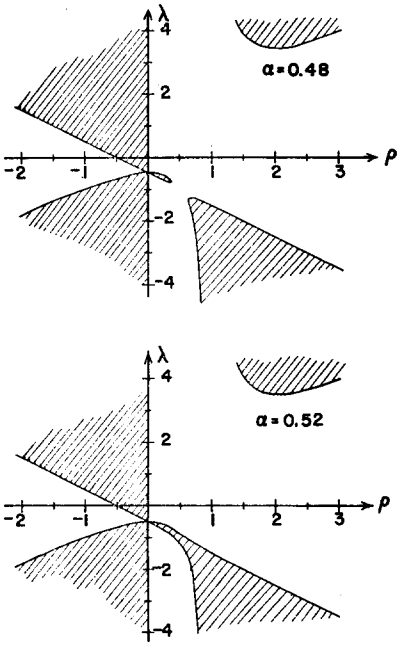

FIG. 8. A diagram of the λ–ρ plane, showing the regions in which light trajectories can exist. Shaded regions are forbidden; their boundaries define the turning points for null rays. Two cases, $\alpha = 0.48$ and $\alpha = 0.52$, are shown; the former contains horizons, at $\rho_- = 0.36$ and $\rho_+ = 0.64$. Note the small forbidden region extending from $\rho = 0$ to $\rho = \rho_-$ when λ is negative in the former case. When $\alpha \geq \frac{1}{2}$ the two forbidden regions for $\lambda < 0$ coalesce.

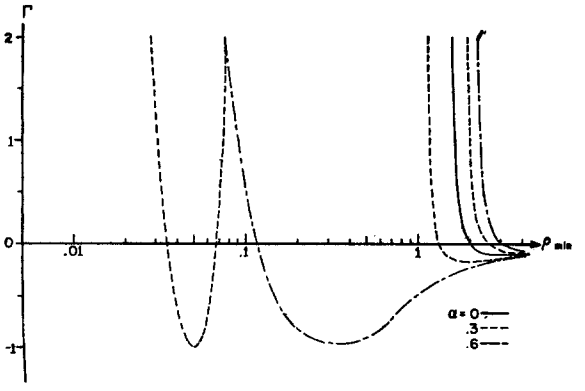

FIG. 9. Minimum pericenter for timelike geodesics as a function of effective total energy Γ, shown for $\alpha = 0, .3, .6$. The graph for $\alpha = 0.3$ has three components, corresponding to $\lambda < 0$, $\rho < \rho_-$, $\lambda < 0$, and $\rho > \rho_+$, $\lambda > 0$; that for $\alpha = 0.6$ has two components, corresponding to $\lambda < 0$ and $\lambda > 0$. Note that the $\rho < \rho_-$ portion (when it exists) covers the entire available energy range $\Gamma \geq -1$. Stable bound orbits in the exterior region occur for $\rho > \rho_{crit}$, where ρ_{crit} is the value of ρ at which the graph has zero slope—see also Fig. 11.

positive energy. At $\rho = 2$ the required total energy is zero, and it decreases to a minimum value of $\Gamma = -\frac{1}{8}$ at $\rho = 3$ (or $r = 6\,m$). These features are reflected in the stability of circular orbits: they are stable if and only if ρ exceeds 3. We therefore recognize in the Schwarzschild problem three characteristic radii—in addition to the famous "singularity" at $\rho = 1$—given by $\rho = \frac{3}{2}$, 2, and 3. Let us try to determine the corresponding characteristic radii in the Kerr metric.

First of all, the minimum pericenters analogous to $\rho = \frac{3}{2}$ are given once more by Eqs. (5.4a) or (5.4b). But the striking feature of the null case persists here as well: at *all* energies (i.e., for all $\Gamma \geq -1$) pericenters are found within the region $0 < \rho \leq \rho_-$, so that these minima are not absolute ones if α differs from zero.

To see how the character of the motion changes as Γ is varied, a diagram of the λ–ρ plane, showing allowed and forbidden regions for various values of Γ, may again be helpful (see Fig. 10). The turning points, which separate these two regions, are still given by the zeros of a cubic polynomial:

$$\bar{\psi}(\rho) \equiv \Gamma\rho^3 + \rho^2 + (\Gamma\alpha^2 - \lambda^2)\rho + (\lambda + \gamma\alpha)^2. \quad (5.6)$$

Solved explicitly for λ, this gives

$$\lambda = (1 - \rho)^{-1}\{-\gamma\alpha \pm [\rho(\rho - \rho_+)(\rho - \rho_-)(\Gamma\rho + 1)]^{\frac{1}{2}}\}, \quad (5.7)$$

analogous to Eq. (5.5). By studying the double zeros of $\bar{\psi}(\rho)$ one constructs the plot of Γ vs minimum pericenter shown in Fig. 9. The critical energy corresponding to Darwin's value $\Gamma_{crit} = -\frac{1}{8}$ splits into

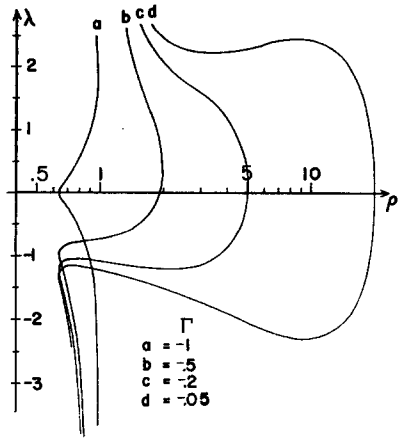

FIG. 10. A portion of the λ–ρ plane (cf. Fig. 8), showing regions in which timelike trajectories can exist. Graphs were plotted for $\alpha = 0.48$ and for several values of $\Gamma < 0$. In each case the allowed region lies to the left of the curve. An additional forbidden region, not shown on the graph, extends from $\rho = 0$ to $\rho = \rho_-$, much like the one shown in Fig. 8. Allowed regions also exist on the negative sheet for $\Gamma > 0$. Graphs (a) and (b) illustrate the case of no bound orbits; (c) has bound orbits for $\lambda < 0$, (d) for $\lambda > 0$ as well.

two values, $\Gamma^{\pm}_{crit}$ say. As Fig. 10 makes clear, stable bound orbits in the exterior region are possible if $\Gamma^{-}_{crit} < \Gamma < 0$ and λ is negative; if $\Gamma^{+}_{crit} < \Gamma < 0$, they are possible for positive λ as well.

It is easy to see that Γ attains these critical values whenever the cubic polynomial $\bar{\psi}(\rho)$ has a triple zero. Consequently,

$$\Gamma_{crit} = -1/(3\rho_{crit}) \tag{5.8a}$$

with ρ_{crit} in turn a real root of the following quartic:

$$\rho^4 - 6\rho^3 + (9 - 6\alpha^2)\rho^2 - 14\alpha^2\rho + 9\alpha^4 = 0. \tag{5.8b}$$

There are in general two such roots, with corresponding critical energies $\Gamma^{\pm}_{crit}$, shown in Fig. 11.

Exact Solutions; The Deflection of Light

The above work is based almost exclusively on the energy integral, Eq. (4.5). This does not differ substantially from the corresponding formula in the Schwarzschild problem—at least, not as long as one restricts his attention to the equatorial plane—so that the solution for $r(\mu)$ should be basically the same. Indeed, on making the standard transformation $u = 1/r$ one gets

$$\dot{u}^2 = u^4 B(u) \tag{5.9a}$$

with $B(u)$ a cubic polynomial:

$$B(u) = \Gamma + \epsilon\beta u + (\Gamma a^2 - l^2)u^2 + \beta(l + \gamma a)^2 u^3. \tag{5.9b}$$

Hence $u(\mu)$ can be expressed in terms of elliptic functions. The other first integrals, for $\dot{\varphi}$ and $\dot{t}$, do not yield so easily; however, the orbit equation is again relatively simple. One finds[20]

$$\frac{d\varphi}{du} = \frac{a}{D(u)} \pm \frac{A(u)}{D(u)[B(u)]^{\frac{1}{2}}} \tag{5.10}$$

with $B(u)$ defined as above, and

$$A(u) = l - \beta(l + \gamma a)u,$$

$$D(u) = 1 - \beta u + a^2u^2 \equiv a^2(u_+ - u)(u_- - u).$$

The sign of the second term in Eq. (5.10) is to be chosen to agree with that of $\dot{u}$. With $a = 0$ this reduces to $d\varphi/du = \pm l[B(u)]^{-\frac{1}{2}}$, whose solution in terms of Jacobi functions is immediate. The general case requires elliptic integrals of the third kind, which are clumsier to deal with and not very illuminating; for this reason we avoided detailed discussions of exact solutions in the previous sections. For purposes of illustration, however, we think it is instructive and not too painful to work out one example in detail, and we accordingly derive here an exact formula for the deflection of a beam of light confined to the equatorial plane in the Kerr field.

The total deflection $\Delta\varphi$ is obtained from Eq. (5.10) on integrating the right-hand side over a contour extending from $u = 0$ to the branch point $u = 1/d$ at pericenter and back to $u = 0$ again. The first term, being analytic, does not contribute

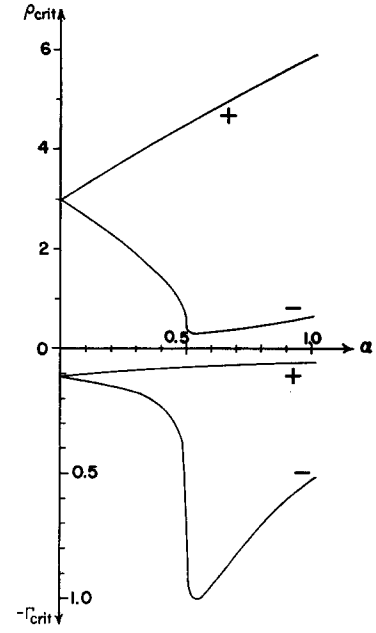

FIG. 11. Values of critical radii and corresponding critical energies, calculated from Eq. (5.8) for $0 \leq \alpha \leq 1$. Curves labeled (+) have $\lambda > 0$; those labeled (−) have $\lambda < 0$. ρ_{crit} gives the minimum radius—in units of $2m$—for which stable bound orbits are possible (excluding those orbits which penetrate the inner horizon and bounce back onto another sheet of the manifold).

to the total path integral, so one is left with

$$\Delta\varphi = 2\int_0^{1/d} \frac{A(u)}{D(u)} \frac{du}{[B(u)]^{\frac{1}{2}}}. \tag{5.11}$$

To bring this to standard form, split the rational portion of the integrand into partial fractions, as

$$\frac{A(u)}{D(u)} = \frac{A_+}{u_+ - u} + \frac{A_-}{u_- - u},$$

and write the cubic $B(u)$ as

$$\beta(l + a)^2(s_1 - u)(s_2 - u)(u - s_3)$$

with $s_1 \geq s_2 \geq 0 \geq s_3$. By definition, $s_2 = 1/d$; the remaining two zeros can be expressed in terms of l, a and d as follows:

$$s_1 = (Q + P - \beta)/2\beta d, \quad s_3 = -(Q - P + \beta)/2\beta d, \tag{5.12}$$

where

$$P = d(l - a)/(l + a), \qquad Q^2 = (P - 2\beta)(P + 3\beta)$$

have been chosen to agree as closely as possible with Darwin's notation.

Introducing the further parameters

$$k^2 = \frac{s_2 - s_3}{s_1 - s_3}, \quad \alpha_\pm^2 = \frac{s_2 - s_3}{u_\pm - s_3}, \quad \sin^2\psi_0 = \frac{-s_3}{s_2 - s_3}$$

one obtains after some manipulation

$$\begin{aligned}\Delta\varphi = {} & 4(l + a)^{-1}(d/Q)^{\frac{1}{2}}\{A_+(u_+ - s_3)^{-1} \\ & \times [\Pi(\alpha_+^2, k) - \Pi(\psi_0, \alpha_+^2, k)] + A_-(u_- - s_3)^{-1} \\ & \times [\Pi(\alpha_-^2, k) - \Pi(\psi_0, \alpha_-^2, k)]\},\end{aligned} \tag{5.13}$$

where $\Pi(\alpha^2, k)$, $\Pi(\psi_0, \alpha^2, k)$ are, respectively, the complete and incomplete elliptic integrals of the third kind. It can be shown that the above result reduces to $\Delta\varphi = \pi$ when $\beta = 0$. This is of course to be expected, since the Kerr metric becomes flat for $\beta = 0$, and serves as a check on the intermediate calculations. In the limit $a \to 0$ one finds $A_+ = 0$, $A_-/(u_- - s_3) = l$, and $\alpha_- = 0$, so

$$\begin{aligned}\Delta\varphi \underset{a\to 0}{\to} {} & 4(P/Q)^{\frac{1}{2}}[K(k) - F(\psi_0, k)] \\ & = 4(P/Q)^{\frac{1}{2}}F(\psi_1, k),\end{aligned} \tag{5.14}$$

with $\cot\psi_1 = (1 - k^2)^{\frac{1}{2}}\tan\psi_0$, in agreement with Darwin's result.

From Eq. (5.13) one readily deduces the corrections to the familiar deflection formula due to rotation of the central body. We assume β/d and a/d small, and keep terms to order $\beta a/d^2$. Setting

$$\Delta\varphi = \pi + \delta,$$

where δ is the deflection angle as usually defined, we find

$$\delta \approx \frac{2\beta}{d} + \frac{2\beta a}{d^2} = \frac{4m}{d}\left(1 + \frac{a}{d}\right). \tag{5.15}$$

This result has also been obtained by Skrotskii[30] using the weak-field metric of Landau–Lifshitz, and from a more general viewpoint by Plebański.[31] The value for the correction term a/d is difficult to estimate in the case of the sun, since its angular momentum is not well known. Assuming essentially uniform rotation throughout its interior, at the rate 14.3°/day observed for sun spots near the equator, one computes $a \simeq 1.9$ km and consequently $a/d \simeq 3 \times 10^{-6}$ for a light ray grazing the sun's disk, which is undetectable by several orders of magnitude. Measurements of the solar oblateness are not inconsistent with a value of a/d several times as large as this[32]; however, this would require a mass quadrupole moment $Q_\odot$ of order $10^{-4}\ M_\odot R_\odot^2$, which in turn would contribute a correction term to δ of about the same magnitude.[33]

Note added in proof: It has been pointed out by C. V. Vishveshwara (University of Maryland Tech. Rept. No. 589) that if one considers "stationary" sources and observers (i.e., those whose world lines are the Killing trajectories $\partial/\partial t$), then the surface of infinite red shift occurs at $g_{tt} = 0$, which does not coincide with the null horizon in the Kerr case (unless $a = 0$).

ACKNOWLEDGMENTS

We are much indebted to Professor A. Schild and members of the Center for Relativity Theory of the University of Texas for their hospitality and many stimulating conversations. We are also grateful to Brandon Carter for many helpful discussions, and for placing several unpublished results at our disposal.

Added in proof: I acknowledge above all my debt to a dear friend, collaborator and co-author, whose penetrating insight and ingenuity were responsible for bringing this work to a successful conclusion. To the extent that this paper conveys these traits, it bears the imprint of his thought, and stands here as a memorial to his accomplishment.—R.W.L.

[30] G. V. Skrotskii, Dokl. Akad. Nauk SSR 114, 73 (1957) [English transl.: Soviet Phys.—Doklady 2, 226 (1957)].

[31] J. Plebański, Phys. Rev. 118, 1396 (1960).

[32] R. H. Dicke, Nature 202, 432 (1964), has suggested that the interior of the sun might have a rotational period as small as 25 h, without leading to an unreasonably large visual oblateness or violating stellar structure theory. This would increase a/d to 6×10^{-5}, still too small to be significant.

[33] The quadrupole field adds to Eq. (5.15) a term $\frac{1}{2}(4m/d)(k/d)^2$, where $Q = mk^2$ is the quadrupole moment. Taking a gravitational oblateness of 5×10^{-5} for the sun, as suggested by Dicke,[32] we calculate $(k/d)^2_\odot = 7 \times 10^{-5}$.

Difficulties in the Kinetic Theory of Dense Gases*

J. R. DORFMAN
University of Maryland, College Park, Maryland
AND
E. G. D. COHEN
The Rockefeller University, New York, New York
(Received 26 April 1966)

For the determination of the transport coefficients of a dense gas, the long-time behavior of the pair distribution function F_2 for small intermolecular distances is obtained from a density expansion in terms of the first distribution function F_1. On the basis of the dynamics of small groups of particles, it is shown that this expansion contains divergences so that it cannot be used for (a) the computation of the long-time behavior of F_2 beyond $O(n)$; (b) the demonstration of the decay of the initial state beyond $O(n^2)$. Similar divergences are encountered in the computation of the transport coefficients from time-correlation functions. The nature of the divergences suggests (a) there is no kinetic stage in the approach of a dense gas to equilibrium, in the sense of Bogoliubov; (b) a weak logarithmic density dependence of the transport coefficients.

1. INTRODUCTION

SINCE Bogoliubov's[1] work in 1946 on the approach to equilibrium of a dense gas with central short-range intermolecular forces, his results have been reproduced by a great number of authors with a great variety of methods.[2-6]

The basic idea was that due to the existence of three widely spaced basic relation times: the duration of a collision $t_{coll}(\sim 10^{-12}$ sec), the mean free time $t_{mfp}(\sim 10^{-9}$ sec), and a macroscopic time t_{macr} $(\sim 10^{-5}$ sec); a dense gas not in equilibrium would, for a large class of initial conditions, approach to equilibrium in three well-separated stages. In each consecutive stage the system would be adequately characterized by fewer variables than in the preceding stage, viz., by those variables that vary slowly during the preceding stage.[1,7,8] This then leads to an approach of the gas to equilibrium first (after a time $t \gg t_{coll}$) through a kinetic stage, where the state of the gas is characterized by the first distribution function F_1 alone and then (after a time $t \gg t_{mfp}$) through a hydrodynamic stage, where the state of the gas is characterized by the five hydrodynamical quantities alone, viz., the local density $n(\mathbf{r}, t)$, the local velocity $\mathbf{u}(\mathbf{r}, t)$, and the local temperature $T(\mathbf{r}, t)$.

For a description of a gas with additive intermolecular forces, the first and second distribution functions F_1 and F_2, respectively, are usually sufficient. In particular, for the computation of the transport coefficients in the hydrodynamical stage, the long-time behavior of F_1 and F_2, i.e., their behavior for times $t \gg t_{mfp} \gg t_{coll}$, is needed. The crucial assertion is then that, for such a gas in the kinetic state, i.e., after a time $t \gg t_{coll}$, F_2 would be a functional $F_2(|F_1)$ of F_1 as far as its time dependence is concerned. This then leads to a *kinetic equation* for F_1 of the form

$$\partial F_1/\partial t = A(|F_1). \tag{1.1}$$

Here A is a time-independent operator acting on F_1, which is obtained in the form of a density expansion. This equation then leads in the hydrodynamic stage to hydrodynamic equations, in which the transport coefficients are obtained in the form of expansions in powers of the (local) density n. In particular, one finds the Navier–Stokes equations, with virial expansions for the viscosity η and the heat-conductivity λ:

$$\eta(n, T) = \eta_0(T) + n\eta_1(T) + n^2\eta_2(T) + \cdots, \tag{1.2}$$

$$\lambda(n, T) = \lambda_0(T) + n\lambda_1(T) + n^2\lambda_2(T) + \cdots, \tag{1.3}$$

where η_0 and λ_0, obtained by Chapman and Enskog, and η_1 and λ_1, first obtained by Choh and Uhlenbeck,[7] are functions of the (local) temperature T.

* This paper is dedicated to Professor G. E. Uhlenbeck for his 65th birthday.

[1] N. N. Bogoliubov, *Studies in Statistical Mechanics I*, J. de Boer and G. E. Uhlenbeck, Eds. (North-Holland Publishing Company, Amsterdam, 1962), p. 5.

[2] M. S. Green, J. Chem. Phys. 25, 836 (1956); Physica 24, 393 (1958).

[3] E. G. D. Cohen, Physica 28, 1025, 1045, 1060 (1962); J. Math. Phys. 4, 143 (1963).

[4] M. S. Green and R. A. Piccirelli, Phys. Rev. 132, 1388 (1963).

[5] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).

[6] H. B. Hollinger and C. F. Curtiss, J. Chem. Phys. 33, 1386 (1960).

[7] S. T. Choh and G. E. Uhlenbeck, "The Kinetic Theory of Dense Gases," University of Michigan Report (1958); E. G. D. Cohen, *Fundamental Problems in Statistical Mechanics* (North-Holland Publishing Company, Amsterdam, 1961), p. 110.

[8] G. E. Uhlenbeck, and G. W. Ford, *Lectures in Statistical Mechanics*, (American Mathematical Society, Providence, Rhode Island, 1963).

It should be stressed that the virial expansions (1.2) and (1.3) which are obtained for the transport coefficients are directly related to the existence of a kinetic equation (1.1), with a density expansion for A. The kinetic equation was obtained by Bogoliubov by seeking special solutions of the BBGKY hierarchy.

This method has two disadvantages: (1) it is difficult in this way to characterize the general term; (2) no estimate of the error is possible. Both these disadvantages are circumvented by using cluster expansions,[2-4] which can be considered as generalizations to non-equilibrium states of the well-known cluster expansions for a dense gas in equilibrium.[9] These cluster expansions enable one to express the probability densities in the many-particle system in terms of those for small groups of particles. In fact, in this way F_2 in an infinite system is expressed identically in terms of F_1 by a power series in the density, where consecutive terms, apart from a dependence on the initial state, depend on the dynamics of isolated groups of an increasing number of particles in infinite space, through time-dependent streaming operators [cf., Eq. (2.14)]. What one would have to show now to prove the existence of a kinetic stage, as proposed by Bogoliubov, is that for a large class of initial conditions after a time $t \gg t_{\text{coll}}(1)$ the influence of the initial state on the time development of F_2 can be neglected, and (2) the streaming operators have attained their asymptotic values in time, so that the whole time dependence of F_2 is then through F_1 *only*. This paper purports to show, however, that this is in general *not* so. In fact, it can only be shown to be true for the first two terms in the density expansion of F_2 in terms F_1. In general, due to the occurrence of divergences, one can neither show that the initial state is forgotten, nor that the whole time dependence of F_2 is through F_1.

Similar divergences are encountered, if one attempts to determine the long-time behavior of time-correlation functions for a dense gas in equilibrium. In particular, it affects a computation of the transport coefficients for a dense gas from equilibrium time-correlation functions.[10]

In Sec. 2 we derive the cluster expansion for F_2 in terms of F_1 using streaming operators. For a study of the existence of a kinetic stage, we rewrite this expansion for F_2 identically as the sum of two expansions, one of which only involves the dynamics of isolated groups of particles, while the other one involves in addition the initial state of the gas. This can be considered as a formal solution to a *statistical problem*, which relates a many-particle property to that of small groups of particles. In Sec. 3 we study a *dynamical problem*, viz., the dynamics of isolated groups of three and four particles insofar as this is relevant for a discussion of the various terms in the cluster expansion of F_2. In Sec. 4 we estimate, on the basis of the results obtained in Sec. 3, the various terms in the two cluster expansions mentioned above. This constitutes an approximate evaluation of the cluster integrals and therefore an approximate solution to an *integral problem*.[11]

It suffices for our purpose (i.e., the calculation of the transport coefficients) to consider F_2 only for interparticle distances within the range of the intermolecular forces. We then find these expansions cannot be used to study the long-time behavior of F_2.[12] In Sec. 5 we discuss some of the consequences of the results obtained in Sec. 4.

2. STATISTICAL PROBLEM: CLUSTER EXPANSIONS OF THE FIRST AND SECOND DISTRIBUTION FUNCTION; BOGOLIUBOV'S RESULTS

The long-time behavior of the first and second distribution function in a dense gas not in equilibrium can be discussed with the help of cluster expansions which can be considered as generalizations of those employed for the computation of the first and second distribution functions in a dense gas in equilibrium. As this has been discussed extensively before, we only repeat the main results.[3] We first consider a gas of N particles in a volume V. One introduces a set of D functions, which obey the Liouville equation and refer to probability densities of isolated groups of 1, 2, $\cdots$, s, $\cdots$, N particles in the volume V:

$$\frac{\partial D_s(x_1 \cdots x_s; t)}{\partial t} = \{H_s(x_1 \cdots x_s),\, D_s(x_1 \cdots x_s, t)\} = -\mathcal{H}_s(x_1 \cdots x_s)\, D_s(x_1 \cdots x_s; t). \qquad (2.1)$$

[9] G. E. Uhlenbeck and G. W. Ford, in *Studies in Statistical Mechanics. I* (North-Holland Publishing Company, Amsterdam, 1962), p. 123.

[10] M. H. Ernst, J. R. Dorfman, and E. G. D. Cohen, Phys. Letters **12**, 314 (1964); Physica **31**, 493 (1965): M. H. Ernst, thesis, University of Amsterdam (1965).

[11] For a dense gas in equilibrium only a statistical and an integral problem, but not a dynamical problem, have to be solved, to obtain the pair distribution function in equilibrium (cf., Ref. 9).

[12] J. R. Dorfman and E. G. D. Cohen, Phys. Letters **16**, 124 (1965). Similar conclusions have been obtained by J. Weinstock, Phys. Rev. **140** A 460 (1965), and by E. A. Frieman and R. Goldman, Bull. Am. Phys. Soc. **10**, 531 (1965). See also K. Kawasaki and I. Oppenheim, Phys. Rev. **139**, A 1763 (1965).

Here

$$H_s(x_1 \cdots x_s) = \sum_{i=1}^{s} \frac{\mathbf{p}_i^2}{2m} + \sum_{\substack{i<j \\ 1}}^{s} \phi(r_{ij}) \tag{2.2}$$

is the Hamilton function of s particles, where $x_i = \mathbf{q}_i, \mathbf{p}_i$, and m is the mass of the particles and $\phi(r_{ij})$ the central repulsive and additive intermolecular potential between molecules i and j, which has a finite range r_0;

$$\mathfrak{IC}_s(x_1 \cdots x_s) = \sum_{i=1}^{s} \frac{\mathbf{p}_i}{m} \cdot \frac{\partial}{\partial \mathbf{q}_i} - \sum_{\substack{i<j \\ 1}}^{s} \theta_{ij}, \tag{2.3}$$

where

$$\theta_{ij} = \frac{\partial \phi(r_{ij})}{\partial \mathbf{q}_i} \cdot \frac{\partial}{\partial \mathbf{p}_i} + \frac{\partial \phi(r_{ij})}{\partial \mathbf{q}_j} \cdot \frac{\partial}{\partial \mathbf{p}_j}. \tag{2.4}$$

The solution of (2.1) can be written in the form

$$D_s(x_1 \cdots x_s; t) = S_{-t}^{(s)}(x_1 \cdots x_s)\, D_s(x_1 \cdots x_s; 0), \tag{2.5}$$

where

$$S_{-t}^{(s)}(x_1 \cdots x_s) = \exp\left[-t\mathfrak{IC}_s(x_1 \cdots x_s)\right] \tag{2.6}$$

is the streaming operator associated with s particles. We normalize the D functions according to[13]:

$$\int D_s(x_1 \cdots x_s; t)\, dx_1 \cdots dx_s = V^s. \tag{2.7}$$

The distribution functions for 1, 2, $\cdots$, s, $\cdots$ particles in an infinite system ($N, V \to \infty, N/V = n$) can be defined by

$$F_s(x_1 \cdots x_s; t) = \lim_{\substack{N,V\to\infty \\ N/V=n}} V^s \int \cdots \int dx_{s+1} \cdots dx_N \frac{D_N(x_1 \cdots x_N; t)}{V^N} \tag{2.8}$$

and can be expanded in terms of D_1 functions.

One obtains the following expansions for F_1 and F_2:

$$nF_1(x_1; t) = n\mathfrak{U}_1(x_1; t)\, D_1(x_1; t) + n^2 \int dx_2\, \mathfrak{U}_2(x_1x_2; t)\, D_1(x_1; t)\, D_1(x_2; t) + \frac{n^3}{2!} \int dx_2 \int dx_3\, \mathfrak{U}_3(x_1x_2x_3; t) \prod_{i=1}^{3} D_1(x_i; t) + \cdots, \tag{2.9}$$

$$n^2F_2(x_1x_2; t) - n^2F_1(x_1; t)F_1(x_2; t) = n^2\mathfrak{U}_2(x_1x_2; t)\, D_1(x_1; t)\, D_1(x_2; t) + n^3 \int dx_3\, \mathfrak{U}_3(x_1x_2x_3; t) \prod_{i=1}^{3} D_1(x_i; t) + \cdots. \tag{2.10}$$

Here the $\mathfrak{U}$ operators are defined in terms of the S operators by

$$\mathfrak{U}_1(x_1; t) = 1, \tag{2.11a}$$

$$\mathfrak{U}_2(x_1x_2; t) = \mathfrak{S}_t^{(2)}(x_1x_2) - 1, \tag{2.11b}$$

$$\mathfrak{U}_3(x_1x_2x_3; t) = \mathfrak{S}_t^{(3)}(x_1x_2x_3) - \mathfrak{S}_t^{(2)}(x_1x_2) - \mathfrak{S}_t^{(2)}(x_1x_3) - \mathfrak{S}_t^{(2)}(x_2x_3) + 2, \tag{2.11c}$$

with

$$\mathfrak{S}_t^{(s)}(x_1 \cdots x_s) = S_{-t}^{(s)}(x_1 \cdots x_s) \times a_s(x_1 \cdots x_s; 0) \prod_{i=1}^{s} S_t^{(1)}(x_i), \tag{2.12}$$

where the $a_s(x_1 \cdots x_s; 0)$ characterize the correlations of the initial state at $t = 0$:

$$D_s(x_1 \cdots x_s; 0) = a_s(x_1 \cdots x_s; 0) \prod_{i=1}^{s} D_1(x_i; 0). \tag{2.13}$$

The expansions (2.9) and (2.10) for F_1 and F_2, respectively, are identical to expansions given before[3] except that in (2.9) and (2.10) the initial state correlations have been taken into account through the $a_s(x_1 \cdots x_s; 0)$.

If there are no correlations between the s molecules $1 \cdots s$ at $t = 0$, then the $a_s(x_1 \cdots x_s; 0) = 1$. We restrict ourselves to initial states, with correlations over distances of the order of the range r_0 of intermolecular forces, i.e., the $a_s(x_1 \cdots x_s; 0)$ ($s = 2, 3, 4, \cdots$) are only unequal to 1 if two or more of the molecules are within a distance of the order of r_0.[14]

The various terms in the expansions (2.9) and (2.10) can be represented by the same (connected) diagrams as the corresponding terms in the activity expansions of the density n and pair distribution function in equilibrium, respectively. The expansions (2.9) and (2.10) cannot be used to obtain the long-time behavior of F_1 and F_2; they can only be used for times $t \ll t_{\mathrm{mfp}}$ due to the occurrence of secular terms, i.e., of terms which grow with time. Thus, for example, the secular behavior of the second

[13] We are indebted to Professor T. Y. Wu for pointing out to us that the normalization previously used (cf. Ref. 3) is not correct in that it leads to F and D functions of different dimensions. The present normalization does not suffer from this objection.

[14] If one would assume a correlation length $l_c > r_0$ at $t = 0$, one would introduce a new length l_c into the problem and one could not expect the existence of a kinetic stage after $t >> t_{\mathrm{coll}}$ for such initial conditions.

term on the right-hand side of (2.9) follows directly from the fact that the integrand of this term is different from zero for all phases of particle 2 at time t such that a binary collision between the particles 1 and 2 occurs in the time interval between 0 and t, due to occurrence of the operator $\tilde{S}_t^{(2)}(x_1x_2)$.[15] This happens whenever particle 2 is within the collision cylinder with respect to particle 1. As the volume of this cylinder grows proportional to t, this implies that also the volume in the phase space of particle 2, which contributes to the term mentioned, grows $\sim t$, so that the term itself in general grows $\sim t$ [cf., Fig. 1(a)]. Similarly, contributions of two successive binary collisions between the particles 1, 2, and 3 in the time interval from 0 to t grow $\sim t^2$, as the volume in the combined phase space of particles 2 and 3 associated with these events grows $\sim t^2$ [cf., Fig. 1(b)].

In general, the coefficients of n^l in the expansions (2.9) and (2.10) contain contributions which grow with powers of t. The leading contribution to the term n^l in nF_1 is $\sim(t/t_{\text{coll}})^{l-1}$, while in n^2F_2 it is $\sim(t/t_{\text{coll}})^{l-2}$, respectively. The expansions can be used up to times t for which $(nr_0^3)(t/t_{\text{coll}}) \ll 1$ or for $t \ll t_{\text{mfp}}$ (cf., Ref.1).[16]

In order to obtain improved expansions for F_1 and F_2 which can be used for times $t \gg t_{\text{mfp}}$, one eliminates the secular contributions mentioned above from the expansions (2.9) and (2.10). This has been achieved in two ways. One way is to use dynamical and combinatorial arguments to classify the secular terms in the D_1 expansions of F_1 and F_2. One then shows that certain terms, which give rise to the secular behavior in the D_1 expansion of F_2, can be grouped together, using the D_1 expansion of F_1. In this way one can obtain an expansion of F_2 in terms of F_1. This method has been followed by Green and Piccirelli.[4] Another way, and the one we use here, is to eliminate the D_1 between the expansions (2.9) and (2.10). This then also leads to an expansion of F_2 in terms of F_1 instead of in terms of D_1.

In both cases one subsequently uses the first hierarchy equation between F_1 and F_2 to obtain a second relation between F_1 and F_2. As has been discussed before,[3] the elimination of the D_1 between the expansions (2.9) and (2.10) of nF_1 and n^2F_2 can be performed in a completely analogous way as the elimination of the activity between the activity expansions of the density n and the pair distribution function in a dense gas in equilibrium.

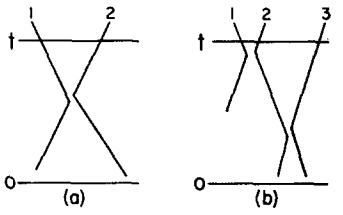

FIG. 1. Secular contributions to F_1 and F_2.

Then, solving nD_1 in terms of nF_1 from (2.9) and substituting this into (2.10) yields the following expansion of n^2F_2 in terms of nF_1:

$$n^2F_2(x_1x_2;\,t) = n^2\tilde{S}_t^{(2)}(x_1x_2)F_1(x_1;\,t)F_1(x_2;\,t) + n^3\int dx_3\,\tilde{\mathfrak{I}}_t^{(3)}(x_1x_2 \mid x_3)\prod_{i=1}^{3}F_1(x_i;\,t) + \cdots, \tag{2.14}$$

where

$$\tilde{\mathfrak{I}}_t^{(3)}(x_1x_2 \mid x_3) = \tilde{S}_t^{(3)}(x_1x_2x_3) - \tilde{S}_t^{(2)}(x_1x_2)\tilde{S}_t^{(2)}(x_1x_3) - \tilde{S}_t^{(2)}(x_1x_2)\tilde{S}_t^{(2)}(x_2x_3) + \tilde{S}_t^{(2)}(x_1x_2). \tag{2.15}$$

The various terms in (2.14) can be classified by the same diagrams as the corresponding terms in the virial expansion of the pair distribution function in equilibrium. $\tilde{\mathfrak{I}}_t^{(4)}(x_1x_2 \mid x_3x_4)$ can be obtained from elsewhere.[3,4] For further investigation of the expansion (2.14) it is convenient to write (2.14) identically as the sum of two contributions, one of which contains the influence of the correlations present in the initial state. Thereto we introduce to the $a_s(x_1 \cdots x_s;\,0)$ cluster functions $u_l(x_1 \cdots x_l;\,0)$ which characterize the correlations at $t = 0$:

$$u_1(x_1;\,0) = a_1(x_1;\,0) = 1, \tag{2.16a}$$

$$u_2(x_1x_2;\,0) = a_2(x_1x_2;\,0) - 1, \tag{2.16b}$$

$$u_3(x_1x_2x_3;\,0) = a_3(x_1x_2x_3;\,0) - a_2(x_1x_2;\,0) - a_2(x_1x_3;\,0) - a_2(x_2x_3;\,0) + 2, \tag{2.16c}$$

etc.

With the help of the u functions we rewrite the expansion (2.14) as

$$n^2F_2(x_1x_2;\,t) = n^2F_2(x_1x_2;\,t \mid F_1) + n^2C_i(x_1x_2;\,t), \tag{2.17}$$

where

$$n^2F_2(x_1x_2;\,t \mid F_1) = n^2F_2^{(0)}(x_1x_2;\,t \mid F_1) + n^3F_2^{(1)}(x_1x_2;\,t \mid F_1) + \cdots, \tag{2.18}$$

[15] Strictly speaking, a dynamical as well as an integral problem already enter here. As their discussion is elementary we postpone a detailed treatment to the next sections.

[16] It should be noted that the coefficients of n^l in the expansions (2.9) and (2.10) are *not* power series in t, due to the occurrence of terms which depend logarithmically on t (see Sec. 4).

with

$$F_2^{(0)}(x_1x_2;\, t \mid F_1) = \mathcal{S}_t^{(2)}(x_1x_2)F_1(x_1;\, t)F_1(x_2;\, t), \tag{2.18a}$$

$$F_2^{(1)}(x_1x_2;\, t \mid F_1) = \int dx_3\, \mathfrak{I}_t^{(2)}(x_1x_2 \mid x_3) \prod_{i=1}^{3} F_1(x_i;\, t), \tag{2.18b}$$

etc., while

$$n^2C_i(x_1x_2;\, t) = n^2C_i^{(0)}(x_1x_2;\, t) + n^3C_i^{(1)}(x_1x_2;\, t) + \cdots, \tag{2.19}$$

with

$$C_i^{(0)}(x_1x_2;\, t) = S_{-t}^{(2)}(x_1x_2)u_2(x_1x_2;\, 0) \prod_{i=1}^{2} S_t^{(1)}(x_i)F_1(x_i;\, t), \tag{2.19a}$$

$$\begin{aligned} C_i^{(1)}(x_1x_2;\, t) = \int dx_3 \Big(\Big\{ & S_{-t}^{(3)}(x_1x_2x_3)u_3(x_1x_2x_3;\, 0) + [S_{-t}^{(3)}(x_1x_2x_3) - S_{-t}^{(2)}(x_1x_2)S_t^{(1)}(x_1)S_{-t}^{(2)}(x_1x_3)]u_2(x_1x_3;\, 0) \\ & + [S_{-t}^{(3)}(x_1x_2x_3) - S_{-t}^{(2)}(x_1x_2)S_t^{(1)}(x_2)S_{-t}^{(2)}(x_2x_3)]u_2(x_2x_3;\, 0) \\ & + [S_{-t}^{(3)}(x_1x_2x_3) - S_{-t}^{(2)}(x_1x_2)S_{-t}^{(1)}(x_3)]u_2(x_1x_2;\, 0)\Big\} \prod_{i=1}^{3} S_t^{(1)}(x_i) \\ & - S_{-t}^{(2)}(x_1x_2)u_2(x_1x_2;\, 0)S_t^{(1)}(x_1)S_t^{(1)}(x_2)\Big[\tilde{\mathcal{S}}_t^{(2)}(x_1x_3) + \tilde{\mathcal{S}}_t^{(2)}(x_2x_3)\Big]\Big) \prod_{i=1}^{3} F_1(x_i;\, t). \end{aligned} \tag{2.19b}$$

Here

$$\mathcal{S}_t^{(s)}(x_1 \cdots x_s) = S_{-t}^{(s)}(x_1 \cdots x_s) \prod_{i=1}^{s} S_t^{(1)}(x_i), \tag{2.20a}$$

while

$$\begin{aligned} \mathfrak{I}_t^{(3)}(x_1x_2 \mid x_3) = {} & \mathcal{S}_t^{(3)}(x_1x_2x_3) - \mathcal{S}_t^{(2)}(x_1x_2)\mathcal{S}_t^{(2)}(x_1x_3) \\ & - \mathcal{S}_t^{(2)}(x_1x_2)\mathcal{S}_t^{(2)}(x_2x_3) + \mathcal{S}_t^{(2)}(x_1x_2). \end{aligned} \tag{2.20b}$$

The contributions of order n^l in C_i follow directly, with (2.16), from the term of order n^l in (2.14). In the expansions (2.17)–(2.19) no approximations have yet been made. In order now to obtain the results of Bogoliubov for the long-time behavior of F_2, one would have to show that, for the class of initial conditions mentioned above, for times $t \gg t_{\text{coll}}$:

(1) the initial condition term $C_i(x_1x_2;\, t)$ can be neglected;

(2) the $F_2(x_1x_2;\, t \mid F_1)$ can be replaced by its asymptotic value, as far as its time dependence through the S operators is concerned, so that its whole time dependence is through $F_1(x;\, t)$ only, viz.,

$$n^2F_2(x_1x_2;\, t \mid F_1) \Rightarrow n^2F_2(x_1x_2 \mid F_1),$$

where

$$\begin{aligned} & n^2F_2(x_1x_2 \mid F_1) \\ & \quad = n^2F_2^{(0)}(x_1x_2 \mid F_1) + n^3F_2^{(1)}(x_1x_2 \mid F_1) + \cdots \\ & \quad = n^2\mathcal{S}_\infty^{(2)}(x_1x_2)F_1(x_1;\, t)F_1(x_2;\, t) \\ & \qquad + n^3 \int dx_3\, \mathfrak{I}_\infty^{(3)}(x_1x_2 \mid x_3)F_1(x_1;\, t) \\ & \qquad \times F_1(x_2;\, t)F_1(x_3;\, t) + \cdots. \end{aligned} \tag{2.21}$$

This implies that one can replace for $t \gg t_{\text{coll}}$ all $\tilde{\mathcal{S}}_t$ operators in (2.14) by $\mathcal{S}_\infty$ operators, thereby tacitly assuming that the $F_2^{(l)}(x_1x_2;\; t \mid F_1)$ no longer contain secular terms.[17] As the *only* characteristic length in each order of n for both $F_2(t \mid F_1)$ and C_i is the range of the forces r_0, one would expect that the characteristic time for the $C_i^{(l)}$ to become zero and for $F_2(t \mid F_1)$ to approach the Bogoliubov functional $F_2(\;\mid F_1)$ would be t_{coll}.

It has been shown that if nF_1 is the *equilibrium* Maxwell–Boltzmann distribution function, the functional $F_2(x_1x_2 \mid F_1)$ given by (2.21) exists and reduces term by term to the known virial expansion of the pair distribution function of a gas in equilibrium.[8,18]

If the functional $F_2(x_1x_2 \mid F_1)$ of (2.21) would *also* exist for a large class of F_1 *outside* equilibrium, then a kinetic equation for F_1 as well as virial expansions for the transport coefficients are obtained. For, introducing $F_2(x_1x_2 \mid F_1)$ of (2.21) for $F_2(x_1x_2;\, t)$ into the right-hand side of the first hierarchy equation:

[17] Previously we wrote $F_2(x_1x_2;\, t|F_1)$ of (2.17) identically as the sum of two terms, one of which is the Bogoliubov functional $F_2(x_1x_2|F_1)$, while the other was called the asymptotic correction term C_2. (a) E. G. D. Cohen, in *Statistical Mechanics of Equilibrium and Non-Equilibrium*, J. Meixner, Ed. (North-Holland Publishing Company, Amsterdam, 1965), p. 140. (b) E. G. D. Cohen, "On the Statistical Mechanics of a Moderately Dense Gas not on Equilibrium," Boulder Lectures (1965) (c) J. R. Dorfman, in *Dynamics of Fluids and Plasmas* (Interscience Publishers, New York, to be published). The sum $C_i + C_2$ is with trivial modification identical to the error term of Green and Piccirelli.[4] The absence of secular terms in $F_2(x_1x_2;\, t|F_1)$ implies then that $C_2 \to 0$ as $t \to \infty$.

[18] R. A. Piccirelli, J. Math. Phys. 7, 922 (1966).

$$\frac{\partial F_1(x_1;\, t)}{\partial t} + \frac{\mathbf{p}_1}{m}\cdot\frac{\partial F_1}{\partial \mathbf{q}_1} = n\int dx_2\ \theta_{12}F_2(x_1x_2;\, t) \quad (2.22)$$

leads to a self-contained equation for F_1, in which the right-hand side is given in the form of a density expansion, viz.,

$$\frac{\partial F_1}{\partial t} + \frac{\mathbf{p}_1}{m}\cdot\frac{\partial F_1}{\partial \mathbf{q}_1} = A(|F_1), \quad (2.23)$$

where

$$A(|F_1) = A_1(x_1 \mid F_1) + nA_2(x_1 \mid F_1) + \cdots, \quad (2.24)$$

with

$$A_1(x_1 \mid F_1) = \int dx_2\ \theta_{12}F_2^{(0)}(x_1x_2 \mid F_1), \quad (2.24a)$$

$$A_2(x_1 \mid F_1) = \int dx_2\ \theta_{12}F_2^{(1)}(x_1x_2 \mid F_1), \quad (2.24b)$$

etc. As A is a time-independent operator on F_1, Eq. (2.24) is called *a kinetic equation*. As $A_1(x_1 \mid F_1)$ can be reduced to the usual Boltzmann collision operator, Eq. (2.23) can also be called a generalized Boltzmann equation. It has been shown[7] that this equation leads for times $t \gg t_{mfp}$ to the Navier–Stokes equations of hydrodynamics, in which, due to the density expansion (2.24) for $A(\ |\ F_1)$, the transport coefficients η and λ are obtained in the form of virial expansions.

The cluster expansions employed in this section can also be applied to obtain the long-time behavior of time-correlation functions in a dense gas in equilibrium. In particular, they can be used to obtain the transport coefficients in a dense gas from the time correlation function expressions for these coefficients.[10,17a] In Appendix A, we briefly indicate the close parallel between the cluster expansions given above for the computation of the long-time behavior of the distribution functions and those appropriate for the corresponding problem for time-correlation functions. As the discussions of the following sections for the distribution functions can be applied with only minor modifications also to the case of the time correlation functions, we mainly restrict the discussion to the distribution functions.

3. DYNAMICAL PROBLEM

The cluster expansions of the distribution functions in an infinite system discussed in the preceding section were all made with the help of streaming operators involving the dynamics of 2, 3, 4, $\cdots$ isolated particles in infinite space. Therefore the discussion of the various terms in these expansions and in particular the discussion of the long-time behavior of $F_2(x_1x_2;\, t \mid F_1)$ and the existence of the Bogoliubov functional $F_2(x_1x_2 \mid F_1)$ depends crucially on the properties of these operators.

The streaming operators transform the phases of a group of particles at a certain time t into those at another time. Thus $S_t^{(s)}(x_1 \cdots x_s)$ transforms the phases $x_1 \cdots x_s$ of the particles $1 \cdots s$ at time t into those a time t earlier, i.e., at $t = 0$, if the particles move in infinite space under their mutual interaction only.

Therefore one has to consider the dynamics of isolated groups of 2, 3, $\cdots$ particles in infinite space. We can restrict ourselves for the discussion of the transport coefficients to a computation of F_2 for times $t \gg t_{coll}$ and for configurations of the particles 1 and 2 at time t such that their distance $r_{12} \leq r_0$. As a result, in this paper we confine ourselves to a discussion of the dynamics of groups of 2, 3, 4, $\cdots$ particles, where the particles 1 and 2 are colliding at time t. The basic dynamical problem is therefore the following[19]: What are the possible dynamical events for two particles 1 and 2 or for three particles 1, 2, and 3, etc. between time $t = 0$ and the time $t \gg t_{coll}$ of interest, such that particles 1 and 2 collide at t? Furthermore, as the u functions in C_i depend on the phases of the particles at time $t = 0$, we are *also* interested, for a discussion of C_i, in the initial phases of the particles at time $t = 0$ that lead to the phases at time t. A collision between two or more particles takes place when the mutual distance between each particle and at least one other of the group is smaller or equal than r_0. A collision is indicated by writing the particles involved between brackets. The dynamical events between 0 and t can then be characterized by writing from right to left the collisions as they occur between 0 and t (cf., Fig. 2). For a discussion of F_2 and C_i not only sequences of real collisions that actually take place are of importance but also imaginary (hypothetical) collisions which would have occurred but for the presence of other particles (see also Refs. 4, 20, and 21, and reference in Footnote 19). These imaginary

[19] We should remark that due to the occurrence of S operators, i.e., $S_t^{(1)}$-operators in addition to $S_{-t}^{(s)}$ operators in the integrands of $F_2^{(1)}$ and $C_i^{(1)}$, more complicated configurations or events are possible than are considered here [J. V. Sengers, Phys. Fluids 9, 1333 (1966)]. Thus, for example, the operator $S_t^{(3)}(x_1x_2x_3)$ applied to a configuration of the particle 1, 2, and 3, where $r_{12} < r_0$ and $r_{23} = O(r_0)$ can transform this configuration into one where $r_{12} < r_0$ and $r_{12} = O(r_0)$. Although such dynamical events contribute to $F_2^{(1)}$ and $C_i^{(1)}$ they are not relevant for a discussion to the existence of $F_2^{(1)}$ and the vanishing of $C_i^{(1)}$ for large t. Therefore we restrict ourselves to dynamical events associated with $S_{-t}^{(s)}$ operators only.

[20] S. Ono and T. J. Shizume, Phys, Soc. Japan 18, 29 (1963).

[21] J. Weinstock, Phys. Rev. 132, 460 (1963).

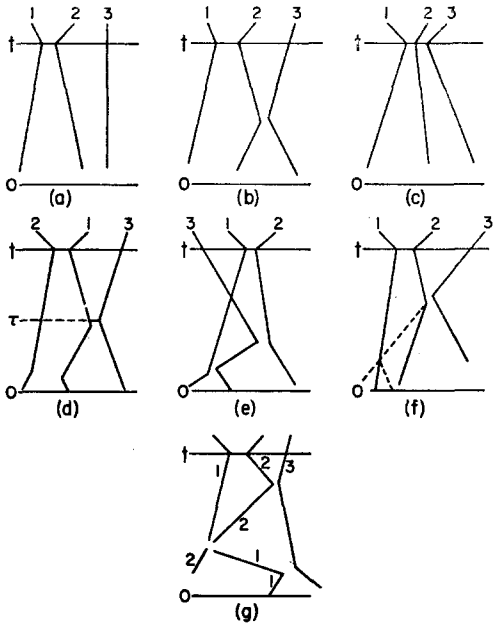

FIG. 2. Basic collisional events of three particles (but for permutations) between 0 and t, leading to a binary collision at t. (a) one binary collision: (12); (b) two successive binary collisions: (12)(23)[22]: (c) genuine triple collision (123); three successive binary collisions: (d) recollision: (12)(13)(12); (e) cyclic collision: (12)(23)(31): (f) imaginary collision: (12)(23)$\overline{(13)}$; (g) four successive (real) binary collisions: (12)(23)(12)(13).

collisions will be distinguished from the real collisions by a bar. Thus we have:

(1) *Case of two particles.* As two isolated particles in infinite space can at most suffer one collision, no (other) collision between the particles 1 and 2 can occur before t. As furthermore $t \gg t_{coll}$, the phases of the particles 1 and 2 at $t = 0$ are such that $r_{12} \gg r_0$.

(2) *Case of three particles.* The possible dynamical events between time 0 and $t \gg t_{coll}$ leading to a (12) collision of the particles 1 and 2 at time t can be classified as follows: (a) one binary collision, i.e., only the collision (12) at t occurs and no other [Fig. 2(a)]; (b) two successive binary collisions, i.e., the collision (12) at t takes place after a previous binary collision (13) or (23) at a time $0 \leq \tau < t$ has finished [Fig. 2(b)][22]; (c) a (genuine) triple collision at t, i.e., particle 3 interacts with particle 1 or particle 2 or both at t [see Fig. 2(c)][22]; (d) three successive binary collisions. The binary collisions before the (12)-collision at t can be real or imaginary (hypothetical) depending on whether a binary collision really takes place [Fig. 2(d), (e)] or would have taken place between the two particles involved but for the presence of a third particle [Fig. 2(f)]; (e) four or more successive binary collisions. The existence of four real successive binary collisions between three hard spheres [Fig. 2(g)] in two and three dimensions has been found by Foch and Cohen[17a,b] and by Thurston and Sandri.[23] That this is also the maximum number of real successive binary collisions between three hard spheres has been mentioned by Sandri[23] and has been proved by Murphy.[17b,24]

(3) *Case of four particles.* The possible collisional events between time 0 and $t \gg t_{coll}$ leading to a (12) collision at t can be classified as follows: (a) One binary collision (12) at t while the particles 3 and 4 do or do not collide between 0 and t; (b) events involving only three out of the four particles [leading to a (12) collision at t], while the fourth particle does not participate in any collision between 0 and t; (c) a (genuine) quadruple collision at t; (d) at least one genuine (real or imaginary) triple collision takes place before (12) at t [Fig. 3(a), (b)]; (e) four or more successive (real or imaginary) binary collisions among four particles. [Fig. 3(c)–(g)].

We distinguish between: (e_1) reducible sequences of collisions, where not all collisions in the sequence from 0 to t are essential for the phases of the four particles at t. A sequence is reducible if upon removal of one of the particles from consideration, the phases of the three remaining particles at time t are consistent with a sequence of collisions between 0 and t which differs from the original sequence by the absence of one collision, involving the removed particle, while all the other collisions remain unchanged [Fig. 3(c), (d)]; (e_2) irreducible sequences, where all collisions in a sequence are essential, in the sense defined above, for the phases of the four particles at t [Fig. 3(e), (f), (g)][25]

We omit a more detailed analysis of the dynamical events between four particles, as this is not essential for the purpose of this paper. We remark, however, that no results as mentioned in (2e) for three particles are at present available for four particles,

[22] There is also the possibility of an imaginary collision $\overline{(13)}$ or $\overline{(23)}$ in addition to the (real) (12)-collision at t. Just as the event sketched in Fig. 2(b) these events do not contribute to $F_2^{(1)}$ and $C_1^{(1)}$. We assume furthermore, in view of the discussions in Secs. 4, that for the case of strongly repulsive intermolecular forces considered in this paper, a genuine triple collision at time t cannot be preceeded by one or more sucessive binary collisions, unless they take place a time t_{coll} before t, so that the whole sequence of events can be considered as an "extended" triple collision. Similar considerations apply to those extra events that can occur due to the presence of $S_t^{(1)}$ operators in the integrands.[19]

[23] W. Thurston and G. Sandri, Bull. Am. Phys. Soc. 9, 387 (1964). See also G. Sandri, R. D. Sullivan, and P. Norem, Phys. Rev. Letters 13, 743 (1964).

[24] T. J. Murphy (to be published).

[25] The two successive binary collisions mentioned before in (2b) are also of the reducible type (e_1).

i.e., the maximum number of successive real binary collisions between four particles is not known.

4. INTEGRAL PROBLEM: ESTIMATES

The dynamical studies of the previous section are needed to discuss the long-time behavior of F_2 on the basis of the cluster expansions of Sec. 2. In particular, they can be used to discuss the long-time behavior of the coefficients $F_2^{(l)}(x_1x_2;\ t \mid F_1)$ in the expansion of $F_2(x_1x_2;\ t \mid F_1)$ and the decay of the initial state correlations with time through the coefficients $C_i^{(l)}(x_1x_2;\ t)$ in the expansion of $C_i(x_1x_2;\ t)$ for fixed x_1 and x_2 with $r_{12} \leq r_0$ at t.

(a) $F_2^{(0)}(x_1x_2;\ t \mid F_1)$ and $C_i^{(0)}(x_1x_2;\ t)$

For $r_{12} \leq r_0$ at t and $t \gg t_{\text{coll}}$ one has

$$F_2^{(0)}(x_1x_2;\ t \mid F_1) = F_2^{(0)}(x_1x_2 \mid F_1) = \mathcal{S}_\infty^{(2)}(x_1x_2)F_1(x_1;\ t)F_1(x_2;\ t), \tag{4.1}$$

while

$$C_i^{(0)}(x_1x_2;\ t) = 0 \tag{4.2}$$

as

$$S_{-t}^{(2)}(x_1x_2)u_2(x_1x_2;\ 0) = 0. \tag{4.3}$$

The equations (4.1) and (4.2) are in agreement with Bogoliubov's results.

(b) $F_2^{(1)}(x_1x_2;\ t|F_1)$ and $C_i^{(1)}(x_1x_2;\ t)$

Using (4.3) and that (cf., Footnote 22)

$$S_{-t}^{(3)}(x_1x_2x_3)u_3(x_1x_2x_3;\ 0) = 0, \tag{4.4}$$

the initial condition term $C_i^{(1)}$ given by (2.18b) can be reduced to

$$C_i^{(1)}(x_1x_2;\ t) = \int dx_3\ \{[S_{-t}^{(3)}(x_1x_2x_3) - S_{-t}^{(2)}(x_1x_2)S_t^{(1)}(x_1)S_{-t}^{(2)}(x_1x_3)]u_2(x_1x_3;\ 0) + [S_{-t}^{(3)}(x_1x_2x_3) - S_{-t}^{(2)}(x_1x_2)S_t^{(1)}(x_2)S_{-t}^{(2)}(x_2x_3)]u_2(x_2x_3;\ 0) + S_{-t}^{(3)}(x_1x_2x_3)u_2(x_1x_2;\ 0)\} \prod_{i=1}^{3} S_t^{(1)}(x_i)F_1(x_i;\ t). \tag{4.5}$$

We now prove the existence of $F_2^{(1)}(x_1x_2;\ t \mid F_1)$ given by (2.18b) as well as the vanishing of $C_i^{(1)}(x_1x_2;\ t)$ for $t \to \infty$ by estimating the volume in the phase space of particle 3 of those phases x_3 of particle 3, that contribute to $F^{(1)}(x_1x_2;\ t \mid F_1)$ or to $C_i^{(1)}(x_1x_2;\ t)$, respectively.

First we remark that the structure of the operator $\mathfrak{I}_i^{(3)}(x_1x_2 \mid x_3)$, is such that of the events listed above under (2), only those of (2c)–(2e) contribute to $F_2^{(1)}$ of (2.18b).[26] This is a consequence of the fact that, for all dynamical events, with the exception of the genuine triple collision, the $\mathcal{S}_t^{(3)}$ operator factorizes into a product of $\mathcal{S}_t^{(2)}$ operators.[17b]

Thus only for such phases x_3 of particle 3 that in the time interval between 0 and t a triple collision or a sequence of 3 or more successive binary collisions occur, is a nonvanishing contribution to $F_2^{(1)}(x_1x_2;\ t \mid F_1)$ obtained.

Similarly, the structure of the integrand of $C_i^{(1)}$ is such that the only contributions to the integral on the right-hand side of (4.5) come from dynamical events as listed under (2d) and (2e) above. In fact, only for such phases x_3 of particle 3 for which, in

[26] In addition to these events, there are also contribution from events mentioned in Ref. 19 but, as stated before, they are not relevant for our considerations (see Ref. 22).

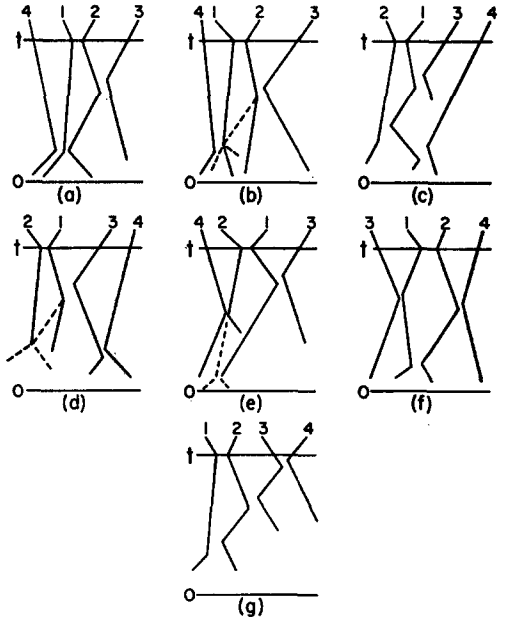

FIG. 3. Basic collisional events of four particles (but for permutations) between 0 and t leading to a binary collision at t. (a) real triple collision before (12): (12)(23)(124); (b) imaginary triple collision before (12): (12)(23)$\overline{(134)}$; four successive binary collisions before (12): reducible; (c) (12)(13)(12)(14); (d) (12)(13)$\overline{(23)}$(34) irreducible; (e) (12)(13)(24)$\overline{(12)}$; (f) (12)(13)(24)(12); (g) (12)(34)(23)(12).

the time interval between 0 and t, 3 or more successive binary collisions occur of which one binary collision occurs at $t = 0$ in addition to the one at t, a nonvanishing contribution to $C_i^{(1)}(x_1x_2;\ t)$ is obtained. That one of the binary collisions has to take place at $t = 0$ is due to the cluster property of the u_2.

On the basis of the preceding considerations one can make the following estimates for $t \gg t_{coll}$:

$$F_2^{(1)}(x_1x_2;\ t \mid F_1) = F_2^{(1)}(x_1x_2 \mid F_1) + \alpha_1(t_{coll}/t), \quad (4.6)$$

$$C_i^{(1)}(x_1x_2;\ t) = \beta_1(t_{coll}/t)^2, \quad (4.7)$$

where α_1 depends on the intermolecular potential and F_1, while β_1 depends in addition on the initial state

The result (4.6) has also been obtained by Green and Piccirelli.[4,27]

The details of the computations that lead to (4.6) and (4.7) are given in Appendix B. The essential features that lead to the given time dependences are simple and are presented here. We restrict ourselves to a discussion of the case of a recollision, with the first (12) collision taking place exactly at $t = 0$ [cf., Fig. 2(d)]. All other basic events sketched in Fig. 2 involving three successive binary collisions allow a similar discussion. The volume of the phase space of 3 associated with more complicated dynamical events as well as with the four-collision event sketched in Fig. 2(f), for example, cannot exceed that associated with the three-collision events, because for these dynamically more complicated events to occur, more severe restrictions are placed upon the phase space of particle 3, than for the three-collision events. Also, the phase space associated with genuine triple collisions is finite, since all three particles have to be within distances of $O(r_0)$ of each other. Therefore for a discussion of $F_2^{(1)}(x_1x_2;\ t \mid F_1)$ and of $C_i^{(1)}(x_1x_2;\ t)$ we can restrict ourselves to the events of (2d). As we are only interested in questions of convergence, it suffices for our purpose to ascertain whether the phase space associated with recollisions (12)(13)(12) to $C_i^{(1)}$ and $F_2^{(1)}$ remain finite. We first consider their contributions to $C_i^{(1)}$.

After the first (12)-collision at $t = 0$, particle 1 moves away from particle 2 with a relative velocity $\mathbf{v}_{21}$. Particle 3 must then hit particle 1, for instance, at a time $\tau < t$ in such a way that particle 1 collides with 2 again at time t after the first (12) collision. For this to occur for fixed velocity $\mathbf{v}_3$ of particle 3 and for *large* $\tau \gg t_{coll}$, which suffices for our purpose, particle 3 must hit particle 1 in such a way that the apseline of the (31) collision is in a solid angle $\sim r_0^2/(v_{21}t)^2$, where $v_{21} = |\mathbf{v}_{21}|$. This implies that for large τ the configuration space of particle 3 is restricted to a volume of the order of magnitude of $\sim[r_0^3/(v_{21}t)^2]\sigma_{tot}$, where σ_{tot} is the total cross section for the (32) collision. Because this result holds for every $\mathbf{v}_3$, integration over $\mathbf{v}_3$ does not change the result. Thus the volume in the phase space of particle 3 associated with nonvanishing contributions to $C_i^{(1)}$ due to recollisions remains finite, as the contributions due to large τ decrease with time $\sim(t_{coll}/t)^2$ for $t \gg t_{coll}$.

On the basis of this discussion we can also estimate the phase space of particle 3 associated with the contributions of recollisions to $F_2^{(1)}(x_1x_2;\ t \mid F_1)$. For a recollision to contribute to $F_2^{(1)}$, the first (12) collision must take place in the time interval between $t = 0$ and t, say at t'. We can now estimate the total volume in the phase space of particle 3 for large $t - t'$. Thereto we simply take, for varying t' and for $t - t' \gg t_{coll}$, the sum of the volumes of those regions of the phase space of particle 3 for which one (12) collision occurs at t' and another a time $t - t'$ later at t. This leads to a volume in the phase space of particle 3 that remains *finite* as the total contribution for large $t - t'$ decreases as the time *integral* of that which occurred in $C_1^{(1)}$, i.e., $\sim(t_{coll}/t)$ for $t \gg t_{coll}$.

As at least the same time dependence is obtained for the volumes in the phase space of particle 3 that contribute to $C_i^{(1)}$ and $F_2^{(1)}$ for all other possible collisional events, and since the number of these events is finite, the conclusions reached above, viz., that $F_2^{(1)}$ remains finite and that $C_2^{(1)}$ vanishes in the limit $t \to \infty$, are generally valid. It is of interest to consider in addition to the three-dimensional case the two-dimensional case. Then although the $C_i^{(1)}$ still goes to zero, the $F_2^{(1)}(x_1x_2;\ t \mid F_1)$ does not exist for $t \to \infty$; in fact $F_2^{(1)}$ diverges logarithmically with t, viz., *in two dimensions*,

$$F_2^{(1)}(x_1x_2;\ t \mid F_1) \sim O[\ln(t/t_{coll})], \quad (4.8)$$

$$C_i^{(1)}(x_1x_2;\ t) = \beta_1'(t_{coll}/t). \quad (4.9)$$

This is also proven in Appendix B. These results can be easily understood on the basis of the simple phase space arguments presented above. For, instead of a solid angle $\sim(r_0/v_{21}t)^2$, one now has only a plane angle $\sim(r_0/v_{21}t)$ to consider, which leads immediately to the results (4.8) and (4.9).

From (4.6) one can conclude that also the

[27] A related result has been obtained by S. Ono and T. J. Shizume (see Ref. 20).

Bogoliubov functional $F_2^{(1)}(x_1x_2 \mid F_1)$ given by (2.21) exists. Bogoliubov's results have thus been shown to be valid up to terms of $O(n)$ in F_2. In the next order, however, and already in this order in two dimensions, the Bogoliubov functional cannot be shown to exist anymore.

(c) $F_2^{(2)}(x_1x_2;\, t \mid F_1)$ and $C_i^{(2)}(x_1x_2;\, t)$

The structure of the operator $\mathfrak{I}_i^{(4)}(x_1x_2 \mid x_3x_4)$ is such that the only contributions to $F_2^{(2)}(x_1x_2;\, t \mid F_1)$ come from the dynamical events listed under (3c), (3d), and ($3e_2$).

Similarly the structure of the integrand of $C_i^{(2)}$ is such that the only contributions to $C_i^{(2)}$ come from (3d) and ($3e_2$), where in both cases, in addition to the (12) collision at t, at least one collision of the sequence occurs at $t = 0$.

On the basis of these considerations one can make the following estimates for $t \gg t_{\text{coll}}$:

$$F_2^{(2)}(x_1x_2;\, t \mid F_1) \sim O[\ln\,(t/t_{\text{coll}})], \tag{4.10}$$

$$C_i^{(2)}(x_1x_2;\, t) = \beta_2(t_{\text{coll}}/t). \tag{4.11}$$

In *two dimensions* these estimates become

$$F_2^{(2)}(x_1x_2;\, t \mid F_1)\ (t/t_{\text{coll}}), \tag{4.12}$$

$$C_i^{(2)}(x_1x_2;\, t) \sim O[\ln\,(t/t_{\text{coll}})]. \tag{4.13}$$

The proof of formulas (4.10)–(4.13) can be found in Appendix B. These results can be understood in the same simple way as those for $F_2^{(1)}$ and $C_2^{(1)}$ before. Thereto we consider, for example, the contributions of the events of Fig. 3(f) to $C_2^{(2)}$. Fixing first the phase of particle 4 (in addition to those of particles 1 and 2), the volume in the phase space of particle 3 for large t contains a contribution which decreases with increasing $t \sim (t_{\text{coll}}/t)^2$. Integrating now over all the phases of particle 4, i.e., over the collision cylinder of 4 with respect to 2, leads to a factor $\sim(t/t_{\text{coll}})$. A similar discussion applies to the event of Fig. 3(g). The total volume of the events mentioned in the combined phase spaces of the particles 3 and 4 decreases therefore $\sim(t_{\text{coll}}/t)$, so that $C_2^{(2)}$ exists. However, from this follows also that $F_2^{(2)}(x_1x_2;\, t \mid F_1)$ contains contributions which grow with time $\sim \ln\,(t/t_{\text{coll}})$ and consequently cannot be shown to exist.[28]

Thus one sees that in three dimensions, although the initial state is still "forgotten" to $O(n^2)$, $F_2^{(2)}(x_1x_2;\, t \mid F_1)$ does not exist for $t \to \infty$, which implies that also the Bogoliubov functional $F_2^{(2)}(x_1x_2 \mid F_1)$ does not exist. In the next order—and already in this order in two dimensions—the initial conditions cannot be shown to be forgotten anymore. In fact, $C_i^{(3)}(x_1x_2;\, t)$ then diverges logarithmically [cf., (4.15)].

(d) $F_2^{(l)}(x_1x_2;\, t \mid F_1)$ and $C_2^{(l)}(x_1x_2;\, t)$

Generalizing the preceding considerations, one easily convinces oneself that one has in d dimensions, for $l \geq d$ $(d = 2, 3)$,

$$F_2^{(l)}(x_1x_2;\, t \mid F_1) \sim O[(t/t_{\text{coll}})^{l-d+1}], \tag{4.14}$$

while for $l = d$,

$$C_i^{(l)}(x_1x_2;\, t) \sim O[\ln\,(t/t_{\text{coll}})], \tag{4.15}$$

and for $l \geq d + 1$

$$C_i^{(l)}(x_1x_2;\, t) \sim O[(t/t_{\text{coll}})^{l-d}]. \tag{4.16}$$

These results can easily be understood on the basis of events like those of Fig. 3(g), where, in general, the (12) recollision is triggered by a sequence of successive binary collisions $(l,\ l-1)$, $(l-2,\ l-3)$, $\cdots$, (4, 3) between the particles $l, l-1, \cdots$, 4, 3 instead of between the particles 3 and 4 only, as in Fig. 3(g). As each extra collision adds a factor (t/t_{coll}), Eqs. (4.14)–(4.16) follow readily from Eqs. (4.10)–(4.13).

Thus the expansion (2.18) for $F_2(x_1x_2;\, t \mid F_1)$ cannot be used to obtain the long-time behavior of F_2, if one wants to restrict oneself to the first few terms only. Similarly the expansion (2.19) for C_i cannot be used to prove the decay of the initial correlations. We remark that phase space estimates alone are not sufficient to establish the divergences mentioned above. In fact, one has assumed that the coefficient of the term which gives the dominant time dependence is not equal to zero. This coefficient depends, apart from on F_1, on the dynamics of isolated groups of particles. The vanishing of these coefficients seems only possible if, for a general intermolecular potential, a special connection exists between the dynamics of isolated groups of 3, 4, $\cdots$ particles and F_1. This is in general not so, as F_1 is determined by the system as a whole. Only for the special case that the system is in local or total equilibrium, does this connection seem to exist.

From the considerations of Appendix A it follows that the results obtained in this section for F_2 can be directly transferred to the time correlation function ψ_2. In fact, formulas (4.1), (4.6), (4.10), and (4.14) remain valid if F_2 is replaced by ψ_2, and πF_1 is linearized with respect to ψ_1 (cf. Appendix A).

[28] Similar results have been found by J. Weinstock[11] and by E. A. Frieman and R. Goldman.[12]

The significance of the conclusions reached in this section is briefly discussed in the next section.

5. DISCUSSION

A few comments on the results of the previous sections seem in order.

(a) The transition from the cluster expansion (2.10) for F_2 in terms of D_1 functions to the expansion (2.17) for F_2 in terms of F_1 functions has eliminated the most secular terms present in (2.10), viz., the coefficient of n^l in (2.17) contains terms growing with time $\sim(t/t_{\text{coll}})^{l-4}$ instead of $\sim(t/t_{\text{coll}})^{l-2}$, as in (2.10). However, the presence of secular terms in (2.17), although weaker than in (2.14), makes the expansion (2.17) still unsuited for the computation of the long-time behavior of F_2, if one wants to restrict oneself to a few terms only. It implies that the Bogoliubov functional $F_2(x_1x_2 \mid F_1)$ does not exist beyond $O(n)$.

(b) The nonexistence of the $F_2^{(l)}(x_1x_2 \mid F_1)$ is related to the absence of a similar cluster property of the integrands, as is the case in equilibrium. Also the impossibility to prove the disappearance of the correlations present in the initial state beyond $O(n^2)$ is related to the fact that the most important contributions to the integrals in F_2 and C_l for large l come from configurations of the particles, 3, 4, $\cdots$, l where they are far from the particles 1 and 2. This implies, for example, that, unlike in equilibrium, the probability density to find particles 1 and 2 colliding at t would depend significantly on the particles 3, 4, $\cdots$, l very far away.

(c) The point raised in part (b) is clearly unphysical, since in a gas the correlations should not extend beyond a few mean free paths. This then means that one cannot expand the pair distribution function of a dense gas in terms of the dynamics of small isolated groups of particles in infinite space alone. Clearly a "cutoff" of the free motion over distances of the order of a mean free path l (or over times of the order of t_{mfp}) should be introduced.[28] This, however, would introduce in the expansions (2.17) and (2.18) a new length (viz., l), characteristic for the collective behavior of the gas as a whole, in addition to the range of the forces r_0, which is characteristic for the behavior of small groups of individual molecules. This then implies that a classification of the approach of a gas to equilibrium in two well-separated stages, characterized by t_{coll} and t_{mfp}, respectively, is *not* possible, so that no kinetic stage in the sense of Bogoliubov exists. Mathematically a resummation of the expansions (2.17) and (2.18) has to be carried out, which would lead in a natural way to a cutoff of the integrals in (2.17) and (2.18)[28,29] (see especially last reference cited in Footnote 11). If one assumes that the result of such a resummation would lead in first approximation to a replacement in the formulas (4.6)–(4.16) of t by t_{mfp}, then one would not only expect a decay of the influence of the initial state in a characteristic time of t_{mfp} but in addition one would expect to obtain contributions to the transport coefficients $\sim(nr_0^3)\ \ln\ (nr_0^3)$ ($\sim nr_0^2\ \ln\ nr_0^2$ in two dimensions).[17b,28,30,31] This logarithmic density dependence of the transport coefficients follows under the same assumption *directly* from the time-correlation function expressions for the transport coefficients. In principle, therefore, the logarithmic divergences in F_2 have direct observational consequences.[32] Of course, in addition to terms in η and λ which have a logarithmic dependence on the density, and which are due to the contributions of recollisions, etc., to F_2 (i.e., which incorporate "memory effects over a mean free path"), there are also present the terms which depend on the powers of n alone and which are obviously due to (convergent) contributions of small groups of isolated particles to F_2. On this basis, one would expect then that the Choh–Uhlenbeck expressions for η_1 and λ_1 still give the correct coefficients of the density proportional corrections to the Chapman–Enskog results η_0 and λ_0, respectively.

(d) Although of great theoretical importance the contributions of the "memory effects over a mean free path" do not seem to be important for a numerical calculation of the transport coefficients. Using a cutoff as mentioned in part (c), Sengers had made a computation of the coefficient of the contribution $\sim nr_0^2\ \ln\ nr_0^2$ to the viscosity of a two-dimensional gas of hard disks.[33] Here the logarithmic term amounted to only a few percent of the estimated value of the term $\sim nr_0^2$. Furthermore, if one neglects these effects altogether and restricts oneself to the contributions of triple, quadruple, etc. collisions only, agreement with experiment is obtained within the present experimental error of a few percent.[34,35]

---

[29] E. G. D. Cohen and J. R. Dorfman (to be published).
[30] J. V. Sengers, Phys. Rev. Letters 15, 515 (1965).
[31] A calculation based on the binary collision expansion leading to the same result as obtained by J. V. Sengers has been given by L. K. Haines, J. R. Dorfman and M. H. Ernst, Phys. Rev. 144, 207 (1966).
[32] In fact, elementary considerations using mean-free-path concepts suggest that the transport coefficients contain contributions $\sim n^r(\ln n)^s$. This has also been suggested by J. M. J. van Leeuwen and A. Weijland, Phys. Letters, 19, 562 (1962).
[33] J. V. Sengers, Phys. Fluids 9, 1685, (1966).
[34] J. V. Sengers, Intern. J. Heat Mass Transfer I, 1103 (1965).
[35] D. K. Hoffman and C. F. Curtiss, Phys. Fluids 8, 890 (1965).

ACKNOWLEDGMENT

The authors would like to express their gratitude to Dr. J. V. Sengers of the National Bureau of Standards for many valuable discussions.

APPENDIX A

In this Appendix we indicate cluster expansions appropriate to the evaluation of the time correlation function expressions for the transport coefficients. These expansions are analogous to those used for the distribution functions in Sec. 2 of this paper. We only outline the main developments so as to exhibit the close analogy with the distribution functions.

As was shown in a previous paper,[10] the transport coefficients in an infinite system can be expressed in terms of the long-time behavior of reduced time-correlation functions, of which in the case of additive intermolecular forces only the first and the second (i.e., $s = 1$ and 2) are needed:

$$n^s\psi_s(x_1 \cdots x_s; t) = \lim_{V\to\infty} \sum_{N\geq s}^{\infty} \frac{z^N}{(N-s)!\, Z_{gr}} \int dx^{N-s}\, S_{-t}^{(N)}(x^N) e^{-\beta H_N} M_N, \tag{A1}$$

where M_N is in general given by

$$M_N = M_N(x^N) = \sum_{i=1}^{N} m_1(x_i) + \sum_{i<j}^{N}\sum_{1} m_2(x_i x_j). \tag{A2}$$

Expressions for $m_1(x_i)$ and $m_2(x_i, x_j)$ have been given elsewhere[10]; we only remark that $m_2(x_i x_j) = 0$ whenever $r_{ij} > r_0$. Furthermore in (A1) and (A2) $x^N = x_1, x_2, \cdots, x_N$, Z_{gr} is the grand canonical partition function, z the activity, and $\beta = 1/kT$, where k is Boltzmann's constant.

The cluster expansions for distribution functions can most conveniently be applied to a computation of the[36] ψ_1 and ψ_2 if one first introduces generating functions $X_s(x_1 \cdots x_s; t; \Lambda)$ defined by[37]

$$n^s X_s(x_1 \cdots x_s; t; \Lambda) = \lim_{V\to\infty} \sum_{N\geq s} \frac{z^N}{(N-s)!\, Z_{gr}(\Lambda)} \times \int dx^{N-s}\, S_{-t}^{(N)}(x^N) e^{-\beta H_N} e^{\Lambda M_N}, \tag{A3}$$

to which the functions of interest are related by

$$\psi_s(x_1 \cdots x_s; t) = \left[\frac{\partial X_s(x_1 \cdots x_s; t; \Lambda)}{\partial \Lambda}\right]_{\Lambda=0}. \tag{A4}$$

In (A3) $Z_{gr}(\Lambda)$ is defined by

$$Z_{gr}(\Lambda) = \lim_{V\to\infty} \sum_{N=0} \frac{z^N}{N!} \int dx^N\, e^{-\beta H_N} e^{\Lambda M_N}. \tag{A5}$$

The X functions can be considered as non-equilibrium distribution functions for a special initial state given by $D_N(x_1 \cdots x_N; 0) = e^{-\beta H_N} e^{\Lambda M_N}$. Therefore their long-time behavior can also be obtained in a way completely analogous to the case of non-equilibrium distribution functions. Thus the same basic expansion exists for $n^2X_2(x_1x_2; t)$ in terms of nX_1 as for $n^2F_2(x_1x_2; t)$ in terms of nF_1 [cf., (2.20)]. One only has to replace the $a_s(x_1 \cdots x_s; 0)$ in the $\mathfrak{F}_t$ operators by

$$a_s(x_1 \cdots x_s; 0) \to W_s(x_1 \cdots x_s; \Lambda) = \exp\left\{-\sum_{i_1<j}^{s} [\beta\phi(r_{ij}) - \Lambda M_2(x_i x_j)]\right\},$$

$$\prod_{i=1}^{l} F_1(x_i; t) \quad \text{by} \quad \prod_{i=1}^{l} X_1(x_i; t; \Lambda).$$

In this expansion the contributions of the initial state can be split off by introducing $u_l(x_1 \cdots x_l; \Lambda)$ functions to the W functions [cf. (2.16)]. We remark that unlike the u functions of (2.16) the $u(\Lambda)$ functions have a cluster property, viz., they vanish as soon as two particles are outside the range r_0 of the intermolecular forces, because the W functions have a product property. Thus the X_2 can be written identically as the sum of two terms, viz.,

$$n^2X_2(x_1x_2; t; \Lambda) = n^2X_2(x_1x_2; t \mid X_1(\Lambda)) + n^2C_i(x_1x_2; t; \Lambda). \tag{A6}$$

Here the initial-state correction term C_i is given by the expansion (2.19), where the a_s and u_s have been replaced by W_s and $u_s(x_1 \cdots x_s; \Lambda)$, respectively, and the $\Pi F_1(x_i; t)$ by $\Pi X_1(x_i; t; \Lambda)$, while the $X_2(t \mid X_1)$ is obtained from the expansion (2.18) for $F_2(t \mid F_1)$ by replacing ΠF_1 by ΠX_1.

From (A6) the corresponding expression for $n^2\psi_2$ in terms of $n\psi_1$ can be obtained by using (A4). We only remark that $n^2\psi_2$ is a *linear* functional of $n\psi_1$ since by (A4), $\Pi X_1(x_i; t; \Lambda)$ is replaced by $\Sigma\psi_1(x_1; t) f_0(p_2) \cdots f_0(p_l)$, where $f_0(p)$ is the equilibrium Maxwell–Boltzmann distribution function and

[36] Alternative cluster expansions have been discussed in Refs. 10 and 17a.

[37] The authors are much indebted to Dr. M. H. Ernst for pointing this out to them.

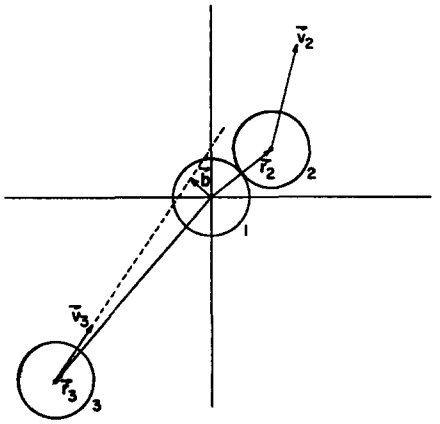

FIG. 4. Phases at $t=0$ for three-particle recollision (12)(13)(12).

the sum is over all permutations of the l molecules $1 \cdots l$.

APPENDIX B

Estimates of Phase Space Volumes

(1) We consider first the case of three particles in infinite space, with strong short-range repulsive interactions of range r_0.[38] We first consider the recollision event. As we are only interested in the question whether a volume in phase space is finite or infinite for large t(i.e., $t \gg t_{coll}$) we restrict ourselves to a discussion of the case $t \gg t_{coll}$. Then we wish to estimate the volume $d\Gamma_3$ in the phase space of particle 3 for given phases x_1 and x_2 of the particles 1 and 2 (and for $r_{12} = |\mathbf{q}_1 - \mathbf{q}_2| \leq r_0$) such that a recollision sequence (12)(13)(12) occurs with a time interval between the first and the second (12)-collision between t and $t + dt$ [cf. Fig. 2(d)].

After the first (12) collision at $t = 0$, particle 1 moves away from particle 2. Particle 3 must then hit particle 1 at a time τ [after the first (12)-collision] in such a way that particle 1 catches up with particle 2 and a second (12)-collision occurs between a time t and $t + dt$ after the first (12)-collision at $t = 0$.

We choose a coordinate system such that particle 1 is at rest at the origin immediately after the first (12)-collision at $t = 0$. In this coordinate system let $\mathbf{r}_2, \mathbf{v}_2$ be the phase of particle 2 and $\mathbf{r}_3, \mathbf{v}_3$ the phase of particle 3, immediately after the first collision (cf. Fig. 4). If $\mathbf{r}_1', \mathbf{v}_1'$ is then the phase of particle 1 immediately after the (13)-collision, $\mathbf{v}_1'$ is given by

$$\mathbf{v}_1' = (\mathbf{v}_3 \cdot \mathbf{k})\mathbf{k}, \tag{B1}$$

where $\mathbf{k}$ is a unit vector in the direction of the apse line of the (13)-collision. We perform the estimate of the volume of the phase space of particle 3 in two steps: first, [part (a)], we consider the dynamical condition for a second (12) collision to occur. Then [part (b)], we estimate the volume mentioned on the basis of this.

(a) The *dynamical condition* that a second (12)-collision takes place between t and $t + dt$ implies that there is a time between t and $t + dt$ for which the distance of the particles 1 and 2 equals r_0, viz,

$$|\mathbf{r}_1(t + \alpha\, dt) - \mathbf{r}_2(t + \alpha\, dt)| = r_0 \tag{B2}$$

for some α with $0 \leq \alpha \leq 1$.

In general Eq. (B2) can refer either to a-just-beginning or to a-just-ending collision at $t + \alpha\, dt$. We consider only the former case.

Let $\mathbf{b}$ be the vector from the origin which is perpendicular to $\mathbf{v}_3$, so that $b = |\mathbf{b}|$ is the impact parameter of the (13)-collision. We are interested, for given $\mathbf{b}$, $\mathbf{v}_3$, and $\mathbf{v}_2$, in those values of τ, that lead to a (12)-collision at $t + \alpha\, dt$, for large t.

First, we remark for later use [part (b)], that $\mathbf{r}_3$ can be expressed in terms of $\mathbf{v}_3$, $\mathbf{b}$, and τ by (cf. Fig. 5)

$$\mathbf{r}_3 = -\left[\tau + \frac{(r_0^2 + b^2)^{\frac{1}{2}}}{v_3}\right]\mathbf{v}_3 + \mathbf{b}. \tag{B3}$$

Now we note that Eq. (B2) can be written in the form

$$|[\mathbf{r}_2 + \mathbf{v}_2(t + \alpha\, dt)] - [\mathbf{r}_2' + \mathbf{v}_1'(t + \alpha\, dt - \tau - t_{13})]| = r_0, \tag{B4}$$

where t_{13} is the duration of the (13)-collision. Using that $|\mathbf{r}_2|$, $|\mathbf{r}_1'|$, and $v_1' t_{13}$ are all of the order of r_0,

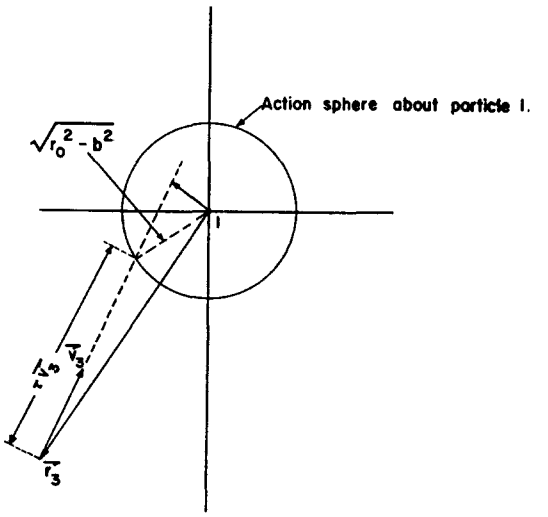

FIG. 5. Collision geometry of (13) collision at τ.

[38] A brief discussion of the case of three hard spheres in two and three dimensions can be found in Refs. 17(b) and 33.

(B4) can be replaced for large $t \gg t_{13}$ by an equation in which the details of the initial positions (i.e., $\mathbf{r}_2$ and $\mathbf{r}_1'$) no longer occur:

$$\left|\mathbf{v}_2 - \mathbf{v}_1'\left(1 - \frac{\tau}{t + \alpha\, dt}\right)\right| = \frac{r_0}{t + \alpha\, dt} + O\left[\left(\frac{t_{coll}}{t}\right)^2\right]. \quad \text{(B5)}$$

Equation (B5) is a quadratic equation for τ, which expresses the condition for the occurrence of a second (12)-collision for large t. Solving (B5) for τ, for fixed $\mathbf{b}$ and $\mathbf{v}_3$ (i.e., $\mathbf{v}_1'$) and $\mathbf{v}_2$, yields neglecting terms of the order $(t_{coll}/t)^2$:

$$\frac{\tau}{t + \alpha\, dt} = \frac{\mathbf{v}_1'\cdot(\mathbf{v}_1' - \mathbf{v}_2) + \left[\left(\frac{r_0}{t + \alpha\, dt}\right)^2 - v_2^2 \sin^2\beta\right]^{\frac{1}{2}} v_1'}{v_1'^2}, \quad \text{(B6)}$$

where $v_1' = |\mathbf{v}_1'|$ and $v_2 = |\mathbf{v}_2|$.

Here β is the angle between $\mathbf{v}_1'$ and $\mathbf{v}_2$ or also between $\mathbf{k}$ and $\mathbf{v}_2$.

It follows from (B5) that for τ to be *real*, one must have up to $(t_{coll}/t)^2$:

$$|\sin\beta| \leq r_0/t\, |v_2|. \quad \text{(B7)}$$

Equation (B7) expresses the condition that particle 3 must hit particle 1 into the solid angle subtended at the origin by the interaction sphere around particle 2 at time t (cf. Fig. 6). Similarly it follows from Eq. (B6) that for τ to be *positive*, one must have

$$\mathbf{v}_1'\cdot(\mathbf{v}_1' - \mathbf{v}_2) \geq 0, \quad \text{(B8)}$$

as the second term in the numerator on the right-hand side of (B5) is always smaller than the first for sufficiently large t.

Equation (B8) expresses the condition that the particles 1 and 2 move towards each other after the (13)-collision. Using (B1), (B8) can, up to terms of $O[(t_{coll}/t)^2]$, be replaced by the condition

$$(\mathbf{v}_3\cdot\mathbf{k}) \geq v_2. \quad \text{(B9)}$$

(b) Equations (B6)–(B9) allow us to estimate *the volume $d\Gamma_3$ of the phase space of particle 3* such that a second (12)-collision takes place between a time t and $t + dt$ after the first (12)-collision:

$$d\Gamma_3 = \int d\mathbf{r}_3 \int d\mathbf{v}_3 = \int d\mathbf{b} \int d\mathbf{v}_3 \int_{\tau_0}^{\tau_1} v_3\, d\tau, \quad \text{(B10)}$$

where τ_0 and τ_1 are the values of τ that lead to a second (12)-collision at t (i.e., $\alpha = 0$) and at $t + dt$ (i.e., $\alpha = 1$), respectively.

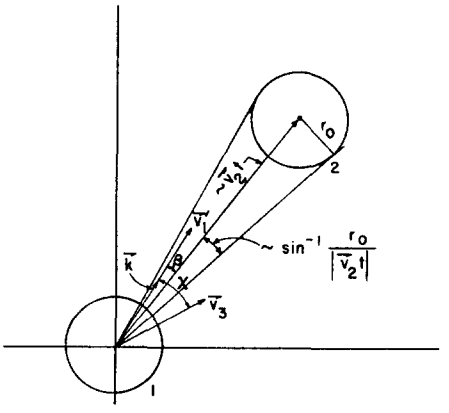

FIG. 6. Geometry for (12)-recollision at t.

Using (B6) and (B9), (B10) can be written in the form

$$d\Gamma_3 = dt \int d\mathbf{b} \int_{\mathbf{v}_3\cdot\mathbf{k} \geq v_2} d\mathbf{v}_3\, v_3 \times \frac{\mathbf{v}_1'\cdot(\mathbf{v}_1' - \mathbf{v}_2) + v_1'(r_0^2/t^2 - v^2 \sin^2\beta)^{\frac{1}{2}}}{v_1'^2}. \quad \text{(B11)}$$

It is now convenient to transform the integration over $\mathbf{b}$ to one over $\mathbf{k}$. Using Eq. (B7) we can easily estimate the $\mathbf{k}$ integral. Writing then with (B1)

$$v_1' = \mathbf{v}_3\cdot\mathbf{k} = v_3 \cos\chi,$$

where χ is the scattering angle for the (13)-collision, setting $\cos\beta = 1$ and $d\mathbf{b} = b\, db\, d\epsilon$, where ϵ is an azimuth angle, neglecting terms of $O(r_0/v_1't)$ and using that the differential cross section for the (13)-collision $\sigma_{13}(\chi, \epsilon) = (b/2 \sin 2\chi)|db/d\chi|$ is bounded by the total cross section σ_T, Eq. (B11) can be replaced by

$$d\Gamma_3 \leq 4\, dt\, \sigma_T \int d\mathbf{k} \int_{v_3 \cos\chi > v_2} d\mathbf{v}_3\, (v_3 \cos\chi - v_2). \quad \text{(B12)}$$

Now, according to Eq. (B7), $\mathbf{k}$ must lie in a solid angle of size $2\pi(r_0/v_2 t)^2$ around $\mathbf{v}_2$. Therefore, if $\mathbf{k}_0$ is a unit vector in the direction of $\mathbf{v}_2$, and if $\mathbf{v}_3\cdot\mathbf{k}_0 = v_3 \cos\chi_0$, one has, after expanding $\mathbf{k}$ around $\mathbf{k}_0$ and integrating:

$$d\Gamma_3 \leq 8\pi\, dt\, \sigma_T \left(\frac{r_0}{v_2 t}\right)^2 \times \int_{v_3 \cos\chi_0 > v_2} d\mathbf{v}_3\, (v_3 \cos\chi_0 - v_2), \quad \text{(B13)}$$

Equation (B13) gives an estimate of the volume in the phase space of particle 3, such that a second (12)-collision occurs, beginning between t and $t + dt$, for large t. The result (4.6) for $F_2^{(1)}(x_1 x_2;\ t \mid F_1)$

follows now immediately by integrating (B13) over t from t to ∞, assuming that the momentum distribution function is integrable.

The result (4.7) is obtained by a slight variation of the calculation that leads to (B13). For, in that case we require that the particles 1 and 2 are undergoing a second collision at the time t. For this to be so, this collision has to *begin* between $t - t_{\text{coll}}$ and t. Thus to obtain (4.7) one has to replace the interval t to $t + dt$ by $t - t_{\text{coll}}$ to t.

The two-dimensional results are obtained by a trivial modification. One merely replaces solid angles by plane angles to obtain

$$d\Gamma_3^{(2)} \leq 2\left(\frac{r_0}{v_2 t}\right) dt\, \sigma_T^{(2)} \times \int_{v_3 \cos \chi_0 > v_2} d\mathbf{v}_3\, (v_3 \cos \chi_0 - v_2), \qquad \text{(B14)}$$

where the superscript (2) refers to the values of the quantities in two dimensions.

We finally remark that phase space estimates can be made for all other three-particle events which contribute to $F_2^{(1)}(x_1 x_2; t \mid F_1)$ and to $C_i^{(1)}$ in a completely similar fashion. They all lead to the results (4.6) and (4.7), respectively.

(2) We now turn to a discussion of the case of four particles with strong short-range repulsive interactions of range r_0 in infinite space. The events illustrated in Fig. 3(a), (b), which involve at least one genuine triple collision, give rise to volumes in the combined phase space of the particles 3 and 4 of the same time dependence as in the case of three particles discussed part (1). This follows immediately from the observation that the time dependence of these events is governed by the volume of the phase space of that particle which participates in the middle (binary) collision, just as in the case of a recollision of three particles.

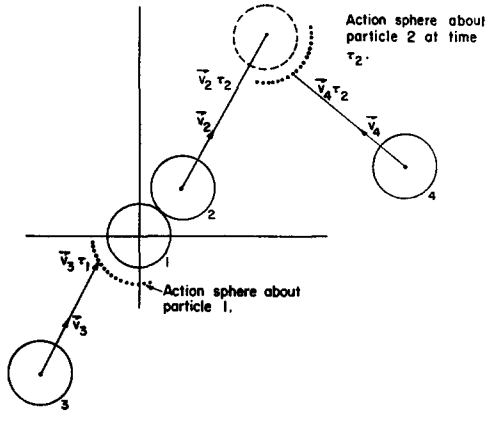

FIG. 7. Phases at $t = 0$ for four particle recollision: (12)(13)(24)(12).

Next we consider the events like those illustrated in Fig. 3(e), (f), which involve four successive binary collisions between the four particles. The time dependence of the phase space associated with the event of Fig. 3(e) follows directly from the previous arguments. In fact one is interested in the volume in the combined phase space of particles 3 and 4 such that the event illustrated in Fig. 3(e) takes place between t and $t + dt$. One sees immediately, that for this to occur, particle 4 need only to lie in the collision cylinder of 4 with respect to 2, so that the (24)-collision takes place between 0 and t. Hence for the volume $d\Gamma_{34}$ in the combined phase space of particles 3 and 4, one readily derives

$$d\Gamma_{34} \leq 8\pi\left(\frac{r_0}{v_2 t}\right)^2 t\, dt\, \sigma_T^2 \times \int_{\mathbf{v}_4 \cdot \mathbf{v}_2 > 0} d\mathbf{v}_4\, |\mathbf{v}_4 - \mathbf{v}_2| \times \int_{v_3 \cos \chi_0 > v_2} d\mathbf{v}_3\, (v_3 \cos \chi_0 - v_2) \qquad \text{(B15)}$$

or

$$d\Gamma_{34} \leq f(\mathbf{v}_2, r_0, \sigma_T)(dt/t), \qquad \text{(B16)}$$

where $f(\mathbf{v}_2, r_0, \sigma_T)$ indicates a function of $\mathbf{v}_2$, r_0, and σ_T.

The results (4.8) and (4.9) for $F_2^{(2)}(x_1 x_2; t \mid F_1)$, and for $C_i^{(2)}(x_1 x_2; t)$ for this event follow immediately from (B16) using similar arguments as before in part (1).

The events illustrated in Fig. 3(f) require a slightly more complicated analysis. Again, one is interested in the volume of the combined phase space of the particles 3 and 4, for which the second (12)-collision takes place between t and $t + dt$ after the first (12)-collision at $t = 0$. This case is similar to the case of the recollision of three particles treated in part (1) except that for a second (12)-collision to occur, particle 4 must now intervene to hit particle 2 to where particle 1 is at t. This extra process can happen at any time τ before t. For $\tau = 0$, the recollision case is regained.

We choose a coordinate system (cf. Fig. 7) such that particle 1 is at rest at the origin immediately after the first (12)-collision at $t = 0$ and that particle 2 is moving away with a velocity $\mathbf{v}_2$. Let τ_1 be the time of the (13)-collision; τ_2 the time of the (24) collision; $\mathbf{v}_1'$ the velocity of particle 1 after τ_1 and $\mathbf{v}_2'$ the velocity of particle 2 after the (24)-collision. Then the *dynamical condition* for the occurrence of

a second (12)-collision between t and $t + dt$ (B2) reads in this case, neglecting terms of $O(t_{\text{coll}}/t)$,

$$|\mathbf{v}_1'(t + \alpha\, dt - \tau_1) - \mathbf{v}_2\tau_2 - \mathbf{v}_2'(t + \alpha\, dt - \tau_2)| = r_0, \qquad \text{(B17)}$$

(B17) is a quadratic equation for τ_1. Solving (B17) for τ_1, one obtains an equation for τ_1 identical to that obtained before for τ as given by Eq. (B6), except that $\mathbf{v}_2(t + \alpha\, dt)$ in (B6) is replaced here by $\mathbf{v}_2\tau_2 + \mathbf{v}_2'(t + \alpha\, dt - \tau_2)$. Also two conditions are obtained analogous to (B7) and (B8). Using this, one obtains, in a similar fashion as before, for the volume $d\Gamma_{34}$ of the combined phase space of particle 3 and 4 such that a second (12)-collision takes place between t and $t + dt$ after the first:

$$\begin{aligned} d\Gamma_{34} &= \int d\mathbf{r}_3 \int d\mathbf{r}_4 \int d\mathbf{v}_3 \int d\mathbf{v}_4 \\ &\leq 2\pi\left(\frac{r_0}{v_2 t}\right)^2 dt\, \sigma_T^2 \int_0^t d\tau_2 \int_{\mathbf{v}_4\cdot\mathbf{v}_2>0} d\mathbf{v}_4 \\ &\times \int_{|v_3\cos\chi_0|>|\mathbf{v}_2\tau_2/t+\mathbf{v}_2'(t-\tau_2)/t|} d\mathbf{v}_3 \\ &\times \frac{|v_3 t\cos\chi_0 - |\mathbf{v}_2\tau_2 + \mathbf{v}_2'(t-\tau_2)||}{|\mathbf{v}_2\tau_2 + \mathbf{v}_2'(t-\tau_2)|^2}. \end{aligned} \qquad \text{(B18)}$$

This expression for $d\Gamma_{34}$ is considerably more complicated than that obtained for $d\Gamma_3$ before. We only extract the behavior for large t. Using the substitution $\tau_2 = \lambda t$ (with $0 < \lambda < 1$), we obtain, neglecting terms of $O[(t_{\text{coll}}/t)^2]$[39]

$$d\Gamma_{34} \leq f(\mathbf{v}_2, r_0, \sigma_T)(dt/t). \qquad \text{(B19)}$$

The results (4.8) and (4.9) for $F_2^{(2)}(x_1x_2;\, t \mid F_1)$ and for $C_i^{(2)}(x_1x_2;\, t)$ for this event follow now immediately from Eq. (B19). The corresponding formulas for the two-dimensional case can be derived again by simply replacing in the present derivation solid angles by plane angles.

One can convince oneself that there are no four particle events contributing to $F_2^{(2)}$ and $C_i^{(2)}$, for which the associated volume in the combined phase space $d\Gamma_{34}$ of the particles 3 and 4 decays for large t slower with t than as given by (B19).

The generalization to dynamical events involving more than four particles and the derivation of the results (4.14)–(4.16) is straightforward.

[39] It will be appreciated that singularities in the integrand of (B18) could only occur when particle 4 collides with particle 2 so that this particle is hit directly back to the origin, thus eliminating the necessity for particle 3 to hit particle 1 very far. In such a case restrictions are placed on the (24)-collision, which when taken into account, do not change the result (B19).

Microscopic Approach to Kinetic Theory

WILLARD R. CHAPPELL*
Smithsonian Institution Astrophysical Observatory and Harvard University, Cambridge, Massachusetts
(Received 2 June 1966)

A microscopic kinetic theory is developed for a plasma by the use of approximate equations of motion for the microscopic "exact" distribution function

$$f(\mathbf{r}, \mathbf{v}, t) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i(t))\, \delta(\mathbf{v} - \mathbf{v}_i(t)).$$

These equations can be solved to obtain asymptotic expressions for $f(\mathbf{r}, \mathbf{v}, t)$ that are then used to calculate correlation functions. These approximate equations are also used to obtain the approximate equations for the correlation functions and the principal results of the test-particle approach. In particular, two sets of equations are constructed. One set describes a homogeneous, slowly varying system and yields the Balescu–Lenard equation. The other set describes a homogeneous, slowly varying system that contains small, inhomogeneous and quickly varying perturbations.

INTRODUCTION

SEVERAL methods for deriving kinetic equations for a plasma have been developed in recent years.[1-6] In general, these methods fall into two classes. One class is composed of those techniques that are based on Bogoliubov's work and the BBKGY hierarchy. The other class follows the lines of development of the Prigogine school. These methods are alike in the sense that they are developed entirely from the Liouville equation, which is a macroscopic equation of motion (where by macroscopic we mean that the individual particle motion has been smeared out by an ensemble average).

A few attempts have been made to use a microscopic point of view. In particular, Klimontovich[7] and Dupree[8] have used the technique that initially refers to the individual particle motion. The singular "exact" distribution functions are exhibited. The basic quantities of the theory consist of averages of products of the singular functions. A hierarchy is then derived for these fluctuation quantities. This hierarchy has a one-to-one correspondence with the BBKGY hierarchy. In a sense, there is nothing microscopic about the approach of Klimontovich and Dupree because the averaging process takes place at a very early stage. On the other hand, in the past few years another approach has been developed that has very striking microscopic features; namely, the test-particle method of Rostoker and Rosenbluth.[9,10] The motion of a single particle and the effect of this particle on the system are considered in this approach. Such a consideration can certainly be described as microscopic. However, there is no obvious connection between this method and the Klimontovich description.

Recently, Dawson and Nakayama[11] have shown how kinetic equations can be developed by starting from approximate equations of motion for the singular "exact" distribution functions. These equations of motion are obtained by an expansion about the straight-line motion of particles. They obtain the basic test-particle results by this approach.

We have approached this problem from a slightly different point of view, consisting of the use of a method developed by Wyld and Fried[12] for the quantum electron gas. This technique is based essentially on the random phase approximation and the Bogoliubov (adiabatic) assumption. In a recent paper[13] we extended this technique to the electron–phonon system. We obtained a kinetic equation for the electrons, an expression for the phonon spectral function, and an expression for the density autocorrelation function. We also indicated the close relationship of the results to the test-particle theory. In this paper we show how the program is carried out in the classical theory. In the first section we

* Present address: University of Colorado, Boulder, Colorado.

[1] N. N. Bogoliubov, "Problems of a Dynamical Theory in Statistical Mechanics," Moscow, 1946 (translated by E. K. Gora, AFCRC-TR-59-235).
[2] I. Prigogine, *Non-Equilibrium Statistical Mechanics* (Interscience Publishers, Inc., New York, 1962).
[3] R. Balescu, Phys. Fluids 3, 52 (1960).
[4] A. Lenard, Ann. Phys. (N.Y.) 10, 390 (1960).
[5] R. L. Guernsey, Phys. Fluids 5, 322 (1962).
[6] T. H. Dupree, Phys. Fluids 4, 696 (1961).
[7] Yu. L. Klimontovich, Zh. Eksperim. i Teor. Fiz. 35, 1276 (1958) [English transl.: Soviet Phys.—JEPT 8, 891 (1959)].
[8] T. H. Dupree, Phys. Fluids 6, 1714 (1963); 7, 923 (1964).
[9] N. Rostoker and M. N. Rosenbluth, Phys. Fluids 3, 1 (1960).
[10] N. Rostoker, Phys. Fluids 7, 479, 491 (1964).
[11] J. Dawson and T. Nakayama, Phys. Fluids 9, 252 (1966); Tech. Rept., Nagoya Plasma Physics Institute (1965).
[12] H. W. Wyld, Jr. and B. D. Fried, Ann. Phys. (N.Y.) 23, 374 (1963).
[13] W. R. Chappell, J. Math. Phys. 7, 1153 (1966).

assume an approximate equation of motion for the singular distribution function. This approximate equation of motion is simply a linearized Vlasov equation for the singular distribution function. We then show that this microscopic equation of motion is equivalent to the approximate macroscopic equation of motion for a homogeneous system. The microscopic equation is then solved as an initial value problem. An asymptotic ($t \to \infty$) expression for the singular distribution function is obtained and then used to calculate correlation functions. The resulting kinetic equation is the well-known Balescu–Lenard equation.

In the next section we show the relationship of this microscopic approach to the test-particle method. The test-particle expression for the two-particle correlation function is derived for a homogeneous system with no external forces and for the more general case of an inhomogeneous system with external forces. The general picture of the plasma as a system of quasi-particles (test particles surrounded by shielding clouds) is seen to arise in a natural way from the microscopic approach. The two-time distribution functions as used in the test-particle theory are introduced and the test-particle results for these are also obtained.

In the last section we obtain a set of approximate microscopic equations that describe a system containing small inhomogeneous and rapidly varying perturbations. The homogeneous and slowly varying part of the system is again described by the microscopic linearized Vlasov equation. These equations are shown to be equivalent to the macroscopic equations for the two-particle correlation functions used by Wu.[14] We consider only a one-component system in this paper. The generalization to a many-component system is straightforward.

MICROSCOPIC VLASOV EQUATION

The singular "exact" one-particle distribution function is

$$\hat{f}_1(\mathbf{r}, \mathbf{v}, t) = \sum_{i=1}^{N} \delta(\mathbf{r} - \mathbf{r}_i(t))\, \delta(\mathbf{v} - \mathbf{v}_i(t)), \qquad (1)$$

where $\mathbf{r}_i(t)$ and $\mathbf{v}_i(t)$ are the position and velocity of the ith particle. This quantity satisfies the equation of motion

$$\frac{\partial}{\partial t} \hat{f}_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{v}\cdot\nabla_{\mathbf{r}} \hat{f}_1(\mathbf{r}, \mathbf{v}, t) - \frac{1}{m} \iint d\mathbf{r}'\, d\mathbf{v}'\, \nabla_{\mathbf{r}} V(\mathbf{r} - \mathbf{r}') \cdot\nabla_{\mathbf{v}} \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \hat{f}_1(\mathbf{r}', \mathbf{v}', t) = 0. \qquad (2)$$

[14] C. S. Wu, J. Math. Phys. 5, 1701 (1964).

We are considering a system of N electrons in a neutralizing, smeared-out background of positive charge. We take N to be very large and neglect effects of order N^{-1}.[15] The above equation is a microscopic Vlasov equation. We obtain macroscopic equations by performing an averaging process. We define the average quantities by removing the carets, i.e.,

$$f_1(\mathbf{r}, \mathbf{v}, t) = \langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \rangle.$$

Thus, the equation of motion for $f_1(\mathbf{r}, \mathbf{v}, t)$ is

$$\frac{\partial}{\partial t} f_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{v}\cdot\nabla_{\mathbf{r}} f_1(\mathbf{r}, \mathbf{v}, t) - \frac{1}{m} \iint d\mathbf{r}'\, d\mathbf{v}'\, \nabla_{\mathbf{r}} V(\mathbf{r} - \mathbf{r}') \cdot\nabla_{\mathbf{v}} \langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \hat{f}_1(\mathbf{r}', \mathbf{v}', t) \rangle = 0, \qquad (3)$$

where $V(\mathbf{r})$ is the coulomb potential e^2/r. We see then that, to obtain a kinetic equation, we want to find an asymptotic ($t \to \infty$) expression for the correlation function

$$\langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \hat{\rho}_1(\mathbf{r}', t) \rangle,$$

where

$$\hat{\rho}_1(\mathbf{r}, t) = \int d\mathbf{v}\, \hat{f}_1(\mathbf{r}, \mathbf{v}, t). \qquad (4)$$

Clearly, a hierarchy can be generated involving correlation functions of the form $\langle \hat{f}_1 \rangle$, $\langle \hat{f}_1 \hat{f}_1 \rangle$, $\langle \hat{f}_1 \hat{f}_1 \hat{f}_1 \rangle$, etc. The relationship between this hierarchy and the BBKGY hierarchy is found by obtaining relations between the correlation functions $\langle \hat{f}_1(1) \hat{f}_1(2) \rangle$, $\langle \hat{f}_1(1) \hat{f}_1(2) \hat{f}_1(3) \rangle$, etc., and the s-particle distribution functions where $s = 1, 2, 3, \cdots$. The first two of these relations are given by

$$\langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \rangle = f_1(\mathbf{r}, \mathbf{v}, t), \qquad (5)$$

$$\langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t) \hat{f}_1(\mathbf{r}', \mathbf{v}', t) \rangle = f_2(\mathbf{r}, \mathbf{v}; \mathbf{r}', \mathbf{v}', t) + \delta(\mathbf{r} - \mathbf{r}')\, \delta(\mathbf{v} - \mathbf{v}') f_1(\mathbf{r}, \mathbf{v}, t), \qquad (6)$$

where

$$\iint d\mathbf{r}\, d\mathbf{v}\, f_1(\mathbf{r}, \mathbf{v}, t) = N,$$

$$\iiiint d\mathbf{r}\, d\mathbf{v}\, d\mathbf{r}'\, d\mathbf{v}'\, f_2(\mathbf{r}, \mathbf{v}; \mathbf{r}', \mathbf{v}', t) = N(N - 1).$$

The approach presented here closely parallels that used by Wyld and Fried,[12] who studied the quan-

[15] Thus, we igonre differences between N and N-1 that arise in the definition of the macroscopic correlation functions in terms of averages of products of the singular distribution functions.[7,11]

tum electron gas. In the quantum case one wants to obtain an asymptotic expression for the correlation function $\langle b_s(\mathbf{k}, \mathbf{p})\rho(-\mathbf{k})\rangle$, where

$$b_s(\mathbf{k}, \mathbf{p}) = c^{\dagger}_{\mathbf{p}-(\frac{1}{2}\mathbf{k}),s}c_{\mathbf{p}+(\frac{1}{2}\mathbf{k}),s}, \tag{7}$$

$$\rho(\mathbf{k}) = \sum_{\mathbf{p},s} b_s(\mathbf{k}, \mathbf{p}). \tag{8}$$

The quantities $c^{\dagger}_{\mathbf{p}s}$ and $c_{\mathbf{p}s}$ are the creation and annihilation operators for the electrons. Wyld and Fried[12] calculated correlation functions by obtaining an approximate equation of motion for the operator $b_s(\mathbf{k}, \mathbf{p})$, which was then solved as an initial value problem. An asymptotic expression was then obtained and used to calculate asymptotic correlation functions. The approximate equation of motion for $b_s(\mathbf{k}, \mathbf{p})$ was obtained by the Random Phase Approximation (RPA). The RPA is a widely used assumption, and practically the only assumption available having a direct connection to the microscopic description.

It is quite simple to construct the analogous classical equations. The operator $b_s(\mathbf{k}, \mathbf{p})$ is the Fourier transform of the Wigner distribution operator, which in turn is the natural (via the Weyl prescription) quantum generalization of the singular function $\hat{f}_1(\mathbf{r}, \mathbf{v}, t)$.[16] Thus, we find that the analogous approximate equation of motion for $\hat{f}_1(\mathbf{r}, \mathbf{v}, t)$ is given by

$$\frac{\partial}{\partial t}\hat{f}_1(\mathbf{r}, \mathbf{v}, t) + \mathbf{v}\cdot\nabla_r\hat{f}_1(\mathbf{r}, \mathbf{v}, t) - \frac{\nabla_v F(\mathbf{v}, t)}{m} \cdot\nabla_r \iint d\mathbf{r}'\, d\mathbf{v}'\; V(\mathbf{r} - \mathbf{r}')\hat{f}_1(\mathbf{r}', \mathbf{v}', t) = 0, \tag{9}$$

where we have assumed the system to be homogeneous so that we can write

$$\langle \hat{f}_1(\mathbf{r}, \mathbf{v}, t)\rangle = F(\mathbf{v}, t) = f_1(\mathbf{r}, \mathbf{v}, t). \tag{10}$$

Equation (9) is basically the linearized Vlasov equation.

The idea of the method is then to use Eq. (9) to obtain an approximate asymptotic expression for $\hat{f}_1(\mathbf{r}, \mathbf{v}, t)$, which we can then use to calculate various correlation functions such as that needed for the kinetic equation. It is difficult to justify this process because we are dealing with highly singular quantities. The approach developed by Dawson and Nakayama[11] has the same problem. On the other hand, it is easy to show the connection of this assumption to the assumptions made in the usual macroscopic treatments by using Eq. (9) to calculate the equation of motion for the two-particle correlation function $g_2(1, 2)$ where

$$g_2(1, 2) = \langle \hat{f}_1(1)\hat{f}_1(2)\rangle - F(\mathbf{v}_1)F(\mathbf{v}_2) - \delta(1 - 2)F(\mathbf{v}_1), \tag{11}$$

where

$$(1) = (\mathbf{r}_1, \mathbf{v}_1).$$

We find that

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_{r_1} + \mathbf{v}_2\cdot\nabla_{r_2}\right)g_2(1, 2) - \frac{1}{m}\nabla_{v_1}F(\mathbf{v}_1)\cdot\int d(3)\;\nabla_{r_1}V(\mathbf{r}_1 - \mathbf{r}_3)g_2(2, 3) - \frac{1}{m}\nabla_{v_2}F(\mathbf{v}_2)\cdot\int d(3)\;\nabla_{r_2}V(\mathbf{r}_2 - \mathbf{r}_3)g_2(1, 3) = \frac{1}{m}\nabla_{r_1}V(\mathbf{r}_1 - \mathbf{r}_2)\cdot(\nabla_{v_1} - \nabla_{v_2})F(\mathbf{v}_1)F(\mathbf{v}_2), \tag{12}$$

where we have used the Bogoliubov assumption, which says that we calculate $g_2(1, 2)$ by assuming $F(\mathbf{v})$ to be time independent. The above equation is the usual equation of motion arrived at in treatments involving truncation of the heirarchy.[4] This equation has been used by several authors to obtain the Balescu–Lenard equation. We see, therefore, that Eq. (9) is completely equivalent to the Balescu–Lenard equation. A similar relationship in the quantum case was found by Wyld and Fried.[12]

We solve Eq. (9) by the use of Fourier–Laplace transforms. Defining the quantity

$$h(\mathbf{k}, \mathbf{v}, \omega) = \int_0^{\infty} dt \int d\mathbf{r}\, e^{-i(\mathbf{k}\cdot\mathbf{r} - \omega t)}\hat{f}_1(\mathbf{r}, \mathbf{v}, t), \tag{13}$$

we obtain

$$-i(\omega - \mathbf{k}\cdot\mathbf{v})h(\mathbf{k}, \mathbf{v}, \omega) = \hat{f}_1(\mathbf{k}, \mathbf{v}, 0) + \frac{iV(\mathbf{k})}{m}\mathbf{k}\cdot\nabla_v F(\mathbf{v})\int d\mathbf{v}'\; h(\mathbf{k}, \mathbf{v}', \omega), \tag{14}$$

where Im $\omega > 0$. We have used the Bogoliubov (adiabatic) assumption by neglecting the time dependence of $F(\mathbf{v})$. From Eq. (6) we obtain

$$\rho(\mathbf{k}, \omega) = \int d\mathbf{v}\; h(\mathbf{k}, \mathbf{v}, \omega) = i\int \frac{d\mathbf{v}\; \hat{f}_1(\mathbf{k}, \mathbf{v}, 0)}{\epsilon(\mathbf{k}, \omega)(\omega - \mathbf{k}\cdot\mathbf{v})}, \tag{15}$$

where

$$\epsilon(\mathbf{k}, \omega) = 1 + \frac{1}{m}V(\mathbf{k})\int \frac{d\mathbf{v}\;\mathbf{k}\cdot\nabla_v F(\mathbf{v})}{\omega - \mathbf{k}\cdot\mathbf{v}}. \tag{16}$$

[16] W. E. Brittin and W. R. Chappell, Rev. Mod. Phys. **34**, 620 (1962).

We can now calculate an asymptotic expression for $\rho(\mathbf{k}, t)$. Clearly, the time dependence arises from the poles in $\epsilon(\mathbf{k}, \omega)$. For a stable plasma all of the poles of $\epsilon(\mathbf{k}, \omega)$ are in the lower half-plane and decay in time. We assume that for sufficiently long times we can neglect these poles. We then find that the only pole is at $\omega = \mathbf{k}\cdot\mathbf{v}$; thus[17]

$$\rho(\mathbf{k}, t) \sim \int \frac{d\mathbf{v}\, \hat{f}_1(\mathbf{k}, \mathbf{v}, 0)e^{-i\mathbf{k}\cdot\mathbf{v}t}}{\epsilon^+(\mathbf{k}, \mathbf{k}\cdot\mathbf{v})}, \tag{17}$$

where

$$\epsilon^+(\mathbf{k}, \mathbf{k}\cdot\mathbf{v}) = \epsilon(\mathbf{k}, \mathbf{k}\cdot\mathbf{v} + i\eta).$$

The quantity η is the positive infinitesimal. We henceforth drop the superscript $+$. We solve for $\hat{f}_1(\mathbf{k}, \mathbf{v}, t)$ in a self-consistent manner by first substituting the above expression for $\rho(\mathbf{k}, t)$ into Eq. (9) and then looking for a solution consistent with Eq. (17). The resulting expression is

$$\hat{f}_1(\mathbf{k}, \mathbf{v}, t) = \hat{f}_1(\mathbf{k}, \mathbf{v}, 0)e^{-i\mathbf{k}\cdot\mathbf{v}t} + \int d\mathbf{v}'\, P^*(\mathbf{k}, \mathbf{v}' \mid \mathbf{v})\hat{f}_1(\mathbf{k}, \mathbf{v}', 0)e^{-i\mathbf{k}\cdot\mathbf{v}'t}, \tag{18}$$

where

$$P^*(\mathbf{k}, \mathbf{v}' \mid \mathbf{v}) = -\frac{V(\mathbf{k})\mathbf{k}\cdot\nabla_{\mathbf{v}}F(\mathbf{v})}{m\epsilon(\mathbf{k}, \mathbf{k}\cdot\mathbf{v}')[\mathbf{k}\cdot(\mathbf{v}' - \mathbf{v}) + i\eta]}. \tag{19}$$

The quantity $P(\mathbf{k}, \mathbf{v} \mid \mathbf{v}')$ was introduced because it is the same quantity that appears in the test-particle theory as the measure of the system's response to the test particle.[10] Clearly, the first term in Eq. (18) represents the straight-line motion of particles. The second term might be considered as the dressing cloud that accompanies the electrons.

In order to obtain a kinetic equation, we need to calculate the quantity

$$\mathrm{Im}\, \langle \hat{f}_1(\mathbf{k}, \mathbf{v}, t)\rho(-\mathbf{k}, t)\rangle$$

since

$$\frac{\partial F(\mathbf{v}, t)}{\partial t} = \frac{1}{m(2\pi)^3}\nabla_{\mathbf{v}}\cdot\int d\mathbf{k}\, V(-\mathbf{k})\mathbf{k} \times \mathrm{Im}\, \langle \hat{f}_1(\mathbf{k}, \mathbf{v}, t)\rho(-\mathbf{k}, t)\rangle. \tag{20}$$

[17] Although we have been ignoring the differences between the finite and infinite systems, it is important at this point that we consider the system to be finite in order to make the singularity at $\omega = \mathbf{k}\cdot\mathbf{v}$ a simple pole. We do this by assuming we have used periodic boundary conditions. After the calculation of the correlation functions we then let the volume become infinite. A rigorous discussion of these limiting procedures is given by G. G. Emch, University of Maryland Tech. Note BN-423 (1965).

With the use of Eqs. (17) and (18) and the identity

$$\mathrm{Im}\, [\epsilon(\mathbf{k}, \omega)]^{-1} = -\frac{\pi V(\mathbf{k})}{m\,|\epsilon(\mathbf{k}, \omega)|^2} \times \int d\mathbf{v}\, \mathbf{k}\cdot\nabla_{\mathbf{v}}F(\mathbf{v})\, \delta(\omega - \mathbf{k}\cdot\mathbf{v}), \tag{21}$$

we obtain

$$\frac{\partial F(\mathbf{v}, t)}{\partial t} = \frac{\pi}{(2\pi)^3 m^2}\int d\mathbf{k}\int d\mathbf{v}'\, \mathbf{k}\cdot\nabla_{\mathbf{v}}\left|\frac{V(\mathbf{k})}{\epsilon(\mathbf{k}, \mathbf{k}\cdot\mathbf{v})}\right|^2 \times \mathbf{k}\cdot(\nabla_{\mathbf{v}} - \nabla_{\mathbf{v}'})F(\mathbf{v}, t)F(\mathbf{v}', t)\, \delta(\mathbf{k}\cdot\mathbf{v} - \mathbf{k}\cdot\mathbf{v}'). \tag{22}$$

The above equation is the familiar Balescu–Lenard equation.[3,4]

We can also use these results to calculate the density autocorrelation function $\langle \rho(\mathbf{k}, t)\rho(-\mathbf{k}, t')\rangle$. We find that

$$\langle \rho(\mathbf{k}, t)\rho(-\mathbf{k}, t')\rangle = \int d\mathbf{v}\, \frac{F(\mathbf{v})e^{i\mathbf{k}\cdot\mathbf{v}(t'-t)}}{|\epsilon(\mathbf{k}, \mathbf{k}\cdot\mathbf{v})|^2}, \tag{23}$$

which is the usual result.

We note that in the process of letting the volume Ω become infinite, the normalized distribution function $\phi(\mathbf{v}, t)$ must be introduced. The relationship between $\phi(\mathbf{v}, t)$ and $F(\mathbf{v}, t)$ is given by

$$\phi(\mathbf{v}, t) = \frac{\Omega}{N}F(\mathbf{v}, t) \underset{N\to\infty}{\longrightarrow} n^{-1}F(v, t).$$

Consequently, $\phi(\mathbf{v}, t)$ satisfies the normalization condition

$$\int d\mathbf{v}\, \phi(\mathbf{v}, t) = 1.$$

TEST PARTICLES

Dawson and Nakayama[11] have indicated the close connection of the microscopic approach and the test-particle method. We note in particular that Eq. (18) lends itself to the interpretation that the system can be described as a superposition of bare particles dressed by screening clouds.

We can make an even closer identification with the test-particle theory by obtaining the test-particle result for the two-particle correlation function. We can use Eq. (18) to calculate the quantity $\langle \hat{f}_1(\mathbf{k}, \mathbf{v}, t)\, \hat{f}_1(-\mathbf{k}, \mathbf{v}', t)\rangle$, which is related to $g_2(\mathbf{k}, \mathbf{v}, \mathbf{v}', t)$. If we neglect the initial correlations, we obtain

$$g_2(\mathbf{k}, \mathbf{v}, \mathbf{v}', t) = P^*(\mathbf{k}, \mathbf{v}' \mid \mathbf{v})F(\mathbf{v}', t) + P(\mathbf{k}, \mathbf{v} \mid \mathbf{v}')F(\mathbf{v}, t) + \int d\mathbf{v}''\, P^*(\mathbf{k}, \mathbf{v}'' \mid \mathbf{v})P(\mathbf{k}, \mathbf{v}'' \mid \mathbf{v}')F(\mathbf{v}'', t). \tag{24}$$

This equation was also obtained by Dawson and Nakayama.[11]

We can also obtain the much more general relationship that applies for inhomogeneous systems in the presence of external fields. We consider a more general linearized microscopic Vlasov equation

$$\frac{\partial}{\partial t}\hat{f}_1(X, t) + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{f}_1(X, t) + \frac{\mathbf{F}_M}{m}\cdot\nabla_{\mathbf{v}}\hat{f}_1(X, t) + \frac{e\mathbf{E}(\mathbf{r}, t)}{m}\cdot\nabla_{\mathbf{v}}f(X, t) = 0, \tag{25}$$

where $X = (\mathbf{r}, \mathbf{v})$,

$$\mathbf{F}_M(X, t) = \mathbf{F}_{ex}(X, t) - e^2\int dX'\,\nabla_{\mathbf{r}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}f_1(X', t), \tag{26}$$

$$\mathbf{E}(\mathbf{r}, t) = -e\int dX'\,\nabla_{\mathbf{r}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}[\hat{f}_1(X') - f_1(X')]. \tag{27}$$

Again we use $\hat{f}_1$ to denote the microscopic distribution and f_1 to denote the macroscopic distribution function.

We next define the quantity $P(X' \mid Xt)$ by the equation

$$\hat{f}_1(X, t) = \hat{f}_1^0(X, t) + \int dX'\,P(X' \mid Xt)[\hat{f}_1^0(X') - f_1(X')], \tag{28}$$

where $\hat{f}_1^0$ obeys the equation of motion

$$[\partial\hat{f}_1^0(X, t)/\partial t] + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{f}_1^0(X, t) + (1/m)\mathbf{F}_M(X, t)\cdot\nabla_{\mathbf{v}}\hat{f}_1^0(X, t) = 0 \tag{29}$$

and the initial condition

$$\hat{f}_1^0(X, 0) = \hat{f}_1(X, 0). \tag{30}$$

If the expression for $\hat{f}_1$ given by Eq. (28) is substituted into Eq. (25), we obtain the following equation of motion for $P(X' \mid Xt)$:

$$\frac{\partial}{\partial t}P(X' \mid Xt) + \mathbf{v}\cdot\nabla_{\mathbf{r}}P(X' \mid Xt) + \frac{\mathbf{F}_M(X, t)}{m}\cdot\nabla_{\mathbf{v}}P(X' \mid Xt) + \mathbf{v}'\cdot\nabla_{\mathbf{r}}P(X' \mid Xt) + \frac{\mathbf{F}_M(X, t)}{m}\cdot\nabla_{\mathbf{v}'}P(X' \mid Xt) - \frac{e^2}{m}\nabla_{\mathbf{v}}f_1(X, t)\cdot\int dX''\,\nabla_{\mathbf{r}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}P(X' \mid X''t) - \frac{e^2}{m}\nabla_{\mathbf{v}}f_1(X, t)\cdot\nabla_{\mathbf{r}}\frac{1}{|\mathbf{r}-\mathbf{r}'|} = 0. \tag{31}$$

We can again calculate the two-particle correlation function by using Eq. (28). We find

$$g_2(X, X', t) = P(X' \mid Xt)f_1(X', t) + P(X \mid X't)f_1(X, t) + \int dX''\,P(X'' \mid Xt)P(X'' \mid X't)f_1(X'', t). \tag{32}$$

Equations (31) and (32) are identical to test-particle results.[10] We also note that we can write

$$\hat{f}_1^0(X, t) = \sum_{i=1}^{N}\delta(\mathbf{r} - \mathbf{r}_i^0(t))\,\delta(\mathbf{v} - \mathbf{v}_i^0(t)), \tag{33}$$

where

$$\dot{\mathbf{r}}_i^0(t) = \mathbf{v}_i(t), \qquad \dot{\mathbf{v}}_i^0(t) = (1/m)\mathbf{F}_M(X, t), \tag{34}$$

$$\mathbf{r}_i^0(0) = \mathbf{r}_i(0), \qquad \mathbf{v}_i^0(0) = \mathbf{v}_i(0). \tag{35}$$

The above formulation is seen to be closely related to the microscopic approach of Dawson and Nakayama,[11] which consists of an expansion about the zero-order orbits. If we substitute the expression given by Eq. (33) into Eq. (28) we find for homogeneous systems

$$\hat{f}_1(X, t) = \hat{f}_1^0(X, t) + \sum_{i=1}^{N}P(X_i^0 \mid Xt). \tag{36}$$

The above form describes the plasma as a superposition of N "bare" particles accompanied by shielding clouds.

It is often useful to consider autocorrelation functions of the form

$$\langle A(\mathbf{r}, t)B(\mathbf{r}', t')\rangle,$$

where $A(\mathbf{r}, t)$ and $B(\mathbf{r}', t')$ are observables that have the form

$$A(\mathbf{r}, t) = \sum_{i=1}^{N}a(X_i \mid \mathbf{r}), \qquad B(\mathbf{r}, t) = \sum_{i=1}^{N}b(X_i \mid \mathbf{r}). \tag{37}$$

An example of such an observable is the density

$$\rho(\mathbf{r}, t) = \sum_{i=1}^{N}\delta(\mathbf{r} - \mathbf{r}_i(t)). \tag{38}$$

Because of the property of the delta function we can write

$$A(X, t) = \int dX'\,a(X \mid X')\hat{f}_1(X', t). \tag{39}$$

Thus we reduce the calculation of all autocorrelation functions to the calculation of

$$\langle\hat{f}_1(X, t)\hat{f}_1(X', t')\rangle.$$

This two-time correlation function can be related to the two-time correlation functions, W_{11} and W_{12}, that are used in the test-particle theory.[10] We do this by noting that the structure of the correlation function $\langle \hat{f}_1(X, t)\hat{f}_1(X', t)\rangle$ is given by

$$\langle \hat{f}_1(X, t)\hat{f}_1(X', t)\rangle = f_1(X, t)f_1(X', t) + g_2(X, X', t) + \delta(X - X')f_1(X, t). \quad (40)$$

The last term in the above equation contains the self-correlations. Such self-correlations are also present in the two-time correlation function. We denote these self-correlations by $nW_{11}(1, t; 2', t')$, where n is the particle density. The remaining correlations are then included in $W_{12}(1, t; 2', t')$ and we write

$$\langle \hat{f}_1(X, t)\hat{f}_1(X', t')\rangle = n^2W_{12}(X, t; X', t') + nW_{11}(X, t; X', t'). \quad (41)$$

Clearly, W_{11} and W_{12} are given by

$$W_{11}(X, t; X', t') = n^{-1}\sum_{i=1}^{N}\langle \delta(X - X_i(t))\,\delta(X' - X_i(t'))\rangle, \quad (42)$$

$$W_{12}(X, t; X', t') = n^{-2}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\langle \delta(X - X_i(t))\,\delta(X' - X_j(t'))\rangle. \quad (43)$$

We can calculate the equation of motion for $W_{12}(X, t; X', t')$ as a function of t' from the equation of motion for $\langle \hat{f}_1(X, t)\hat{f}_1(X', t')\rangle$ by keeping only those terms that have no self-correlations in the variables Xt and $X't'$. Thus we write

$$(\partial/\partial t')W_{12}(X, t; X', t') = n^{-2}(\partial/\partial t')\langle \hat{f}_1(X, t)\hat{f}_1(X', t')\rangle|_{\text{no self-correlation}}. \quad (44)$$

If we use the microscopic linearized Vlasov equation [Eq. (25)], we obtain for a homogeneous system

$$\left(\frac{\partial}{\partial t'} + \mathbf{v}'\cdot\nabla_{\mathbf{r}'}\right)W_{12}(X, t; X't') + \frac{\mathbf{F}_M(X', t')}{m}\cdot\nabla_{\mathbf{v}'}W_{12}(X, t; X', t') = \frac{1}{n^2m}\nabla_{\mathbf{v}'}F(\mathbf{v}')\cdot\nabla_{\mathbf{r}'} \times \int dX''\, V(\mathbf{r}' - \mathbf{r}'')\langle \hat{f}_1(X, t)\hat{f}_1(X'', t')\rangle. \quad (45)$$

Making use of Eq. (41), we obtain

$$\left(\frac{\partial}{\partial t'} + \mathbf{v}'\cdot\nabla_{\mathbf{r}'}\right)W_{12}(X, t; X't') + \frac{\mathbf{F}_M}{m}(X', t')\cdot\nabla_{\mathbf{v}'}W_{12}(X, t; X', t') - \frac{1}{m}\nabla_{\mathbf{v}'}F(\mathbf{v}')\cdot\nabla_{\mathbf{r}'} \times \int dX''\, V(\mathbf{r}' - \mathbf{r}'')W_{12}(X, t; X'', t'') = \frac{1}{nm}\nabla_{\mathbf{v}'}F(\mathbf{v}')\cdot\nabla_{\mathbf{r}'} \times \int dX''\, V(\mathbf{r}' - \mathbf{r}'')W_{11}(X, t; X'', t'). \quad (46)$$

The initial condition is clearly given by

$$W_{12}(X, t; X', t) = f_1(X, t)f_1(X', t) + g_2(X, X', t). \quad (47)$$

The equation of motion for $W_{11}(X, t; X', t')$ can be calculated in a similar manner. In this case we obtain

$$\left[\frac{\partial}{\partial t'} + \mathbf{v}'\cdot\nabla_{\mathbf{r}'} + \frac{F_M(X', t')}{m}\cdot\nabla_{\mathbf{v}'}\right] \times W_{11}(X, t; X', t') = 0. \quad (48)$$

Because of the self-correlation in Xt and $X't'$, the initial condition becomes

$$W_{11}(X, t; X', t) = \delta(X - X')f_1(X, t). \quad (49)$$

HIGHER-ORDER MICROSCOPIC EQUATION

We have seen that for a homogeneous system the microscopic linearized Vlasov equation given by Eq. (9) is equivalent to the usual approximate equation assumed for the two-particle correlation function [Eq. (12)]. In this section we present a microscopic equation describing a system that can have spatial inhomogeneities and that can vary on a fast time scale. We assume the system to be composed of two parts (this is the same model as discussed by Wu[14]): a homogeneous and slowly varying main body and small perturbations that can be inhomogeneous and quickly varying.

We assume that we can write

$$\hat{f}_1(\mathbf{r}, \mathbf{v}, t) = \hat{F}(\mathbf{r}, \mathbf{v}, t) + \hat{f}'(\mathbf{r}, \mathbf{v}, t), \quad (50)$$

where $\hat{F}$ corresponds to the main body of the system and $\hat{f}'(\mathbf{r}, \mathbf{v}, t)$ corresponds to the small perturbations.

Thus, we assume that relative to $\hat{F}$ the quantity $\hat{f}'$ has order ϵ, where ϵ is a small number characterizing the perturbation. If we then substitute Eq. (37) into Eq. (2), we obtain, to order ϵ, the following equations of motion:

$$\frac{\partial \hat{F}(\mathbf{r}, \mathbf{v}, t)}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{F}(\mathbf{r}, \mathbf{v}, t) - \frac{1}{m}\nabla_{\mathbf{v}}\hat{F}(\mathbf{r}, \mathbf{v}, t)\cdot\iint d\mathbf{r}'\, d\mathbf{v}' \times \nabla_{\mathbf{r}}V(\mathbf{r}-\mathbf{r}')\hat{F}(\mathbf{r}', \mathbf{v}', t) = 0, \quad (51)$$

$$\frac{\partial \hat{f}'(\mathbf{r}, \mathbf{v}, t)}{\partial t} + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{f}'(\mathbf{r}, \mathbf{v}, t) = -\frac{1}{m}\iint d\mathbf{r}'\, d\mathbf{v}'\, \nabla_{\mathbf{r}}V(\mathbf{r}-\mathbf{r}') \cdot\nabla_{\mathbf{v}}[\hat{F}(\mathbf{r}, \mathbf{v}, t)\hat{f}'(\mathbf{r}', \mathbf{v}', t) + \hat{f}'(\mathbf{r}, \mathbf{v}, t)\hat{F}(\mathbf{r}', \mathbf{v}', t)]. \quad (52)$$

We again linearize these equations by assuming that we can replace $\hat{F}(\mathbf{r}, \mathbf{v}, t)\hat{F}(\mathbf{r}', \mathbf{v}', t)$ by

$$\hat{F}(\mathbf{r}, \mathbf{v}, t)F(\mathbf{v}', t) + F(\mathbf{v}, t)\hat{F}(\mathbf{r}', \mathbf{v}', t)$$

and $\hat{F}(\mathbf{r}, \mathbf{v}, t)\hat{f}'(\mathbf{r}', \mathbf{v}', t)$ by

$$F(\mathbf{v}, t)\hat{f}'(\mathbf{r}', \mathbf{v}', t) + \hat{F}(\mathbf{v}, t)f'(\mathbf{r}', \mathbf{v}', t).$$

By linearization we mean that a product of two singular distribution functions is replaced by sums of products of only one singular distribution function and a smooth distribution function. The quantities $F(\mathbf{v}, t)$ and $f'(\mathbf{r}, \mathbf{v}, t)$ are the macroscopic distribution functions for the main body and the perturbations, respectively. They are defined as the averages of $\hat{F}(\mathbf{r}, \mathbf{v}, t)$ and $\hat{f}'(\mathbf{r}, \mathbf{v}, t)$, respectively. We are, in essence, doing an expansion in the discreteness parameter (which gives the linearization) and an expansion in the perturbation parameter ϵ.

The fact that $f'(\mathbf{r}, \mathbf{v}, t)$ can have a spatial variation introduces a macroscopic electric field into the problem. We can generalize the problem to include a small external electric field, $\mathbf{E}_{ex}(\mathbf{r}, t)$ (this allows us to calculate the high-frequency conductivity). We then find the approximate microscopic equations of motion to be given by

$$\frac{\partial}{\partial t}\hat{F}(\mathbf{r}, \mathbf{v}, t) + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{F}(\mathbf{r}, \mathbf{v}, t) - \frac{1}{m}\nabla_{\mathbf{v}}F(\mathbf{v}, t)\cdot\iint d\mathbf{r}'\, d\mathbf{v}' \times \nabla_{\mathbf{r}}V(\mathbf{r}-\mathbf{r}')\hat{F}(\mathbf{r}', \mathbf{v}', t) = 0, \quad (53)$$

$$\frac{\partial}{\partial t}\hat{f}'(\mathbf{r}, \mathbf{v}, t) + \mathbf{v}\cdot\nabla_{\mathbf{r}}\hat{f}'(\mathbf{r}, \mathbf{v}, t) + \frac{e}{m}\mathbf{E}(\mathbf{r}, t)\cdot\nabla_{\mathbf{v}}\hat{F}(\mathbf{r}, \mathbf{v}, t) - \frac{1}{m}\nabla_{\mathbf{v}}F(\mathbf{v}, t)\cdot\iint d\mathbf{r}'\, d\mathbf{v}' \times \nabla_{\mathbf{r}}V(\mathbf{r}-\mathbf{r}')\hat{f}'(\mathbf{r}', \mathbf{v}', t) = \frac{1}{m}\nabla_{\mathbf{v}}f'(\mathbf{r}, \mathbf{v}, t)\cdot\iint d\mathbf{r}'\, d\mathbf{v}' \times \nabla_{\mathbf{r}}V(\mathbf{r}-\mathbf{r}')\hat{F}(\mathbf{r}', \mathbf{v}', t), \quad (54)$$

where

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_{ex}(\mathbf{r}, t) - e\iint d\mathbf{r}'\, d\mathbf{v}'\, \nabla_{\mathbf{r}}\frac{1}{|\mathbf{r}-\mathbf{r}'|}f'(\mathbf{r}', \mathbf{v}', t). \quad (55)$$

We then define the two-particle correlation functions for the main body and perturbations as $G_2(\mathbf{r}-\mathbf{r}', \mathbf{v}, \mathbf{v}', t)$ and $g'_2(\mathbf{r}, \mathbf{v}; \mathbf{r}', \mathbf{v}', t)$, where

$$G_2(\mathbf{r}-\mathbf{r}', \mathbf{v}, \mathbf{v}', t) = \langle\hat{F}(\mathbf{r}, \mathbf{v}, t)\hat{F}(\mathbf{r}', \mathbf{v}', t)\rangle - F(\mathbf{v}, t)F(\mathbf{v}', t) - \delta(\mathbf{r}-\mathbf{r}')\,\delta(\mathbf{v}-\mathbf{v}')F(\mathbf{v}, t), \quad (56)$$

$$g'_2(\mathbf{r}, \mathbf{v}; \mathbf{r}', \mathbf{v}', t) = \langle\hat{F}(\mathbf{r}, \mathbf{v}, t)\hat{f}'(\mathbf{r}', \mathbf{v}', t)\rangle + \langle\hat{f}'(\mathbf{r}, \mathbf{v}, t)\hat{F}(\mathbf{r}', \mathbf{v}', t)\rangle. \quad (57)$$

The approximate equations assumed for $\hat{F}$ and $\hat{f}'$ can be used in conjunction with the above equations to obtain the corresponding equations of motion for G_2 and g_2. These equations are given by

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_{\mathbf{r}_1} + \mathbf{v}_2\cdot\nabla_{\mathbf{r}_2}\right)G_2(1, 2) - \frac{1}{m}\nabla_{\mathbf{v}_1}F(\mathbf{v}_1)\cdot\int d(3)\,\nabla_{\mathbf{r}_1}V(\mathbf{r}_1-\mathbf{r}_3)G_2(2, 3) - \frac{1}{m}\nabla_{\mathbf{v}_2}F(\mathbf{v}_2)\cdot\int d(3)\,\nabla_{\mathbf{v}_2}V(\mathbf{r}_2-\mathbf{r}_3)G_2(1, 3) = \frac{1}{m}\nabla_{\mathbf{r}_1}V(\mathbf{r}_1-\mathbf{r}_2)\cdot(\nabla_{\mathbf{v}_1}-\nabla_{\mathbf{v}_2})F(\mathbf{v}_1)F(\mathbf{v}_2), \quad (58)$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_1\cdot\nabla_{\mathbf{r}_1} + \mathbf{v}_2\cdot\nabla_{\mathbf{r}_2}\right)g'_2(1, 2) - \frac{1}{m}\nabla_{\mathbf{v}_1}F(\mathbf{v}_1)\cdot\int d(3)\,\nabla_{\mathbf{r}_1}V(\mathbf{r}_1-\mathbf{r}_3)g'_2(2, 3) - \frac{1}{m}\nabla_{\mathbf{v}_2}F(\mathbf{v}_2)\cdot\int d(3)\,\nabla_{\mathbf{r}_2}V(\mathbf{r}_2-\mathbf{r}_3)g'_2(1, 3) + \frac{e}{m}\mathbf{E}(\mathbf{r}_1, t)\cdot\nabla_{\mathbf{v}_1}G_2(1, 2) + \frac{e}{m}\mathbf{E}(\mathbf{r}_2, t)\cdot\nabla_{\mathbf{v}_2}G(1, 2)$$

$$= \frac{1}{m} \nabla_{v_1} f'(1) \cdot \int d(3)\, \nabla_{r_1} V(\mathbf{r}_1 - \mathbf{r}_3) G_2(2, 3)$$

$$+ \frac{1}{m} \nabla_{v_2} f'(2) \cdot \int d(3)\, \nabla_{r_2} V(\mathbf{r}_2 - \mathbf{r}_3) G_2(1, 3)$$

$$+ \frac{1}{m} \nabla_{r_1} V(\mathbf{r}_1 - \mathbf{r}_2) \cdot (\nabla_{v_1} - \nabla_{v_2})$$

$$\times [f'(1) F(2) + F(1) f'(2)]. \tag{59}$$

The above equations are identical to the set of equations used by Wu[14] (when generalized to a multicomponent system) to obtain Guernsey's results and an expression for the high-frequency conductivity. Equations (38) and (39) are to be used in conjunction with the following equations for F and f':

$$\frac{\partial}{\partial t} F(1) = \frac{1}{m} \int d(2)\, \nabla_{r_1} V(\mathbf{r}_1 - \mathbf{r}_2) \cdot \nabla_{v_1} G_2(1, 2), \tag{60}$$

$$\frac{\partial}{\partial t} f'(1) + \mathbf{v}_1 \cdot \nabla_{r_1} f'(1) + \frac{e}{m} \mathbf{E}(1) \cdot \nabla_{v_1} F(\mathbf{v}_1)$$

$$= \frac{1}{m} \int d(2)\, \nabla_{r_1} V(\mathbf{r}_1 - \mathbf{r}_2) \cdot \nabla_{v_1} g_2'(1, 2). \tag{61}$$

DISCUSSION

We have presented a method for deriving asymptotic expressions for correlation functions by constructing approximate equations of motion for the singular distribution functions. This approach is essentially the classical analog of a method used by Wyld and Fried[12] for the quantum electron gas. We have shown that the Balescu–Lenard equation and the test-particle picture are obtained very simply by the microscopic approach. We have also generalized the method to obtain higher-order equations that describe inhomogeneities and high-frequency effects.

The central approximation in the lower-order equations is the random phase approximation in the sense of replacing $\hat{f}_1(1)\hat{f}_1(2)$ by $f_1(1)\hat{f}_1(2) + \hat{f}_1(1)f_1(2)$. This well-known approximation suffers from an almost complete lack of rigor. Of course, any approximation made in equations for highly singular quantities like $\hat{f}_1(1)$ are difficult to justify. This problem must be considered as the most severe criticism of a microscopic approach such as the one presented here and that developed by Dawson and Nakayama.[11] The fact that the microscopic linearized Vlasov equation [Eq. (9)] and the higher-order equation [Eq. (39)] are equivalent to well-known macroscopic equations [Eqs. (12) and (44)] indicates that the approximations made here and those made in the macroscopic developments are equivalent. It is hoped that a better understanding of the assumptions used in both approaches will soon be obtained. In a future paper we will introduce the full Maxwell equations into the problem and obtain kinetic equations for a plasma with radiation.

Note added in proof: In the derivation of Eq. (32) we have dropped terms of the order of the square of the plasma parameter. The author would like to express his gratitude to Dr. C. S. Wu for pointing out an error in the original expression for Eq. (25). The author would also like to mention two relevant papers that appeared while this manuscript was in press. These are D. A. Tidman, T. J. Birmingham, J. Dawson, and T. Nakayama, Phys. Fluids **9**, 1881 (1966), and J. Price, Ph.D. Dissertation, University of California (1966).

ACKNOWLEDGMENTS

The author is grateful to Professor M. Krook of Harvard University and Dr. E. H. Klevans of the Jet Propulsion Laboratory for helpful discussions.

Translational Invariance Properties of a Finite One-Dimensional Hard-Core Fluid*

HARVEY S. LEFF and MICHAEL H. COOPERSMITH
Department of Physics, Case Institute of Technology, Cleveland, Ohio
(Received 10 June 1966)

A formalism is developed for expressing the n-particle distribution functions $D_n(x_1 \leq x_2 \leq \ldots \leq x_n)$ explicitly in terms of the configurational partition function for one-dimensional fluids with hard-core repulsive and nearest-neighbor attractive forces. The translational invariance properties of the D_n functions are investigated for the case of no attractive forces when the system is *finite*. When the number density is less than half the close packing density, there exists a central region in which $D_1(x)$ is constant and all the D_n functions, $n \geq 2$, are functions of the $(n-1)$ nearest-neighbor separation distances. Several relevant theorems are proved and limiting cases are investigated.

I. INTRODUCTION

A MAIN objective in the theory of equilibrium statistical mechanics is the determination of equations of state for systems of physical interest. The common procedures for achieving this entail either the evaluation of an appropriate partition function (canonical or grand canonical) or the determination of the two-particle distribution function and subsequent use of the virial theorem of statistical mechanics.[1] These methods are completely equivalent and both involve making a connection between mathematical entities of statistical mechanics and the functions of classical thermodynamics for bulk materials. Therefore, *the systems under consideration must be uniform.* For liquids and gases, in the limit of very large systems and in the absence of external fields, this property is usually assumed. However, if the system is of *finite* size, the container walls introduce external fields and the degree of uniformity must evidently depend on the density of particles. Furthermore, due to the presence of attractive interparticle forces, this uniformity may also depend on the temperature.

Statistical mechanics provides a direct method of checking the degree of uniformity, namely, by examining the translational invariance properties of the 1, 2, $\cdots$, n-particle distribution functions. However, except for the trivial case of the ideal gas, one seldom sees an exact calculation of these distribution functions for specific examples, and a rigorous proof of uniformity does not appear to exist. The standard text and reference books on statistical mechanics pass over this point rather hastily. They usually state as factual, without proof, that the single-particle distribution function is a constant except for a negligible region near the walls of the container.[2] Similarly, the n-particle functions are stated as being functions of $(n-1)$ vectors which specify the positions of $(n-1)$ of the particles relative to the nth particle. This again depends on being far from the container walls, but the precise meaning of "far" is lacking. Fisher[3] has recently emphasized that boundary effects persist even in the thermodynamic limit, if one or more particles go to infinity with the walls. Since the existence of uniformity is crucial in making the usual connection between statistical mechanics and thermodynamics, it is desirable to find concrete examples which are amenable to calculation.

A possible approach is to derive and ultimately solve integral equations for the distribution functions.[4] Unfortunately, practical considerations in carrying this out force one to make assumptions which are no more acceptable than assuming uniformity itself. We therefore abandon this type of approach and adopt the philosophy of seeking models which may be oversimplified with respect to reality, but which are mathematically tractable. Toward this end, we consider what is probably the most simple, yet nontrivial, model in equilibrium statistical mechanics: A gas of hard-core particles in a one-dimensional container. The canonical partition function for this system can be found exactly and has been investigated by a number of authors.[5]

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

[1] J. O. Hirshfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (John Wiley & Sons, Inc., New York, 1954), pp. 132–137.

[2] See, for example, T. L. Hill, *Statistical Mechanics* (McGraw-Hill Book Company, Inc., New York, 1956), p. 182; R. Kubo, *Statistical Mechanics* (John Wiley & Sons, Inc., New York, 1965), p. 312; I. Z. Fisher, *Statistical Theory of Liquids* (University of Chicago Press, Chicago, 1964), p. 42; J. E. Mayer, *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1958), Vol. 12, p. 153.

[3] M. E. Fisher, J. Math. Phys. **6**, 1643 (1965).

[4] J. G. Kirkwood and E. Monroe, J. Chem. Phys. **9**, 514 (1941).

[5] K. F. Herzfeld and M. Goeppert-Mayer, J. Chem. Phys. **2**, 38 (1934); L. Tonks, Phys. Rev. **50**, 955 (1936); H. Takahasi, Proc. Phys. Math. Soc. (Japan) **24**, 60 (1942); G. Rushbrooke and H. Ursell, Proc. Cambridge Phil. Soc. **44**, 263 (1948); F. Gursey, *Ibid.*, **46**, 182 (1950).

Salsburg, Zwanzig, and Kirkwood[6] have presented a formalism for evaluating the n-particle distribution functions for one-dimensional fluids with nearest-neighbor forces. In particular, they have shown that the one-particle distribution function $D_1(x)$ approaches the number density if $x \to \infty$ *after* the thermodynamic limit has been taken. Similarly, they have shown the two-particle function to be a function of $|x_2 - x_1|$ for $x_1, x_2 \to \infty$ *after* the thermodynamic limit has been taken.

It is the purpose of this paper to investigate the translational invariance properties of the n-particle distribution functions for the pure hard-core system (no attractive forces). The investigation is carried out for a *finite* number of particles N in a container of *finite* length L. We show that the effects of the container persist up to a distance $(N-1)d$ from each end of the interval, where d is the hard-core diameter. If $L > 2(N-1)d$, then there is a *central* region in which $D_1(x)$ is constant and each n-particle distribution is a function of the $(n-1)$ nearest-neighbor separation distances. These translational invariance properties do not hold outside the central region and also do not hold if $L < 2(N-1)d$.

In Sec. II, we develop a formalism which relates the n-particle distribution functions directly to the canonical configurational partition function for one-dimensional hard-core systems with nearest-neighbor attractive forces. The method differs from that of Salsburg, Zwanzig, and Kirkwood.[6] In Sec. III, this formalism is applied to the one-particle function $D_1(x)$ for the pure hard-core system. Two general theorems are proved and several limiting cases are discussed. In Sec. IV, the results for $D_1(x)$ are used to establish the translational invariance properties of the n-particle functions $D_n(x_1, \cdots, x_n)$. Major results are stated in the form of theorems. Section V consists of a summary and discussion of the results.

II. n-PARTICLE DISTRIBUTION FUNCTIONS AND THE PARTITION FUNCTION

In this section, it is shown that the n-particle distribution functions for a one-dimensional gas with nearest-neighbor forces are simply related to the canonical partition function. Specifically, we assume a two-body interaction potential energy $w(x)$ which consists of a hard-core repulsive part $w_h(x)$ plus a nearest-neighbor attractive part $w_a(x)$, x is the separation distance between the two particles.

$$w(x) = w_h(x) + w_a(x), \tag{1a}$$

$$w_h(x) = \begin{cases} \infty, & x \le d, \\ 0, & \text{otherwise.} \end{cases} \tag{1b}$$

$$-A \le w_a(x) \le 0, \qquad A \ge 0, \tag{1c}$$

$$w_a(x) = 0, \qquad x \ge 2d. \tag{1d}$$

Equation (1c) ensures the existence of the integrals occurring in the partition function and (1d) is the restriction of nearest-neighbor forces. The canonical partition function for this system, contained in the interval $[0, L]$, is

$$Q(L, \beta, N) = \left(\frac{2\pi m}{\beta}\right)^{\frac{1}{2}N} \times \int_0 \cdots \int dx_1 \cdots dx_N \, W_N(x_1, \cdots, x_N). \tag{2}$$

Here, $\beta = 1/kT$, k is Boltzmann's constant, T is the absolute (Kelvin) temperature, x_i is the position coordinate for particle i, $\mathcal{O}$ is the domain of integration $0 \le x_1 \le x_2 \le \cdots \le x_N \le L$, and

$$W_N(x_1, \cdots, x_N) = \prod_{i<j} \exp\left[-\beta w(|x_i - x_j|)\right]. \tag{3}$$

Equation (2) can also be written with an extra $1/N!$ factor and with each x-integral ranging from zero to L, since W_N is symmetric under pairwise interchange of x_i and x_j for all $i \ne j$.

The one-particle distribution function $D_1(x)$ is defined as

$$D_1(x) = \left\langle \sum_{i=1}^{N} \delta(x - y_i) \right\rangle, \tag{4}$$

where $\langle\ \rangle$ denotes an average using the multivariate distribution function

$$P_N(y_1, \cdots, y_N) = (2\pi m/\beta)^{\frac{1}{2}N}[W_N(y_1, \cdots, y_N)/N!\, Q(L, \beta, N)] \equiv W_N(y_1, \cdots, y_N)/Z(L, N). \tag{5}$$

$Z(L, N) \equiv (2\pi m/\beta)^{-\frac{1}{2}N} Q(L, \beta, N) N!$ is the configurational partition function. The β dependence has been suppressed for notational convenience. In carrying out the average, each y_i integral runs from zero to L. Because of the symmetry of $P_N(y_1, \cdots, y_N)$, (4) can also be written as $D_1(x) = N\langle\delta(x - y_1)\rangle$. The two-particle distribution function is defined as

$$D_2(x_1, x_2) = \langle \sum_{i \ne j} \delta(x_1 - y_i)\, \delta(x_2 - y_j)\rangle, \tag{6}$$

and a straightforward generalization leads to the definition

$$D_n(x_1, \cdots, x_n) = \langle \sum_{\{k_i\}}{}' \delta(x_1 - y_{k_1}) \cdots \delta(x_n - y_{k_n})\rangle \tag{7}$$

[6] Z. W. Salsburg, R. W. Zwanzig, and J. G. Kirkwood, J. Chem. Phys. 21, 1098 (1953).

for $1 \leq n \leq N$. The primed summation in (7) indicates a sum over $k_1, \cdots, k_n$ excluding all terms for which $k_i = k_j$, $i \neq j$. Combining (2), (5), and (7), it is clear that D_n is normalized to $N!/(N-n)!$. Since D_n is symmetric under permutations of the $\{x_i\}$, it is sufficient to examine

$$D_n(x_1 \leq x_2 \leq \cdots \leq x_n) \equiv D_n^0(x_1, \cdots, x_n). \tag{8}$$

The superscript 0 implies that the variables are ordered. D_n^0 is normalized to $N!/(N-n)!n!$ over the ordered domain $0 \leq x_1 \leq x_2 \cdots \leq x_n \leq L$.

One may write

$$D_n^0(x_1, \cdots, x_n) = \frac{N!}{(N-n)!\,n!} \int_0^L \cdots \int_0^L dy_{n+1} \cdots dy_N \times \frac{W_N(x_1, \cdots, x_n, y_{n+1}, \cdots, y_N)}{Z(L, N)}. \tag{9}$$

Now, divide the interval $[0, L]$ into subintervals $R_1 = [0, x_1]$, $R_2 = [x_1, x_2]$, $\cdots$, $R_{n+1} = [x_n, L]$ and apply a theorem which is proved in Appendix A. Let $N_1, \cdots, N_{n+1}$ denote the numbers of particles in $R_1, \cdots, R_{n+1}$, respectively. Then

$$D_n^0(x_1, \cdots, x_n) = \frac{1}{n!\,Z(L, N)} \sum_{\{N_i\}}{}'' \frac{N!}{\Pi N_i!} \times \int_{N_i \epsilon R_i} \cdots \int dy_{n+1} \cdots dy_N \times W_N(x_1, \cdots, x_n, y_{n+1}, \cdots, y_N). \tag{10}$$

The doubly primed summation signifies that $N_i = 0, 1, \cdots, N$ for $j = 1, 2, \cdots, n+1$, with $\sum N_i = N - n$.

The usefulness of (10) lies in the fact that the integral appearing therein factors into a product of $(n+1)$ configurational partition functions. This is so because the forces are effectively between nearest neighbors only, and the different intervals R_i are separated by fixed particles at $x_1, \cdots, x_n$. The first factor in the product is $\tilde{Z}(x_1, N_1)$, the configurational partition function for N_1 particles in an interval of length x_1, with one extra particle fixed at x_1. The last factor is $\tilde{Z}(L - x_n, N_{n+1})$, where the tilde refers to one extra particle being fixed at x_n. The remaining $(n-1)$ factors are $\prod_{k=2}^{n} \bar{Z}(x_k - x_{k-1}, N_k)$. $\bar{Z}$ denotes a configurational partition function for the case of two extra fixed particles present, one at each end of the interval. Thus,

$$D_n^0(x_1, \cdots, x_n) = \frac{1}{n!\,Z(L, N)} \sum_{\{N_i\}}{}'' \frac{N!}{\Pi N_i!} \tilde{Z}(x_1, N_1) \times \bar{Z}(L - x_n, N_{n+1}) \prod_{k=2}^{N} \bar{Z}(x_k - x_{k-1}, N_k). \tag{11}$$

A similar result was previously derived by Salsburg, Zwanzig, and Kirkwood[6] using Laplace transformation techniques. It is both interesting and useful because it explicitly represents the distribution functions in terms of the configurational partition functions Z, $\tilde{Z}$, and $\bar{Z}$.[7]

III. $D_1(x)$ FOR THE PURE HARD-CORE FLUID

We now concentrate on the special case for which $w_a(x)$ is identically zero, and the forces are entirely due to the hard-core repulsion. For this case (see Appendix B),

$$Z(\tau, m) = [\tau - (m-1)d]^m \theta(\tau - md + d), \tag{12a}$$

$$\tilde{Z}(\tau, m) = [\tau - md]^m \theta(\tau - md), \tag{12b}$$

$$\bar{Z}(\tau, m) = [\tau - (m+1)d]^m \theta(\tau - md - d), \tag{12c}$$

where

$$\theta(s) = \begin{cases} 1, & \text{for } s \geq 0, \\ 0, & \text{for } s < 0. \end{cases} \tag{13}$$

Letting

$$N - 1 \equiv M \tag{14a}$$

and

$$L - Md \equiv \xi, \tag{14b}$$

(11) and (12) yield

$$D_1(x) = \frac{M+1}{\xi^{M+1}} \sum_{n=0}^{M} \binom{M}{n} [x - nd]^n \times [\xi - (x - nd)]^{M-n} \theta(x - nd)\,\theta(\xi - x + nd). \tag{15}$$

The product of θ functions in (15) may be called $\epsilon_n(x)$, having the property

$$\epsilon_n(x) = \begin{cases} 1, & \text{for } nd \leq x \leq \xi + nd, \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

If $L \geq 2Md$ and $Md \leq x \leq L - Md$, $\epsilon_n(x) = 1$ in every term of Eq. (15). If $L \geq 2Md$ and $x < Md$, the upper limit of the summation is less than M. Similarly, if $x > L - Md$, the lower limit of the summation is greater than zero. Specifically, one can write

$$D_1(x) = \frac{M+1}{\xi^{M+1}} \sum_{\sigma(x)}^{\gamma(x)} \binom{M}{n} (x - nd)^n \times [\xi - (x - nd)]^{M-n}, \tag{17}$$

[7] If two extra particles were initially fixed at the ends of [0,L] to provide the "walls," then only $\bar{Z}$ functions would appear in (11).

where

$$\gamma(x) = k, \qquad \sigma(x) = 0,$$
$$\text{for} \quad kd \leq x \leq (k+1)d, \tag{18a}$$
$$\gamma(x) = M, \qquad \sigma(x) = k,$$
$$\text{for} \quad L - Md + (k-1)d \leq x \leq L - Md + kd, \tag{18b}$$

with $k = 0, 1, \cdots, M$, and $L \geq 2Md$. For $L < 2Md$, the $\epsilon_n(x)$ functions can make $\sigma(x) > 0$ and $\gamma(x) < M$ *simultaneously*. At the end of this section, we discuss the close packing limit, but otherwise restrict ourselves to the case $L \geq 2Md$.

We now show that the above results lead to a simple physical interpretation. For $L \geq 2Md$, the wall effects of the container manifest themselves up to distances Md from each wall. These *outer* regions can in principle be filled with particles (not simultaneously, of course). In the remainder of the interval the *central* region is completely free of edge effects. For $L < 2Md$ the entire interval $[0, L]$ may be thought of as an *outer* region and edge effects are present throughout. The absence of edge effects in the central region is made explicit by the following theorem.

Theorem I: For $N = M + 1$ hard-core particles of length d in the one-dimensional interval $[0, L]$, where $L \geq 2Md$, $D_1(x)$ is independent of x for $Md \leq x \leq L - Md$ (called the *central* region).

Proof: From (14)–(16),

$$D_1(x) = \frac{M+1}{\xi^{M+1}} \sum_{n=0}^{\infty} \binom{M}{n} (x - nd)^n \sum_{j=0}^{\infty} \binom{M-n}{j} \times (-1)^j (x - nd)^j \xi^{M-n-j}. \tag{19}$$

The second summation is simply $[\xi - (x - nd)]^{M-n}$ and both of the summations are automatically cut off by the binomial coefficients. Changing summation variables, $(j, n) \to (n + j = l, n)$, and using

$$\binom{M}{n}\binom{M-n}{l-n} = \binom{l}{n}\binom{M}{l}, \tag{20}$$

we have

$$D_1(x) = \frac{M+1}{\xi^{M+1}} \sum_{l=0}^{\infty} \binom{M}{l} (-1)^l \xi^{M-l} \times \sum_{n=0}^{l} \binom{l}{n} (-1)^n (x - nd)^l. \tag{21}$$

In order to obtain a power series expansion in x, we write

$$(x - nd)^l = \mathbf{D}_s^l \exp[s(x - nd)]|_{s=0}, \tag{22}$$

where $\mathbf{D}_s$ is the differentiation operator d/ds. It follows that

$$\sum_{n=0}^{l} \binom{l}{n} (-1)^n (x - nd)^l = \mathbf{D}_s^l \{[1 - \exp(-sd)]^l \exp(sx)\}|_{s=0} = d^l l!, \tag{23}$$

and

$$D_1(x) = \frac{M+1}{\xi} \sum_{l=0}^{\infty} \binom{M}{l} l! \left(-\frac{d}{\xi}\right)^l \tag{24}$$

for $Md \leq x \leq \xi$, $L \geq 2Md$. Equation (24) explicitly demonstrates the x independence of $D_1(x)$ in the *central* region, completing the proof of theorem I.

It is a remarkable fact that, when the θ functions of (15) are replaced by unity, the summation is independent of x[8]. We now examine (24) in the thermodynamic limit ($\lim_T$) $M \to \infty$, $L \to \infty$, $M/L = D$, which is finite.

Corollary: For the problem considered in theorem I, $\lim_T D_1(x) = D$ for all values of x in the *central* region.

Proof: Denoting the M dependence of $D_1(x)$ by the notation $D_1^{(M)}(x)$, (24) can be written as

$$D_1^{(M)}(x) \equiv [(M+1)/\xi] G^{(M)}(\eta) = \left(\eta + \frac{1}{\xi}\right) \sum_{l=0}^{M} \frac{M!}{(M-l)!\, M^l} (-\eta d)^l, \tag{25}$$

where $\eta d = Md/\xi = dD(1 - dD)^{-1} < 1$. In the thermodynamic limit, η is fixed and $\lim_T = \lim_{M\to\infty, \eta \text{ fixed}}$. From (25) one has the recursion formula

$$G^{(M+1)}(\eta) = 1 - \eta d \sum_{l=0}^{M} \frac{M!}{(M-l)!\,(M+1)^l} (-\eta d)^l = 1 - \eta d G^{(M)}(\eta) + \eta d \sum_{l=0}^{M} \frac{M!\,(-\eta d)^l}{(M-l)!\, M^l} \times [1 - (1 + M^{-1})^{-l}]. \tag{26}$$

In the thermodynamic limit, the summation vanishes (see Appendix C). Since $\lim_{M\to\infty} G^{(M+1)}(\eta) = \lim_{M\to\infty} G^{(M)}$, one has[9] $\lim_{M\to\infty} G^{(M)}(\eta) = (1 + \eta d)^{-1}$. It follows that

$$\lim_T D_1(x) = D = \text{number density} \tag{27}$$

in the *central* region.

[8] The naive application of the binomial theorem in (15) would yield $D_1(x) = (M+1)/\xi$. This is also x-independent, but is incorrect since the square brackets of (15) contain n.

[9] Alternately, one can approximate $M!/(M-l)!$ by M^l, and let the summation run from zero to infinity in order to get (27). The *rigorous* proof of (27) necessitates the statement that $\lim_{M\to\infty} G^{(M)}(\eta)$ exists. This follows from the fact that $G^{(M)}(\eta)$, for any M, is an alternating series whose successive terms decrease monotonically in absolute value.

We now prove that, in the *outer* region, $D_1(x)$ is *not* constant. From (15), it follows that $D_1(x) = D_1(L - x)$ for all x in $[0, L]$. Thus, it suffices to discuss only the *left outer* region $0 \leq x \leq Md$.

Theorem II: Consider the one-dimensional hard-core system of Theorem I with $L \geq 2Md$. $D_1(x)$ is a continuous function in the *left outer* region $0 \leq x \leq Md$ and has discontinuous lth derivatives at $x = kd$ for $k \leq l \leq M$, $k = 1, 2, \cdots, M$.

Proof: Define subintervals $\mathbf{0}_k$ for which $kd \leq x \leq (k+1)d$, $k = 0, 1, \cdots, M$.[10] Denote $D_1(x)$ by $D_1(x, k)$ for x in $\mathbf{0}_k$. From (17) and (18), the lth derivative of $D_1(x, k)$ can be compared with the lth derivative of $D_1(x, k-1)$ at $x = kd$.

$$\frac{d^l}{dx^l}\left[D_1(x,k) - D_1(x,k-1)\right]\Big|_{x=kd}$$
$$= \frac{M+1}{\xi^{M+1}}\binom{M}{k}\frac{d^l}{dx^l}\left[(x-kd)^k(\xi+kd-x)^{M-k}\right]\Big|_{x=kd}$$
$$= \frac{(M+1)!\,l!}{k!\,(M-l)!\,(l-k)!}(-1)^{l+k}\frac{1}{\xi^{l+1}}. \tag{28}$$

Equation (28) vanishes automatically for $l > M$ and $l < k$. In particular, for $l = 0$, (28) vanishes for $k = 1, 2, \cdots, M$. Thus D_1 is continuous at the boundaries of the $\mathbf{0}_k$ regions. Since it is a polynomial within every such region, it is continuous everywhere. The discontinuous derivatives of $D_1(x)$ in the *outer* regions preclude the possibility of this function being constant. This completes the proof.

Remarks: (a) At $x = kd$, $k = 1, \cdots, M$, the lth derivatives are discontinuous for $k \leq l \leq M$. As one moves further from the walls, the first discontinuity appears in higher derivatives. The wall effects become less severe and the D_1 function becomes less "nonanalytic".

(b) For fixed k and l, the thermodynamic limit of (28) is

$$\lim_T \frac{d^l}{dx^l}\left[D_1(x,k) - D_1(x,k-1)\right]\Big|_{x=kd}$$
$$= \frac{(-1)^{l+k}}{d^{l+1}}\binom{l}{k}\left(\frac{dD}{1-dD}\right)^{l+1}. \tag{29}$$

Thus, the discontinuity in the lth derivative of $D_1(x)$ remains *finite* in the *left outer* region at integral values of x/d greater than or equal to l.[11]

(c) For $(M-1)d \leq x \leq Md$,

$$D_1(x) = \tilde{D}_1(x) - [(M+1)/\xi^{M+1}] \times (Md - x)^M(-1)^M, \tag{30}$$

where $\tilde{D}_1(x)$ refers to the value of $D_1(x)$ in the *central* region, as given by (24). Since $(Md - x)$ is positive and is bounded from above by d, the second term vanishes in the thermodynamic limit. An extension of this argument shows that, in the thermodynamic limit, $D_1(x) = D$ when x is a *finite* distance from the boundary of the *central* region as well as when x is *in* the *central* region.

(d) In the thermodynamic limit, $D_1(x)$ is *not* equal to D for all values of x. For example, for $x = 0$

$$\lim_T D_1(0) = D/(1 - dD) \geq D. \tag{31}$$

The equality holds only for $D = 0$, i.e., in the limit of zero density. The limiting form of $D_1(0)$ given by (31) serves as a specific example of the "wall theorem" of Reiss, Frisch, and Lebowitz.[12,13] This states that the equation of state of the system is $\beta P = D_{\text{wall}}$. If $kd \leq x \leq (k+1)d$ the generalization of (31) is[14]

$$\lim_T D_1(x) = \frac{D}{1-dD}\sum_{n=0}^{k}\frac{1}{n!}\left[\frac{D(x-nd)}{1-dD}\right]^n \times \lim_{M\to\infty}\left[1 - \frac{(x-nd)D}{M(1-dD)}\right]^M$$
$$= \frac{D}{1-dD}\sum_{n=0}^{k}\frac{1}{n!}\left[\frac{D(x-nd)}{1-dD}\right]^n \times \exp\left[-\frac{(x-nd)D}{1-dD}\right]$$
$$\neq D. \tag{32}$$

In general, $D_1(x)$ may be greater than or less than D, depending on x.

To close this section, consider the high-density region $\xi < d$ or $1/d - 1/L < D < 1/d$. The nonzero portions of $\epsilon_n(x)$ and $\epsilon_{n+1}(x)$ do not overlap and (15) predicts that $D_1(x)$ consists of a sequence of

[10] The point $x = kd$ belongs to both $\mathbf{0}_k$ and $\mathbf{0}_{k-1}$, $k = 1, \ldots, M$.

[11] $\lim_T D_1(x)$ could still be constant near these points. For example consider the function $M^{-1}\sin M^2x$, which is zero in the limit $M \to \infty$ but whose first derivative is unbounded in this limit.

[12] H. Reiss, H. L. Frisch, and J. L. Lebowitz, J. Chem. Phys. 31, 369 (1959).

[13] One of us (H.S.L.) would like to thank Professor A. J. F. Siegert for pointing out Ref. 12. See also A. J. F. Siegert and E. Meeron, J. Math. Phys. 7, 741 (1966).

[14] Note that this argument breaks down for $x > (M-s)d$, where s is fixed. Here, the number of terms in the summation increases as the thermodynamic limit is taken, and x must also increase in this limit. This case is treated in the preceding remark (c).

$M + 1$ "spikes", one localized near each value $x = kd, k = 0, 1, \cdots, M$. In the limit of close packing, these spikes become Dirac delta functions, which can be seen as follows. Consider an arbitrary bounded continuous function $f(x)$ which is integrable in $[0, \infty]$. Multiply $f(x)$ by a typical term of (15) and integrate from zero to infinity.

$$\frac{M+1}{\xi^{M+1}}\binom{M}{k}\int_0^\infty dx\, f(x)(x-kd)^k(\xi-x+kd)^{M-k}\epsilon_k(x)$$

$$\underset{\xi\to 0}{\sim}\frac{(M+1)}{\xi^{M+1}}\binom{M}{k}f(kd)\int_0^\xi dy\, y^k(\xi-y)^{M-k}=f(kd). \tag{33}$$

Hence, each term of (15) behaves as a delta function at close packing and

$$D_1(x)=\sum_{k=0}^{M}\delta(x-kd), \tag{34}$$

a result which is intuitively obvious.

IV. D_n FOR THE PURE HARD-CORE FLUID

From (11) and (12), the ordered two-particle distribution function is

$$D_2^0(x_1,x_2)=\frac{M+1}{2!\,\xi^{M+1}}\sum{}''\, M!\,\frac{(x_1-nd)^n}{n!}$$

$$\times\frac{(L-x_2-md)^m}{m!}\,\frac{[x_2-x_1-(l+1)d]^l}{l!}\,\theta_1\theta_2\theta_3. \tag{35}$$

Here, $\theta_1=\theta(x_1-nd)$, $\theta_2=\theta(L-x_2-md)$, $\theta_3=\theta(x_2-x_1-ld-d)$ and the doubly primed summation means n, m, l run from zero to M with $m=M-1-n-l$. This can be rewritten as

$$D_2^0(x_1,x_2)=\frac{(M+1)M}{2!\,\xi^{M+1}}\sum_{l=0}^{\infty}\binom{M-1}{l}$$

$$\times[x_2-x_1-(l+1)d]^l\theta_3\sum_{n=0}^{\infty}\binom{M-1-l}{n}$$

$$\times(x_1-nd)^n[\xi_l^{(2)}-(x_1-nd)]^{M-1-l-n}\theta_1\theta_2, \tag{36}$$

where

$$\xi_l^{(2)}\equiv L-(x_2-x_1)-(M-l-1)d. \tag{37}$$

The summation limits have formally been extended to infinity, but the binomial coefficients provide an automatic cutoff. The second summation can be interpreted as being proportional to a single-particle distribution function for $(M-l)$ particles in an interval $[0, L-(x_2-x_1)]$. According to theorem I, this function is independent of x_1 (but is dependent on x_2-x_1) for fixed l, if

$$L-(x_2-x_1)\geq 2(M-l-1)d \tag{38a}$$

and

$$(M-l-1)d\leq x_1$$

$$\leq L-(x_2-x_1)-(M-l-1)d. \tag{38b}$$

When $x_2-x_1>d$, the $l=0$ term of Eq. (36) is nonzero, and Eqs. (38) must hold for $l=0$ if D_2^0 is to be translationally invariant. Furthermore, if Eqs. (38) hold for $l=0$, they hold for all $l>0$. For $x_2-x_1\geq 0$ and $l=0$, Eqs. (38) reduce to $L\geq 2(M-1)d$ and $(M-1)d\leq x_1\leq x_2\leq L-(M-1)d$. It is therefore possible for D_2^0 to be translationally invariant when the density is somewhat greater than one-half the close packing density, i.e., when D_1 is *not* translationally invariant.

It is clear that a similar argument can be made for $D_3, D_4, \cdots, D_N$. This is true because of the structure of (11). D_n is always expressible as a function of $(x_2-x_1), \cdots, (x_n-x_{n-1})$ and either x_1 or x_n. The latter dependence occurs solely through the two $\tilde{Z}$ factors. One single sum involving these factors can always be isolated, as in the case of D_2 above, and conditions under which this sum depends only on (x_n-x_1) are readily found. If one or more of the x_i coordinates lies outside the resulting central region (which depends on n), then using Theorem II one can show that $(\partial^m/\partial x_i^m)D_n$, holding the nearest-neighbor separations fixed, has discontinuities for certain values of x_i and m. These remarks are made precise by the following theorems.

Theorem III: Consider $(M+1)$ hard-core particles of length d in the interval $[0, L]$. If

$$L\geq 2(M+1-n)d$$

and

$$(M+1-n)d\leq x_1\leq x_2\leq\cdots$$

$$\leq x_n\leq L-(M+1-n)d,$$

then the ordered distribution function $D_n^0(x_1,\cdots,x_n)$ is expressible as a function of $(x_2-x_1), (x_3-x_2), \cdots, (x_n-x_{n-1})$. If

$$L\geq 2\,Md$$

and

$$Md\leq x_1\leq x_2\leq\cdots\leq x_n\leq L-Md,$$

then *all* $M+1$ distribution functions $D_1^0, D_2^0, \cdots, D_{M+1}^0$ are translationally invariant.

Theorem IV: Consider the system of Theorem III, with $L\geq 2(M+1-n)d$. Suppose $0\leq x_1<$

$(M + 1 - n)d$. The ordered distribution function $D_n^0(x_1, \cdots, x_n)$ is expressible as a function of $(x_k - x_{k-1})$, $k = 2, \cdots, n$, and x_1. For fixed values of the $\{x_k - x_{k-1}\}$, D_n^0 is a continuous, nonconstant function of x_1. Similarly, if $x_n > L - (M + 1 - n)d$, D_n^0 is a continuous, nonconstant function of x_1 for fixed values of the $\{x_k - x_{k-1}\}$.

Remarks: (a) In analogy with the discussion of D_2^0, the relevant summation in the proof of Theorem III is

$$\sum_{s=0}^{\infty} \binom{M + 1 - n - l}{s} (x_1 - sd)^s \times [\xi_l^{(n)} - (x_1 - sd)]^{M+1-n-l-s} \theta_1 \theta_2 \qquad (39)$$

with

$$N_1 = s, \quad N_{n+1} = M + 1 - n - \sum_{i=2}^{n} N_i - s, \quad \sum_{i=2}^{n} N_i \equiv l,$$

and

$$\xi_l^{(N)} \equiv L - (x_n - x_1) - (M + 1 - n - l)d.$$

The necessary conditions for translational invariance come from Theorem I, applied to a system of $(M + 2 - n - l)$ particles in an interval $[0, L - (x_n - x_1)]$. As in the case of D_2^0, these conditions must hold for $l = 0$ and thus for all $l > 0$, yielding the conditions stated in Theorem III. The summation in expression (39) will then be a function of $(x_n - x_1)$ which is fully determined by $(x_2 - x_1)$, $(x_3 - x_2)$, $\cdots$, $(x_n - x_{n-1})$.

(b) Suppose $x_1 < (M+1-n)d$ and $L \geq 2(M + 1 - n)d$. The sth derivative of expression (39) with respect to x_1, holding $(x_n - x_1)$ fixed, is discontinuous at $x = kd$ for $k = 1, 2, \cdots, l$ when $s \geq k$. The sth partial derivative of D_n^0 with respect to x_1, holding all nearest-neighbor separation distances fixed, will then have discontinuities at $x = kd$ for $k = 1, 2, \cdots, (M + 1 - n)$ (unless D_n^0 is identically zero, because 2 nearest neighbors are separated by a distance less than d). The discontinuities are expressible as mixed polynomials in $(x_2 - x_1)$, $(x_3 - x_2)$, $\cdots$, $(x_n - x_{n-1})$ and are in general nonzero. This illustrates the nonconstant nature of D_n^0 as a function of x_1, as stated in Theorem IV. If $x_n > L - (M + 1 - n)d$, a similar argument can be made for D_n^0 as a function of x_1 and the $(n - 1)$ nearest-neighbor distances.

V. DISCUSSION

The main results of this paper are the following: (1) $D_n^0(x_1, \cdots, x_n)$ is explicitly expressible in terms of the configurational partition function for a one-dimensional system with hard-core repulsive and nearest-neighbor attractive forces. This result follows directly from the decomposition of a multiple integral and does not require complicated Laplace transform considerations.[6] (2) For the pure hard-core case, $D_1(x)$ is a constant if $L \geq 2Md$ and $Md \leq x \leq L - Md$. This constant approaches the number density in the thermodynamic limit. Outside the latter *central* region $D_1(x)$ is continuous but has derivative discontinuities at $x = kd$, $k =$ integer. These discontinuities occur in higher derivatives far from the walls and remain finite in the thermodynamic limit. (3) In the close packing limit, $D_1(x)$ becomes a sum of Dirac delta functions. (4) The ordered distribution functions $D_n^0(x_1, \cdots, x_n)$ are functions of the nearest-neighbor separation distances, if $L \geq 2(M + 1 - n)d$ and $(M + 1 - n)d \leq x_1 \leq \cdots \leq x_n \leq L - (M + 1 - n)d$. Therefore, D_{n+1}^0 may be translationally invariant, although D_n^0 is not. However, in the *central* region, where $D_1(x)$ is constant, all the $D_n^0(x_1, \cdots, x_n)$ are translationally invariant. This statement is nontrivial for any particular value of n when the *central* region is of length $> (n - 1)d$. These results suggest that finite two- and three-dimensional hard-core systems have translational invariance properties in some interior region.

The general formalism of Sec. II is applicable to all nearest-neighbor one-dimensional fluids with hard-core repulsion. The proof of Theorem I, however, depends very specifically on the fact that $w_a = 0$ and cannot be generalized in an obvious way. It is, of course, of interest to know the effects of attractive forces on the translational invariance properties and, in particular, their dependence on the temperature. An investigation of these points will be contained in a subsequent paper.

APPENDIX A

Consider the integral

$$K_m = \int_0^L \cdots \int_0^L dy_1 \cdots dy_m \, F_m(y_1, \cdots, y_m), \qquad (A1)$$

where F_m is symmetric under permutations of the $\{y_i\}$. Subdivide $[0, L]$ into $(n + 1)$ disjoint subintervals $R_1 = [0, x_1]$, $R_2 = [x_1, x_2]$, $\cdots$, $R_{n+1} = [x_n, L]$, where $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq L$.

$$\int_0^L dy_k = \sum_{i=1}^{n+1} \int_{R_i} dy_k, \qquad k = 1, 2, \cdots, m. \qquad (A2)$$

Denoting the integral operator $\int_{R_i} dy_k$ by $\mathbf{I}_i^k$, we have

$$K_m = \prod_{k=1}^{m} \left(\sum_{j=1}^{n+1} \mathrm{I}_j^k \right) F_m. \tag{A3}$$

The latter product can be expanded in a way which is formally analogous to the multinomial expansion, a typical term being $(\prod_{k=1}^{m} \mathrm{I}_{j_k}^k)$. Here, N_1 of the $\{j_k\}$ are equal to 1, N_2 are equal to 2, and so forth, with $\sum_{i=1}^{n+1} N_i = m$. Due to the symmetry of F_m, each such term appears $(m!/N_1! \cdots N_{n+1}!)$ times, and

$$K_m = \sum_{\{N_i\}}{}'' \frac{m!}{\Pi N_i!} \int \cdots \int_{N_i \in R_i} dy_1 \cdots dy_m \, F_m. \tag{A4}$$

The doubly primed summation takes into account the restriction $\sum N_i = m$. Taking $m = N - n$ and

$$F_m(y_1, \cdots, y_m) = \frac{N!}{(N-n)!\,n!} \frac{W_N(x_1, \cdots, x_n, y_1, \cdots, y_m)}{Z(L, \beta, N)}, \tag{A5}$$

one effectively has $K_m = D_n^0(x_1, \cdots, x_n)$. Application of (A4) then yields Eq. (10) of Sec. II.

APPENDIX B

Consider m hard-core particles of length d, with $w_a \equiv 0$, confined to $[0, \tau]$. The total potential energy is taken as

$$V = \sum_{i<j}^{m} w_h(x_{ij}) + a \sum_{i=1}^{m} w(x_i) + b \sum_{i=1}^{m} w_h(\tau - x_i) + abw_h(\tau). \tag{B1}$$

$a = (1, 0)$ if (one, zero) extra particle is fixed at $x = 0$. $b = (1, 0)$ if (one, zero) extra particle is fixed at $x = \tau$. For any nontrivial case, the last term of (B1) is identically zero. Define

$$Y(\tau, a, b) = m! \int_{\mathcal{O}} \cdots \int dx_1 \cdots dx_m \exp(-\beta V). \tag{B2}$$

$\mathcal{O}$ denotes the ordered domain of integration $0 \leq x_1 \leq \cdots \leq x_m \leq L$. The Laplace transform of $Y(\tau, a, b)$ is

$$\Phi(s, a, b) = \int_0^\infty d\tau \exp(-s\tau) Y(\tau, a, b). \tag{B3}$$

Define the $(m + 1)$ new variables $\sigma_1 = x_1$, $\sigma_2 = x_2 - x_1$, $\cdots$, $\sigma_m = x_m - x_{m-1}$, $\sigma_{m+1} = \tau - x_m$. Clearly, $\tau = \sigma_1 + \sigma_2 + \cdots + \sigma_{m+1}$. The Jacobian of the transformation is unity and each σ_i ranges over $[0, \infty]$. Due to the nearest-neighbor nature of the hard-core potential,

$$\begin{aligned}\Phi(s, a, b) &= m! \int_0^\infty d\sigma_1 \exp[-s\sigma_1 - \beta a w_h(\sigma_1)] \\ &\quad \times \int_0^\infty d\sigma_{m+1} \exp[-s\sigma_{m+1} + \beta b w_h(\sigma_{m+1})] \\ &\quad \times \left\{ \int_0^\infty d\sigma \exp[-s\sigma - \beta w_h] \right\}^{m-1} \\ &= \frac{m!}{s^{m+1}} \exp[-(a+b)sd] \exp[-(m-1)sd].\end{aligned} \tag{B4}$$

Laplace transform inversion gives

$$Y(\tau, a, b) = [\tau - (m-1) - a - b]^m \times \theta(\tau - md + d - a - b). \tag{B5}$$

The connections

$$Y(\tau, 0, 0) = Z(\tau, m), \quad Y(\tau, 1, 1) = \bar{Z}(\tau, m), \quad Y(\tau, 1, 0) = Y(\tau, 0, 1) = \tilde{Z}(\tau, m) \tag{B6}$$

yield Eqs. (12a)–(12c).

APPENDIX C

Although the individual terms in the summation on the right-hand side of Eq. (26) vanish in the limit of large M, this does not guarantee that the whole sum vanishes, since the number of terms increases with M. We now show that the limit of the summation for $M \to \infty$ is actually zero. The sum is first rewritten as follows:

$$\sum_{l=0}^{M} a_l^{(M)} (-x)^l = \sum_{l=0}^{l_0-1} a_l^{(M)} (-x)^l + \sum_{l=l_0}^{M} a_l^{(M)} (-x)^l, \tag{C1}$$

where $x = \eta d$, and the coefficients are given by

$$a_l^{(M)} = [M!/(M-l)!\,M^l][1 - (1 + M^{-1})^{-l}]. \tag{C2}$$

We show that l_0 may be chosen so that the terms in the second sum decrease in magnitude as l increases. Since the terms alternate in sign, the sum is bounded by the magnitude of its first term. In the limit $M \to \infty$, we show that a finite l_0 with the above property exists, and that the bound on the second term vanishes in this limit. The first sum has a finite number of terms each of which vanishes for $M \to \infty$.

In order to find l_0, we look at the ratio of the

absolute values of two successive terms in the sum. Denoting this by $xf_M(l)$, we have

$$f_M(l) = \frac{a_{l+1}^{(M)}}{a_l^{(M)}} = 1 - \frac{l}{M} + \frac{M - l}{M(M+1)} \frac{1}{(1 + M^{-1})^l - 1}. \quad \text{(C3)}$$

It is evident that $f_M(l)$ is infinite at $l = 0$ and decreases monotonically to zero for $l = M$. Thus, there exists an l_0 such that for $l > l_0$, $xf_M(l) < 1$. Also,

$$\lim_{M \to \infty} f_M(l) \equiv f(l) = 1 + 1/l. \quad \text{(C4)}$$

Therefore, in the limit $M \to \infty$, l_0 may be found by setting $xf_M(l_0) = 1$, which gives the result

$$l_0 = x/(1 - x). \quad \text{(C5)}$$

To be precise, l_0 should be chosen as the nearest integer greater than $x/1 - x$. This is finite for $0 \leq x < 1$. The first term of the second sum on the right-hand side of Eq. (C1) is then bounded in magnitude by

$$l_0 x^{l_0}/M,$$

which vanishes for $M \to \infty$. This completes the proof.

Temperateness of the Absorptive Part of the Scattering Amplitude in the Proof of Dispersion Relations

PETER N. DOBSON, JR.
University of Hawaii, Honolulu, Hawaii
(Received 21 February 1966)

A proof is given within the context of the Bremermann, Oehme, and Taylor proof of dispersion relations that the absorptive part of the elastic scattering amplitude remains tempered throughout the relevant region of analytic continuation. The technique of this proof, which avoids the use of the Jost–Lehmann–Dyson representation, may be applicable to some field theories which do not assume local commutativity and for which such a representation cannot, therefore, be constructed.

I. STATEMENT OF THE PROBLEM

IN the Bremermann, Oehme, and Taylor[1] proof of dispersion relations in the framework of axiomatic quantum field theory, we begin with the reduction formulas,

$$M_{r,a}(k+k';p',p) = \pm i\int d^4x \exp\{i\tfrac{1}{2}(k+k')\cdot x_\theta\}(\pm x^0)\langle p'|\times[j(\tfrac{1}{2}x), j(-\tfrac{1}{2}x)]\,|p\rangle, \quad (1)$$

where the $M_{r,a}$ are related to the elastic scattering amplitude for the process $k+p\to k'+p'$ involving scalar particles of equal mass and $j(x)$ is the usual current operator. We introduce the "brick wall" coordinate system in which

$$\tfrac{1}{2}(k+k') = \omega, \{(\omega^2 - E_\Delta^2)^{\frac{1}{2}}\mathbf{e}\}; \qquad p' = p_\Delta, \quad p = p_{-\Delta},$$

where

$$p_\Delta = \{E_\Delta, \boldsymbol{\Delta}\}, \quad \boldsymbol{\Delta}\cdot\mathbf{e} = 0, \qquad E_\Delta = (m^2+\Delta^2)^{\frac{1}{2}}, \quad |\mathbf{e}| = 1.$$

Then (1) becomes

$$M_{r,a}(\omega,\Delta^2) = \pm i\int d^4x \exp\{i\omega x^0 - i(\omega^2 - E_\Delta^2)^{\frac{1}{2}}\mathbf{e}\cdot\mathbf{x}\}\times\theta(\pm x^0)\langle p_\Delta|\,[j(\tfrac{1}{2}x), j(-\tfrac{1}{2}x)]\,|p_{-\Delta}\rangle.$$

Now this relation does not immediately imply any analyticity for the M's. So BOT introduce a new variable β and write

$$M_{r,a}(\omega,\beta,\Delta^2) = \pm i\int d^4x \exp\{i\omega x^0 - i(\omega^2-\beta)^{\frac{1}{2}}\mathbf{e}\cdot\mathbf{x}\}\times\theta(\pm x^0)\langle p_\Delta|\,[j(\tfrac{1}{2}x), (j-\tfrac{1}{2}x)]\,|p_{-\Delta}\rangle.$$

[1] H. J. Bremermann, R. Oehme, and J. G. Taylor, Phys. Rev. 109, 2178 (1958); hereafter referred to as BOT.

Then we see that $M_{r,a}(\omega,\beta,\Delta^2)$ may be continued in ω into the entire upper (for M_r) or lower (for M_a) half-planes, provided we take $\beta<0$ and Δ^2 fixed. Furthermore, for $\beta<-\Delta^2$, M_r and M_a have the same limit as ω goes to a segment of the real axis from the respective half-planes. Under these assumptions, BOT derive the representation

$$M(\omega,\beta,\Delta^2) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{A(\omega',\beta,\Delta^2)\,d\omega'}{\omega'-\omega}, \quad (2)$$

where

$$A(\omega,\beta,\Delta^2) = (1/2i)[M_r(\omega,\beta,\Delta^2) - M_a(\omega,\beta,\Delta^2)]$$

is the absorptive part of the amplitude. In order to get a dispersion relation for the physical amplitude, it is necessary to make an analytic continuation of the absorptive part in β back to the physical value of $\beta=+E_\Delta^2$. In BOT it is shown that this is equivalent to the proof of their Theorem 1, which we paraphrase here as follows.

Theorem 1: We are given four distributions of the four-vector variables y_1, y_2, and y_3.

$$\tilde{g}_{ij}(y_1,y_2,y_3), \qquad i=r,a, \qquad j=r,a.$$

The $\tilde{g}_{ij}$ are assumed to be tempered, invariant under the transformations of the inhomogeneous, orthochronous Lorentz group, and as retarded or advanced in y_1 or y_2, as denoted by the subscripts i, j. The Fourier transforms of the $\tilde{g}_{ij}$,

$$g_{ij}(q_1q_2q_3) = \int d^4y_1\,d^4y_2\,d^4y_3 \times\exp[i(q_1\cdot y_1 + q_2\cdot y_2 + q_3\cdot y_3)]\tilde{g}_{ij}(y_1y_2y_3),$$

are assumed to have the properties,

$$g_{rj} - g_{aj} = 0 \quad \text{for} \quad (q_1+q_3)^2 < 4m^2, \quad (q_1-q_3)^2 < 4m^2, \qquad j=r,a,$$

$$g_{ir} - g_{ia} = 0 \quad \text{for} \quad (q_2+q_3)^2 < 4m^2, \quad (q_2-q_3)^2 < 4m^2, \qquad i=r,a,$$

$$g_{ij} \equiv 0 \quad \text{for} \quad q^2 < 4m^2, \ \text{or} \ q_3^0 < 0, \qquad i,j=r,a.$$

Then there exists a function $\chi(z_1, z_2, z_3, z_4, z_5, z_6)$ of the complex variables $z_1, \cdots, z_5$ and the real variable z_6 with the properties:

(i) For each real z_6, $\chi(z_1, z_2, z_3, z_4, z_5, z_6)$ is holomorphic in $z_1, \cdots, z_5$ in the domain D,

$$D = \{z_1, \cdots, z_5 | \, |z_{1,2} - m^2| < \delta, \quad |z_{3,4} - \gamma| < \delta, \, |z_5 + 4\Delta^2| < \delta\},$$

where $-R \leq \gamma \leq m^2$, for any positive number R, and where δ is some small positive number. We further find:

(ii) For q_1, q_2, q_3 real, and $q_3^0 > 0$, with

$$z_1 = (q_1 + q_3)^2 = p'^2,$$
$$z_2 = (q_2 + q_3)^2 = p^2,$$
$$z_3 = (q_1 - q_3)^2 = k'^2,$$
$$z_5 = (q_1 - q_2)^2 = (p' - p)^2,$$
$$z_6 = 4q_3^2 = (p' + k')^2,$$

and $z_1, \cdots, z_5$ in D, we have the representation

$$g_{ij}(q_1, q_2, q_3) = \chi(z_1, \cdots, z_6).$$

(iii) $\chi \equiv 0$ for $z_6 < 4m^2$.

In the proof of this theorem it is shown that there exists some $g(q_1, q_2, q_3)$ analytic for each fixed real q_3 in the region $W \times W$, where W is the wedge domain,

$$W = \{q | \, |\text{Im } q^0| > |\text{Im } \mathbf{q}|, \quad |\text{Re } q^0| < \infty, \, |\text{Re } \mathbf{q}| < \infty\},$$

and represented by one of the g_{ij} in every part of the region. Furthermore, for Im q in a bounded subset of its region of variation in W, g is of at most polynomial increase in Re q. The proof consists of constructing a portion of the holomorphy envelope of $W \times W$, which is then shown to contain the relevant points to map onto D. However, it is never proven in BOT that the function $g(q_1, q_2, q_3)$ retains its polynomial increase property, its temperateness, at all the new points, and this property is crucial in order for the representation (2) to remain valid during the continuation in β.

In modern treatments of dispersion relations which make use of the Jost–Lehmann–Dyson[2,3] representation, the necessary temperateness follows directly from the representation.[4] However, in discussions of field theories for which local commutativity does not hold, it may be difficult to find an extension of the Jost–Lehmann–Dyson technique, while arguments analogous to those of BOT may still be useful.[5] For this reason, it seems desirable to construct a proof of the extension of temperateness of the absorptive part within the context of the BOT proof.

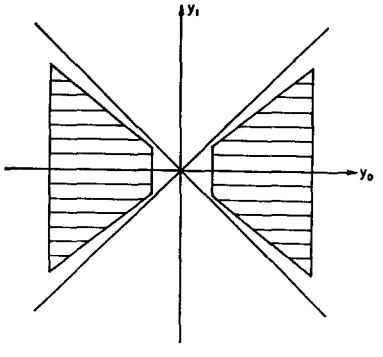

FIG. 1. The wedge domain as seen in the $y_0 - y_1$ plane. The shaded region is an example of a truncated wedge of the type used in this paper.

II. A LEMMA ON TRUNCATED WEDGE DOMAINS

In this section we prove the extension of temperateness into a region resulting from the partial analytic completion of a truncated wedge domain. The basic result is contained in the following lemma, in which for the sake of clarity we deal with a function of only two complex variables.

Lemma: We are given a function $g(z_0, z_1)$ of the two complex variables,

$$z_0 = x_0 + iy_0, \qquad z_1 = x_1 + iy_1,$$

holomorphic in the region,

$$W = \{(z_0, z_1) | \, |y_1| < |y_0|; \, |x_0| < \infty; \, |x_1| < \infty\}.$$

A section of W in the $y_0 - y_1$ plane is shown in Fig. 1. This region is the union of two disconnected domains of holomorphy, in each of which we assume that g can be represented as the Fourier transform of a tempered distribution. g is further assumed to have tempered distributions for boundary values as $y^0 \to 0$. A connection between the portions of W is provided by the hypothesis that there exists a set,

$$E = \{(z_0, z_1) | \, y_0 = y_1 = 0; \quad |x_1| < \infty; \, |x_0| < \eta, \, \eta > 0\},$$

such that if $(z_0^{(n)}, z_1^{(n)}) \in W$ and

$$\lim_{n \to \infty} (z_0^{(n)}, z_1^{(n)}) \in E,$$

[2] R. Jost and H. Lehmann, Nuovo Cimento 5, 1598 (1957).
[3] F. J. Dyson, Phys. Rev. 110, 1460 (1958).
[4] K. Hepp, Helv. Phys. Acta 37, 639 (1964).
[5] For example, see P. N. Dobson, Ph.D. dissertation, University of Maryland (1965).

then

$$\lim_{n\to\infty} g(z_0^{(n)}, z_1^{(n)})$$

exists and depends only on the limit point in E. We now assert that $g(z_0, z_1)$ may be analytically continued into a region H, to be defined in the proof, which contains points not in W and where $|g|$ is bounded by a polynomial in x_0 and x_1. In the next section, we show that the region H is large enough to allow the desired extension of temperateness of the absorptive part of the scattering amplitude.

Proof: By virtue of the hypothesis that g is the Fourier transform of a tempered distribution in W, we have

$$|g(z_0, z_1)| < P(x_0, x_1)$$

as y_0 and y_1 vary over any compact subset of the region $|y_1| < |y_0|$, where $P(x_0, x_1)$ is some polynomial in x_0 and x_1.[6] In particular, it is so bounded in the truncated wedge,

$$W_T(a, b, c) = \{(z_0, z_1) |\ a \leq |y_1| < b\ |y_0| < c;\ |x_0| < \infty, |x_1| < \infty\}, \quad (3)$$

where a, b, and c are positive numbers such that $b < 1$. An example, of such a domain is indicated by the shaded region in Fig. 1.

We now employ the techniques used by BOT in their discussion of the edge-of-the-wedge theorem.[7] Let us consider an arbitrary point in E, say (X_0, X_1), such that $|X_0| < \eta - d$. We now define new variables,

$$w_0 = z_0 - X_0, \qquad w_1 = z_1 - X_1,$$

and the function $h(w_0, w_1) = g(z_0, z_1)$. Consider the analytic plane: $z_0 = \lambda$, $z_1 = \alpha\lambda$, with $\alpha < b < 1$. Then in the λ plane, h is holomorphic everywhere except for cuts which approach no closer than d to the origin. If we expand h in a power series about the origin,

$$h(w_0, w_1) = \sum c_n \lambda^n,$$

the series will converge absolutely and uniformly for $|\lambda| < d$. The number d, it should be noted, was chosen independent of the value of X_1. In the smaller region $|\lambda| < bd$, h is not only holomorphic, but tempered in the sense that if we write

$$|h(w_0, w_1)| < M \quad \text{for} \quad |\lambda| < bd,$$

then the bound M can grow at most as a polynomial in X_1 as we go from one starting point to another. Now the edge-of-the-wedge theorem permits a continuation of h into

$$|w_0| < \tfrac{1}{3}bd, \qquad |w_1| < \tfrac{1}{3}b^2 d.$$

Furthermore, in this region we have

$$|h(w_0, w_1)| < M \sum nd^n = Md/(1 - d)^2$$

for $d < 1$. The right-hand side is just M times a factor independent of X_1. This argument implies that we may certainly continue $g(z_0, z_1)$ into the domain

$$N = \{(z_0, z_1) |\ |x_1| < \infty,\ |x_0| < \eta - \delta;\ |y_0|, |y_1| < \tfrac{1}{4}\delta\}$$

such that for sufficiently small δ, $|g|$ is bounded in N by a polynomial in x_1.

We now consider the domain formed from the union of N and $W_T(\frac{1}{5}\delta, b, c)$, in which $|g|$ is bounded by some polynomial in x_0 and x_1. We take the analytic plane $z_1 = X_1$ (real). Then, in the z_0 plane, the domain of analyticity resulting from $W_T \cup N$ will be the strip B,

$$B = \{z_0 \mid -c < y_0 < c\} - \{z_0 \mid -\tfrac{1}{5}\delta < y_0 < \tfrac{1}{5}\delta;\ \eta - \delta < |x_0| < \infty\}. \quad (4)$$

Using the method of Bremermann,[8] we take a circle centered on x_0, with $|x_0| < \eta - \delta$, and of radius $r < \eta - |x_0| - \delta$. We are concerned with the Euclidean distance from any point on this circle and the boundary of the domain $W_T \cup N$. It is clear that if we choose c large enough, say greater than 2η, this distance will be the same as if the wedge were not truncated. Making this choice of c, we may proceed as in BOT to continue g into the region

$$N' = \{(z_0, z_1) |\ |x_1| < \infty;\ |x_0| < \eta - \delta;\ |y_0|, |y_1| < \tfrac{1}{4}b(\eta - |x_0| - \delta)\}.$$

We assert that g will be tempered in N'. For, in $W_T \cup N$, $|g|$ was bounded by a polynomial of, say, order n in x_0 and order m in x_1. If we take

$$h(z_0, z_1) = g(z_0, z_1)/(z_0 - \alpha)^n (z_1 - \beta)^m \quad (5)$$

with Im α and Im β sufficiently large, we have a function holomorphic in $W_T \cup N$ and bounded by a constant there. The function h may be continued into N', and has the same bound there by virtue

[6] R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964), p. 53.

[7] Reference 6, pp. 74–84.

[8] H. J. Bremermann, Rev. Mat. Hispano—Am. **27**, 175 (1957).

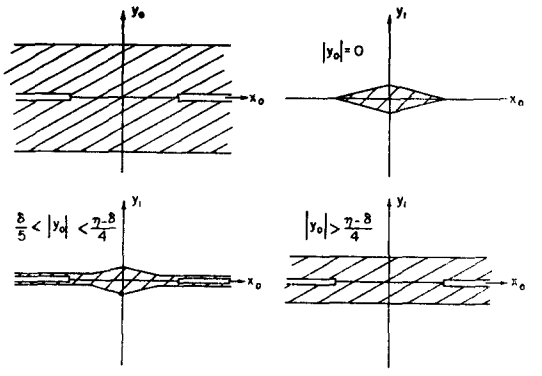

FIG. 2. The semi-tube region $W_T \cup N'$, shown in the $z_0 - y_1$ space, holding x_1 fixed.

of the fact that a holomorphic function of several complex variables takes on no new values in the analytic completion of its domain of holomorphy.[9] The relation (5) holds in N' by analytic continuation, which gives the desired bound on $|g|$. We now have g holomorphic and tempered in the semi-tube $W_T \cup N'$. This region (Fig. 2) may be expressed in the form

$$\{(z_0, z_1) | z_0 \in B;\ |x_1| < \infty,\ m(z_0) \leq |y_1| < M(z_0),$$

where

$$m(z_0) = \begin{cases} 0, & |x_0| < \eta - \delta, \\ \frac{1}{8}\delta, & |x_0| \geq \eta - \delta, \end{cases}$$

$$M(z_0) = b \max \{|y_0|, \tfrac{1}{4}(\eta - |x_0| - \delta)\}.$$

Using standard methods,[10] we may further enlarge this region to

$$H = \{(z_0, z_1) | z_0 \in B;\ |x_1| < \infty;\ m(z_0) \leq |y_1| < \tilde{M}(z_0), \quad (6)$$

where

$$\tilde{M}(z_0) = b\ \mathrm{Im}\,[z_0^2 - (\eta - \delta)^2]^{\frac{1}{2}}$$

By the same reasoning as used in going from N to N', g will still be tempered in H.

III. EXTENSION OF TEMPERATENESS TO THE STRIP D

The lemma developed in the preceding section may now be generalized to the case of a function $g(q_1, q_2, q_3)$ considered as a function of eight complex variables for each fixed real q_3. The extension is not trival, but differs in no significant respect from the procedure outlined in BOT. We simply state the result here.

Under the assumptions of Theorem 1, $g(q_1, q_2, q_3)$ may be analytically continued into the region $H \times H$ in the variables q_1 and q_2, where

$$H = \{q^0 \mid q^0 \in B;\ |\mathrm{Re}\ \mathbf{q}| < \infty;\ m(q^0) \leq |\mathrm{Im}\ \mathbf{q}| < \tilde{M}(q^0)\},$$

which is the generalization of the region (6) for a four-vector variable, and $|g|$ is bounded in $H \times H$ by polynomials in the Re q^α.

We now examine the extent to which the points of D, as defined in Theorem 1, are images of points in $H \times H$. First, with BOT we denote $(q_3)^2 = t$ (real), and note that the parameter η appearing in the definition of H may be taken as $2m - t$. Now, put

$$q_1^0 = q_2^0 = (m^2 - \gamma)/4t,$$

$$\mathbf{q}_1 = \rho(t, \gamma)\mathbf{e}_1 + \Delta\mathbf{e}_2,$$

$$\mathbf{q}_2 = \rho(t, \gamma)\mathbf{e}_1 - \Delta\mathbf{e}_2,$$

where $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$ and

$$\rho^2(t, \gamma) = \{t + (m^2 - \gamma)/4t\}^2 - E_\Delta^2.$$

We are interested in all $\gamma \leq m^2$ and $t \geq m$. If $\rho^2 > 0$, the point (q_1, q_2) is real, and furthermore lies in the set connecting the various parts of $W \times W$. g is holomorphic in this set by the edge-of-the-wedge theorem and tempered there by hypothesis. Thus we need consider only the shaded region of Fig. 3. The requirement that q_1^0 and q_2^0 be in B is simply

$$(m^2 - \gamma)/4t < 2m - t - \delta. \quad (7)$$

Note that this condition is met for any strip width c in the definition of H and B, an indication that truncation of the wedge will prove immaterial for

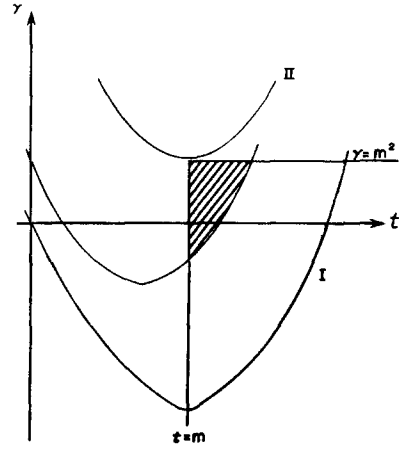

FIG. 3. Region of the $t - \gamma$ plane covered by the domain $H \times H$.

[9] H. Behnke and P. Thullen, *Theorie der Funktionen Mehrerer Komplexer Veränderlichen* (Chelsea Publishing Company, New York, 1934), p. 74.

[10] H. J. Bremermann, Math. Ann. **127**, 406 (1954).

the points in D. Points satisfying (7) lie above the curve I in Fig. 3, which is actually drawn for $\delta = 0$. For sufficiently small δ, all points in the shaded region will satisfy the requirement. The remaining condition that (q, q) lie in $H \times H$ is

$$m(q^0) \leq |\mathrm{Im}\ \mathbf{q}| < \tilde{M}(q^0).$$

For the q^0 under consideration, $|\mathrm{Re}\ q^0| < \eta - \delta$, so $m(q^0) = 0$. Thus we need only satisfy

$$|\mathrm{Im}\ \mathbf{q}| < b\ \mathrm{Im}\ [(q^0)^2 - (2m - t - \delta)^2]^{\frac{1}{2}}$$

or

$$|\mathrm{Im}\ \rho(t, \gamma)| < b\ \mathrm{Im}\ \{[(m^2 - \gamma)/4t]^2 - (2m - t - \delta)^2\}^{\frac{1}{2}}.$$

The relevant points lie below the curve,

$$\left|E_\Delta^2 - \left(t + \frac{m^2 - \gamma}{4t}\right)^2\right| = b^2\left\{(2m - t - \delta)^2 - \left(\frac{m^2 - \gamma}{4t}\right)^2\right\},$$

such as that marked II in Fig. 3. If we put $b = 1/(1 + \epsilon)$, then for ϵ and δ small, and neglecting second order, this curve has its minimum for

$$\gamma_{\min} \geq 3m^2 - 2\Delta^2 - 4m\delta - 4m^2\epsilon.$$

Thus, if $\Delta^2 = m^2 - d^2$ and we choose ϵ and δ so that

$$d^2 \gg 2m\delta + 2m^2\epsilon,$$

then $\gamma_{\min} > m^2$. For any fixed $\Delta^2 < m$, then, we can find some truncated wedge such that all relevant q_1 and q_2 lie in the domain $H \times H$ so defined. At these q_1 and q_2, the function g is not only holomorphic but tempered.

The basic element in the discussion given here is that the points of the strip D may be reached by continuation from a portion of the wedge which is bounded in the imaginary directions. In such a portion, we have polynomial bounds on the absorptive part which may be extended to D, permitting the analytic continuation of the dispersion relation (2), or some subtracted form of it.

ACKNOWLEDGMENTS

The author is greatly indebted to K. Hepp, who not only pointed out the existence of this problem, but also supplied the suggestion on which the arguments of this paper are based.

This work was supported in part by the United States Atomic Energy Commission under Contract AT-(04-3)-511.

Consistency Conditions on Models for High-Energy Scattering*

R. J. EDEN
Cavendish Laboratory, Cambridge, England
(Received 23 July 1966)

From unitarity, analyticity in t, and analyticity in E, a variety of conditions can be derived for the high-energy behavior of a crossing symmetric two-body scattering amplitude $F(E,t)$. These conditions are studied as consistency conditions for a class of models for high-energy scattering based on smoothly varying functions. For this class of models they lead to: (1) the Froissart bound without making explicit use of Martin's enlargement of the Lehmann ellipse, (2) conditions on the phase of the amplitude and its derivative with respect to t in the forward direction. As an example of the use of the phase conditions, it is shown how the Lehmann ellipse can be enlarged to give analyticity in the circle $|t| < R$, where R is fixed. Although the method is different, this result is closely related to the general results of Martin, but with our smoothness assumptions a little more detail can be stated about the behavior of $F(E,t)$.

1. INTRODUCTION

THE basic assumptions in this paper are that the scattering amplitude $F(E, t)$ has the unitarity and analyticity properties that have been derived from quantum field theory.[1,2] From these properties a number of inequalities are derived for the imaginary part of $F(E, t)$ and its derivatives with respect to t at $t = 0$, for large values of the laboratory energy E. Combining these inequalities with a dispersion relation for the derivatives of F at $t = 0$, which has been proved by Martin,[3] and using a smoothness assumption on the total cross section, one obtains a direct derivation of the Froissart bound[4] without explicit use of Martin's enlargement[3] of the Lehmann ellipse.[5]

The main purpose of this paper is to develop methods for studying the consistency of models for high-energy scattering with known or desirable analyticity and unitarity properties. For this purpose we limit our main discussion to a general class of models for which the high-energy behavior of $F(E, t)$ is dominated by terms that do not contain explicitly the normal threshold singularities as $E \to \infty$. In particular, models are considered that are explicitly crossing symmetric and analytic, and for which the high-energy behavior of $F(E, t)$ and its derivatives at $t = 0$ is dominated (or bracketed) by powers of E or (log E). This class of model has been studied by Van Hove,[6] and it is sufficiently general to fit existing experiments and to contain most existing detailed models for high-energy scattering. It is possible to generalize some of the methods presented for smooth functions to include functions with oscillations by use of the theorems of Khuri and Kinoshita.[7]

A condition is obtained on the phase of the coefficient of t^n in the power series expansion of $F(E, t)$ for large E. From this phase condition it is shown that any model of this type must satisfy the Froissart bound and, further, that for large E the amplitude $F(E, t)$ must be analytic in $|t| < R$, where R is a constant. Although these results have been obtained generally by Martin,[3] the approach here is somewhat different from his and it is hoped that it will add to our understanding of the way in which conditions on the amplitude F can be combined to extend our knowledge of its behavior.

The basic assumptions are essentially the same as those used by Martin[3] for enlarging the Lehmann ellipse, and used earlier by Martin[8] and by the author[9] for improving the Greenberg–Low bound.[10] These assumptions are

(1) analyticity of $F(E, t)$ for t in the Lehmann ellipse,

(2) analyticity of $F(E, t)$ for E in the complex plane cut along part of the real axis,

(3) polynomial boundedness for small values of t in the Lehmann ellipse

$$|F(E, t)| < |E|^N, \quad \text{as} \quad E \to \infty,$$

(4) unitarity, particularly Im $f_l(E) \geq 0$ for every partial wave.

* Supported in part by the Air Force Office of Scientific Research, OAR, under grant No. AF EOAR 65-36, with the European Office of Aerospace Research, U. S. Air Force.

[1] J. Bros, H. Epstein, and V. Glaser, Commun. Math. Phys. 1, 240 (1965).
[2] K. Hepp, Helv. Phys. Acta 37, 639 (1964).
[3] A. Martin, CERN Report (1965).
[4] M. Froissart, Phys. Rev. 123, 1053 (1961).
[5] H. Lehmann, Nuovo Cimento Suppl. 14, 153 (1959).
[6] L. Van Hove, Rev. Mod. Phys. 36, 655 (1964).
[7] N. Khuri, and T. Kinoshita, Phys. Rev. 140, B706 (1965).
[8] A. Martin, CERN Report (1965).
[9] R. J. Eden, Phys. Letters 19, 695 (1966); see also R. J. Eden, University of Maryland Physics Department Report 536, (1965).
[10] O. W. Greenberg and F. Low, Phys, Rev. 124, 2047 (1961).

From these assumptions a number of inequalities are derived in Sec. 2 or quoted from earlier work. In Sec. 3 these inequalities are used to derive the Froissart bound, with the extra assumption that σ(total) is smooth (not oscillatory), from a dispersion integral whose convergence has been established by Martin.[3]

In Sec. 4 the form of crossing-symmetric models to be studied is given and phase conditions are investigated. These phase conditions are closely related to the convergence conditions used in Sec. 3. As an illustration of their use with these models it is shown how the Froissart bound $\sigma(\text{total}) \leq (\log E)^2$ can be derived and how the Lehmann ellipse can be enlarged to include $|t| < R$, where R is constant.

2. CONDITIONS ON THE SCATTERING AMPLITUDE

Given analyticity in t of $F(E, t)$ for t inside the Lehmann ellipse, the partial wave expansion is convergent, and

$$\text{Im } F(E, t) = \sum_0^\infty (2l + 1) \text{ Im } f_l(E) P_l(\cos \theta). \quad (2.1)$$

For large E we have

$$\cos \theta \sim 1 + t/E, \quad (2.2)$$

where the unit of mass is taken to be one. From (2.1) we have, for $|\cos \theta| < 1$, $t < 0$,

$$\text{Im } F(E, t) \leq \text{Im } F(E, 0), \quad (2.3)$$

since from unitarity Im $f_l \geq 0$. From polynomial boundedness of F for large E when $t = (t_0/E) > 0$, it follows from (2.1) that

$$\text{Im } f_l(E) < \exp (N \log E - l t_0^{\frac{1}{2}}/E). \quad (2.4)$$

For $L = CE \log E$, we can by choice of C terminate the series (2.1) after L terms with an error less than $E^{-N'}$ for any given N',

$$\text{Im } F(E, t) \sim \sum_0^L (2l + 1) \text{ Im } f_l(E) P_l(\cos \theta). \quad (2.5)$$

We do not make any special use of the choice of L, and could, for example, allow C to tend to infinity as $E \to \infty$. Then the derivatives of Im F would also be approximated by the derivatives of the series (2.5) for large E. Rearranging we get

$$\text{Im } F(E, t) \sim \sum_0^L a_n(E) \frac{t^n}{(n!)^2}. \quad (2.6)$$

From the optical theorem (omitting a constant factor),

$$\text{Im } F(E, 0) \sim E\sigma(E), \quad (2.7)$$

where $\sigma(E)$ denotes the total cross section. We use the symbol $\sim$ to denote the same order of magnitude for large E so that constant factors are in general omitted except for those that change significantly with n.

Let $g(E)$ denote the logarithmic derivative

$$g(E) = (d/dt) \log [\text{Im } F(E, t)]|_{t=0}. \quad (2.8)$$

From (2.1) (2.2) and

$$a_n(E) = (n!)(d^n/dt^n) \text{ Im } F(E, 0) \quad (2.9)$$

we obtain

$$a_n(E) \sim \frac{1}{2^n E^n} \sum_0^L l^{2n+1} \text{ Im } f_l \quad (2.10)$$

provided $n \ll L$. In particular,

$$a_0 \sim E\sigma(E) \sim \sum_0^L l \text{ Im } f_l, \quad (2.11)$$

$$a_1 \sim E\sigma(E)g(E) \sim \frac{1}{2E} \sum_0^L l^3 \text{ Im } f_l. \quad (2.12)$$

Using Cauchy's inequality for the product of two series, we obtain

$$a_n(E) \geq E\sigma[g(E)]^n. \quad (2.13)$$

Writing

$$a_n = E\sigma g^n a_n'(E), \quad (2.14)$$

we obtain from Cauchy's inequality

$$a_{n+1}'/a_n' \geq a_n'/a_{n-1}' \geq 1. \quad (2.15)$$

Following Martin and MacDowell[11] [or directly from (2.11) and (2.12)] a bound can be obtained on $\sigma = \sigma(\text{total})$ for given $g(E)$ by minimizing $a_1(E)$ when σ and $\sigma(\text{el. } im)$ are regarded as given fixed quantities, where

$$\sigma(\text{el. im}) = \sum_0^L l(\text{Im } f_l)^2. \quad (2.16)$$

This gives for large E

$$\sigma = \sigma(\text{total}) \leq 2g(E), \quad (2.17)$$

and from (2.13)

$$a_n(E) = (n!)(d^n/dt^n) \text{ Im } F(E, 0) \geq E[\sigma(E)]^n 2^{-n}. \quad (2.18)$$

An upper bound can also be obtained directly from (2.5) using Im $f_l \leq 1$,

$$E(2E \log^2 E)^{n+1} \geq a_n(E). \quad (2.19)$$

[11] A. Martin and S. W. MacDowell, Phys. Rev. **135**, B960 (1965).

The factors 2^n are not normally required in our discussion of orders of size for large E and so are omitted in the following. If the Martin[3] enlargement of the Lehmann ellipse was assumed one would lose the factor E^{n+1} on the left of (2.19).

The inequalities (2.13) and (2.18) provide a general condition on the polynomial expansion of Im $F(E, t)$ which must be satisfied by any scattering amplitude. We combine this with other conditions which are of several types. One is the simple use of unitarity and the properties of Legendre polynomials as in (2.3). Writing

$$F_n(E, t) = (d^n/dt^n)F(E, t) \tag{2.20}$$

we have $a_n = (n!)$ Im $F_n(E, 0)$. Differentiating (2.1) we obtain for $-E < t < 0$,

$$\text{Im } F_n(E, t) < \text{Im } F_n(E, 0), \qquad n = 0, 1, 2, \cdots . \tag{2.21}$$

We also have, since Im $f_l \geq 0$,

$$0 < \text{Im } F_n(E, 0), \qquad n = 0, 1, 2, \cdots . \tag{2.22}$$

This condition is stated more strongly in our inequality (2.13).

Another type of condition that we use is a consequence of (2.21) and (2.22) combined with a dispersion relation for $F(E, t)$. This condition was derived by Martin[3] as an intermediate step in his enlargement of the Lehmann ellipse, and it is stated and used in the next section. In the following section we use analyticity and crossing symmetry to obtain conditions on the phase of $F_n(E, 0)$.

3. MARTIN CONDITION AND THE FROISSART BOUND

In the forward direction, $F(E, 0)$ satisfies a dispersion relation with no more than two subtractions.[8,9]

$$F(E, 0) = A + BE + \frac{E^2}{\pi}\int_m^\infty \frac{\text{Im } F(E', 0)\, dE'}{E'^2(E' - E)} + \frac{E^2}{\pi}\int_{-\infty}^{-m} \frac{\text{Im } F(E', 0)\, dE'}{E'^2(E' - E)}. \tag{3.1}$$

From (2.3) it follows for E in $(-m, m)$ that $F(E, t)$ satisfies a similar dispersion relation with only two subtractions when $t < 0$. Using (1) the holomorphy of $F(E, t)$ inside a circle of finite radius $|t| < R$, when $-m < E < m$, (2) the regularity in t of $F(E', t)$ at $t = 0$ for finite E', (3) the inequalities (2.21) and (2.22), Martin[3] shows that there is a similar dispersion relation for the nth derivative of $F(E, t)$ at $t = 0$. In particular he shows that

$$\frac{n!}{R^n} > \frac{E^2}{\pi}\int_m^\infty \frac{dE'\ \text{Im } F_n(E', 0)}{E'^2(E' - E)}, \tag{3.2}$$

where F_n is defined by (2.20). The left-hand side of this inequality is obtained from holomorphy of $F(E, t)$ for $|t| < R$. For simplicity we have written Martin's result in the form (3.2) with only one integral on the right-hand cut. A closely analogous inequality holds[3] if F is an amplitude that is crossing symmetric; for example, it could be the sum of the π^+, p and the π^-, p scattering amplitudes. Properties of symmetric amplitudes are stated in more detail in the next section.

We now use the inequality (2.13) in (3.2) to give

$$\frac{n!}{R^n} > \frac{E^2}{\pi}\int_m^\infty \frac{dE'\ \sigma(E')[g(E')]^n}{(n!)E'(E' - E)}. \tag{3.3}$$

If $\sigma(E')$ and $g(E')$ are smooth functions as $E' \to \infty$, one obtains the Froissart bound directly from (3.3). Thus for $g(E) \sim E^\alpha$, $0 < \alpha < 1$, the integral is not even convergent when $(n + 1)\alpha \geq 1$ so the inequality is clearly violated. For

$$g(E) \sim (\log E)^\beta \quad \text{as} \quad E \to \infty, \tag{3.4}$$

we consider the value of n that maximizes,

$$(\log E)^{\beta n}/(n!)^2 \sim [(\log E)^\beta/n^2]^n. \tag{3.5}$$

This is maximum near $n \sim \log E$, and one finds that the corresponding integral (3.3) is dominated by E' near e^n. The inequality $\beta \leq 2$ follows, giving for smooth functions

$$\sigma(\text{total}) < g(E) \leq (\log E)^2, \tag{3.6}$$

which is the Froissart bound.

The above argument does not exclude possible oscillations that exceed the bound, even for large E. This possibility is, of course, excluded by Martin's enlargement of the Lehmann ellipse, and it can probably be excluded also by an extension of the phase discussion in the next sections.

4. PHASE CONDITIONS USING SMOOTH FUNCTIONS

In this section we develop the analog of Martin's convergency condition (3.2) in terms of the phase of derivatives of the scattering amplitude. We use a crossing symmetric amplitude F and limit our discussion to smooth functions whose rate of growth as $E \to \infty$ obeys a power law in E or in $\log E$. For more general situations, including the possibility of oscillatory terms, a number of theorems have

been established by Khuri and Kinoshita[7] which relate bounds on the phase to bounds on the rate of growth of crossing symmetric functions. At least some of these more general situations can be studied in the present context but, although they are important for completeness, they complicate the discussion and are not included here.

The use of the Phragmen–Lindeloff theorems for determining the phase of (smooth) crossing symmetric amplitudes has been developed by Van Hove[6] following earlier work by Meiman.[12] We only outline the main results that we require here. Crossing symmetry for the forward amplitude, combined with Hermitian analyticity gives

$$F(E \exp i\pi, 0) = F^*(E, 0), \tag{4.1}$$

where both amplitudes are evaluated above the branch cuts along the real E axis. This relation determines the phase of $F(E, 0)$ for any given rate of growth with E, by using the Phragmen–Lindeloff theorem for a function regular in the upper half-plane, continuous on the real axis and bounded. Thus if

$$|F(E, 0)/E^\alpha| \to C \quad \text{as} \quad E \to \infty, \tag{4.2}$$

then from (4.1) for large E,

$$F(E, 0) \sim \pm C(-iE)^\alpha. \tag{4.3}$$

Conversely, given that the phase of $F(E, 0)$ is $(-\frac{1}{2}\alpha\pi)$ with $0 < \alpha < 2$ and that for large E,

$$\text{const} < F(E, 0) < E^2 \tag{4.4}$$

it follows that F has the form given in (4.3). The ambiguity in the sign of $F(E, 0)$ can be resolved by using the inequality $\text{Im } F > 0$.

The corresponding form of a crossing symmetric function with a logarithmic power growth together with a single power of E is

$$F(E, 0) \sim iE(\log E - \tfrac{1}{2}i\pi)^\beta. \tag{4.5}$$

It should be noted in particular that, from a given rate of growth for the imaginary part of F, we can deduce the phase and hence obtain the form for the real part. There are analogous results when non-smooth functions are considered.[7] From bounds on $\text{Im } F$ one can obtain bounds on $\text{Re } F$.

In the nonforward direction, crossing symmetry for fixed t has to be expressed in terms of a different variable from E. For equal masses this crossing variable is

$$W = 2mE + \tfrac{1}{2}t - 2m^2. \tag{4.6}$$

[12] N. N. Meĭman, Zh. Eksperim i Teor. **43**, 2277 (1962) [English transl.: Soviet Phys.—JETP **16**, 1609 (1963)].

Writing

$$F(E, t) = G(W, t), \tag{4.7}$$

crossing symmetry gives

$$G(-W + i0, t) = G^*(W + i0, t), \tag{4.8}$$

where we write

$$W \exp i\pi = -W + i0. \tag{4.9}$$

This relation is exact, like (4.1), and not merely asymptotic. It is valid for each fixed t. We can use it to construct models that are crossing symmetric [through satisfying (4.8)] for each t and analytic in W. These models are then considered in relation to the inequalities obtained in Sec. 2.

The simplest generalization of (4.3) for $t \neq 0$, in terms of $G(W, t)$ is

$$G(W, t) \sim (-iW)^{\alpha(t)}, \tag{4.10}$$

where $\alpha(0) = \alpha$, $G(W, 0) = F(E, 0)$. From (2.7) and (2.8), for this model

$$\sigma \sim W^{\alpha-1} \sim E^{\alpha-1}, \tag{4.11}$$

$$g \sim \alpha' \log W \sim \alpha' \log E. \tag{4.12}$$

Then the inequality of Martin and MacDowell (2.17), gives for large E

$$E^{\alpha-1} \leq \alpha' \log E. \tag{4.13}$$

Hence $\alpha \leq 1$, which is a well-known result. Similarly if we generalize (4.5) in the form

$$G(W, t) \sim -(iW)^{\alpha(t)}(\log W - \tfrac{1}{2}i\pi)^{\beta(t)} \tag{4.14}$$

we obtain for $t = 0$, if $\alpha(0) = 1$,

$$\sigma \sim (\log E)^\beta, \qquad g \sim \alpha' \log E. \tag{4.15}$$

The inequality (2.17) now gives $\beta \leq 1$. This generalization was first noted by Martin using another method.

So far we have only made limited use of crossing symmetry through the phases of F or G. It is apparent that the models indicated by the assumed forms (4.10) and (4.14) are too simple to allow the Froissart bound to be attained.

A much more general class of model is one based on the series expansion of $F(E, t)$ in powers of t, for which we have investigated conditions on the dominant terms of $\text{Im } F$ in Sec. 2. We can readily make this series expansion for the dominant part of $\text{Im } F(E, t)$ into a crossing-symmetric series for $G(W, t)$ provided we assume that each term in $\text{Im } F_n(E, 0)$ has a dominant part which is a smooth function of E. The explicit simplification that is

required for our method is the absence (by assumption) of the normal threshold singularities from the dominant terms in the series expansion for Im $F(E, t)$ for large values of E. This assumption is likely to be made in any tractable model for high-energy scattering and it may even be correct rigorously (the general method used by Martin[3] might be used to determine whether this is the situation). Alternatively it may be possible to use the inequalities obtained for Im $F_n(E, 0)$ to obtain inequalities on the phase of $F_n(E, 0)$ without explicit use of the smoothness assumption. We recall from (2.20) that

$$F_n(E, t) = (d^n/dt^n)F(E, t). \tag{4.16}$$

We now limit our discussion to models for which the dominant terms in Im $F_n(E, 0)$ are smooth. For these terms we can obtain, by substitution in the series for Im $F(E, t)$, a series for Im $G(W, t)$. This leads to a crossing symmetric form of $G(W, t)$ if we use the condition (4.8) for each term $G_n(W, 0)$, the nth derivative of G. However, so long as we restrict ourselves to the range where $|t|$ is small compared with E, the dominant terms of $F(E, t)$ and $G(E, t)$ are the same. In particular we can obtain the phase of $F_n(E, 0)$ by the condition for large E,

$$F_n(-E + i0, 0) \sim F_n^*(E + i0, 0) \tag{4.17}$$

provided we know its rate of growth with E.

We have an inequality on the rate of growth of Im $F_n(E, 0)$ in (2.13) or (2.18). One can readily verify using (4.17) that this gives an inequality on the phase of $F_n(E, 0)$, for example by using the theorems of Khuri and Kinoshita.[7] To obtain an upper bound on $g(E)$ and hence on σ(total), it is sufficient to assume that the inequality (2.13) is in fact an equality. We therefore consider typical situations using this assumption.

First we consider

$$\sigma \sim g(E) \sim E^\alpha, \qquad 0 < \alpha < 1. \tag{4.18}$$

Using (2.13) as an equality gives

$$(n!) \text{ Im } F_n(E, 0) \sim E^{(1+n)\alpha+1}. \tag{4.19}$$

From (4.17) and (4.19) we obtain, using a uniqueness condition analogous to (4.2),

$$(n!)F_n(E, 0) \sim E^{(1+n)\alpha+1}\{i + \tan [\tfrac{1}{2}(1 + n)\alpha\pi]\} \tag{4.20}$$

provided $(1 + n)\alpha + 1$ is not an even integer. If it were even, say $2m$, we would obtain instead of (4.20)

$$E^{2m}[i - (2/\pi) \log E]. \tag{4.21}$$

In principle, an arbitrary power of E^2 could be added to (4.20) without affecting its crossing symmetry or analyticity, just as in a dispersion relation, unless one has a uniqueness condition like (4.2). However, this lack of uniqueness is removed by the fact that $F_n(E, 0)$ is the derivative of $F_{n-1}(E, t)$ at $t = 0$ and that we have some inequalities like (2.21). We use this inequality to discuss the form (4.20) for F_{n-1} and F_n. It gives

$$\text{Im } F_{n-1}(E, t) \leq \text{Im } F_{n-1}(E, 0), \quad \text{for} \quad t < 0. \tag{4.22}$$

Let us assume that $0 < (n\alpha + 1) < 2$, so that from (4.20),

$$\text{Re } F_{n-1}(E, 0) > 0. \tag{4.23}$$

Because of the inequality (4.22), for $t < 0$ the phase of $F_{n-1}(E, t)$, as well as its rate of growth, is bounded by its phase and rate of growth at $t = 0$. It is not necessary to assume that the phase depends on t as in (4.10), but one can just use continuity and the inequalities (4.22), (4.23) to give

$$\text{Re } F_{n-1}(E, t) < \text{Re } F_{n-1}(E, 0), \quad \text{for} \quad t < 0. \tag{4.24}$$

Hence by differentiation at $t = 0$, (4.24) gives

$$\text{Re } F_n(E, 0) > 0. \tag{4.25}$$

But with our assumption (4.18) it would be possible to choose n large enough so that

$$(n\alpha + 1) < 2 < (n + 1)\alpha + 1. \tag{4.26}$$

This contradicts the result (4.25) and the inequality (4.22), and shows that a power law for $\sigma(E)$ or $g(E)$ is not consistent with analyticity and unitarity. This result is the analog of the divergence of the dispersion integral discussed following the inequality (3.3).

A logarithmic rate of growth as considered in the dispersion method of Sec. 3 involves a more delicate condition. We see also that the phase condition used here gives more information than the dispersion method. We assume that

$$\sigma \sim g(E) \sim (\log E)^\beta \tag{4.27}$$

and we use (2.18) as an equality to give a bound on the phase and rate of growth of F_n for large E,

$$(n!) \text{ Im } F_n(E, 0) \sim E(\log E)^{(1+n)\beta}. \tag{4.28}$$

From the crossing symmetry conditions

$$(n!)F_n(E, 0) \sim E(\log E)^{(1+n)\beta} \times \{i + \tan [(1 + n)\beta\pi/2 \log E]\}. \tag{4.29}$$

As $E \to \infty$ this term always satisfies the condition

(4.25) when (4.23) is given, that is to say, when $\beta > 0$.

The question that we wish to consider now is whether the term (4.28) can remain the dominant term in Im $F_n(E, 0)$ when E is finite. The analogous problem in the dispersion method arose in the integral in (3.3), which was dominated by values near $\log E \sim n$, and the inequality (3.3) could not be satisfied unless $\beta \leq 2$.

It is evident that (4.28) must not be the only term in Im $F_n(E, 0)$ since for values of E such that

$$(1 + n)\beta \sim \log E \tag{4.30}$$

the real part of $F_n(E, 0)$ becomes arbitrarily large and then changes sign. This contradicts the result (4.25) and can be prevented only by the existence of another term in Im $F_n(E, 0)$ which has a smaller rate of growth with E than in (4.28), and which becomes comparable with Im $F_n(E, 0)$ near the value in (4.30). The smaller rate of growth is necessary so that this other term does not itself contradict the phase condition. It is clear that by successive steps one finds that, near (4.30), Im $F_n(E, 0)$ must be dominated by a term independent of E. We consider the relation of this term to the term (4.28). Denote the corresponding series by F_1 and F_2, both being parts of F.

$$\text{Im } F_1(E, t) = E \sum_0^L C_n t^n, \tag{4.31}$$

$$\text{Im } F_2(E, t) = E \sum_0^L (\log E)^{(n+1)\beta} \frac{t^n}{(n!)^2}, \tag{4.32}$$

where C_n is bounded by a term independent of E. Each term in (4.32) is the smallest term compatible with (4.27), their size being determined by $g(E)$. We require "phase interference" between the terms F_{1n} and F_{2n} that correspond to the coefficients of t^n in (4.31) and (4.32). For this phase interference to prevent contradiction of (4.22) we require, near $n \sim \log E$,

$$C_n > (\log E)^{n\beta}/(n!)^2. \tag{4.33}$$

The largest allowed value of C_n is of order unity for (4.31) to be convergent; more precisely

$$C_n \leq 1/R^n, \tag{4.34}$$

where R is a radius of convergence that does not depend on E for large E. The Froissart bound follows from (4.33) and (4.34), giving $\beta \leq 2$. Under this condition the real part of $F_n(E, 0)$ obtained from $(F_{1n} + F_{2n})$ cannot change sign by going through infinity but only by going through zero. This is called the "phase condition."

As a further application of the phase condition assume that Im $F_2(E, t)$ is represented by a series of dominant terms that is more general than (4.32), namely,

$$\text{Im } F_2(E, t) \sim E\sigma \sum_0^L (\log E)^{n\beta} \frac{t^n a_n'}{(n!)^\gamma}, \tag{4.35}$$

where $a_n' \sim 1$ for large n, and from (2.6) and (2.13), $\gamma \leq 2$. Note that we are not asserting for negative t that Im F_2 is dominant, but only that for each coefficient of t^n, with n fixed and $E \to \infty$ the individual terms in (4.35) dominate over corresponding terms in Im F the full amplitude. For the sum itself there are complicated cancellations when $t < 0$, but our discussion has avoided the need to consider these.

With (4.31) and (4.35) the inequality (4.33) is replaced by

$$1/R^n \geq C_n \geq (\log E)^{n\beta}/n^{n\gamma} \tag{4.36}$$

when $n > \log E$. This gives

$$\beta \leq \gamma \leq 2. \tag{4.37}$$

For positive values of t, if we let $L \to \infty$ in (4.35), it does represent a bound on Im $F(E, t)$ as $E \to \infty$, given that $g \sim (\log E)^\beta$, since each term is now positive so there is no cancellation. The inequalities (4.36) and (4.37) ensure that Im $F(E, t)$ given by (4.35) must be analytic in the region

$$|t| < R, \tag{4.38}$$

where R is fixed as $E \to \infty$. This result for our model is evidently closely related to Martin's enlargement of the Lehmann ellipse, but here we have not made any direct use of the analyticity domain for low energy.

Conversely, if the result (4.38) is assumed, then we can derive the condition (4.37) from (4.35). For large E, we have (for $t > 0$) from (4.35).

$$\text{Im } F(E, t) \sim E\sigma \exp [t^{1/\gamma}(\log E)^{\beta/\gamma}]. \tag{4.39}$$

If this is to be polynomial bounded for fixed $t > 0$, we must have $\beta \leq \gamma$, and from (2.6), (2.13), we have $\gamma \leq 2$.

ACKNOWLEDGMENTS

The author is indebted to Professor J. R. Oppenheimer for hospitality at the Princeton Institute for Advanced Study, where part of this work was completed, and to Professor M. L. Goldberger, Professor S. B. Treiman, and Professor K. Hepp for valuable discussions.

Homogeneous Lichnerowicz Universes

István Ozsváth
Southwest Center for Advanced Studies, Dallas, Texas
(Received 4 April 1966)

Lichnerowicz has developed a general-relativistic theory of an electrically charged fluid with infinite conductivity. No exact solutions of the corresponding Einstein–Lichnerowicz equations have previously been given. The present paper establishes homogeneous solutions of these equations.

1. INTRODUCTION

In a recent series of lectures[1] given at the Southwest Center for Advanced Studies, Lichnerowicz developed a general-relativistic theory of an electrically charged fluid with infinite conductivity. The relevant equations are as follows:

$$R_{ik} = -\chi\{(rf + \mu\,|h|^2)u_iu_k - \tfrac{1}{2}[rf + \mu\,|h|^2 - 2(p/c^2)]g_{ik} - \mu h_ih_k\}, \tag{1.1}$$

$$u_iu^i = 1, \qquad u_ih^i = 0, \tag{1.2}$$

$$(u^ih^k - u^kh^i)_{;i} = 0, \tag{1.3}$$

where $|h|^2 = -h_ih^i$, χ is the relativistic constant of gravitation, μ is the permeability, $rf(>0)$ is the density of the fluid, $p(\geq 0)$ is the pressure, and c is the velocity of light; the semicolon denotes covariant derivation.

Lichnerowicz proved the *existence* of the solutions; i.e., he solved the Cauchy problem connected with (1.1)–(1.3). Our aim is to find homogeneous solutions[2] of the Einstein–Lichnerowicz equations, i.e., space–times allowing four parametric simply transitive transformation groups. We give at the beginning a few theorems concerning homogeneous Lichnerowicz universes.

Suppose that the vector field

$$\xi^i = \xi^i(x^k) \tag{1.4}$$

generates a one-parametric group of transformations, and denote the Lie derivative with respect to (1.4) by the symbol

$$\underset{\xi}{\pounds}. \tag{1.5}$$

The operator (1.5) has the following formal properties:

$$\underset{\xi}{\pounds} a = a_{,i}\xi^i, \qquad \underset{\xi}{\pounds}\, bd = (\underset{\xi}{\pounds}\, b)d + b(\underset{\xi}{\pounds}\, d), \qquad \underset{\xi}{\pounds}\,(b + d) = \underset{\xi}{\pounds}\, b + \underset{\xi}{\pounds}\, d, \tag{1.6}$$

where a is an arbitrary scalar and b and d are arbitrary tensors. The Killing equation

$$\underset{\xi}{\pounds}\, g_{ik} = \xi_{(i;k)} = 0 \tag{1.7}$$

expresses the fact that the metric is invariant under the transformations of the group generated by (1.4). From (1.7) it follows that

$$\underset{\xi}{\pounds}\, R_{ik} = 0, \tag{1.8}$$

where R_{ik} is the Ricci tensor corresponding to g_{ik}. In order to avoid possible troubles, we mention our definitions:

$$u_{i;k;l} - u_{i;l;k} = -u^iR_{ikl}, \tag{1.9}$$

$$R_{jl} = R^k{}_{jkl}, \tag{1.10}$$

concerning the Riemann and the Ricci tensors.

Using the notations

$$a^2 = \chi(rf + \mu\,|h|^2),$$

$$b = \tfrac{1}{2}\chi[rf + \mu\,|h|^2 - 2(r/c^2)], \qquad a^2l^2 = \chi\mu\,|h|^2$$

$$h_i = |h|\,\hat{h}_i, \qquad l^2 < 1, \qquad \mu = \text{const}, \tag{1.11}$$

we can write (1.1) in the following form:

$$R_{ik} = a^2(-u_iu_k + l^2\hat{h}_i\hat{h}_k) + bg_{ik}. \tag{1.12}$$

From (1.8) and (1.12) it then follows that

$$\underset{\xi}{\pounds}\,\{a^2(-u_iu_k + l^2\hat{h}_i\hat{h}_k) + bg_{ik}\} = 0. \tag{1.13}$$

Using (1.6) and (1.7) we see that

$$\underset{\xi}{\pounds}\, a^2 = 0, \qquad \underset{\xi}{\pounds}\, l^2 = 0, \qquad \underset{\xi}{\pounds}\, b = 0, \qquad \underset{\xi}{\pounds}\, u_i = 0, \qquad \underset{\xi}{\pounds}\, \hat{h}_i = 0. \tag{1.14}$$

In a homogeneous space–time, we have four linearly independent generators

$$\xi_a^i = \xi_a^i(x^k), \qquad a = 0, 1, 2, 3, \tag{1.15}$$

satisfying the equations

$$\xi_a^i\xi_b^k{}_{,i} - \xi_b^i\xi_a^k{}_{,i} = C_{ab}{}^c\xi_c^k \tag{1.16}$$

[1] A. Lichnerowicz, *Lecture Notes* (W. A. Benjamin, Inc., New York, to be published).
[2] I. Ozsvath, J. Math. Phys. 4, 590 (1965).

and the condition

$$\det(\xi_a^i) \neq 0, \tag{1.17}$$

where $C_{ab}{}^e = -C_{ba}{}^e$ are the structure constants of the group satisfying the Jacobi identities

$$C_{ab}{}^f C_{fc}{}^d + C_{bc}{}^f C_{fa}{}^d + C_{ca}{}^f C_{fb}{}^d = 0. \tag{1.18}$$

All the four generators (1.15) satisfy (1.7). Therefore, using (1.14) and (1.17), we have the following.

Theorem 1: In a homogeneous Lichnerowicz universe, the scalars a^2, l^2, and b are constant and the vectors u_i and $\hat{h}_i$ are invariant vectors of the group.

Theorem 2: In a homogeneous Lichnerowicz universe,

$$u^i{}_{;i} = 0, \qquad \hat{h}^i{}_{;i} = 0 \tag{1.19}$$

always hold.

Using (1.12), Theorem 1, and the twice-contracted Bianchi identities, we obtain the equations

$$(u^i u_k - l^2 \hat{h}^i \hat{h}_k)_{;i} = 0. \tag{1.20}$$

Equation (1.3) takes the form

$$(u^i \hat{h}_k - \hat{h}^i u_k)_{;i} = 0. \tag{1.21}$$

Using the equations

$$u^i u_i = 1, \qquad \hat{h}^i \hat{h}_i = -1, \qquad u^i \hat{h}_i = 0, \tag{1.22}$$

the statement follows.

One defines the acceleration vector by the following equation:

$$\dot{u}_i = u_{i;k} u^k. \tag{1.23}$$

Using this definition, one sees that, in the case of a homogeneous space–time, equations (1.3) and (1.20) can be replaced by (1.19) and the following equations:

$$u^i \hat{h}^k{}_{;i} - \hat{h}^i u^k{}_{;i} \equiv u^i \hat{h}^k{}_{,i} - \hat{h}^i u^k{}_{,i} = 0, \tag{1.24}$$

$$\dot{u}_k = l^2 \hat{h}_{k;i} \hat{h}^i. \tag{1.25}$$

From (1.22) and (1.25) immediately follows

$$\dot{u}_k \hat{h}^k = 0; \tag{1.26}$$

i.e., we have

Theorem 3: In a homogeneous Lichnerowicz universe, the magnetic unit vector $\hat{h}^i$ and the acceleration vector are mutually perpendicular.

We later use the mutually orthogonal vectors

$$u^i, \hat{h}^i, \dot{u}^i$$

to fix our tetrad. We mention for later use these further definitions:

$$\omega^i = \tfrac{1}{2} \eta^{ijkl} u_{j;k} u_l \tag{1.27}$$

is the rotation vector and

$$\sigma_{jk} = u_{(j;k)} - \tfrac{1}{3}\theta(g_{jk} - u_j u_k) - \dot{u}_{(j} u_{k)} \tag{1.28}$$

is the shear tensor, where

$$\theta = u^i{}_{;i} \tag{1.29}$$

is the expansion scalar.

The following statements also hold:

Theorem 4: In a homogeneous Lichnerowicz universe, the rotation and the acceleration vector are perpendicular; i.e., the equation

$$\omega_i \dot{u}^i = 0 \tag{1.30}$$

always holds.

Theorem 5: There exist only two families of homogeneous Lichnerowicz universes with geodesic and shear-free motion given by

$$ds^2 = R^2 \Big\{ (dx^0 + e^{x^3}\, dx^3)^2 - \frac{1}{2}\left(\frac{1 - 2l^2}{1 - l^2}\right) e^{2x^3} (dx^2)^2 - (dx^1)^2 - (dx^3)^2 \Big\}, \tag{1.31}$$

where $0 \leq l^2 < \frac{1}{2}$ ($l^2 = 0$ is the Gödel cosmos); and

$$ds^2 = R^2 \{ (dx^0 + x^3\, dx^2)^2 - \exp[-(2l^2 - 1)^{\frac{1}{2}} x^1](dx^2)^2 - (dx^1)^2 - \exp[(2l^2 - 1)^{\frac{1}{2}} x^1](dx^3)^2 \}, \tag{1.32}$$

where $\frac{1}{2} \leq l^2 < 1$ and $R^2 \neq 0$ is an arbitrary constant. The proofs of these statements emerge in the text. [See (9.20) and (9.21).]

After these general remarks, we turn to the question of how to construct homogeneous Lichnerowicz universes. The basic idea is to apply the usual tetrad formalism using basic tetrads invariant under the transformations of the group.

Tetrads, satisfying these conditions, can be built out of the invariant vectors

$$e_a^i = e_a^i(x^k), \tag{1.33}$$

and the reciprocal vectors

$$e^a{}_i = e^a{}_i(x^k) \tag{1.34}$$

of the group, where $a = 0, 1, 2, 3$ labels the vectors and $j = 0, 1, 2, 3$ labels the coordinates.

The invariant vectors (1.33) are defined by the equations

$$\xi^i_a e^k_{b,i} - e^i_b \xi^k_{a,i} = 0, \tag{1.35}$$

which must hold for each a and b, with the proviso that

$$\det (e^i_a) \neq 0. \tag{1.36}$$

If we impose the conditions

$$e^i_a(x^k_0) = \xi^i_a(x^k_0), \tag{1.37}$$

where x^k_0 are the coordinates of some arbitrarily chosen point, we have for (1.33) the equations

$$e^i_a e^k_{b,i} - e^i_b e^k_{a,i} = -c_{ab}{}^c e_c{}^k, \tag{1.38}$$

as a consequence of (1.16) and (1.35). The reciprocal vectors (1.34) are defined by

$$e_a{}^i e^b{}_i = \delta^b_a \quad \text{or} \quad e_a{}^i e^a{}_k = \delta^i_k, \tag{1.39}$$

and then satisfy the equations

$$e^a{}_{i,k} - e^a{}_{k,i} = -c_{bc}{}^a e^b{}_i e^c{}_k, \tag{1.40}$$

and the condition

$$\det (e^a{}_i) \neq 0. \tag{1.41}$$

The transformations of the group carry the vector fields $e_a{}^i$ and $e^a{}_i$ into themselves, therefore, we have an *invariant tetrad* of contravariant vectors (1.33) and an *invariant tetrad* of covariant vectors (1.34) in each point, and it is possible to assign unique tetrad components to each tensor, e.g., $T^i{}_k$

$$T^i{}_k = e^i_a T^a{}_b e^b_k, \qquad T^a{}_b = e^a{}_i T^i{}_k e^k_b. \tag{1.42}$$

The *invariant tensors* (tensor fields going into themselves under the transformations of the group) have *constant* tetrad components. For example, according to Theorem 1, the tetrad components of u_i and $\hat{h}_i$

$$u_a = u_i e^i_a, \qquad \hat{h}_a = \hat{h}_i e^i_a \tag{1.43}$$

are constant. The tetrad components of the metric

$$g_{ab} = e_a{}^i g_{ik} e^k_b, \qquad g^{ab} = e^a{}_i g^{ik} e^b{}_k \tag{1.44}$$

are constant and can be chosen to be

$$g_{ab} = g^{ab} = \operatorname{diag}(+1, -1, -1, -1), \tag{1.45}$$

which fixes our tetrad up to Lorentz transformations. The metric is given by

$$g_{ik} = e^a{}_i g_{ab} e^b{}_k, \qquad g^{ik} = e^i_a g^{ab} e^k_b. \tag{1.46}$$

The Ricci rotation coefficients defined by

$$A^a{}_{bc} = e^a{}_{i;k} e^i_b e^k_c \tag{1.47}$$

are constant and are given by

$$A_{abc} = -\tfrac{1}{2}(C_{bca} + C_{cab} - C_{abc}), \tag{1.48}$$

as one can see easily. From (1.48) follows

$$C_{abc} = A_{cba} - A_{cab}. \tag{1.49}$$

The covariant derivative of an invariant vector is given by

$$v_{i;k} = v_{a;b} e^a{}_i e^b{}_k, \tag{1.50}$$

where

$$v_{a;b} = v^f A_{fab}. \tag{1.51}$$

From all these, it is obvious that the Einstein–Lichnerowicz equations for a homogeneous space–time reduce to an algebraic system. This system reads as follows:

$$A^f{}_{ag} A^g{}_{bf} - k^f A_{fab} = a^2(-u_a u_b + l^2 \hat{h}_a \hat{h}_b) + b g_{ab}, \tag{1.52}$$

$$u_a u^a = 1, \qquad \hat{h}_a \hat{h}^a = -1, \qquad u_a \hat{h}^a = 0, \tag{1.53}$$

$$u_a k^a = 0, \qquad \hat{h}_a k^a = 0, \tag{1.54}$$

$$u^f \hat{h}^g C_{fga} \equiv u^f \hat{h}^g (A_{agf} - A_{afg}) = 0, \tag{1.55}$$

$$\dot{u}_a = u^f u^g A_{fag} = l^2 \hat{h}^f \hat{h}^g A_{fag}, \tag{1.56}$$

where

$$k_a = C_{fa}{}^f = A_{fa}{}^f. \tag{1.57}$$

These equations correspond to (1.12), (1.22), (1.19), (1.24), and (1.25), respectively. In addition to these equations, we have the Jacobi identities (1.18). In this system, we regard the constants A_{abc}, a^2, l^2, b, u_a, $\hat{h}_a$, and k_a as unknowns. We listed Eq. (1.56) separately, in spite of the fact that it is a consequence of the other equations, because it is linear in A_{abc} and can be used to simplify our later calculations.

The construction of a homogeneous Lichnerowicz universe goes in two steps:

(i) Solution of the algebraic equations (1.52)–(1.57), (1.18).

(ii) Explicit construction of the tetrad.

Concerning (i), we remark that, according to (1.54), we distinguish two main cases:

$$\text{Case 1,} \qquad k_a \neq 0, \tag{1.58}$$

$$\text{Case 2,} \qquad k_a = 0, \tag{1.59}$$

which are developed separately in the next sections.

To conclude this section, we make a technical remark concerning step (ii), which is of some use

later. Suppose we solved the algebraic equations; i.e., we know the structure constants of the group. In order to obtain the tetrads we would have to integate Eqs. (1.40). Any set of solutions satisfying (1.41) could be picked. Using different solutions only means different choice of coordinates. But, in order to obtain coordinate systems where the operations of the group "look simplest," it is useful to apply the following procedure:

Knowing the structure constants of the group, we have the commutator relations

$$(X_a X_b) = C_{ab}{}^c X_c, \tag{1.60}$$

where $X_a = \xi_a^i(\partial/\partial X^i)$ and $(X_a X_b) = X_a X_b - X_b X_a$, which correspond to (1.16). Introducing new linear operators by

$$Y_b = X_a A^a{}_b, \tag{1.61}$$

where $A^a{}_b$ are constant and $\det(A^a{}_b) \neq 0$, we get the commutator relations

$$(Y_a Y_b) = A^d{}_a A^e{}_b C_{de}{}^f A^{-1c}{}_f Y_c = \bar{C}_{ab}{}^c Y_c \tag{1.62}$$

of a group isomorphic to (1.60). By suitable choice of $A^a{}_b$, one can arrange that $\bar{C}_{ab}{}^c$ has one of the normal forms first given by Lie.[3] Then integrating the corresponding equations (1.16) under the condition (1.17), one finds a set of infinitesimal generators η_a^i of (1.62). Here, the remark that different solutions only mean different coordinates applies equally. After that, one integrates the corresponding Eqs. (1.35) under the conditions (1.36) and (1.37), and finds the invariant vectors $\bar{e}_a^i$ and [using (1.39)] the reciprocal vectors $\bar{e}^a{}_i$ of (1.62). The reciprocal vectors of (1.60) are then finally given by

$$e^a{}_i = A^a{}_b \bar{e}^b{}_i, \tag{1.63}$$

as one easily sees.

2. ALGEBRAIC SYSTEM IN CASE $k_a \neq 0$

In this case, we can choose our tetrad uniquely in such a way that $u_i = e^0{}_i$, $\hat{h}_i = -e^1{}_i$, $k_i = -e^3{}_i$, i.e.,

$$u_a = \delta_a^0, \qquad \hat{h}_a = -\delta_a^1, \qquad k_a = -\delta_a^3. \tag{2.1}$$

Equations $k_a = -\delta_a^3$ imply that $k_i k^i = -1$, which is no restriction of generality, since the length of k^i is just a constant conform factor on the metric as one easily sees. By this normalization, Eqs. (1.53) and (1.54) are automatically satisfied, and the remaining equations can be written as

$$A^f{}_{ag} A^g{}_{bf} - A_{3ab} = a^2\{-\delta_a^0\,\delta_b^0 + l^2\,\delta_a^1\,\delta_b^1\} + b g_{ab}, \tag{2.2}$$

$$A_{fa}{}^f = -\delta_a^3, \tag{2.3}$$

$$C_{10a} = A_{a01} - A_{a10} = 0, \tag{2.4}$$

$$\dot{u}_a = A_{0a0} = l^2 A_{1a1}, \tag{2.5}$$

$$C_{ab3} = A_{3ba} - A_{3ab} = 0, \quad a, b = 0, 1, 2, 3, \tag{2.6}$$

$$C_{3\alpha}{}^{\varphi} C_{\varphi\beta\gamma} + C_{\alpha\beta}{}^{\varphi} C_{\varphi 3\gamma} + C_{\beta 3}{}^{\varphi} C_{\varphi\alpha\gamma} = 0, \quad \alpha, \beta, \gamma, \varphi = 0, 1, 2, \tag{2.7}$$

where (2.3) is the consequence of the normalization, (2.6) is the contracted Jacobi identity, and (2.7) are the rest of the Jacobi identities. Geometrically, (2.6) indicates that the vector field $k^i = k^a e_a{}^i$ is hypersurface orthogonal ($k^f A_{f[ab]} = 0 \leftrightarrow k_{[i;k]} = 0$). One can replace the C's by the A's in (2.7) using (1.49).

We now proceed to set up our equations explicitly. We regard a^2 and l^2 (the parameters and the coefficients of A_{abc}) as unknowns which satisfy the linear equations (2.3)–(2.6). The following table gives the results of this operation listing the coefficients of A_{abc} and defining our notations

ab \ c	0	1	2	3
23	B	C	$1 + (1 - l^2)A$	0
31	D	A	$-C$	0
12	E	0	0	G
10	0	0	F	H
20	0	$-E$	0	K
30	$l^2 A$	D	$-B$	0 .

(2.8)

We write now Eqs. (2.7) and (2.2) explicitly

$$\begin{aligned}
&(B - K)(E + F) = 0, \qquad (C + G)(E - F) = 0,\\
&(1 - 2l^2 A)(E + F) = 0, \qquad (1 + 2A)(E - F) = 0,\\
&(D + H)(E + F) - (D - H)(E - F) = 0,\\
&-2EF + 2DH - 2BK - l^2 A = -a^2 + b,\\
&-2EF + 2DH + 2CG - A = a^2 l^2 - b,\\
&-2E^2 - 2BK - 2CG + (1 - l^2)A + 1 = -b,\\
&(1 + l^2)AH + BG - CK - D = 0,\\
&(2l^2 - 1)AK - CH + DG + B - K = 0,\\
&C(E - F) = 0,
\end{aligned} \tag{2.9, 2.10}$$

[3] S. Lie und G. Scheffers, *Vorlesungen über continuirliche* Gruppen (B. G. Teubner, Leipzig, 1893).

$$(2 - l^2)AG - BH + DK + C + G = 0,$$
$$B(E + F) = 0,$$
$$ED = 0,$$
$$2\{(l^4 - l^2 + 1)A^2 + (1 - l^2)A - B^2 + C^2 - D^2\} + 1 = -b.$$

The tetrad components of the acceleration and the rotation vector [see (1.25) and (1.27)] and the nonvanishing tetrad components of the shear tensor [see (1.28)] are given, respectively, by

$$\dot{u}_a = [0, 0, 0, -l^2A], \tag{2.11}$$
$$\omega_a = [0, -\tfrac{1}{2}(B + K), -\tfrac{1}{2}(D - H), -\tfrac{1}{2}(E + F)], \tag{2.12}$$
$$\sigma_{12} = \sigma_{21} = \tfrac{1}{2}(E - F),$$
$$\sigma_{13} = \sigma_{31} = -\tfrac{1}{2}(D + H), \tag{2.13}$$
$$\sigma_{23} = \sigma_{32} = \tfrac{1}{2}(B - K).$$

3. ALGEBRAIC SYSTEM IN CASE $k_a = 0$

Using the normalization

$$u_a = \delta^0_a, \qquad \hat{h}_a = -\delta^1_a, \tag{3.1}$$

as before, we can write the Einstein–Lichnerowicz equations in the form

$$A^f{}_{ag}A^g{}_{bf} = a^2\{-\delta^0_a\,\delta^0_b + l^2\,\delta^1_a\,\delta^1_b\} + bg_{ab}, \tag{3.2}$$
$$A_{fa}{}^f = 0, \tag{3.3}$$
$$C_{10a} = A_{a01} - A_{a10} = 0, \tag{3.4}$$
$$\dot{u}_a = A_{0a0} = l^2A_{1a1}, \tag{3.5}$$
$$C_{ab}{}^fC_{fcd} + C_{bc}{}^fC_{fad} + C_{ca}{}^fC_{fbd} = 0. \tag{3.6}$$

In order to fix our tetrad completely, we use the acceleration and the rotation vector [See (1.25), (1.27)]

$$\dot{u}_a = [0, 0, -A_{200}, -A_{300}], \tag{3.7}$$
$$\omega_a = [0, -\tfrac{1}{2}C_{230}, -\tfrac{1}{2}C_{310}, -\tfrac{1}{2}C_{120}], \tag{3.8}$$

and the Jacobi identity

$$C_{12}{}^fC_{f30} + C_{23}{}^fC_{f10} + C_{31}{}^fC_{f20} \equiv 4\dot{u}_a\omega^a = 0 \tag{3.9}$$

[Eq. (3.9) proves Theorem 4 in case $k_a = 0$], and proceed as follows:

If $\dot{u}^i \neq 0$, then we can choose $e^i{}_3$ in the direction of $\dot{u}^i$, i.e.,

$$A_{200} = 0, \tag{3.10}$$

and since $A_{300} \neq 0$ from (3.9), it then follows

$$C_{120} \equiv A_{102} - A_{201} = 0. \tag{3.11}$$

If $\dot{u}^i = 0$, then (3.10) is satisfied and we can choose our tetrad in such a way that (3.11) holds.

Therefore, we can always assume (3.10) and (3.11) without restriction of generality, and this assumption fixes our tetrad uniquely, except in the case, where $\dot{u}_a = 0$ holds, and ω^a lies in the plane of u^a and $\hat{h}^a$. In this special case, which in fact occurs later, we fix our tetrad by other means.

After solving the linear equations (3.3)–(3.5), (3.10), and (3.11), we have for the coefficients of A_{abc} the following table:

ab \ c	0	1	2	3
23	F	G	$(1 - l^2)A$	0
31	C	A	H	E
12	$-B$	0	E	K
10	0	0	B	M
20	0	B	D	N
30	l^2A	C	P	$-D$.

(3.12)

Equations (3.6) have the form

$$AB = 0,$$
$$B(C - M) = 0,$$
$$l^2AE + D(C - M) = 0,$$
$$l^2A(G + H) - (C - M)(F + P) = 0,$$
$$AD - 2B(G + K) + E(C + M) = 0,$$
$$A(F + P) + 2BE - (G + H)(C + M) = 0,$$
$$D(G + K) + E(F - N) = 0,$$
$$D(G + H) + E(F + P) = 0,$$
$$(F + P)(G + K) - (F - N)(G + H) = 0,$$

(3.13)

and the field equations are given by

$$B^2 + D^2 + CM + NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$
$$B^2 + E^2 + CM - HK = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b,$$
$$B^2 + GK + FN = \tfrac{1}{2}b,$$
$$(l^4 - l^2 + 1)A^2 - C^2 + FP - GH = -\tfrac{1}{2}b,$$
$$(1 + l^2)AM - 2DE + NH - KP = 0,$$
$$(2l^2 - 1)AN + BE + CK - GM = 0,$$
$$(1 - 2l^2)AD + B(G - H) - CE = 0,$$
$$(2 - l^2)AK + BD + CN - FM = 0,$$
$$(2 - l^2)AE - B(F + P) + CD = 0,$$
$$-BC + DF + EG = 0.$$

(3.14)

The tetrad components of the acceleration and the rotation vector [see (3.7) and (3.8)] and the nonvanishing tetrad components of the shear tensor [see (1.28)] are given by

$$\dot{u}_a = [0, 0, 0, -l^2 A], \tag{3.15}$$

$$\omega_a = [0, -\tfrac{1}{2}(N - P), -\tfrac{1}{2}(C - M), 0], \tag{3.16}$$

$$\sigma_{12} = \sigma_{21} = -B, \quad \sigma_{13} = \sigma_{31} = -\tfrac{1}{2}(C + M),$$
$$-\sigma_{22} = \sigma_{33} = D, \quad \sigma_{23} = \sigma_{32} = -\tfrac{1}{2}(N + P). \tag{3.17}$$

The first step in finding a homogeneous Lichnerowicz universe is to solve Eqs. (2.9), (2.10), (3.13), and (3.14). These equations look very complicated at first sight, but, as we see in the next sections, one can reduce them easily to simpler systems.

4. DISCUSSION OF THE CASE $k_a \neq 0$. EQUATIONS (2.9) AND (2.10)

It follows from the equations

$$(1 - 2l^2A)(E + F) = 0 \text{ and } (1 + 2A)(E - F) = 0$$

that

$$(E + F)(E - F) = 0, \tag{4.1}$$

which leads to three different possibilities,

$$F = E = 0, \tag{4.2}$$

$$F = -E \neq 0, \tag{4.3}$$

$$F = E \neq 0. \tag{4.4}$$

We now show that (4.4) is not possible, since, in this case, Eqs. (2.9) and (2.10) reduce to

$$\begin{aligned} &F = E \neq 0, \quad B = 0, \quad D = 0, \quad H = 0, \\ &K = 0, \quad A = \frac{1}{2l^2}, \quad C = -[(2 + l^2)/2l^2]G, \\ &-(2E^2 + \tfrac{1}{2}) = -a^2 + b, \\ &-b = (1/2l^2)[2(2 + l^2)G^2 - 4l^2E^2 + 1 + l^2], \\ &-(1/2l^2)[2(2 + l^2)G^2 + 4l^2E^2 + 1] = a^2l^2 - b, \\ &-b = (1/2l^4)[(2 + l^2)^2G^2 + l^4 + l^2 + 1], \end{aligned} \tag{4.5}$$

which are not consistent, since the last equation tells that $-b > 0$, therefore $a^2l^2 - b > 0$, which contradicts second to the last equation.

Equations (2.9) and (2.10) reduce in the case (4.2) to the following system:

$$\begin{aligned} &E = 0, \quad F = 0, \\ &2DH - 2BK - l^2A = -a^2 + b, \\ &2CG + 2DH - A = a^2l^2 - b, \\ &-2BK - 2CG + (1 - l^2)A + 1 = -b, \\ &(1 + l^2)AH + BG - CK - D = 0, \\ &(2l^2 - 1)AK - CH + DG + B - K = 0, \\ &(2 - l^2)AG - BH + DK + C + G = 0, \\ &2[(l^4 - l^2 + 1)A^2 + (1 - l^2)A \\ &\quad - B^2 + C^2 - D^2] + 1 = -b. \end{aligned} \tag{4.6}$$

Table (2.8) then takes the form

ab \ c	0	1	2	3
23	B	C	$1 + (1 - l^2)A$	0
31	D	A	$-C$	0
12	0	0	0	G
10	0	0	0	H
20	0	0	0	K
30	l^2A	D	$-B$	0 .

(4.7)

The tetrad components of the acceleration and the rotation vector [see (2.11) and (2.12)] and the nonvanishing tetrad components of the shear tensor [see (2.13)] are given, respectively, by

$$\dot{u}_a = [0, 0, 0, -l^2A], \tag{4.8}$$

$$\omega_a = [0, -\tfrac{1}{2}(B + K), -\tfrac{1}{2}(D - H), 0], \tag{4.9}$$

$$\sigma_{13} = \sigma_{31} = -\tfrac{1}{2}(D + H), \quad \sigma_{23} = \sigma_{32} = \tfrac{1}{2}(B - K). \tag{4.10}$$

The commutator relations [see (1.60)] of the corresponding group are

$$\begin{aligned} &(X_0X_1) = (X_1X_2) = (X_2X_0) = 0, \\ &(X_0X_3) = -l^2AX_0 \\ &\quad + (D + H)X_1 - (B - K)X_2, \\ &(X_1X_3) = -(D - H)X_0 + AX_1 - (C + G)X_2, \\ &(X_2X_3) = (B + K)X_0 \\ &\quad - (C - G)X_1 - [1 + (1 - l^2)A]X_2. \end{aligned} \tag{4.11}$$

Equations (2.9) and (2.10) reduce in the case of (4.3) to the following system:

$$\begin{aligned} &F = -E \neq 0, \quad A = -\tfrac{1}{2}, \\ &C = D = G = H = 0, \quad B = \tfrac{1}{2}(1 + 2l^2)K, \\ &2E^2 - (1 + 2l^2)K^2 + \tfrac{1}{2}l^2 = -a^2 + b, \\ &2E^2 + \tfrac{1}{2} = a^2l^2 - b, \\ &2E^2 + (1 + 2l^2)K^2 - \tfrac{1}{2}(1 + l^2) = b \\ &l^4 + l^2 + 1 - (1 + 2l^2)K^2 = -2b. \end{aligned} \tag{4.12}$$

Table (2.8) then has the form

ab \ c	0	1	2	3
23	B	0	$\frac{1}{2}(1+l^2)$	0
31	0	$-\frac{1}{2}$	0	0
12	E	0	0	0
10	0	0	$-E$	0
20	0	$-E$	0	K
30	$-\frac{1}{2}l^2$	0	$-B$	0 .

(4.13)

The tetrad components of the acceleration and the rotation vector [see (2.11) and (2.12)] and the nonvanishing tetrad components of the shear tensor [see (2.13)] are given, respectively, by

$$\dot{u}_a = [0, 0, 0, \tfrac{1}{2}l^2], \tag{4.14}$$

$$\omega_a = [0, -\tfrac{1}{2}K, 0, 0], \tag{4.15}$$

$$\sigma_{12} = \sigma_{21} = E, \qquad \sigma_{23} = \sigma_{32} = -\tfrac{1}{2}K. \tag{4.16}$$

The commutator relations (1.60) of the corresponding group are given by

$$\begin{aligned}
&(X_0X_1) = 0, \qquad (X_1X_2) = 0, \qquad (X_2X_0) = 2EX_1,\\
&(X_0X_3) = \tfrac{1}{2}l^2X_0 \qquad -(B-K)X_2,\\
&(X_1X_3) = -\tfrac{1}{2}X_1,\\
&(X_2X_3) = (B+K)X_0 \qquad -\tfrac{1}{2}(1+l^2)X_2.
\end{aligned} \tag{4.17}$$

[If $E \to 0$, (4.17) goes into a special case of (4.11).]

One sees from (4.8), (4.9) and (4.14), (4.15) that the equation

$$\dot{u}_a\omega^a = 0 \tag{4.18}$$

also holds in case $k_a \neq 0$ [see (3.9)], which completes the proof of Theorem 4.

One sees that Eqs. (4.12) have no solutions in case $K = 0$; i.e., the cases (4.3) models always have shear and rotation and the motion of the matter is never geodesic. One finds the case $k_a \neq 0$ models solving Eqs. (4.6) and (4.12) in later sections.

5. DISCUSSION OF THE CASE $k_a = 0$. EQUATIONS (3.13), (3.14)

Dealing with Eqs. (3.13) and (3.14), we want to distinguish two different cases [see (3.15)].

$$\text{Nongeodesic case:} \quad A \neq 0, \tag{5.1}$$

$$\text{Geodesic case:} \quad A = 0. \tag{5.2}$$

We now develope the equations for the nongeodesic case. We obviously have

$$B = 0, \tag{5.3}$$

and prove that

$$D = E = 0 \tag{5.4}$$

also hold.

Consider the following four equations:

$$\begin{aligned}
(C-M)D + l^2AE &= 0,\\
AD + (C+M)E &= 0,\\
(2l^2-1)AD + CE &= 0,\\
CD + (2-l^2)AE &= 0,
\end{aligned}$$

as linear equations in D and E.

If one of the determinants of this system does not vanish, our statements follow. We now show that, under the assumption (5.1), the determinants $A\{(2-l^2)(C-M) - l^2C\}$, $A\{C + (1-2l^2)(C+M)\}$, $l^2A^2 - C^2 + M^2$, cannot vanish simultaneously, since the equations $2(1-l^2)C = (2-l^2)M$, $2(1-l^2)C = (2l^2-1)M$, $l^2A^2 - C^2 + M^2 = 0$, would imply $A = 0$ under the condition that $l^2 \neq 1$, as one easily sees. Equations (3.13) and (3.14) reduce therefore, in the nongeodesic case, to the following system:

$$\begin{aligned}
A = 1, \quad B = 0, \quad D = 0, \quad E &= 0,\\
l^2(G+H) - (C-M)(F+P) &= 0,\\
(C+M)(G+H) - (F+P) &= 0,\\
(F-N)(G+H) - (G+K)(F+P) &= 0,\\
CM + NP &= -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,\\
CM - HK &= \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b,\\
GK + FN &= \tfrac{1}{2}b,\\
(l^4 - l^2 + 1) - C^2 + FP - GH &= -\tfrac{1}{2}b,\\
(1+l^2)M + NH - KP &= 0,\\
(2l^2-1)N + CK - GM &= 0,\\
(2-l^2)K + CN - FM &= 0.
\end{aligned} \tag{5.5}$$

Table (3.12), coefficients of A_{abc}, has the form

ab \ c	0	1	2	3
23	F	G	$(1-l^2)A$	0
31	C	A	H	0
12	0	0	0	K
10	0	0	0	M
20	0	0	0	N
30	l^2A	C	P	0 .

(5.6)

The tetrad components of the acceleration and the rotation vector [see (3.15) and (3.16)] and the

nonvanishing tetrad components of the shear tensor [see (3.17)] are given, respectively, by

$$\dot{u}_a = [0, 0, 0, l^2 A], \tag{5.7}$$

$$\omega_a = [0, -\tfrac{1}{2}(N - P), -\tfrac{1}{2}(C - M), 0], \tag{5.8}$$

$$\begin{aligned} \sigma_{13} &= \sigma_{31} = -\tfrac{1}{2}(C + M), \\ \sigma_{23} &= \sigma_{32} = -\tfrac{1}{2}(N + P). \end{aligned} \tag{5.9}$$

We now want to set up the equations of the geodesic case: $A = 0$. Equations (3.13) and (3.14) now read

$$\begin{gathered} A = 0, \\ B(C - M) = 0, \\ D(C - M) = 0, \\ (F + P)(C - M) = 0, \\ -2B(G + K) + E(C + M) = 0, \\ 2BE - (G + H)(C + M) = 0, \\ D(G + K) + E(F - N) = 0, \\ D(G + H) + E(F + P) = 0, \\ (F + P)(G + K) - (F - N)(G + H) = 0, \\ B^2 + D^2 + CM + NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \\ B^2 + E^2 + CM - HK = \tfrac{1}{2}a^2 l^2 - \tfrac{1}{2}b, \\ B^2 + GK + FN = \tfrac{1}{2}b, \\ C^2 + GH - FP = \tfrac{1}{2}b, \\ 2DE + KP - NH = 0, \\ BE + CK - GM = 0, \\ B(G - H) - CE = 0, \\ BD + CN - FM = 0, \\ B(F + P) - CD = 0, \\ -BC + DF + EG = 0. \end{gathered} \tag{5.10}$$

We now show that these equations imply

$$B = C = M = D = E = 0. \tag{5.11}$$

We prove this statement step by step. The first step is to show that (5.3) holds. Suppose, on the contrary,

$$B = 1, \tag{5.12}$$

then after trivial calculations (5.10) reduces to

$$\begin{gathered} A = 0, \quad B = 1, \quad M = C, \quad K = -H, \\ CE = G - H, \\ E = C(G + H), \\ D(G - H) + E(F - N) = 0, \\ D(G + H) + E(F + P) = 0, \\ (F + P)(G - H) - (F - N)(G + H) = 0, \\ 1 + D^2 + C^2 + NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \\ 1 + E^2 + C^2 + H^2 = \tfrac{1}{2}a^2 l^2 - \tfrac{1}{2}b, \\ 1 - GH + FN = \tfrac{1}{2}b, \\ C^2 + GH - FP = \tfrac{1}{2}b, \\ 2DE = H(N + P), \\ D = C(F - N), \\ F + P = CD, \\ C = DF + EG. \end{gathered} \tag{5.13}$$

From equations

$$D(G - H) + E(F - N) = 0,$$

$$E = C(G + H), \quad \text{and} \quad D = C(F - N)$$

it follows that

$$\begin{aligned} 0 &= D(G - H) + E(F - N) \\ &= D(G - H) + (G + H)C(F - N) \\ &= D(G - H) + D(G + H) = 2DG, \end{aligned}$$

i.e.,

$$DG = 0. \tag{5.14}$$

Multiplying equations

$$G - H = CE, \qquad E = C(G + H)$$

by D, and using (5.14), we see that

$$DH = 0, \qquad DE = 0 \tag{5.15}$$

follows. Using (5.14) and (5.15), Eqs. (5.13) simplify to the system

$$\begin{gathered} A = 0, \quad B = 1, \quad M = C, \quad K = -H, \\ H = \frac{1 - C^2}{1 + C^2} G, \qquad E = \frac{2CG}{1 + C^2}, \\ G(F - N) = 0, \\ DG = DH = DE = 0, \\ 1 + C^2 + D^2 + NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \\ 1 + C^2 + E^2 + H^2 = \tfrac{1}{2}a^2 l^2 - \tfrac{1}{2}b, \\ 1 - GH + FN = \tfrac{1}{2}b, \\ C^2 + GH - FP = \tfrac{1}{2}b, \\ H(N + P) = 0, \\ F + P = CD, \\ C = DE + EG, \end{gathered} \tag{5.16}$$

which has no solutions, as we prove immediately:

Consider the case $C \neq 0$ and $G \neq 0$, then the relevant equations are

$$D = 0, \quad N = F, \quad P = -F,$$

$$H = \frac{1 - C^2}{1 + C^2} G, \quad 2G^2 = 1 + C^2,$$

$$2G^2 - F^2 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad \text{and} \quad G^2 + F^2 = \tfrac{1}{2}b.$$

Using (1.11), it follows, from the last two equations, that $(Xp/c^2) = \frac{1}{2}a^2 - b = -3G^2 < 0$, which contradicts $p \geq 0$.

Consider the case $C \neq 0$, $G = 0$. The relevant equations are then

$$C \neq 0, \quad G = 0, \quad H = 0, \quad E = 0,$$

$$1 + C^2 + D^2 + NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$

$$1 + FN = \tfrac{1}{2}b, \quad C^2 - FP = \tfrac{1}{2}b,$$

$$F + P = CD, \quad C = DF.$$

Using the last two equations to eliminate C and P, we are left with the following equations:

$$1 + D^2(1 + F^2) + (D^2 - 1)FN = -\tfrac{1}{2}a^2 + b,$$

$$1 + FN = \tfrac{1}{2}b, \quad F^2 = \tfrac{1}{2}b.$$

From the last two equations, we see that $FN = F^2 - 1$; therefore we have the equations

$$2(1 + D^2F^2) - F^2 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad \tfrac{1}{2}b = F^2,$$

and the argument with the negative pressure applies.

Consider the case $C = 0$, $G \neq 0$. The relevant equations are then

$$C = 0, \quad H = G \neq 0, \quad D = 0,$$

$$E = 0, \quad N = F, \quad P = -F,$$

$$1 - F^2 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad 1 + F^2 = \tfrac{1}{2}b,$$

and the argument with the negative pressure applies.

Consider the case $C = 0$, $G = 0$. The relevant equations are

$$C = 0, \quad G = 0, \quad H = 0, \quad E = 0, \quad P = -F,$$

$$1 + D^2 - FN = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad 1 + FN = \tfrac{1}{2}b,$$

and the argument with the negative pressure applies, which completes the first step in proving (5.11).

In order to prove

$$C = 0, \tag{5.17}$$

we consider (5.10) under the assumption

$$B = 0, \quad C = 1. \tag{5.18}$$

Equations (5.10) read

$$A = 0, \quad B = 0, \quad C = 1, \quad D = 0,$$

$$E = 0, \quad M \neq 0, \quad K = GM, \quad N = FM,$$

$$(F + P)(1 - M) = 0, \quad (G + H)(1 + M) = 0,$$

$$(F + P)(1 + M)G - (G + H)(1 - M)F = 0,$$

$$FH - GP = 0, \quad M(1 + FP) = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$

$$M(1 - GH) = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b, \quad M(F^2 + G^2) = \tfrac{1}{2}b,$$

$$1 + GH - FP = \tfrac{1}{2}b. \tag{5.19}$$

We distinguish now the three different cases

$$M = 1, \quad M = -1, \quad M \neq -1, 0, +1.$$

If $M = 1$, the relevant equations are $H = -G$,

$$G(F + P) = 0, \quad 1 + FP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$

$$F^2 + G^2 = \tfrac{1}{2}b, \quad 1 - G^2 - FP = \tfrac{1}{2}b.$$

If $G \neq 0$ then $P = -F$ and

$$1 - F^2 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad F^2 + G^2 = \tfrac{1}{2}b,$$

and the argument with the negative pressure applies.

If $G = 0$, then the relevant equations are

$$1 + FP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b \quad \text{and} \quad 1 - FP = \tfrac{1}{2}b,$$

and the argument with the negative pressure applies, therefore the possibility of $M = 1$ is excluded.

If $M = -1$, the relevant equations are

$$P = -F, \quad F(G + H) = 0,$$

$$F^2 - 1 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \quad GH - 1 = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b,$$

$$-(F^2 + G^2) = \tfrac{1}{2}b, \quad GH + 1 + F^2 = \tfrac{1}{2}b.$$

If $F \neq 0$ then $H = -G$, and the relevant equations are $-(1 + G^2) = \frac{1}{2}a^2l^2 - \frac{1}{2}b$, $-(F^2 + G^2) = \frac{1}{2}b$, which lead to the impossible equation

$$\tfrac{1}{2}a^2l^2 = -(1 + F^2 + 2G^2).$$

If $F = 0$ then

$$-1 = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$

$$GH - 1 = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b, \quad GH + 1 = \tfrac{1}{2}b,$$

and therefore

$$GH = -\tfrac{1}{2}a^2 + b = -xp/c^2 \leq 0,$$

$$2GH = \tfrac{1}{2}a^2l^2 > 0,$$

should hold simultaneously, which is not possible.

The case $M \neq -1, 0, +1$ is easily excluded. The relevant equations are

$$P = -F, \quad H = -G, \quad M(1 - F^2) = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$
$$M(1 + G^2) = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b, \qquad M(F^2 + G^2) = \tfrac{1}{2}b.$$

Therefore

$$M(1 + G^2) = -\tfrac{1}{2}a^2 + b = -\chi p/c^2 < 0,$$
$$M(1 + F^2 + 2G^2) = \tfrac{1}{2}a^2l^2 > 0,$$

i.e., $M < 0$ and $M > 0$ should hold simultaneously, which completes the second step in proving (5.11).

To prove that

$$M = 0 \tag{5.20}$$

is now very easy, since the assumptions

$$B = C = 0, \qquad M = 1$$

applied to (5.10) leads to the equations

$$C = 0, \qquad D = 0, \qquad P = -F = 0,$$
$$E = 0, \qquad H = -G = 0, \qquad -\tfrac{1}{2}a^2 + \tfrac{1}{2}b = 0,$$
$$\tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b = 0, \qquad \tfrac{1}{2}b = 0,$$

which violate the conditions

$$a^2 > 0, \qquad p \geq 0.$$

We are left now to show that the assumptions

$$D = 0, \qquad E = 0 \tag{5.21}$$

can be made without restriction of generality.

First we point out that in case

$$A = C = M = 0 \tag{5.22}$$

our tetrad is not fixed, since the acceleration vector vanishes and the rotation vector lies in the plane spanned by the velocity and the magnetic vectors [see (5.7) and (5.8)]. We therefore have the freedom to perform rotations in the 2–3 plane, i.e., make Lorentz transformations of the form

$$\mathfrak{L}^a{}_b = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\varphi & \sin\varphi \\ 0 & 0 & -\sin\varphi & \cos\varphi \end{pmatrix} \tag{5.23}$$

on the tetrad components of our tensors.

Note: Whenever in this paper we exhibit a matrix it is understood that the first or upper index denotes the row and the other the column.

For example, we can apply (5.23) to A_{abc}:

$$\tilde{A}_{abc} = A_{def}\mathfrak{L}^d{}_a\mathfrak{L}^e{}_b\mathfrak{L}^f{}_c, \tag{5.24}$$

and try to simplify them.

Using (5.22), Table (3.12) takes the form

ab \ c	0	1	2	3
23	F	G	0	0
31	0	0	H	E
12	0	0	E	K
10	0	0	0	0
20	0	0	D	N
30	0	0	P	$-D$.

(5.25)

The Jacobi identities [see (5.10)]

$$D(G + K) + E(F - N) = 0,$$
$$D(G + H) + E(F + P) = 0,$$

give the equation

$$D(K - H) + E(N + P) = 0. \tag{5.26}$$

It is very easy to see that the transformed $\tilde{A}_{abc}$ has the same table of coefficients as (5.25) with $\sim$ on the top of the corresponding letters, and, in particular, it holds for

$$\tilde{E} = E\cos 2\varphi + \tfrac{1}{2}(H - K)\sin 2\varphi,$$
$$\tilde{D} = D\cos 2\varphi - \tfrac{1}{2}(N + P)\sin 2\varphi. \tag{5.27}$$

If $E = 0$, $D = 0$ should already hold, then we do not have to do anything (i.e., choose $\varphi = 0$).

If $E = 0$ while $D \neq 0$, then $H - K = 0$ follows from (5.26), and $\tilde{E} = E = 0$ from (5.27), i.e., $E = 0$ is invariant under (5.27), and φ can be chosen in such a way that $\tilde{D} = 0$ holds.

If $DE \neq 0$, then we choose φ such that either $\cot 2\varphi = -(H - K)/2E$ or $\cot 2\varphi = (N + P)/2D$ hold, but according to (5.26), these two choices are identical, i.e., we can always arrange that

$$\tilde{D} = 0, \qquad \tilde{E} = 0$$

hold, in other words, assumption (5.21) can always be made without restriction of generality. This remark finishes the proof of (5.11), and using it we can reduce (5.10) to the following system:

$$A = 0, \qquad B = 0, \qquad C = 0,$$
$$M = 0, \qquad D = 0, \qquad E = 0,$$
$$(F + P)(G + K) - (F - N)(G + H) = 0,$$
$$NH - KP = 0, \qquad NP = -\tfrac{1}{2}a^2 + \tfrac{1}{2}b,$$
$$-HK = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b, \qquad GK + FN = \tfrac{1}{2}b,$$
$$G(K - H) + F(N + P) = 0, \tag{5.28}$$

which is really simple.

Table (3.12) reduces to

ab \ c	0	1	2	3
23	F	G	0	0
31	0	0	H	0
12	0	0	0	K
10	0	0	0	0
20	0	0	0	N
30	0	0	P	0

(5.29)

The tetrad components of the rotation vector [see (3.16)] and the nonvanishing tetrad components of the shear tensor [see (3.17)] are given, respectively, by

$$\omega_a = [0, -\tfrac{1}{2}(N - P), 0, 0], \tag{5.30}$$

$$\sigma_{23} = \sigma_{32} = -\tfrac{1}{2}(N + P). \tag{5.31}$$

The commutator relations of the corresponding group are

$$\begin{aligned}
(X_2X_3) &= (N - P)X_0 + (H + K)X_1,\\
(X_3X_1) &= (G + K)X_2,\\
(X_1X_2) &= (G + H)X_3,\\
(X_1X_0) &= 0,\\
(X_2X_2) &= -(F + P)X_3,\\
(X_3X_0) &= (F - N)X_2.
\end{aligned} \tag{5.32}$$

Our algebraic problem for finding homogeneous Lichnerowicz universes is finally reduced to the solution of Eqs. (4.6), (4.12), (5.5), and (5.28), which we discuss in the next sections.

6. SOLUTION OF THE EQUATIONS (4.6). CASE $k_a \neq 0$, ABELIAN SUBCASE

Equations (4.6) can still be simplified. To this end we look into the following three equations:

$$\begin{aligned}
2DH - 2BK - l^2A &= -a^2 + b,\\
-2DH - 2CG + A &= -a^2l^2 + b,\\
2BK + 2CG - (1 - l^2)A &= b + 1.
\end{aligned} \tag{6.1}$$

By adding these equations, we obtain

$$b = \tfrac{1}{3}[a^2(1 + l^2) - 1], \tag{6.2}$$

which we can use instead of the third equation. The first two equations can be written in the form

$$\begin{aligned}
-l^2A - 2KB + 2HD + \tfrac{1}{3}a^2(2 - l^2) &= -\tfrac{1}{3},\\
A - 2GC - 2HD - \tfrac{1}{3}a^2(1 - 2l^2) &= -\tfrac{1}{3},
\end{aligned}$$

or, equivalently,

$$\tfrac{1}{2}(1 - l^2)A - KB - GC + \tfrac{1}{6}a^2(1 + l^2) = -\tfrac{1}{3}, \tag{6.3}$$

$$(l^4 - l^2 + 1)A + (2l^2 - 1)KB - (2 - l^2)GC - (1 + l^2)HD = -\tfrac{1}{2}(1 - l^2), \tag{6.4}$$

and the rest of (4.6) has the form

$$(1 + l^2)HA + GB - KC - D = 0, \tag{6.5}$$

$$(2l^2 - 1)KA + B - HC + GD = K, \tag{6.6}$$

$$(2 - l^2)GA - HB + C + KD = -G, \tag{6.7}$$

$$(l^4 - l^2 + 1)A^2 + (1 - l^2)A - B^2 + C^2 - D^2 + \tfrac{1}{6}a^2(1 + l^2) = -\tfrac{1}{3}. \tag{6.8}$$

Multiplying (6.3), (6.4), (6.5), (6.6), and (6.7) by 1, A, D, $-B$, and C, respectively, and summing up we obtain (6.8). This means that (6.8) is a consequence of the other equations, therefore (4.6) is equivalent to the following system:

$$\tfrac{1}{2}a^2 - b = \tfrac{1}{6}\{a^2(1 - 2l^2) + 2\} = (\chi p/c^2) \quad (\geq 0),$$

[see (1.11)],

$$\begin{aligned}
&\tfrac{1}{6}a^2(1 + l^2)\\
&\quad = -\tfrac{1}{2}(1 - l^2)A + KB + GC - \tfrac{1}{3} \quad (>0),\\
&(l^4 - l^2 + 1)A + (2l^2 - 1)KB - (2 - l^2)GC\\
&\quad - (1 + l^2)HD = -\tfrac{1}{2}(1 - l^2),\\
&(1 + l^2)HA + GB - KC - D = 0,\\
&(2l^2 - 1)KA + B - HC + GD = K,\\
&(2 - l^2)GA - HB + C + KD = -G.
\end{aligned} \tag{6.9}$$

The form of the equations suggests the following procedure: Regard l^2, G, H, K as parameters, solve the four linear equations for A, B, C, D, and substitute into the first equation to obtain a^2 and p as functions of the parameters, which have certain ranges of variation subject to conditions

$$l^2 < 1, \qquad a^2 > 0, \quad \text{and} \quad p \geq 0. \tag{6.10}$$

Instead of carrying out these elementary, but tedious calculations, we remark that the general solution is a four parametric family of homogeneous Riemann spaces allowing (4.11) as group of motions, where the parameters are subject to the conditions (6.10). Equations (4.8), (4.9), and (4.10) show that, in these models, the motion of the matter is no geodesic, in general there are rotation and shear present.

Naturally, one can obtain easily explicit solutions for special values of the parameters. For example, take

$$K = 0, \qquad G = 0. \tag{6.11}$$

A solution of (6.9) is immediately

$$\begin{aligned} &B = C = 0, \\ &A = -[a^2(1 + l^2) + 2]/3(1 - l^2) \qquad (<0), \\ &D = (1 + l^2)HA, \qquad K = G = 0, \\ &H^2 = \frac{(l^4 - l^2 + 1)A + \frac{1}{2}(1 - l^2)}{(1 + l^2)^2 A} \\ &\quad \equiv \frac{(1 + l^2)A - a^2(1 - l^2)}{4(1 + l^2)A} \qquad (>0), \end{aligned} \tag{6.12}$$

where the expressions of A and H^2 are manifestly negative or positive, respectively. Using (1.11) we see that

$$\begin{aligned} \chi rf &= a^2(1 - l^2), \qquad \chi\mu\,|h|^2 = a^2 l^2, \\ \chi p/c^2 &= \tfrac{1}{6}[a^2(1 - 2l^2) + 2], \end{aligned} \tag{6.13}$$

where the parameters a^2 and l^2 are subject to the conditions

$$a^2 \neq 0, \qquad l^2 < 1, \qquad a^2(1 - 2l^2) \geq -2, \tag{6.14}$$

corresponding to (6.10).

The tetrad components of the acceleration and the rotation vector [see (4.8) and (4.9)] and the nonvanishing tetrad components of the shear tensor [see (4.10)] are, respectively, given by

$$\dot{u}_a = [0, 0, 0, l^2 A], \tag{6.15}$$

$$\omega_a = [0, 0, -\tfrac{1}{2}(D - H), 0], \tag{6.16}$$

$$\sigma_{13} = \sigma_{31} = -\tfrac{1}{2}(D + H). \tag{6.17}$$

One sees from (6.12), (6.15), (6.16), and (6.17) that, if $l^2 \neq 0$, the motion of the matter is nongeodesic, and rotation and shear are always present, and if $l^2 = 0$, the motion of the matter is geodesic, and we have a solution with dust and Λ term, or fluid under pressure (the sign of the Λ term allows both interpretations) given in Ref. 2. Formula (4.14) of Ref. 2 contains *all* the groups characterized by $k_a \neq 0$, which allows solutions with dust and Λ term.

The group of this solution is given by [see (4.11)]

$$\begin{aligned} (X_0X_1) &= (X_1X_2) = (X_2X_0) = 0, \\ (X_0X_3) &= -l^2AX_0 + (D + H)X_1, \\ (X_1X_3) &= -(D - H)X_0 + AX_1, \\ (X_2X_3) &= -[1 + (1 - l^2)A]X_2, \end{aligned} \tag{6.18}$$

where A, D, and H are given by (6.12). These commutator relations give a two-parametric family of groups. Within the ranges of the parameters, (6.18) is of three different types, according to the Jordan normal forms of the matrix

$$C^{\alpha}{}_{\beta} = \begin{bmatrix} -l^2A & -(D - H) \\ (D + H) & A \end{bmatrix}, \tag{6.19}$$

which are essentially given by the roots of the characteristic polynomial

$$\lambda^2 - (1 - l^2)A\lambda - l^2A^2 + D^2 - H^2 = 0. \tag{6.20}$$

We want to discuss here only the type which corresponds to real and different roots of (6.20)

$$\begin{aligned} \lambda_0 &= \tfrac{1}{2}(1 - l^2)A + \tfrac{1}{2}\beta \\ &\neq \lambda_1 = \tfrac{1}{2}(1 - l^2)A - \tfrac{1}{2}\beta, \end{aligned} \tag{6.21}$$

where

$$\begin{aligned} \beta &= [(1 + l^2)^2A^2 - 4(D^2 - H^2)]^{\frac{1}{2}} \\ &\equiv [4H^2 + a^2(1 + l^2)(1 - l^2)A]^{\frac{1}{2}}. \end{aligned} \tag{6.22}$$

We now want to demonstrate, on this simple example, how to complete the second step indicated in Sec. 1 in constructing a homogeneous Lichnerowicz universe.

We introduce new operators

$$\begin{aligned} Y_0 &= X_0 + \nu_0 X_1, \\ Y_1 &= X_0 + \nu_1 X_1, \\ Y_2 &= X_2, \qquad Y_3 = X_3, \end{aligned} \tag{6.23}$$

where

$$\nu_0 = -\frac{l^2A + \lambda_0}{D - H}, \qquad \nu_1 = -\frac{l^2A + \lambda_1}{D - H}, \tag{6.24}$$

which are of the form of (1.61) with

$$A^a{}_b = \begin{bmatrix} 1 & 1 & 0 & 0 \\ \nu_0 & \nu_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.25}$$

and compute the new commutator relations according to (1.62), which are given by

$$\begin{aligned} &(Y_0Y_1) = (Y_1Y_2) = (Y_2Y_0) = 0, \\ &(Y_\alpha Y_3) = \lambda_\alpha Y_\alpha, \qquad \alpha = 0, 1, 2, \end{aligned} \tag{6.26}$$

where we introduced the notation

$$\lambda_2 = -[1 + (1 - l^2)A]. \tag{6.27}$$

Following the lines indicated in Sec. 1, we find the infinitesimal generators of (6.26) to be given by

$$\eta^i{}_a = \begin{bmatrix} 1 & 0 & 0 & \lambda_0 x^0 \\ 0 & 1 & 0 & \lambda_1 x^1 \\ 0 & 0 & 1 & \lambda_2 x^2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.28}$$

and the invariant vectors of (6.27) are given by

$$\tilde{e}^i{}_a \begin{bmatrix} e^{\lambda_0 x^3} & 0 & 0 & 0 \\ 0 & e^{\lambda_1 x^3} & 0 & 0 \\ 0 & 0 & e^{\lambda_2 x^3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.29}$$

[integrating (1.35), we used $X_0^k = (0, 0, 0, 0)$ in (1.37)]. The reciprocal vectors of (6.26) are then

$$\tilde{e}^a{}_i = \begin{bmatrix} e^{-\lambda_0 x^3} & 0 & 0 & 0 \\ 0 & e^{-\lambda_1 x^3} & 0 & 0 \\ 0 & 0 & e^{-\lambda_2 x^3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.30}$$

[see (1.39)]. The reciprocal vectors of (6.18) are then finally given according to (1.63), (6.25), and (6.30) by the following expressions:

$$e^a{}_i = \begin{bmatrix} e^{-\lambda_0 x^3} & e^{-\lambda_1 x^3} & 0 & 0 \\ \nu_0 e^{-\lambda_0 x^3} & \nu_1 e^{-\lambda_1 x^3} & 0 & 0 \\ 0 & 0 & e^{-\lambda_2 x^3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.31}$$

and the line element has the form

$$\begin{aligned} ds^2 &= (1 - \nu_0^2) e^{-2\lambda_0 x^3} (dx^0)^2 \\ &\quad + 2(1 - \nu_0 \nu_1) e^{(1-l^2)Ax^3}\, dx^0\, dx^1 \\ &\quad + (1 - \nu_1^2) e^{-2\lambda_1 x^3} (dx^1)^2 - e^{-2\lambda_2 x^3} (dx^2)^2 - (dx^3)^2, \end{aligned} \tag{6.32}$$

$$u_i = (e^{-\lambda_0 x^3}, e^{-\lambda_1 x^3}, 0, 0), \tag{6.33}$$

$$\hat{h}_i = (\nu_0 e^{-\lambda_0 x^3}, \nu_1 e^{-\lambda_1 x^3}, 0, 0), \tag{6.34}$$

and the scalars are given by (6.12) and (6.13).

Instead of (6.31), we want now to choose another set of reciprocal vectors for (6.18), i.e., we go into another coördinate system. Substituting into (1.40), one easily sees that the following set of vectors are indeed reciprocal vectors of (6.18):

$$e^a{}_i = \begin{bmatrix} 1 & -q e^{-\lambda_1 x^3} & 0 & -l^2 A x^0 \\ 0 & e^{-\lambda_1 x^3} & 0 & (D + H) x^0 \\ 0 & 0 & e^{-\lambda_2 x^3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.35}$$

where

$$q = (l^2 A + \lambda_0)/(D + H). \tag{6.36}$$

The line element in this coördinate system has the following form:

$$\begin{aligned} ds^2 &= (dt - q e^{-\lambda_1 x^3}\, dx^1 - l^2 A t\, dx^3)^2 \\ &\quad - (e^{-\lambda_1 x^3}\, dx^1 + (D + H) t\, dx^3)^2 \\ &\quad - e^{-2\lambda_2 x^3} (dx^2)^2 - (dx^3)^2, \end{aligned} \tag{6.37}$$

where

$$t = x^0 \quad \text{and} \quad u^i = \delta_0^i. \tag{6.38}$$

The significance of this choice is the following. Expression (6.37) is of the following form:

$$\begin{aligned} ds^2 &= [dt + p_\alpha(t) e^\alpha{}_\mu(x^\lambda)\, dx^\mu]^2 \\ &\quad - e^\alpha{}_\mu(x^\lambda) \gamma_{\alpha\beta}(t) e^\beta{}_\nu(x^\lambda)\, dx^\mu\, dx^\nu, \\ \alpha, \beta, \cdots &; \mu, \nu, \lambda, \cdots = 1, 2, 3, \cdots, \end{aligned} \tag{6.39}$$

where the covariant vectors $e^\alpha{}_\mu$ satisfy the equations

$$e^\alpha{}_{\mu,\nu} - e^\alpha{}_{\nu,\mu} = -c_{\beta\gamma}{}^\alpha e^\beta{}_\mu e^\gamma{}_\nu, \tag{6.40}$$

and the condition

$$\det (e^\alpha{}_\mu) \neq 0, \tag{6.41}$$

i.e., they are reciprocal vectors of some three-parametric group acting simply transitively in some three-dimensional hypersurfaces and

$$u^i = \delta_0{}^i, \qquad u_i = [1, p_\alpha(t) e^\alpha{}_\mu(x^\lambda)]. \tag{6.42}$$

Elementary calculations show that, in case (6.37), we have the following expressions:

$$p_\alpha = (-q, 0, -l^2 A t), \tag{6.43}$$

$$-\gamma_{\alpha\beta}(t) = \begin{bmatrix} 1 & 0 & (D + H) t \\ 0 & 1 & 0 \\ (D + H) t & 0 & 1 + (D + H)^2 t^2 \end{bmatrix}, \tag{6.44}$$

$$e^\alpha{}_\mu = \begin{bmatrix} e^{-\lambda_1 x^3} & 0 & 0 \\ 0 & e^{-\lambda_2 x^3} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.45}$$

and the group in question is given by

$$(X_1X_2) = 0, \quad (X_1X_3) = \lambda_1 X_1, \quad (X_2X_3) = \lambda_2 X_2. \tag{6.46}$$

If $l^2 = 0$, u_i is geodesic and we have our dust solution as indicated.

Instead of looking into solutions at random, we ask whether we can find solutions which, in some intrinsic sense, are the simplest. Can we find solutions where the motion of the fluid is geodesic and shear free, i.e., can we find solutions of (6.9) under the condition

$$\dot{u}_a = 0, \quad \sigma_{ab} = 0, \tag{6.47}$$

or under the assumption that

$$A = 0, \quad H = -D, \quad K = B \tag{6.48}$$

[see (4.8) and (4.10)]. The answer to this question is affirmative, and we construct this solution explicitly, proving with it a part of Theorem 5.

Using (6.48), we see that (6.9) reduces to the following system:

$$\begin{aligned} \frac{\chi p}{c^2} &= \tfrac{1}{6}[a^2(1 - 2l^2) + 2], \\ \tfrac{1}{6}a^2(1 + l^2) &= B^2 + CG - \tfrac{1}{3} \quad (>0), \\ D &= -B(C - G), \quad D(C + G) = 0, \\ C + G &= -2BD, \\ (1 - 2l^2)B^2 + (2 - l^2)CG & \\ + (1 + l^2)D^2 &= \tfrac{1}{2}(1 - l^2). \end{aligned} \tag{6.49}$$

We remark that

$$B \neq 0 \tag{6.50}$$

must hold, since $B = 0$ implies $G = -C$ and therefore $\frac{1}{6}a^2(1 + l^2) = -(C^2 + \frac{1}{3})(>0)$, which is not possible, therefore, it follows from the second line of (6.49) that

$$D = C = G = 0. \tag{6.51}$$

The solution of the remaining equations is readily given by

$$B^2 = \frac{1 - l^2}{2(1 - 2l^2)}, \quad a^2 = \frac{1}{1 - 2l^2} \ (>0), \quad \frac{\chi p}{c^2} = \tfrac{1}{2}. \tag{6.52}$$

Using (1.11), (4.9), and (4.11), we see that

$$\chi r f = \frac{1 - l^2}{1 - 2l^2}, \quad \chi\mu\,|h|^2 = \frac{l^2}{1 - 2l^2}, \quad \frac{\chi p}{c^2} = \tfrac{1}{2}, \tag{6.53}$$

$$\omega_a = (0, -B, 0, 0), \tag{6.54}$$

$$(X_0X_1) = (X_1X_2) = (X_2X_0) = 0, \tag{6.55}$$

$$(X_0X_3) = 0, \quad (X_1X_3) = 0, \quad (X_2X_3) = 2BX_0 - X_3,$$

and the parameter l^2 is subject to the condition

$$l^2 < \tfrac{1}{2}. \tag{6.56}$$

Repeating the procedure explained in Sec. 1 and demonstrated on the previous example, we can easily construct the line element. Using the new operators

$$\begin{aligned} Y_0 &= X_0, \quad Y_1 = X_1, \\ Y_2 &= X_0 - \frac{1}{2B}X_2, \quad Y_3 = X_3, \end{aligned} \tag{6.57}$$

we see that (6.55) is isomorphic to

$$\begin{aligned} &(Y_0Y_1) = (Y_1Y_2) = (Y_2Y_0) = 0, \\ &(Y_2Y_3) = 0, \ (Y_1Y_3) = 0, \ (Y_2Y_3) = -Y_2. \end{aligned} \tag{6.58}$$

The infinitesimal generators of (6.58) are given by

$$\eta^i{}_a = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -x^2 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.59}$$

the reciprocal vectors of (6.55) are

$$e^a{}_i = \begin{bmatrix} 1 & 0 & e^{x^2} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\dfrac{e^{x^2}}{2B} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6.60}$$

the line element is

$$\begin{aligned} ds^2 &= (dx^0 + e^{x^2}\,dx^2)^2 \\ &\quad - (1/4B)^2 e^{2x^2}(dx^2)^2 - (dx^1)^2 - (dx^3)^2, \end{aligned} \tag{6.61}$$

and

$$u^i = \delta_0{}^i, \quad \hat{h}^i = \delta_1{}^i. \tag{6.62}$$

If $l^2 = 0$ then $B^2 = \frac{1}{2}$, and we have

$$\begin{aligned} ds^2 &= (dx^0 + e^{x^2}\,dx^2)^2 \\ &\quad - \tfrac{1}{2}e^{2x^2}(dx^2)^2 - (dx^1)^2 - (dx^3)^2, \end{aligned} \tag{6.63}$$

which is Gödel's line element as expected.[4]

[4] K. Gödel, Rev. Mod. Phys. 21, 447 (1949).

Integrating the Killing equations

$$g_{ik,l}\xi^l + g_{lk}\xi^l{}_{,i} + g_{il}\xi^l{}_{,k} = 0 \qquad (6.64)$$

for (6.61), we find, in addition to (6.59), a fifth Killing vector

$$\eta^i{}_4 = \begin{pmatrix} -4B^2e^{-x^3} \\ 0 \\ 2B^2e^{-2x^3} - \frac{1}{2}(x^2)^2 \\ x^2 \end{pmatrix} \qquad (6.65)$$

exactly as in Gödel's case. The five parametric maximal group of (6.61) is given by

$$\begin{aligned} &(Y_0Y_1) = (Y_1Y_2) = (Y_2Y_0) = 0, \\ &(Y_0Y_3) = 0, \quad (Y_1Y_3) = 0, \quad (Y_2Y_3) = -Y_2, \\ &(Y_0Y_4) = 0, \qquad (Y_1Y_4) = 0, \\ &(Y_2Y_4) = Y_3, \qquad (Y_3Y_4) = -Y_4. \end{aligned} \qquad (6.66)$$

7. SOLUTION OF THE EQUATIONS (4.12). CASE $k_0 \neq 0$ NON-ABELIAN SUBCASE

After trivial manipulations, we see that the solution of (4.12) is given by

$$\begin{aligned} A = -\tfrac{1}{2}, \qquad B &= \tfrac{1}{2}(2l^2+1)K, \\ C = D = G = H &= 0, \\ E^2 &= \tfrac{1}{16}[2a^2(2l^2-1)-1] \qquad (\geq 0), \\ (2l^2+1)K^2 &= \tfrac{1}{2}[a^2 + \tfrac{1}{2}(2l^2+1)], \\ b &= \tfrac{1}{8}[2a^2(2l^2+1) - 3]. \end{aligned} \qquad (7.1)$$

Using (1.11) we see that

$$\chi p/c^2 = \tfrac{1}{2}a^2 - b = \tfrac{1}{8}[3 - 2a^2(1-2l^2)], \qquad (7.2)$$

which gives the following ranges for the parameters a^2 and l^2:

$$\begin{aligned} &a^2 > \tfrac{1}{2}, \\ &\tfrac{1}{2}(1 + 1/2a^2) < l^2 < \inf\{1, \tfrac{1}{2}(1 + 3/2a^2)\}. \end{aligned} \qquad (7.3)$$

We introduce the new parameter s by

$$l^2 + \tfrac{1}{2} = s. \qquad (7.4)$$

Conditions (7.3) then read

$$\begin{aligned} &a^2 > \tfrac{1}{2}, \\ &(1 + 1/4a^2) < s < \inf\{(1+\tfrac{1}{2}), (1 + 3/4a^2)\}. \end{aligned} \qquad (7.5)$$

From (4.17) and (7.4), it follows that the groups of these solutions are given by

$$\begin{aligned} &(X_0X_1) = 0, \qquad (X_1X_2) = 0, \qquad (X_2X_0) = 2EX_1, \\ &(X_0X_3) = \tfrac{1}{2}(s - \tfrac{1}{2})X_0 - (B - K)X_2, \\ &(X_1X_3) = -\tfrac{1}{2}X_1, \\ &(X_2X_3) = (B + K)X_0 - \tfrac{1}{2}(s + \tfrac{1}{2})X_2, \end{aligned} \qquad (7.6)$$

where

$$\begin{aligned} E &= \tfrac{1}{4}[4a^2(s-1) - 1]^{\frac{1}{2}}, \\ K &= \tfrac{1}{2}[(a^2 + s)/s)]^{\frac{1}{2}}, \qquad B = sK, \end{aligned} \qquad (7.7)$$

[see (7.1) and (7.4)], and using (1.11), we see that

$$\begin{aligned} \chi\rho f &= a^2(\tfrac{3}{2} - s), \qquad \chi_\mu |h|^2 = a^2(s - \tfrac{1}{2}), \\ \chi p/c^2 &= \tfrac{1}{8}[3 - 4a^2(s-1)], \end{aligned} \qquad (7.8)$$

where the parameters a^2 and s are subject to (7.5).

From (4.8), (4.9), and (4.10), it follows that, in these models, the motion of the matter is never geodesic, and shear and rotation are always present.

The commutator relations (7.6) give a two-parametric family of groups of types I, II, and III.[5] We give the line element explicitly in case of type I; i.e., if the roots of the characteristic polynomial

$$[\lambda - \tfrac{1}{2}(s - \tfrac{1}{2})][\lambda + \tfrac{1}{2}(s + \tfrac{1}{2})] + B^2 - K^2 = 0 \qquad (7.9)$$

given by

$$\begin{aligned} \lambda_0 &= -\tfrac{1}{4} + \tfrac{1}{2}\{1 - [(s+1)(s-1)/s]a^2\}^{\frac{1}{2}}, \\ \lambda_2 &= -\tfrac{1}{4} - \tfrac{1}{2}\{1 - [(s+1)(s-1)/s]a^2\}^{\frac{1}{2}} \end{aligned} \qquad (7.10)$$

are real and different.

Repeating the procedure explained in Sec. 1 and demonstrated on an explicit example in Sec. 6, we can easily construct the line element.

Using the new operators

$$\begin{aligned} Y_0 &= X_0 + \nu_0 X_2, \qquad Y_1 = X_1, \\ Y_2 &= X_0 + \nu_2 X_2, \qquad Y_3 = X_3, \end{aligned} \qquad (7.11)$$

where

$$\begin{aligned} \nu_0 &= [\lambda_0 - \tfrac{1}{2}(s - \tfrac{1}{2})]/(B + K), \\ \nu_2 &= [\lambda_2 - \tfrac{1}{2}(s - \tfrac{1}{2})]/(B + K), \end{aligned} \qquad (7.12)$$

we see that (7.6) is isomorphic to

$$\begin{aligned} &(Y_0Y_1) = 0, \qquad (Y_1Y_2) = 0, \\ &(Y_2Y_0) = -2E(\nu_0 - \nu_2)Y_1, \\ &(Y_0Y_3) = \lambda_0 Y_0, \qquad (Y_1Y_3) = -\tfrac{1}{2}Y_1, \\ &(Y_2Y_3) = \lambda_2 Y_2. \end{aligned} \qquad (7.13)$$

[5] A. Z. Petrov, *Einstein Spaces* (State Publishers of Physical Mathematical Literature, Moscow, 1961).

The infinitesimal generators of (7.13) are given by

$$\eta^i{}_a = \begin{bmatrix} 1 & 0 & 0 & \lambda_0 x^0 \\ 0 & 1 & 2E(\nu_0 - \nu_2)x^0 & \lambda_1 x^1 \\ 0 & 0 & 1 & \lambda_2 x^2 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7.14}$$

The reciprocal vectors of (7.6) are given by

$$e^a{}_i = \begin{bmatrix} e^{-\lambda_0 x^3} & 0 & e^{-\lambda_2 x^3} & 0 \\ -2E(\nu_0 - \nu_2)x^2 e^{\frac{1}{2}x^3} & e^{\frac{1}{2}x^3} & 0 & 0 \\ \nu_0 e^{-\lambda_0 x^3} & 0 & \nu_2 e^{-\lambda_2 x^3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{7.15}$$

and the line element takes the form

$$\begin{aligned} ds^2 &= (1 - \nu_0^2)e^{-2\lambda_0 x^3}(dx^0)^2 \\ &+ 2(1 - \nu_0\nu_2)e^{\frac{1}{2}x^3}\, dx^0\, dx^3 + (1 - \nu_2^2)e^{-2\lambda_2 x^3}(dx^2)^2 \\ &- [-2E(\nu_0 - \nu_2)x^2 e^{\frac{1}{2}x^3}\, dx^0 + e^{\frac{1}{2}x^3}\, dx^1]^2 - (dx^3)^2, \end{aligned} \tag{7.16}$$

$$u_i = [e^{-\lambda_0 x^3}, 0, e^{-\lambda_2 x^3}, 0], \tag{7.17}$$

$$\hat{h}_i = [-2E(\nu_0 - \nu_2)x^2 e^{\frac{1}{2}x^3}, e^{\frac{1}{2}x^3}, 0, 0]. \tag{7.18}$$

The line element (7.16) goes to a line element of Sec. 6 if $E \to 0$. This is already obvious from (4.11) and (4.17).

8. SOLUTION OF THE EQUATIONS (5.5). CASE $k_a = 0$ NON-GEODESIC SUBCASE

Looking at the Jacobi identity, part of Eqs. (5.5), i.e., into the equations

$$\begin{aligned} l^2(G + H) - (C - M)(F + P) &= 0, \\ (C + M)(G + H) - (F + P) &= 0, \\ (F - N)(G + H) - (G + K)(F + P) &= 0, \end{aligned} \tag{8.1}$$

we see that there are two possibilities,

$$G + H = 0, \qquad F + P = 0, \tag{8.2}$$

$$G + H \neq 0, \qquad F + P \neq 0. \tag{8.3}$$

We want to develop the corresponding equations separately. In case (8.2), Eqs. (5.5) reduce to the following system:

$$\begin{aligned} A = 1, \quad B = 0, \quad D = 0, \\ E = 0, \quad H = -G, \quad P = -F, \end{aligned} \tag{8.4a}$$

$$\begin{aligned} CM - FN &= -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \\ -CM - GK &= -\tfrac{1}{2}a^2 l^2 + \tfrac{1}{2}b, \\ GK + FN &= \tfrac{1}{2}b, \end{aligned} \tag{8.4b}$$

$$\begin{aligned} FK + (1 + l^2)M - GN &= 0, \\ -CK + GM + (1 - 2l^2)N &= 0, \\ (2 - l^2)K - FM + CN &= 0, \end{aligned} \tag{8.4c}$$

$$l^4 - l^2 + 1 - C^2 - F^2 + G^2 = -\tfrac{1}{2}b. \tag{8.4d}$$

In case (8.3), Eqs. (5.5) take the form

$$l^2 = (C + M)(C - M) \qquad (\neq 0), \tag{8.5a}$$

$$F + P = (G + H)(C + M) \qquad (\neq 0), \tag{8.5b}$$

$$\begin{aligned} F - N &= (G + K)(C + M), \\ CM + NP &= -\tfrac{1}{2}a^2 + \tfrac{1}{2}b, \\ CM - HK &= \tfrac{1}{2}a^2 l^2 - \tfrac{1}{2}b, \\ GK + FN &= \tfrac{1}{2}b, \end{aligned} \tag{8.5c}$$

$$\begin{aligned} l^4 - l^2 + 1 - C^2 + FP - GH &= -\tfrac{1}{2}b, \\ (1 + l^2)M + NH - KP &= 0, \\ (2l^2 - 1)N + CK - GM &= 0, \\ (2 - l^2)K + CN - FM &= 0. \end{aligned} \tag{8.5d}$$

We now solve Eqs. (8.4). By summing Eqs. (8.4b), we obtain

$$b = \tfrac{1}{3}a^2(1 + l^2), \tag{8.6}$$

which can be used instead of, say, the third equation. The other two take the form

$$\begin{aligned} CM - FN &= -\tfrac{1}{6}a^2(2 - l^2), \\ CM + GK &= -\tfrac{1}{6}a^2(1 - 2l^2). \end{aligned} \tag{8.7}$$

Equations (8.4b) can be replaced by (8.6) and (8.7). From (1.11) and (8.6) follows

$$\frac{\chi\rho}{c^2} = \tfrac{1}{2}a^2 - b = \tfrac{1}{6}a^2(1 - 2l^2) \qquad (\geq 0); \tag{8.8}$$

therefore

$$l^2 \leq \tfrac{1}{2} \tag{8.9}$$

must hold.

The determinant d of (8.4c), considered as linear equations or the unknowns K, M, and N, is given by

$$\begin{aligned} d &= (1 + l^2)C^2 + (2 - l^2)G^2 \\ &+ (1 - 2l^2)[F^2 + (1 + l^2)(2 - l^2)], \end{aligned} \tag{8.10}$$

therefore $d \geq 0$ holds [see (8.9)].

If $d > 0$, we have $K = M = N = 0$, which implies $a^2 = 0$ via (8.7), therefore we have to assume that

$$d = 0 \tag{8.11}$$

holds. Using (8.10), we see that (8.11) requires

$$C = 0, \qquad G = 0, \qquad l^2 = \tfrac{1}{2}. \tag{8.12}$$

Using all that, Eqs. (8.4) simplify to the following system:

$$\begin{aligned} &A = 1, \quad B = C = D = E = G = H = 0, \\ &P = -F, \quad l^2 = \tfrac{1}{2}, \quad \chi p/c^2 = 0, \\ &FN = \tfrac{1}{4}a^2, \quad F^2 = \tfrac{1}{4}(a^2 + 3), \\ &FK + \tfrac{3}{2}M = 0, \quad \tfrac{3}{2}K - FM = 0. \end{aligned} \tag{8.13}$$

The last two equations imply

$$K = M = 0. \tag{8.14}$$

Collecting our results, we can say that the nonvanishing components of A_{abc} [see (5.6)] are given by

$$\begin{aligned} A = 1, \quad F = -P = \tfrac{1}{2}(a^2 + 3)^{\frac{1}{2}}, \\ N = \tfrac{1}{2}a^2(a^2 + 3)^{-\frac{1}{2}}. \end{aligned} \tag{8.15}$$

Using (1.11) we see that

$$\chi\tau f = \tfrac{1}{2}a^2, \qquad \chi p/c^2 = 0, \qquad \chi_\mu |h|^2 = \tfrac{1}{2}a^2. \tag{8.16}$$

The tetrad components of the acceleration, the rotation vector [see (5.7) and (5.8)], and the nonvanishing tetrad components of the shear tensor are given, respectively, by

$$\dot{u}_a = [0, 0, 0, \tfrac{1}{2}], \tag{8.17}$$

$$\omega_a = \left[0, -\frac{2a^2 + 3}{4(a^2 + 3)^{\frac{1}{2}}}, 0, 0\right], \tag{8.18}$$

$$\sigma_{23} = \sigma_{32} = \tfrac{3}{4}(a^2 + 3)^{-\frac{1}{2}}. \tag{8.19}$$

This model therefore always has shear and rotation. The commutator relations of the corresponding group are given by

$$\begin{aligned} &(X_0X_1) = (X_1X_2) = (X_2X_0) = 0, \\ &(X_0X_3) = -\tfrac{1}{2}X_0 - (F - N)X_2, \\ &(X_1X_3) = X_1, \\ &(X_2X_3) = (F + N)X_0 - \tfrac{1}{2}X_2, \end{aligned} \tag{8.20}$$

where F and N are given by (8.15).

Substituting into (1.40), one sees that a set of reciprocal vectors of (8.20) is given by

$$e^a{}_i = \begin{bmatrix} e^{\frac{1}{2}x^3}\cos\tfrac{1}{2}\beta x^3 & 0 & e^{\frac{1}{2}x^3}\sin\tfrac{1}{2}\beta x^3 & 0 \\ 0 & e^{-x^3} & 0 & 0 \\ \dfrac{\beta}{2(F+N)}e^{\frac{1}{2}x^3}\sin\tfrac{1}{2}\beta x^3 & 0 & -\dfrac{\beta}{4(F+N)}e^{\frac{1}{2}x^3}\cos\tfrac{1}{2}\beta x^3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{8.21}$$

and the line element has the form

$$ds^2 = e^{x^3}\left[(\cos\tfrac{1}{2}\beta x^3\, dx^0 + \sin\tfrac{1}{2}\beta x^3\, dx^2)^2 - \frac{\beta^2}{4(F+N)^2}(\sin\tfrac{1}{2}\beta x^3\, dx^0 - \cos\tfrac{1}{2}\beta x^3\, dx^2)^2\right] - e^{-2x^3}(dx^1)^2 - (dx^3)^2, \tag{8.22}$$

where

$$\beta^2 = 4(F^2 - N^2). \tag{8.23}$$

Using the notations

$$\begin{aligned} &\lambda_0 = -\tfrac{1}{2} + \tfrac{1}{2}i\beta, \qquad \lambda_2 = -\tfrac{1}{2} - \tfrac{1}{2}i\beta, \\ &\nu_0 = \frac{\frac{1}{2} + \lambda_0}{F + N}, \qquad \nu_2 = \frac{\frac{1}{2} + \lambda_2}{F + N}, \end{aligned} \tag{8.24}$$

$$z^0 = \tfrac{1}{2}(x^0 + ix^2), \qquad z^2 = \tfrac{1}{2}(x^0 - ix^2), \tag{8.25}$$

we can write (8.22) in the following form:

$$\begin{aligned} ds^2 = {} & (1 - \nu_0^2)e^{-2\lambda_0 x^3}(dz^0)^2 \\ & + 2(1 - \nu_0\nu_2)e^{-(\lambda_0+\lambda_2)x^3}\, dz^0\, dz^2 \\ & + (1 - \nu_2^2)e^{-2\lambda_2 x^3}(dz^2)^2 \\ & - e^{-2x^3}(dx^1)^2 - (dx^3)^2. \end{aligned} \tag{8.26}$$

The formal similarity to (6.32) is the consequence of the "formal" similarity between (6.18) and (8.20).

If $a^2 = 0$, then

$$\begin{aligned} &\beta = \sqrt{3}, \qquad F = \tfrac{1}{2}\sqrt{3}, \qquad N = 0, \\ &\lambda_0 = -\tfrac{1}{2} + i\tfrac{1}{2}\sqrt{3}, \qquad \lambda_2 = -\tfrac{1}{2} - i\tfrac{1}{2}\sqrt{3}, \\ &\nu_0 = i, \qquad \nu_2 = -i, \end{aligned} \tag{8.27}$$

and we have Petrov's type I vacuum solution[6]

$$ds^2 = 2e^{-2\lambda_0 x^3}(dz^0)^2 - e^{-2x^3}(dx^1)^2 + 2e^{-2\lambda_2 x^3}(dz^2)^2 - (dx^3)^2 \tag{8.28}$$

[see Ref. 2 Eq. (6.20)].

The line element (8.22) gives all the solutions

[6] A. Z. Petrov, in *Recent Development in General Relativity* (Pergamon Press, Inc., New York, 1962).

corresponding to Eqs. (8.4). Concerning Eqs. (8.5), we remark that it has no "simple" solutions—if it has a solution at all. Looking at Eqs. (5.8) and (5.9), one sees that the vanishing of shear or the vanishing of rotation would require

$$(C + M) = 0 \quad \text{or} \quad (C - M) = 0, \tag{8.29}$$

respectively, which contradicts (8.5a).

9. SOLUTION OF THE EQUATIONS (5.28). CASE $k_a = 0$, GEODESIC SUBCASE

To find the solutions of (5.28) is not hard at all. In this case, the motion is always geodesic and we only give all the shear-free solutions in order to prove Theorem 5. Using (5.31) one sees that the shear-free condition requires

$$P = -N. \tag{9.1}$$

[Equations (5.28) have no solutions with vanishing rotation, a fact easily obtained by using (5.30).] We remark that

$$N \neq 0 \tag{9.2}$$

must hold, otherwise $\frac{1}{2}a^2 = b$ would follow, and [via (1.11)] the condition $p \geq 0$ would be violated.

Using (9.1) and (9.2), Eqs. (5.28) can be reduced to the following system

$$\begin{gathered} A = B = C = D = E = M = 0, \\ P = -N, \quad K = -H, \quad H(F - N) = 0, \\ GH = 0, \quad N^2 = \tfrac{1}{2}a^2 - \tfrac{1}{2}b, \\ H^2 = \tfrac{1}{2}a^2l^2 - \tfrac{1}{2}b, \quad FN = \tfrac{1}{2}b, \end{gathered} \tag{9.3}$$

and we have to find all solutions of this system.

We distinguish two cases

$$H = 0, \tag{9.4}$$

$$H \neq 0. \tag{9.5}$$

In case (9.4), we have the following solution:

$$\begin{gathered} A = B = C = D = E = H = K = 0, \\ G \text{ arbitrary}, \quad b = a^2l^2, \quad P = -N, \\ N = a[\tfrac{1}{2}(1 - l)^2]^{\frac{1}{2}}, \quad F = al^2/[2(1 - l^2)]^{\frac{1}{2}}. \end{gathered} \tag{9.6}$$

Using (1.11), we have

$$\chi\rho f = a^2(1 - l^2), \ \frac{\chi p}{c^2} = \tfrac{1}{2}a^2(1 - 2l^2), \ \chi_\mu |h|^2 = a^2l^2 \tag{9.7}$$

$$l^2 \leq \tfrac{1}{2}, \qquad a[\tfrac{1}{2}(1 - l)^2]^{\frac{1}{2}} \tag{9.8}$$

and from (5.30) it follows that

$$\omega_a = [0, -a[\tfrac{1}{2}(1 - l)^2]^{\frac{1}{2}}, 0, 0), \tag{9.9}$$

and the corresponding group is given by [see (5.32)]

$$\begin{gathered} (X_2X_3) = a[2(1 - l^2)]^{\frac{1}{2}}X_0, \\ (X_3X_0) = -\frac{a(1 - 2l^2)}{[2(1 - l^2)]^{\frac{1}{2}}}X_2, \\ (X_0X_2) = -\frac{a(1 - 2l^2)}{[2(1 - l^2)]^{\frac{1}{2}}}X_3, \end{gathered} \tag{9.10}$$

$$(X_0X_1) = 0, \quad (X_1X_2) = GX_3, \quad (X_3X_1) = GX_2.$$

In case (9.5), we have the following solution:

$$\begin{gathered} A = B = C = D = G = M = 0, \\ K = -H, \quad N = F, \quad P = -F, \\ F = \tfrac{1}{2}a, \quad H = \tfrac{1}{2}a(2l^2 - 1)^{\frac{1}{2}}, \quad l^2 \geq \tfrac{1}{2}. \end{gathered} \tag{9.11}$$

Using (1.11), we have

$$\chi\rho f = a^2(1 - l^2), \quad \chi p/c^2 = 0, \quad \chi\mu |h|^2 = a^2l^2; \tag{9.12}$$

from (5.30), it follows that

$$\omega_a = (0, -\tfrac{1}{2}a, 0, 0), \tag{9.13}$$

and the corresponding group is given by [see (5.32)]

$$\begin{gathered} (X_2X_3) = aX_0, \quad (X_3X_0) = 0, \\ (X_0X_2) = 0, \quad (X_0X_1) = 0, \\ (X_1X_2) = \tfrac{1}{2}a(2l^2 - 1)^{\frac{1}{2}}X_3, \\ (X_3X_1) = -\tfrac{1}{2}a(2l^2 - 1)^{\frac{1}{2}}X_2. \end{gathered} \tag{9.14}$$

We now give the line elements corresponding to (9.10) and (9.14). By substituting into Eqs. (1.40), one sees that

$${}^a{}_i = \frac{1}{a(1 - 2l^2)^{\frac{1}{2}}} \begin{bmatrix} 1 & 0 & e^{x^3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (1/2B)e^{x^3}\cos\tau & \sin\tau \\ 0 & 0 & (1/2B)e^{x^3}\sin\tau & -\cos\tau \end{bmatrix}, \tag{9.15}$$

where

$$B = [(1 - l^2)/2(1 - 2l^2)]^{\frac{1}{2}}, \qquad l^2 < \tfrac{1}{2}, \tag{9.16}$$

$$\tau = -(x^0/2B) + [Gx^1/a(1 - 2l^2)^{\frac{1}{2}}]; \tag{9.17}$$

and

$$e^a{}_i = \begin{bmatrix} 1 & 0 & -ax^3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \dfrac{1}{\sqrt{2}}e^{-\lambda x^1} & \dfrac{1}{\sqrt{2}}e^{\lambda x^1} \\ 0 & 0 & -\dfrac{1}{\sqrt{2}}e^{-\lambda x^1} & \dfrac{1}{\sqrt{2}}e^{\lambda x^1} \end{bmatrix}, \tag{9.18}$$

where

$$\lambda = \tfrac{1}{2}a(2l^2 - 1)^{\frac{1}{2}}, \qquad l^2 \geq \tfrac{1}{2} \tag{9.19}$$

are a set of reciprocal vectors corresponding to (9.10) and (9.14), respectively. The line elements are then given by

$$ds^2 = [1/a^2(1 - 2l^2)][(dx^0 + e^{x^3}\,dx^2)^2 - (1/4B^2)e^{2x^3}(dx^2)^2 - (dx^1)^2 - (dx^3)^2], \qquad l^2 < \tfrac{1}{2} \tag{9.20}$$

$$ds^2 = (dx^0 - ax^3\,dx^2)^2 - (dx^1)^2 - e^{-2\lambda x^1}(dx^2)^2 - e^{2\lambda x^1}(dx^3)^2, \qquad l^2 \geq \tfrac{1}{2}. \tag{9.21}$$

The line element (9.20) coincides with (6.61), that is because the maximal group (6.66) of (6.61) is five parametric having a four-parametric subgroup belonging to the case $k_a = 0$, therefore, the appearance of (9.20) is natural. The maximal group of (9.21) is (9.14), as one easily sees by integrating the relevant Killing equations.

ACKNOWLEDGMENTS

The author is indebted to Professor A. Lichnerowicz and Professor W. Rindler for discussions.

This work was supported in part by the National Aeronautics and Space Administration under Grant No. NSG 269-62, by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under Grant No. AF-AFOSR-903-65, and the Office of Naval Research under Grant No. N00014-66-C0169-A03.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 8. NUMBER 2 FEBRUARY 1967

Twistor Algebra

R. PENROSE
Department of Mathematics, Birkbeck College, London, England
(Received 16 May 1966)

A new type of algebra for Minkowski space–time is described, in terms of which it is possible to express any conformally covariant or Poincaré covariant operation. The elements of the algebra (twistors) are combined according to tensor-type rules, but they differ from tensors or spinors in that they describe locational properties in addition to directional ones. The representation of a null line by a pair of two-component spinors, one of which defines the direction of the line and the other, its moment about the origin, gives the simplest type of twistor, with four complex components. The rules for generating other types of twistor are then determined by the geometry. One-index twistors define a four-dimensional, four-valued ("spinor") representation of the (restricted) conformal group. For the Poincaré group a skew-symmetric metric twistor is introduced. Twistor space defines a complex projective three-space C, which gives an alternative picture equivalent to the Minkowski space–time M (which must be completed by a null cone at infinity). Points in C represent null lines or "complexified" null lines in M; lines in C represent real or complex points in M (so M, when complexified, is the Klein representation of C. Conformal transformations of M, including space and time reversals (and complex conjugation) are discussed in detail in twistor terms. A theorem of Kerr is described which shows that the complex analytic surfaces in C define the shear-free null congruences in the real space M. Twistors are used to derive new theorems about the real geometry of M. The general twistor description of physical fields is left to a later paper.

I. INTRODUCTION

IN the study of Lorentz covariant local field theories, field quantities (with spin) are normally described by vectors, tensors, or, more generally, by spinors, all finite-dimensional representations of the Lorentz group[1] being expressible in terms of spinors (vectors and tensors being regarded as effectively special cases). A local (restricted) Lorentz transformation then takes the form

$$\psi_{E\cdots J'\cdots}^{A\cdots G'\cdots} \to \psi_{Q\cdots V'\cdots}^{M\cdots S'\cdots} t_M^A \cdots \bar{t}_{S'}^{G'} t^{(-1)Q}{}_E \cdots \bar{t}^{(-1)V'}{}_{J'}, \tag{1.1}$$

(t_M^A) being a complex unimodular (2×2) matrix, with inverse $(t^{(-1)A}{}_M)$. Except where general relativity is involved, normally it is further required that physical quantities be suitably covariant under the full Poincaré group.[1] However, this covariance is expressed (locally) in a totally different way from the local Lorentz covariance. Whereas the dependence on space–time *direction* is expressed *algebraically* [cf., (1.1)], the dependence on *position* is described by *differential* equations (so that the analog of (1.1) involves integrals). The invariance of all expressions under the translation

$$x^i \to x^i + a^i \tag{1.2}$$

is generally ensured by the fact that the (Minkowskian) coordinates x^i must, themselves, never enter into the field equations explicitly, only the operators $\partial/\partial x^i$ being permitted to occur.

It is curious that the invariance under two types of transformation (1.1) and (1.2), each of which simply refers to symmetries of the Minkowski space–time, should find mathematical expression in so different a way. This is, moreover, *not* just a property of a difference in group structure between the translation and rotation elements of the Poincaré group, but rather a consequence of *what is meant by a local field theory*. In fact, mathematical formalisms do exist[2] in which *all* the Poincaré transformations are represented according to an algebraic "tensor type" law similar to (1.1). It is the purpose of the present paper to exhibit and discuss in some detail one such formalism, the basic elements of which are re-

[1] Throughout this paper, "Lorentz group" always refers to to the six-parameter *homogeneous* group and "Poincaré group" to the ten-parameter *inhomogeneous* group. "Conformal group" refers to the fifteen-parameter group of transformations which preserve the local conformal structure of Minkowski space–time. If in any particular context it is important to exclude the space reversing and time reversing transformations, then this is made explicit, for example, by the use of the term "restricted".

[2] The use of homogeneous coordinates in space–time would afford a simple example of such a formalism (but apparently one of limited physical interest). More significant, of course, is the representation of physical quantities in terms of Hilbert space, which has the advantage that the (infinite demensional) analogs of $(t_M{}^A)$ are unitary [cf., particularly E. P. Wigner, Ann. Math. 40, 149 (1939) V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. 34, 211 (1948)]. However, it will be essential here to preserve the *finite* dimensionality of the operations at *this* stage, so that geometrical questions can be kept in the forefront.

ferred to here as *twistors*. It will turn out that the twistor algebra will have the same type of universality, in relation to the *conformal* group,[1] that the well-known and highly effective two-component spinor algebra of van der Waerden[3] has, in relation to the Lorentz group. Twistors are, in fact, the "spinors" which are relevant to the six-dimensional space whose (pseudo-) rotation group is isomorphic with the conformal group of ordinary Minkowski space–time.[4] The simplest (non-scalar) twistors constitute a *four-dimensional, four-valued representation of the restricted conformal group*[5] (eight-dimensional if reflections are included). The general twistor is then a many-index quantity constructible from the above basic twistors by means of the usual "tensor type" rules. Any finite-dimensional representation of the conformal group is thus expressible as a direct sum of twistor representations.[4]

The emphasis here will be on the *geometrical* aspects of twistors. It will, in fact, be possible to give a fairly complete geometrical picture of twistors and of their basic operations. Ordinary space–time concepts can then be translated into twistor terms. However, the geometrical expressions of the most immediate twistor concepts have a somewhat nonlocal character. Thus, the primary geometrical object will not be a point in Minkowski space–time, but rather a null straight line or, more generally, a twisting congruence of null lines. Points do, in fact, emerge, but only at a secondary stage. (It also turns out that a natural description of physical fields in twistor terms is given by quantities having a nonlocal space–time interpretation.) However, any vector, tensor, or spinor operation *can* be translated into twistor terms, if desired, and *vice versa*.

All the basic operations of the formalism are conformally invariant. This is of relevance in the study of zero rest-mass fields, since, for each spin, the free-field equations are all effectively conformally invariant.[6,7] This latter fact has particular importance in the asymptotic analysis of such fields,[7] and it would appear that a "manifestly conformally invariant" formalism should be valuable to this type of analysis. Furthermore, although the presence of *mass* breaks conformal invariance, the conformal group appears to have relevance as an approximate symmetry in very high energy physics.[8] When the energies of the particles are high enough, their rest masses can be neglected.

However, the applicability of the twistor formalism is not restricted to situations in which the conformal group is relevant. Since the Poincaré group is a subgroup of the conformal group, it follows that Poincaré covariant operations can be expressed in twistor terms. It merely becomes necessary to introduce a fixed (skew-symmetric) "metric twistor" $I^{\alpha\beta}$ which singles out a particular Minkowski structure (consistent with the given conformal structure). By including operations involving $I^{\alpha\beta}$ with the basic twistor operations, a formalism invariant under the Poincaré group is obtained, in terms of which, any Poincaré covariant operation is, in principle, expressible. Furthermore, by making slightly different alternative choices for $I^{\alpha\beta}$, the corresponding formalisms for a class of space–times including the de Sitter and Einstein cosmologies are obtained.

Perhaps the most obvious drawback of the twistor formalism, however, from the point of view of a possible *fundamental* applicability in physics, is that it is so tied to the idea of a *conformally flat* background space–time, that it is difficult to conceive of how to incorporate the formalism (as it stands) completely into general relativity. It is not impossible that some modification of twistor algebra might be applicable to general curved background space–times. Indeed, some of the basic ideas of the formalism, namely those concerning the null line congruences (cf., the results of Robinson and Kerr referred to in Sec. VIII) have, as their origins, some researches into general relativity. However, any such modification would have to be of a different order from the comparatively straightforward adaption of spinor algebra into general relativity.[9] The point of view one must apparently take, is that whereas the vector–tensor–spinor algebra refers to the im-

[3] B. L. van der Waerden, Nachr. Ges. Wiss. Gottingen **100**, 1 (1929).

[4] F. Klein, *Gesammelte Mathem. Abhandlungen* (J. Springer, Berlin, 1921); cf. also H. Weyl, *The Classical Groups* (Princeton University Press, Princeton, New Jersey, 1939); R. Brauer and H. Weyl, Am. J. Math. **57**, 425 (1935).

[5] Quantities which are essentially twistors have been described by W. A. Hepner, Nuovo Cimento **26**, 351 (1962). See also Y. Murai, Nucl. Phys. **6**, 489 (1958); Progr. Theoret. Phys. (Kyoto) **9**, 147 (1953); **11**, 441 (1954). Similar quantities have also recently gained prominence in, for example, the work of A. Salam, R. Delbourgo, and J. Strathdee, Proc. Roy. Soc. (London) **A284**, 147 (1965).

[6] E. Cunningham, Proc. London Math. Soc. **8**, 77 (1910); H. Bateman, *ibid.* **8**, 223 (1910); P. A. M. Dirac, Ann. Math. **37**, 429 (1936); J. A. McLennan, Jr., Nuovo Cimento **10**, 1360 (1956); H. A. Buchdahl, *ibid.* **11**, 496 (1959).

[7] R. Penrose, Proc. Roy. Soc. (London) **A284**, 159 (1965).

[8] H. A. Kastrupp, Phys. Letters (Amsterdam) **3**, 78 (1962); Phys. Rev. **142**, 1060 (1966). A recent additional suggestion is that the mass splittings of strong interaction physics may be derivable from conformal group symmetry: D. Bohm, M. Flato, D. Sternheimer, and J. P. Vigier, Nuovo Cimento **38**, 1941 (1965). For a discussion of the relevance of the conformal group in physics, see T. Fulton, F. Rohrlich, and L. Witten, Rev. Mod. Phys. **34**, 442 (1962).

[9] L. Infeld and B. L. van der Waerden, S. B. Preuss. Akad. **9**, 380 (1933).

mediate neighborhood of a "point" in space–time, the twistor algebra refers to a more extended "nonlocal" region. It seems to be mathematically more difficult to piece together such regions into a curved space–time, than to build a space–time out of points.

Finally, one of the most important initial motivations for considering a formalism of this kind should be briefly mentioned, although implications of this idea are not discussed in this paper. It concerns possible applications to quantum field theory. One of the most significant quantum mechanical operations is the splitting of field amplitudes into their positive and negative frequency parts. Although this is generally expressed in terms of Fourier transforms, there is the alternative description in terms of singularity-free analytic extensions into the upper or lower complex half-planes.[10] In the present formalism, the idea is to use this second approach and to combine it with global properties of complexified Minkowski space–time. Similarly to the way that the real axis divides the complex plane into two disconnected halves, the space of "null" twistors (describing real Minkowskian null lines) divides twistor space into two disconnected halves, namely the "right-handed" and the "left-handed" twistors (which may be thought of as describing "complexified" null lines).

Free fields of zero rest mass and arbitrary spin can be described particularly conveniently in twistor terms. Moreover, it turns out that they can be generated in a remarkably simple way as contour integrals of arbitrary analytic functions defined in twistor space. (A *null* field is generated when the contour surrounds a simple pole of the function. This leads to a certain generalization of a theorem of Robinson.[11]) If the relevant singularities of this function are associated with one half, rather than the other, of the twistor space, this ensures that the field is of (say) positive frequency. Other properties of zero rest-mass fields also find a natural expression in twistor terms, notably the interrelation with certain types of potential fields[7] and the structure of their total energy-momentum–angular-momentum (which give ten of the fifteen components of a trace-free "Hermitian" twistor E^{α}_{β}). The discussion of these matters is left to a later paper.

Care has been taken here to illustrate that twistor algebra is not merely an abstract formalism, but that the algebraic operations have well-defined meanings in terms of space–time geometry. For whether or not twistors have a significant role to play in future physical theory, the results so far obtained suggest strongly that the formalism ought at least to be thoroughly explored from both geometric and analytic points of view.

[10] See any standard work on quantum field theory. The essential matters referred to here can be found in R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (W. A. Benjamin, Inc., New York, 1964).

[11] I. Robinson, J. Math. Phys. **2**, 290 (1961).

II. TWISTORS AND NULL LINES

As a starting point for the geometrical description of a twistor, consider a null straight line L in Minkowski space–time M. Choose a set of Minkowski coordinates[12] (x^i) for M with origin O. Let l^i be the position vector of some point P on L and let n^i be a future-pointing tangent vector to L (see Fig. 1). If we wish to assign a set of coordinates to the line L, we may do this (Plücker–Grassmann coordinates) by selecting n^i and the moment

$$m^{ij} = l^i n^j - n^i l^j \tag{2.1}$$

of the vector n^i (acting at P) about O. Then, the ratios of the ten quantitites (n^i, m^{ij}) will uniquely define L. (Note that this is independent of the choice of P on L.) This is, however, a highly redundant representation. In addition to the requirement here that n^i be null

$$n^i n_i = 0, \tag{2.2}$$

there are the consistency relations for (2.1)

$$\epsilon_{ijkl} m^{ij} n^k = 0 \tag{2.3}$$

[and also $\epsilon_{ijkl} m^{ij} m^{kl} = 0$, which is implied by (2.3)]. Equation (2.3) in fact represents just three independent conditions which, together with (2.2), reduce the set of nine ratios in (n^i, m^{ij}) to just *five* independent real numbers. This is consistent with

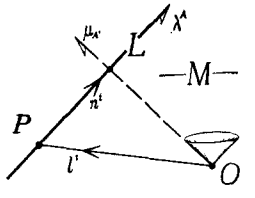

FIG. 1. The representation of a null line.

[12] World vectors and tensors are labeled by lower case Latin indices running over 0,1,2,3, the Minkowski metric being given by $(g^{ij}) = \mathrm{diag}(1,-1,-1,-1)$. Capital Latin index letters, primed or unprimed, denote spinor indices and run over 0,1 or 0′,1′. Greek indices are twistor indices and run over 0,1,2,3. The summation convention applies to each of these four types of index separately. Thus, in particular, no summation takes place between primed and unprimed spinor indices even when the same letter is used, e.g., J and J' are regarded as distinct letters in $x_J{}^{J'}$. This allows us to write the tensor–spinor correspondence in a definite way by simply using the two capital versions of a tensor index as its spinor translation: $x^i \leftrightarrow x^{JJ'}$, etc. (Rindler's convention).

the fact that the null lies in M form an ∞^5 system (for, choosing P as the intersection of L with a fixed spacelike hyperplane, we have ∞^3 choices for P and ∞^2 choices for the null direction at P).

So far, the fact that L is null has not played an essential part. But null directions are very conveniently represented in terms of *two-component spinors*. It will emerge that, by using spinors, we can greatly simplify the description of the null line L; the set of four conditions (2.2), (2.3) being effectively replaced by a *single* condition. Since the following discussion depends essentially on the use of spinors, a very brief review of the ideas required will now be given.[13] The translation from world tensors to spinors is achieved using a quantity[12] $\sigma_i^{JJ'}$ [a Hermitian (2×2) matrix for each j] and its inverse $\sigma^i_{JJ'}$ subject to $\sigma_i^{JJ'}\sigma_k^{KK'}\epsilon_{JK}\epsilon_{J'K'} = g_{ik}$, $\sigma_i^{JJ'}\sigma^i_{JJ'} = \delta^i_j$. The ϵ_{JK}, $\epsilon_{J'K'}$ are skew-symmetric Levi-Civita symbols and are used for raising and lowering spinor indices: $\xi^A\epsilon_{AB} = \xi_B$, $\epsilon^{AB}\xi_B = \xi^A$, i.e., $\xi^0 = \xi_1$, $\xi^1 = -\xi_0$, and similarly for primed indices. Any tensor (e.g., $X^{ij}{}_k$) has a spinor translation, which is written using the same base symbol, but with each tensor index replaced by the corresponding pair of spinor indices, e.g.,

$$X^{ij}{}_k \leftrightarrow X^{II'JJ'}{}_{KK'} = X^{ij}{}_k\sigma_i^{II'}\sigma_j^{JJ'}\sigma^k_{KK'}.$$

Under complex conjugation, the roots of primed and unprimed indices are interchanged, so that reality of tensors is expressed as Hermiticity of spinors. Since M is flat, we can here choose $\sigma_i^{JJ'}$ constant and equal to $2^{-\frac{1}{2}}$ times the unit matrix and Pauli matrices. Then, say,

$$(x^0, x^1, x^2, x^3) \leftrightarrow \begin{pmatrix} x^{00'} & x^{01'} \\ x^{10'} & x^{11'} \end{pmatrix} = 2^{-\frac{1}{2}}\begin{pmatrix} x^0 + x^1 & x^2 + ix^3 \\ x^2 - ix^3 & x^0 - x^1 \end{pmatrix}. \tag{2.4}$$

The spinor translation of a complex null vector α^i (i.e., $\alpha^i\alpha_i = 0$) has the form

$$\alpha^i \leftrightarrow \alpha^{JJ'} = \beta^J\gamma^{J'}. \tag{2.5}$$

If α^i is real and future pointing, then we can take $\gamma^{J'}$ to be the complex conjugate of β^J, i.e.,

$$\alpha^i \leftrightarrow \beta^J\bar{\beta}^{J'}. \tag{2.6}$$

Equation (2.6) implies a geometrical realization of a spinor β^J, up to a phase multiplier, namely as a future-pointing null vector. Moreover, β^J can be completely realized geometrically[14] up to *sign*, in terms of the *bivector*

$$\pi^{ik} \leftrightarrow \beta^J\beta^K\epsilon^{J'K'} + \epsilon^{JK}\bar{\beta}^{J'}\bar{\beta}^{K'}. \tag{2.7}$$

This is real (and null), and it determines a *half-plane element* (the "flag plane") tangent to the light cone along the vector α^i. However, the sign of β^J cannot be realized *locally*[15] geometrically, since a continuous rotation through 2π changes β^J into $-\beta^J$. If we are interested in β^J only up to *proportionality*, then the complete geometrical realization is simply as a null *direction*.

Let us now represent, in spinor terms, the quantities n^i, m^{ij}, which define the null line L. We have

$$n^i \leftrightarrow n^{JJ'} = \lambda^J\bar{\lambda}^{J'} \tag{2.8}$$

and, from (2.1),

$$m^{ik} \leftrightarrow l^{JJ'}\lambda^K\bar{\lambda}^{K'} - \lambda^J\bar{\lambda}^{J'}l^{KK'} = i\epsilon^{JK}\mu^{(J'}\bar{\lambda}^{K')} - i\mu^{(J}\lambda^{K)}\epsilon^{J'K'}, \tag{2.9}$$

where

$$\mu_{A'} = -i\lambda^A l_{AA'}. \tag{2.10}$$

(Round brackets denote symmetrization. The factor $-i$ is for later convenience.) Thus λ^A and $\mu_{A'}$ together determine l^i and m^{ik}. We may think of λ^A as defining the *direction* of L and $\mu_{A'}$ as effectively giving us the *moment* of λ^A ("acting" at P) about O. It is clear from (2.10) that, if λ^A is multiplied by any complex factor, then L is unchanged if $\mu_{A'}$ is multiplied by the same factor. Furthermore, (2.10) implies that $\mu_{A'}$ is independent of the choice of P on L (i.e., if $l_{AA'} \to l_{AA'} + a\lambda_A\bar{\lambda}_{A'}$, then $\mu_{A'}$ is unchanged since $\lambda^A\lambda_A = 0$).

Note a particular choice of P which is of interest, namely the intersection of L with the null cone of O. (This intersection exists uniquely provided L does not lie in any null hyperplane through O.) Then l^i or $-l^i$ has the form (2.6) and it follows from (2.10) that

$$l_{AA'} = i(\lambda^B\bar{\mu}_B)^{-1}\bar{\mu}_A\mu_{A'}. \tag{2.11}$$

[13] For more complete accounts see Refs. 3 and 9 and, for example, W. L. Bade and H. Jehle, Rev. Mod. Phys. **25**, 714 (1953); E, M. Corson, *An Introduction to Tensors, Spinors and Relativistic Wave Equations* (Blackie & Son Ltd., London, 1953); F. A. E. Pirani, in *Brandeis Summer Institute in Theoretical Physics*, 1964, *Lectures on General Relativity* (Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1965), Vol. 1.

[14] E. T. Whittaker, Proc. Roy. Soc. (London) **A158**, 38 (1937); W. T. Payne, Am. J. Phys. **20**, 253 (1952); R. Penrose, "Null Hypersurface Initial Data", in P. G. Bergmann's Aeronautical Research Lab. Tech. Documentary Rept. 63–56 (Office of Aerospace Research, U. S. Air Force, 1963).

[15] To give a rigorous definition of a spinor which takes into account its sign, it is usual to appeal to the theory of fibre bundles. This is not essential, however, and an elementary (nonlocal) geometrical description will be given in an appendix to a forthcoming book by R. Penrose and W. Rindler on the applications of spinors in relativity.

Thus, the null direction defined by $\mu_{A'}$[cf., (2.6)] is that of the null line through O which meets L (Fig. 1). The exceptional case $\lambda^E\bar{\mu}_B = 0$ corresponds to L lying in a null hyperplane through O. [This follows from (2.10), since n^i would be necessarily orthogonal to any choice of l_i.] In this case, λ^A and $\bar{\mu}^A$ are proportional, so the null direction of $\mu_{A'}$ is that of L. More exceptionally still, L might pass through O. This is the case $\mu_{A'} = 0$. On the other hand, for the present, we do not allow $\lambda^A = 0$ (although this will be reconsidered shortly).

The null line L can now be assigned, for coordinates, the three complex ratios of the four complex quantities

$$L^0 = \lambda^0, \qquad L^1 = \lambda^1, \qquad L^2 = \mu_{0'}, \qquad L^3 = \mu_{1'}, \tag{2.12}$$

which we write

$$(L^\alpha) = (\lambda^A, \mu_{A'}). \tag{2.13}$$

These three complex ratios are equivalent to *six* real parameters, so we must expect to find one real relation connecting λ^A and $\mu_{A'}$. This is obtained from (2.10), since the *reality* of the vector l^i implies $l_{AA'}$ is Hermitian, whence

$$\mathrm{Re}\,(\lambda^A\bar{\mu}_A) = 0. \tag{2.14}$$

Provided $\lambda^A \neq 0$, condition (2.14) is also *sufficient* to ensure the existence of a null line L associated with λ^A and $\mu_{A'}$. For, if $\lambda^A\bar{\mu}_A$ is pure imaginary and nonvanishing, then (2.11) gives us a point P, through which we choose L with direction given by λ^A. On the other hand, if $\lambda^A\bar{\mu}_A = 0$, we can easily solve (2.10) to obtain P, with OP spacelike ($l_{AA'}$ of the form $\lambda_A\bar{\nu}_{A'} + \nu_A\bar{\lambda}_{A'}$).

A quantity with components (L^α) given as in (2.13) will be called a *null twistor* (of valence $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$) if (2.14) holds.[16] We shall also be interested in such quantities when (2.14) does *not* hold (and also when λ^A is allowed to vanish). We may regard such a quantity as describing, in some sense, a "complexified" null line [when (2.14) fails], since if we allow l^i to be a *complex* vector, then (2.14) would, in general, be violated. [This is not a straightforward complexification of the null lines in M in the ordinary sense, however, since the real dimensionality of the system of lines is only increased from five to six by dropping (2.14); but cf., Sec. VI.] If $\mathrm{Re}\,(\lambda^A\bar{\mu}_A) > 0$, we refer to L^α as a *right-handed twistor* (valence $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$) and if $\mathrm{Re}\,(\lambda^A\bar{\mu}_A) < 0$, a *left-handed twistor*. The reason for this terminology will become evident later. When $\lambda^A = 0$, we shall still refer to L^α as a *null* twistor even though no finite null line L is represented. The role of such twistors will emerge shortly.

III. INCIDENCE OF NULL LINES IN TWISTOR TERMS

The algebraic rules for the manipulation of twistors will have, as their basis, the idea of *incidence* between null lines. Let us consider, in addition to L, a second null line X, described by the null twistor X^α:

$$(X^\alpha) = (\xi^A, \eta_{A'}), \tag{3.1}$$

where ξ^A and $\eta_{A'}$ are the analogs of λ^A, $\mu_{A'}$ above. That is, ξ^A defines the direction of X and we have

$$\eta_{A'} = -i\xi^A x_{AA'}, \tag{3.2}$$

where x^i is the position vector of a point on X (see Fig. 2). Suppose now that X and L do intersect. Then we may choose $l^i = x^i$ for the coordinate vector of this intersection point, whence by (2.10) and (3.2),

$$\xi^A\bar{\mu}_A = i\xi^A l_{AA'}\bar{\lambda}^{A'} = i\xi^A x_{AA'}\bar{\lambda}^{A'} = -\eta_{A'}\bar{\lambda}^{A'}. \tag{3.3}$$

Let us define the *complex conjugate* of a twistor L^α to be $\bar{L}_\alpha$ (valence $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$), where

$$(\bar{L}_\alpha) = (\bar{\mu}_A, \bar{\lambda}^{A'}) \tag{3.4}$$

when L^α is given by (2.13) [irrespective of the condition (2.14), or whether or not λ^A vanishes]. In component form[17]:

$$\bar{L}_0 = \overline{L^2}, \qquad \bar{L}_1 = \overline{L^3}, \qquad \bar{L}_2 = \overline{L^0}, \qquad \bar{L}_3 = \overline{L^1}. \tag{3.5}$$

Then, (3.3) tells us that a necessary condition for

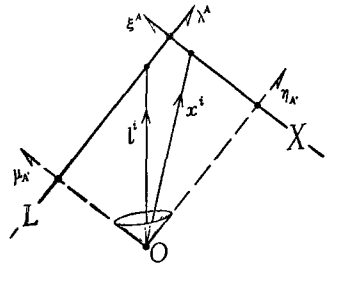

FIG. 2. Incidence of two null lines.

[16] Any temptation to identify the twistor (2.13) with a Dirac spinor should be rejected here, since their transformation properties are quite different [cf., for example, (7.13), (7.17)].

[17] The operation of twistor (or spinor) complex conjugation is denoted by a bar extending only over the base symbol involved and not over the indices. If the bar extends also over the indices, this denotes simply the complex conjugate of the complex number that the symbol represents.

the null lines L and X to meet, in terms of their corresponding twistors, is

$$X^\alpha \bar{L}_\alpha = 0, \tag{3.6}$$

since

$$X^\alpha \bar{L}_\alpha = \xi^A \bar{\mu}_A + \eta_{A'} \bar{\lambda}^{A'}. \tag{3.7}$$

Note, further, that the condition (2.14) for L^α to represent a real null line is simply

$$L^\alpha \bar{L}_\alpha = 0. \tag{3.8}$$

We can think of (3.8) as a special case of (3.6) since L intersects itself! We also have, here,

$$X^\alpha \bar{X}_\alpha = 0. \tag{3.9}$$

Condition (3.6) is, conversely, also *sufficient* for the null lines L and X to intersect, where we assume, for the moment, that λ^A and ξ^A are not proportional, so that L and X are not parallel. For, we then have $\xi^A \lambda_A \neq 0$ and we can construct the (complex) vector

$$p_i \leftrightarrow p_{JJ'} = (i/\xi^A \lambda_A)(\lambda_J \eta_{J'} - \xi_J \mu_{J'}). \tag{3.10}$$

We observe that

$$\mu_{A'} = -i\lambda^A p_{AA'}, \qquad \eta_{A'} = -i\xi^A p_{AA'}.$$

Thus, when p_i is *real*, we can satisfy (2.10) and (3.2) by putting $l_i = p_i = x_i$, whence L and X must intersect. (In fact, it is valid to regard the "complex point" with position vector p^i as the "intersection" of L and X even in the cases when p_i is not real, as we shall see later.) Now p_i is real if and only if $p_{JJ'}$ is Hermitian. Since λ^A and ξ^A are not proportional, we can test the Hermiticity of $p_{JJ'}$ by taking components with respect to λ^A, ξ^A:

$$\lambda^A \bar{\lambda}^{A'}(p_{AA'} - \bar{p}_{A'A}) = i\mu_{A'} \bar{\lambda}^{A'} + i\lambda^A \bar{\mu}_A = iL^\alpha \bar{L}_\alpha,$$

$$\xi^A \xi^{A'}(p_{AA'} - \bar{p}_{A'A}) = i\eta_{A'} \xi^{A'} + i\xi^A \bar{\eta}_A = iX^\alpha \bar{X}_\alpha,$$

$$\xi^A \bar{\lambda}^{A'}(p_{AA'} - \bar{p}_{A'A}) = i\eta_{A'} \bar{\lambda}^{A'} + i\xi^A \bar{\mu}_A = iX^\alpha \bar{L}_\alpha.$$

Thus, the reality of p_i is a consequence of (3.8), (3.9), and (3.6), so that in the case of twistors representing nonparallel null lines, we can state the condition that the lines meet simply as the orthogonality condition (3.6) between the twistors.

We may further ask what is the geometrical meaning of this condition $X^\alpha \bar{L}_\alpha = 0$ when X and L are real *parallel* null lines. Perhaps the simplest way to see the result is to use a limiting argument. Keep L_α fixed but vary X^α, so that $X^\alpha \bar{L}_\alpha = 0$ throughout the motion. Suppose L and X are initially not parallel, but become parallel in the limit. The intersection point recedes to infinity along L. The possible positions of X, for each *finite* position of the intersection point, are the generators of the null cone of this point. As the point recedes to infinity, the null cone becomes, in the limit, the null hyperplane through X. Thus, when L and X are parallel, the condition $X^\alpha \bar{L}_\alpha = 0$ becomes the condition that *L and X lie in the same null hyperplane.*

In fact, it is convenient to regard all the null lines of a null hyperplane as intersecting in a single *point at infinity.* Thus, we adjoin to our space M a set of ∞^3 such points at infinity, one for each of the null hyperplanes in M. The geometrical role played by the twistors L_α with $\lambda^A = 0$ (but $\mu_{A'} \neq 0$) now emerges. For, if $\lambda^A = 0$, the condition $X^\alpha \bar{L}_\alpha = 0$ becomes $\xi^A \bar{\mu}_A = 0$ [cf., (3.7)]. That is to say, *the direction of X coincides with the null direction represented by $\mu_{A'}$*, so that for *fixed* L^α the corresponding lines X are all parallel. Each null hyperplane of parallel lines X gives rise to *one* point at infinity; the aggregate of all these points at infinity, corresponding to all these parallel hyperplanes, gives an ∞^1 system of points at infinity, namely the points of the *null line at infinity* L. In fact, to complete the picture, we must add one further point at infinity, not lying on any finite line, which we call I. The point I is common to all the lines L at infinity. The fact that any two null lines at infinity must, for consistency here, be considered to intersect, follows from (3.7), since if $\lambda^A = 0$ and also $\xi^A = 0$, then $X^\alpha \bar{L}_\alpha = 0$ automatically follows. The point I then plays the part of the *vertex* of a *null cone at infinity* the generators of which are the lines at infinity considered above (see Fig. 3).

The structure of the completed ("compactified") Minkowski space arrived at in this way, by adding a (closed) null cone at infinity, is one which has been considered by a number of authors.[18,7] We henceforth use the symbol M, here, to refer to the *entire* completed space and not just to the set of finite points. The set of finite points is instead denoted by $M\{I\}$ to indicate that the null cone of the

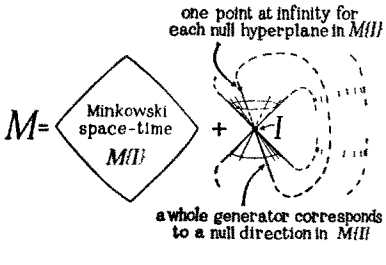

FIG. 3. The construction of the compact manifold M.

[18] N. H. Kuiper, Ann. Math. 50, 916 (1949); H. Rudberg, dissertation, University of Uppsala, Uppsala, Sweden (1958); R. Penrose, in *Proceedings of the 1962 Conference on Relativistic Theories of Gravitation, Warsaw* (Polish Academy of Science, Warsaw, 1965) A. Uhlmann, Acta Phys. Polon. 24, 293 (1963). Rudberg also mentions the four-valuedness of spinors.

particular point I has been removed from M. The geometry of M is briefly as follows: M is a compact nonsingular manifold without boundary which has a well-defined conformal structure everywhere (i.e., a metric defined up to proportionality). It has a transitive ∞^{15} group of motions preserving this conformal structure, every point being on an equal footing with every other point. M contains ∞^5 null "straight" lines (geodesics) each of which is topologically a circle. Through each point P of M pass ∞^2 of these lines generating the (closed up) null cone of P. The removal of any one of these null cones leaves a space with Euclidean topology which can be assigned a Minkowskian metric (unique up to dilatations) consistent with the given conformal structure. The entire space M has the topology $S^3 \times S^1$ and can be realized as a real projective quadric fourfold with signature $(++-----)$.

The null lines in M are in one-to-one correspondence with the proportionality classes of twistors L^α (i.e., κL^α represents the same line as L^α, $\kappa \neq 0$) which satisfy $L^\alpha \bar{L}_\alpha = 0$ and $L^\alpha \neq 0$. Incidence between null lines in M is expressed as orthogonality ($X^\alpha \bar{L}_\alpha = 0$) between the corresponding twistors.

IV. ROBINSON CONGRUENCES

Thus far, the geometrical description of a twistor given here has been restricted to the case of null twistors (of valence $[^1_0]$). As mentioned earlier, if on the other hand $L^\alpha \bar{L}_\alpha \neq 0$, then we may think of L^α as representing, in some sense, a kind of "complexified" null line L. This is, in fact, the way we tend to view "geometrically" a general twistor of valence $[^1_0]$. However, it is significant that a much more precise realization can be given in terms of congruences[19] of null lines. In order to fix our ideas, we regard such a congruence as representing, rather, a twistor of valence $[^0_1]$ (e.g., $\bar{L}_\alpha$), although this makes little difference in practice.

Now, any (real) null line L in M can be completely characterized by *the system of all null lines which meet it.* In twistor terms, that is, if we know the set of all X^α satisfying $X^\alpha \bar{L}_\alpha = 0$, $X^\alpha \bar{X}_\alpha = 0$ (where $L^\alpha \bar{L}_\alpha = 0$, $L^\alpha \neq 0$), then we know L^α up to proportionality. Thus, while we think of L^α as describing the null line L in M, we can think of $\bar{L}_\alpha$ as describing the *congruence* L of null lines (i.e., a three-dimensional system[19] of lines in M) which meet L. In exactly the same way we can represent a general twistor R_α of valence $[^0_1]$, in terms of the congruence R of lines X whose twistors X^α (with $X^\alpha \bar{X}_\alpha = 0$) satisfy $X^\alpha R_\alpha = 0$. If $R_\alpha \bar{R}^\alpha \neq 0$, such a congruence is referred to here as a *Robinson congruence.*[20] [If $R^\alpha \bar{R}^\alpha = 0$, then we denote by $\bar{R}$ the line which is met by all the lines of the congruence R. On the other hand, if L denotes the line, then $\bar{L}$ denotes the corresponding congruence. In general, we adopt the convention that the operation of applying a bar to any aymbol is its own inverse. From this it follows that the definition of $\bar{R}^\alpha$, given R_α, exactly mirrors the definition (3.5) of $\bar{L}_\alpha$, given L^α, since we could put $L^\alpha = \bar{R}^\alpha$. Thus,[17] $(\bar{R}^\alpha) = (\bar{R}^0, \bar{R}^1, \bar{R}^2, \bar{R}^3) = (\overline{R_2}, \overline{R_3}, \overline{R_0}, \overline{R_1})$.] If $R_\alpha \bar{R}^\alpha > 0$, we refer to the Robinson congruence R as right-handed. If $R_\alpha \bar{R}^\alpha < 0$, it is left-handed.

Since Robinson congruences occupy such a basic position in the geometry of twistors, it seems worthwhile to examine a particular such congruence in some detail here. (In fact, any particular case is completely representative of the general case, since it turns out that any two Robinson congruences can be transformed one into the other by a Poincaré transformation, perhaps involving a reflection.)

Choose a real number ϵ and put

$$(R_\alpha) = (\epsilon/\sqrt{2}, 0, 1, 0) \tag{4.1}$$

so that

$$R_\alpha \bar{R}^\alpha = \epsilon\sqrt{2}. \tag{4.2}$$

Let X^α be as in (3.1): $(X^\alpha) = (\xi^0, \xi^1, \eta_{0'}, \eta_{1'})$, so that the condition for the line X to belong to the conrgruence R is

$$\xi^0 \epsilon/\sqrt{2} + \eta_{0'} = 0. \tag{4.3}$$

Write the position vector x^i of a point on X as

$$x^0 = t, \qquad x^1 = z, \qquad x^2 = x, \qquad x^3 = y.$$

Equation (3.2), with (2.4), gives

$$(\eta_{0'}, \eta_{1'}) = -(i/\sqrt{2})(\xi^0, \xi^1)\begin{pmatrix} t-z & -x+iy \\ -x-iy & t+z \end{pmatrix}.$$

Using (4.3), we then get $-\epsilon\xi^0 = i\xi^0(-t+z) + i\xi^1(x+iy)$, i.e.,

$$\xi^0 : \xi^1 = x + iy : t - z + i\epsilon. \tag{4.4}$$

Now $\xi^J \bar{\xi}^{J'}$ corresponds to a tangent vector to X, so that

$$\begin{pmatrix} \xi^0\bar{\xi}^{0'} & \xi^0\bar{\xi}^{1'} \\ \xi^1\bar{\xi}^{0'} & \xi^1\bar{\xi}^{1'} \end{pmatrix} \propto \begin{pmatrix} dt+dz & dx+i\,dy \\ dx-i\,dy & dt-dz \end{pmatrix}.$$

[19] The term "congruence" is used to denote a system of curves (or surfaces, etc.) for which there is just one member of the system (or at most a discrete number) through a general point in the space.

[20] First investigated by I. Robinson (private communication).

Combining this with (4.4) we get the differential equations for the line X

$$dt + dz : dx - i\,dy = dx + i\,dy : dt - dz = x + iy : t - z + i\epsilon. \quad (4.5)$$

The solution of this is

$$\begin{aligned} t - z + i\epsilon &= (x + iy)\alpha, \\ (x - iy) - (t + z)\alpha &= \beta, \end{aligned} \quad (4.6)$$

where α and β are complex constants defining the particular line X of R, for consistency satisfying

$$\mathrm{Im}\,(\bar{\beta}\alpha) = \epsilon. \quad (4.7)$$

In order to get a more visualizable picture of the situation, let us consider the intersection with a spacelike hyperplane $t = \tau$ (τ const). Each null line X meets this hyperplane in one point (τ, x, y, z). The direction of X there can be represented by considering the projection into the hyperplane of the tangent vector to X at this point. These projected tangent vectors are then tangents to a series of curves in the hyperplane which should give some kind of a picture of the structure of the Robinson congruence. To find the differential equations of these curves, we simply replace t by τ in (4.5) and dt by $ds = (dx^2 + dy^2 + dz^2)^{\frac{1}{2}}$. Then we get

$$(x + iy)(ds - dz) = (\tau - z + i\epsilon)(dx + i\,dy).$$

The solutions of this equation are given (apart from the spurious $x + iy = \text{const}$) by

$$\left.\begin{aligned} x^2 + y^2 + (z - \tau)^2 - 2\epsilon(x \sin\varphi \\ + y\cos\varphi)\tan\theta = \epsilon^2, \\ z - \tau = (x\cos\varphi - y\sin\varphi)\tan\theta, \end{aligned}\right\} \quad (4.8)$$

where θ and φ are constants defining the different curves. These curves are evidently circles, being intersections of spheres with planes. They twist around one another in such a way that every pair of circles is linked. The twisting has a *positive* screw sense if $\epsilon > 0$, i.e., if R_α is right-handed.[21] They lie on the set of coaxial tori[22] obtained by eliminating φ between the two equations. [These tori are the rotations about the z axis of a system of coaxial circles in the (x, z) plane.]

From the point of view of the completed space–time M, we should regard the hyperplane $t = \tau$ as being completed (conformally) by a point (namely I) at infinity. It then becomes topologically a three-dimensional sphere S^3 (of which the hyperplane $t = \tau$ may be regarded as the stereographic projection). The vector field on S^3 is everywhere non-singular and nowhere vanishing. The circles constitute an example of what is known as the Hopf fibring of S^3 (one example being "Clifford parallels" on S^3).

Note that all the circles in the hyperplane thread through the particular (smallest) circle of radius $|\epsilon|$ and centre $z = \tau$, $x = y = 0$ given when $\theta = 0$. If ϵ is small, this circle describes, as τ increases, a path approximating that of a null line. If ϵ is zero, the path is exactly the null line $z = t$, $x = y = 0$. For small ϵ, we may think of the lines of the Robinson congruence as defining an approximate null line, but the lines twist around one another and never quite meet. The twisting has a positive or negative screw sence according as R_α is right- or left-handed. [In the limit $\epsilon \to 0$, the circles (4.8) all touch the z axis at $z = \tau$. The tangents to these circles are orthogonal to the spheres touching the (x, y) plane at $z = \tau$, these spheres being the intersections of $z = \tau$ with the null cones with vertices on $z = t$, $x = y = 0$. The lines of the congruence are then just the generators of these null cones in this case, as expected.]

V. THE ASSOCIATED SPINOR FIELD OF A TWISTOR

We saw above that the tangent vectors to the lines of the Robinson congruence constituted a field of null vectors which was regular and nonvanishing everywhere on M. The general Robinson congruence R can be written compactly in spinor terms in a way which exhibits this (and other) facts very simply. This leads to an interpretation of twistors in terms of certain spinor fields.

Put $R_\alpha = \bar{L}_\alpha$, so we can use the notation of Secs. II and III. The condition $X^\alpha \bar{L}_\alpha = 0$, for X to belong to the congruence $\bar{L}$ is then, by (3.7), (3.2),

$$0 = \xi^A \bar{\mu}_A + \eta_{A'} \bar{\lambda}^{A'} = \xi^A(\bar{\mu}_A - i x_{AA'} \bar{\lambda}^{A'}),$$

from which it follows that ξ_A is proportional to the expression in the bracket. Since X defines X^α only up to proportionality, we are at liberty to choose the scale factor for ξ^A so that

$$\xi_A = \bar{\mu}_A - i x_{AA'} \bar{\lambda}^{A'}. \quad (5.1)$$

We can think of (5.1) as defining a *spinor field*, since ξ_A is a function of x_i, the associated null vectors $v^i \leftrightarrow \xi^J \bar{\xi}^{J'}$ being tangents to the lines of

[21] The screw sense arising here depends, of course, on the "handedness" of the choice of matrices in (2.4).

[22] I am grateful to J. Terryl and J. E. Reeve for this observation concerning the Hopf fibring.

the congruence $\bar{L}$. Since $\lambda^A x_{AA'}\bar{\lambda}^{A'}$ is real, it follows that

$$\mathrm{Re}\,(\lambda^A\xi_A) = \mathrm{Re}\,(\lambda^A\bar{\mu}_A) = \tfrac{1}{2}L^a\bar{L}_a, \qquad (5.2)$$

so that ξ_A cannot vanish unless L^a is null. The field of null directions defined by (5.1), i.e., by the Robinson congruence $\bar{L}$, is thus *well defined and regular* throughout $M\{I\}$. The regularity at infinity also follows from the conformal transformation properties of (5.1) as we are to see shortly.

Now write

$$\nabla_{BB'} \equiv \partial/\partial x^{BB'}, \qquad (5.3)$$

so that $\nabla_{BB'}x_{AA'} = \epsilon_{AB}\epsilon_{A'B'}$. Then we have, from (5.1),

$$\nabla_{BB'}\xi_A = i\epsilon_{BA}\bar{\lambda}_{B'}. \qquad (5.4)$$

Thus, the field ξ_A satisfies the differential equation

$$\nabla^{B'}_{(B}\xi_{A)} = 0. \qquad (5.5)$$

Conversely, from this equation we can reconstruct the form of the expression (5.1). For (5.5) implies that for *some* spinor $i\bar{\lambda}_{B'}$, the derivative $\nabla_{BB'}\xi_A$ has the form given in (5.4). Furthermore, this $\bar{\lambda}_{B'}$ must be constant because $\nabla^{C'}_{C}\nabla^{B'}_{B}\xi_A = 0$ [as follows from (5.5), since skew-symmetry in B, A *and* in C, A implies that the expression vanishes]. Integration of (5.4) leads straight to (5.1), where $\bar{\mu}_A$ is another constant.

We now have the important result that *there is a one-to-one relation between twistors of valence* $[^0_1]$ *and spinor fields satisfying* (5.5). This gives us a more complete representation of a twistor than just as a Robinson congruence, since the *factor of proportionality* of L^a is now also represented. Note that $L^a\bar{L}_a$ is also expressible simply in terms of ξ_A as

$$L^a\bar{L}_a = \mathrm{Im}\,(\bar{\xi}^{B'}\nabla^{A}_{B'}\xi_A). \qquad (5.6)$$

But since we are interested in twistors in relation to the *whole* of M, rather than just $M\{I\}$, we should also ask how the field ξ_A is to be defined at infinity. Essentially all that is required here is to verify that (5.5) is invariant under conformal transformation [e.g., under inversions, since then I becomes a finite point,[7] cf., (7.18)]. This invariance is achieved[23] by specifying that ξ_A transform as a *conformal density of weight* $\frac{1}{2}$. Then the field equation (5.5) is also satisfied at infinity, in the sense that it holds with respect to some metric which is regular on the null cone of I. This, in particular, defines the Robinson congruence $\bar{L}$ regularly at infinity. (In fact, there is one line of $\bar{L}$ totally at infinity, namely that corresponding to the direction of λ^A.) The quantity (5.6) is also invariant[23] under conformal transformation (ξ_A of weight $\frac{1}{2}$), so we can refer to $L^a\bar{L}_a$ as the *conformal invariant* of the spinor field ξ_A or of the twistor L^a.

Some subtleties remain, however, concerning the extension of the field ξ_A across infinity. The possibility of defining two-valued (spinor) fields in $M\{I\}$ is well known. But here we require a spinor field defined *globally* on M. The multiple connectedness of M is such as to cause a slight difficulty in this respect. But in any case, we can think, instead, in terms of $\xi_A\xi_B$ or the associated bivector [cf., (2.7)], since these are not two-valued. Then we might expect to find a representation of a twistor up to sign. However, it turns out [cf., (7.25)] that twistors of odd total valence are *four*-valued under the conformal group. Thus, any such conformally invariant representation of a twistor (of odd valence), cannot distinguish the twistor from i times the twistor. (This is analogous to the fact that spinors of odd valence cannot be completely represented in terms of tensors, in a Lorentz covariant way, whereby the spinors are distinguished from their negatives.) In the present case, when we extend the field ξ^A across infinity, we find that the field on the two sides of the null cone of I cannot be exactly matched, but ξ^A on one side must be matched with $i\xi^A$ (or $-i\xi^A$) on the other side. [We can see this explicitly in terms of the inversion of (7.18), for which the conformal factor is $\Omega = x^i x_i/2a^2$. The null cone of I is transformed from the null cone, $\Omega = 0$, of the origin, across which Ω changes sign. The transformed ξ^A field picks up the factor $\Omega^{-\frac{1}{2}}$.] Thus, the spinor field (5.1), when defined over the *whole* of M, must be thought of as *four-valued*; if ξ_A is one value, then $i\xi_A$, $-\xi_A$, $-i\xi_A$ are the other three.

Since spinors can be represented geometrically (up to sign), this implies a geometrical realization of any twistor of valence $[^0_1]$ up to a multiple of 1, i, -1, or $-i$. The phase of ξ_A defines, as we have seen [cf., (2.7)], a half-plane element tangent to the light cone; $-\xi_A$ defines the same half-plane element; but $i\xi_A$ and $-i\xi_A$ define the opposite half-plane element. Thus, a complete unoriented null plane element is defined at each point by the phase of $\bar{L}_a$. As we follow one of the lines X belonging to the Robinson congruence $\bar{L}$, we find that this plane element rotates about X in the opposite direction from the neighboring lines of the congruence and twice as fast. Owing to the way connections are

[23] See Ref. 7, Eqs. (10.1), (10.7), and (10.8) for the relevant formulas.

made at infinity, this results in the aggregate of plane elements at points of X, effectively constituting a Möbius strip, which explains why the plane elements cannot be consistently oriented along X.

To see how these plane elements rotate about X, consider $(\xi^B\bar{\xi}^{B'}\nabla_{BB'})\xi_A = i\bar{\xi}^{B'}\bar{\lambda}_{B'}\xi_A$ [see (5.4)]. The real part of $2i\bar{\xi}^{B'}\bar{\lambda}_{B'}$ measures the rate of extension of the vector represented by ξ_A and the imaginary part measures the rate of rotation of the half-plane element about the null direction of ξ_A. By (5.2),

$$\text{Im}\,(2i\bar{\xi}^{B'}\bar{\lambda}_{B'}) = -2\,\text{Re}\,(\xi_B\lambda^B) = -L^a\bar{L}_a,$$

whence the half-plane element rotates in a negative or positive sense according as the twistor L^a is right or left handed. (It might appear that the rate of rotation is constant here but this is misleading since the scaling is different at different points of X.) Note that if L^a is null, there is no rotation of the half-plane element. We still get a Möbius strip, however, since a half-twist always occurs, in effect, at infinity.

To see how the neighboring lines of the congruence rotate about X, consider

$$\xi^A\bar{\xi}^{B'}\nabla_{BB'}\bar{\xi}_A = -i\bar{\xi}^{B'}\bar{\lambda}_{B'}\xi_B. \tag{5.7}$$

Here, the real part of $-i\bar{\xi}^{B'}\bar{\lambda}_{B'}$ measures the convergence of the null lines and the imaginary part measures their rotation about one another.[24] The rotation is in the opposite sense from that of the half-plane elements and half as fast.

The expansion of the null lines and the change in the magnitude of ξ_A are not conformally covariant concepts, whereas the rotations are. Another conformally covariant concept for a congruence of null lines is the *shear* of these lines.[11,24] In the present case we have

$$\xi^B\xi^A\nabla_{BB'}\xi_A = 0, \tag{5.8}$$

which states that the shear *vanishes*.

VI. THE COMPLEX PROJECTIVE SPACE C

We have seen that the null lines in M form an ∞^5 system which can be extended to an ∞^6 system by including "complexified" null lines (the latter being representable in terms of certain related structures called Robinson congruences). The members of this ∞^6 system can be given complex projective coordinates $(L^a) = (L^0, L^1, L^2, L^3)$. That is, it is just the three complex ratios $L^0 : L^1 : L^2 : L^3$ which are significant, the only restriction on $L^0, \cdots, L^3$ being that they must not *all* vanish. This ∞^6 system we may think of as constituting a *three-dimensional complex projective space* which we denote by the letter C. The "points" of C are just the "complexified" null lines (and the null lines) of M.

In fact, we have two alternative pictures of any given situation, namely the one in terms of M and the one in terms of C. For example, we may think of an object L, with projective coordinates (L^a) *either* as, say, a "complexified" null line of the real space–time M (M picture) *or* simply as a point in a certain projective three space (C picture). These are just two different ways of *visualizing* what is the same physical situation in each case. In order that the two pictures be completely equivalent, however, we need to be able to interpret, in C, the condition of "*reality*" of a null line in M, of *incidence* between null lines in M, and finally of *points* in M. In effect, this requires that the conjugation relation $L^a \leftrightarrow \bar{L}_a$ should have a meaning with regard to the C picture. Now, we have seen that a twistor L^a (valence $\begin{bmatrix}1\\0\end{bmatrix}$) refers to a point L of C; a twistor R_a (valence $\begin{bmatrix}0\\1\end{bmatrix}$) therefore refers to the dual concept of a *plane* R in C, namely the plane of all points X for which $X^aR_a = 0$. (This is, of course, a plane in the *complex* sense. As a *real* manifold it is a four-dimensional subset of the six-real-dimensional manifold C.) The conjugation relation $L^a \leftrightarrow \bar{L}_a$ therefore describes a point $\leftrightarrow$ plane correspondence in C, which we may refer to as a *Hermitian correlation* of signature $(++--)$. The signature here refers to the Hermitian form

$$X^a\bar{X}_a = X^0\overline{X^2} + X^1\overline{X^3} + X^2\overline{X^0} + X^3\overline{X^1}. \tag{6.1}$$

We regard this Hermitian correlation as being an *intrinsic* part of the geometric structure of C.

The (real) null lines in M are the points of the five-real-dimensional subset N (with topology $S^3 \times S^2$) of C defined by the equation $X^a\bar{X}_a = 0$. Thus, N is a hypersurface if we regard C as a real six-dimensional manifold, but it is not a hypersurface in the sense of the *complex* structure of C. We refer to the subset of C for which $X^a\bar{X}_a > 0$ holds as C^+ and the part for which $X^a\bar{X}_a < 0$ as C^-. The two sets C^+ and C^- are then disconnected from one another and have N as their common boundary (Fig. 4). If L is any point of C, we may regard the plane $\bar{L}$ as the *polar plane* of L with respect to N, since "polarizing" $X^a\bar{X}_a$ with L^a yields $X^a\bar{L}_a = 0$ (or $L^a\bar{X}_a = 0$), the equation of the plane $\bar{L}$. The Robinson congruence associated with L is now the intersection (topology S^3) of the plane $\bar{L}$ with N. (A slight inconsistency of notation arises here in that

[24] P. Jordan, J. Ehlers, and R. Sachs, Akad. Wiss. Lit. Mainz, no. 1 (Mainz 2) (1961); E. T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).

it is the Robinson congruence $\bar{L} \cap N$ that had been previously labeled $\bar{L}$, rather than the entire plane. The Robinson congruence includes only "real" null lines of M, by definition. In cases where there is possibility of confusion, we refer to "the Robinson congruence $\bar{L}$" or "the plane $\bar{L}$" as the case may be, but in general $\bar{L}$ refers to an entire plane in C.) When L lies on N, the plane $\bar{L}$ can be thought of as the *complex tangent plane to N at L*. This is also just the case when L *lies on its polar plane.*

Now, if we wish to represent a *point of M* in terms of the C picture, we may do this using incidence properties of null lines in M. Any point P in M can be uniquely represented by an ∞^2 system of null lines in M, namely by the generators of the null cone of P. Let K and L be two null lines in M through P. The generators of the null cone of P are then the null lines common to both $\bar{K}$ and $\bar{L}$ (i.e., the generators must meet both K and L). In the C picture this is an ∞^2 system[25] of lines on N which lie on the intersection of the two planes $\bar{K}$ and $\bar{L}$. This intersection is simply a complex projective (straight) line in C. Thus, we have the result that any point of M is represented, in the C picture, by a complex projective line which lies entirely on N. (A complex projective line is topologically a sphere S^2, in terms of its real structure. This agrees with the topology of the set of null lines through P, i.e., of the null directions at P.) We do not distinguish, notationally, here, between the point P in M and the system of null lines in M through P. Then we are at liberty to denote also by P the *line* in the C picture which represents this *point* P in the M picture.

Conversely, *any* line P in the C picture, which lies entirely on N, represents some point P in the M picture. (In the terminology of projective geometry, in the C picture, "line" always implies "complex projective straight line.") To see this, consider the C picture and let the line P lie entirely on N. Let K and L be two points on P. Then we have $K^a\bar{K}_a = 0$, $L^a\bar{L}_a = 0$, and, more generally $(K^a + \beta L^a)(\bar{K}_a + \bar{\beta}\bar{L}_a) = 0$ for all complex β (the general point on the line P having a twistor of the form $K^a + \beta L^a$). Hence $L^a\bar{K}_a = 0$. Thus, in the M picture, the null lines L and K must intersect. This holds for *any* pair of null lines belonging to the ∞^2 system represented by P. These null lines must therefore all meet in a *point* and, in fact, must be the generators of the null cone of this point. We thus label this point P, and the situation is as before.

[25] A symbol ∞^r indicates r dimensionality in the *real* sense.

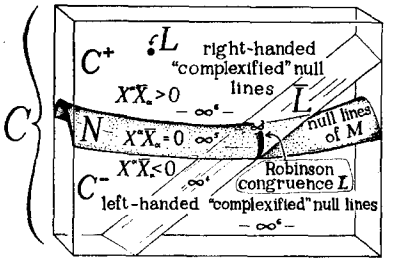

FIG. 4. The C picture.

Note that, in the C picture, the condition that a point L lies on a line P, both L and P lying on N, is interpreted in the M picture as the condition that the null line L passes through the point P. We have, in fact, a kind of duality correspondence between M and N. To sum up, we have the following correspondence between the M picture and the subset N in the C picture:

There is a one-to-one relation between the null lines in M and the points in N. (6.2)

There is a one-to-one relation between the points in M and the complex lines in N. (6.3)

The condition for a point to lie on a null line in M is interpreted, in N, as the condition for the corresponding line to pass through the corresponding point. (6.4)

From (6.4), we see that the condition that two points in M have a null separation is simply the condition that their corresponding lines in N should intersect; the condition that two null lines in M should meet is the condition that the join of the corresponding points in N should lie entirely in N.

We may ask how we should interpret the lines of C which do *not* lie entirely on N. Lines in a projective three space form a four-dimensional system, so that in terms of real dimensions these lines describe an ∞^8 system of objects in M. This suggests that a general line in C represents a *complexified* point in M, where now the straightforward doubling of dimensions suggests that these are complexified points in the usual sense, i.e., their position vectors are allowed to have imaginary parts. To see that this is indeed a consistent interpretation, consider a general line in the C picture as a join of two general points L, X of C. Using the notation of Secs. II and III, we can form the expression (3.10). We may regard this *complex vector* p_i as the position vector, the M picture, of the complex point of intersection of the two null lines ("complexified" or otherwise) in M, defined by L and X. (Note that in this sense *any*

two distinct null lines in M have a complex point in common.)

Any line P of C falls into one of six classes. If q^i is the *imaginary part* of the position vector p^i for the corresponding complex point P in M, then [from (3.10)]

$$P \subset C^+ \Leftrightarrow q^i \text{ timelike, future-pointing;} \tag{6.5}$$

$$P \subset C^+ \cup N,\ P \not\subset N,\ P \not\subset C^+ \Leftrightarrow q^i \text{ null, future-pointing;} \tag{6.6}$$

$$P \text{ intersects all of } C^-,\ N,\ C^+ \Leftrightarrow q^i \text{ spacelike, nonzero;} \tag{6.7}$$

$$P \subset C^- \cup N,\ P \not\subset N,\ P \not\subset C^- \Leftrightarrow q^i \text{ null, past-pointing;} \tag{6.8}$$

$$P \subset C^- \Leftrightarrow q^i \text{ timelike, past-pointing;} \tag{6.9}$$

$$P \subset N \Leftrightarrow q^i = 0. \tag{6.10}$$

In cases (6.6) and (6.8), the line P in C *touches* M. The set (6.5) (in the M picture) is sometimes referred to as a "future-tube."[10]

There is a representation of lines in projective three-space, well known to geometers, called the Klein (or Grassmann) representation.[26] This is a four- (complex-) dimensional quadric, the points of which correspond to the lines of complex projective three-space. Thus, in our case, we may regard the fully complexified version M^* of M as the Klein representation of the lines in C. We noted in Sec. III that M itself was essentially a real $(+\,+\,-\,-\,-\,-)$ quadric fourfold. M^* is the complexified version of this quadric. There are two systems of planes on M^*, called α planes and β planes. The α planes correspond to points in C and the β planes to planes in C. We may regard M as a submanifold of M^*. The α planes which meet M intersect M in the null lines of M. Similarly with the β planes. Thus, when properly complexified, the null lines in M become, in effect, pairs of complex projective planes, namely one α plane and one β plane for each fully complexified null line. The "complexification" of the null lines achieved here by the use of twistors amounts to selecting only the α planes to represent the null lines.

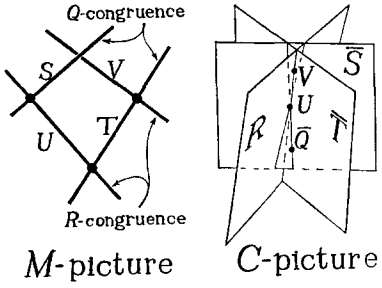

FIG. 5. The condition for $\bar{Q}^\alpha R_\alpha = 0$ is that, in the M picture, V and S should meet, and that in the C picture, $V \in \bar{S}$, whence $\bar{Q} \in R$.

Much of the earlier discussion could have been carried out in terms of the manifold M^*. The emphasis here, however, has been to try and represent twistors as far as possible in terms of "real" structures in the space–time M. One final condition remains to be represented, however, in order that the geometry of C can be completely realized in terms of M. This is that we must be able to interpret the vanishing of a twistor scalar product in geometrical terms[27] in M.

Consider the condition

$$\bar{Q}^\alpha R_\alpha = 0. \tag{6.11}$$

If $\bar{Q}$ and $\bar{R}$ are both null lines in M, we have seen that this is the condition for the lines to intersect. If $\bar{Q}$ is a null line and R a Robinson congruence, then (6.11) is simply the condition that $\bar{Q}$ should belong to the congruence R. It remains to consider the case when *both* Q and R are Robinson congruences in M. Let S and T be two null lines belonging to the Q congruence and let U and V be null lines belonging to the R congruence. Suppose the three pairs of lines (S, U), (U, T), (T, V) intersect. *Then a necessary and sufficient condition for* (6.11) *to hold is that the pair* (V, S) *also intersect* (Fig. 5). To see this, consider the points V, U, $\bar{Q}$ and the planes $\bar{S}$, $\bar{T}$, R in the C picture. We have $S \in Q$, $T \in Q$, whence $\bar{Q} \in \bar{S}$, $\bar{Q} \in \bar{T}$. Also $U \in R$, $V \in R$ and $U \in \bar{S}$, $U \in \bar{T}$, $V \in \bar{T}$. Thus, the line UV must be the intersection of the planes R and $\bar{T}$ (Fig. 5). If (6.11) holds, then Q also lies on this line, whence $V \in \bar{S}$. Conversely if $V \in \bar{S}$ then $\bar{Q} \in R$. Thus, in the M picture, the condition for (6.11) to hold is that the null lines V and S should meet.

This establishes the geometrical equivalence of the M picture with the C picture, whereby the conformal geometry of Minkowski space is completely represented in twistor terms. The *metric* geometry of Minkowski space, on the other hand, requires the introduction of a fixed metric twistor. This is done in Sec. X.

[26] See any standard work on classical algebraic geometry, for example, J. A. Todd, *Projective and Analytical Geometry* (I. Pitman, London, 1947); J. G. Semple and L. Roth, *Algebraic Geometry* (Clarendon Press, Oxford, England, 1949).

[27] Strictly, we should also show that the concept of a Robinson congruence is "geometrical" in M. An explicit construction in terms of incidence of null lines is given in Sec.IX, but in any case, the geometric (and conformally invariant) nature of a Robinson congruence is already implied by Sec.IV and V.

VII. TWISTOR TRANSFORMATIONS

Although the transformation properties of twistors have not been given explicitly as yet, they are implicit in the preceding discussion. Any continuous transformation of M into itself, which preserves its conformal structure (and, hence, the null cone and null line structure of M) must, by (6.4), correspond to a continuous transformation of N into itself which preserves its linearity structure. Furthermore, since the concept of a Robinson congruence is conformally invariant (cf., Sec. V), this transformation of N extends uniquely to a transformation of the whole space C. We have seen that the orthogonality relation (6.11) between points and planes in C can be stated in terms of (conformally invariant) incidence properties in M. Hence, collinearity of points in C has a conformally invariant interpretation in M. It follows that the conformal transformations of M are represented, in the C picture, by continuous point transformations of C, which preserve its linearity structure, and for which the submanifold N is invariant.

If we restrict ourselves to conformal transformations of M which are continuous with the identity, then the corresponding transformations of C is also continuous with the identity. Such transformations, preserving the linearity structure of C, must therefore be *projective point transformations*[26] of C (in the complex sense). If, on the other hand, we allow transformations of M which involve space or time reflections, then we must also consider *anti-projective* transformations of C, that is, transformations which combine a complex conjugation operation with a projective transformation (so that cross ratios become complex conjugated). We may also consider transformations of C which are not point transformations but *correlations*[26] (duality correspondences) in which points of C are mapped to planes, and planes to points. These would not strictly correspond to transformations of M, because of the way C has been constructed here, but it is convenient to think of them as representing transformations of the *complexified* version M^* of M. The *natural* (analytic) extension to M^* of a transformation defined on M sometimes turns out to correspond to a correlation in C.

The basic operation $\mathcal{C}$ of *complex conjugation* in M^*, which leaves the real space M invariant, corresponds, in the C picture, simply to the *Hermitian correlation* given by $X^\alpha \leftrightarrow \bar{X}_\alpha$. A transformation of M equivalent to a *space reflection* corresponds, in the C picture, to a point transformation of C, which is antiprojective and interchanges C^- and C^+. (The twist orientation of Robinson congruences is reversed under space reflection.) However, if we wish to extend this to a space reflection $\mathcal{P}$ of the whole of M^* we must, since $\mathcal{P}$ is analytic, remove the antiprojective nature of the transformation of C and consider, instead, a *correlation* which sends points X of $C^+[C^-]$ into planes $\bar{Y}$ for which Y is in $C^-[C^+]$. The antiprojective point transformation just considered, which interchanges C^+ and C^- would then correspond to $\mathcal{CP}$. Similarly, a point transformation of C which represents a *time reflection* of M is antiprojective, here transforming each of C^-, C^+ into itself. It therefore really represents $\mathcal{C}\mathfrak{I}$, where $\mathfrak{I}$ describes the (analytic) extension to M^* of this time reflection of M. The transformation of C representing $\mathfrak{I}$ is a *correlation* sending each point X of $C^+[C^-]$ into a plane $\bar{Y}$ with $Y \in C^+[C^-]$. The transformation $\mathcal{P}\mathfrak{I}$, being the product of two projective correlations, is a projective point transformation of C. It interchanges C^- with C^+, and is not continuous with the identity unless we widen the group of transformations of M^* to include *complex* conformal transformations. (These would not be point transformations of M, but we may view them as transformations on the line systems in M which transform Robinson congruences into one another. They are represented, in the C picture, as projective transformations of C which do not preserve N.) The transformation $\mathcal{PC}\mathfrak{I}$ is an antiprojective correlation in C.

Let us examine more explicitly the effect of an allowable transformation on a general twistor $A_{\rho\cdots\tau}^{\alpha\beta\cdots\delta}$ of valence $\left[{}^{r}_{s}\right]$. (The indices $\alpha, \beta, \cdots, \delta$ are r in number and $\rho, \cdots, \tau$ are s in number, each ranging over four values 0, 1, 2, 3.) We may define a general twistor, in terms of twistors of valence $\left[{}^{1}_{0}\right]$ and $\left[{}^{0}_{1}\right]$, in any of the standard ways that "tensors" may be built up from "vectors," e.g., as linear combinations of outer products, or in terms of multilinear mappings of one-index twistors into the scalars. Alternatively, we may simply use the transformation properties to define a twistor $A_{\rho\cdots\tau}^{\alpha\cdots\delta}$ of valence $\left[{}^{r}_{s}\right]$ (considering for the moment only transformations continuous with the identity):

$$\tilde{A}_{\rho\cdots\tau}^{\alpha\beta\cdots\delta} = A_{\varphi\cdots\psi}^{\kappa\lambda\cdots\nu} t_\kappa^\alpha t_\lambda^\beta \cdots t_\nu^\delta T_\rho^\varphi \cdots T_\tau^\psi. \tag{7.1}$$

The matrices (t_β^α), (T_β^α) are inverses[28] of each other

$$t_\beta^\alpha T_\gamma^\beta = \delta_\gamma^\alpha = T_\beta^\alpha t_\gamma^\beta, \tag{7.2}$$

[28] In fact, the transformations of C would be the same if we specified only that $(T_\beta{}^\alpha)$ be *proportional* to the inverse of $(t_\beta{}^\alpha)$ since the X^α are *projective* coordinates for C. However, the stronger requirement (7.2) is adopted here since the factor of proportionality of a twistor is required when the more complete representation in accordance with Sec. V is used.

where twistors X^α, Y_α, of respective valence $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$, $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, transform as

$$\tilde{X}^\alpha = t^\alpha_\beta X^\beta, \qquad \tilde{Y}_\alpha = T^\beta_\alpha Y_\beta \tag{7.3}$$

under a conformal transformation of M continuous with the identity (i.e., a projective transformation of C preserving C^+). But since $\bar{X}_\alpha$ is a twistor of valence $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$, it follows that

$$T^\alpha_\beta = \bar{t}^\alpha_\beta, \quad \text{i.e.,} \quad t^\alpha_\beta \bar{t}^\beta_\gamma = \delta^\alpha_\gamma, \tag{7.4}$$

where the convention used earlier for twistor indices under complex conjugation [cf., (3.5)] is being employed here, namely that, under complex conjugation, upper and lower index positions are interchanged and also, the pairs 0, 1 and 2, 3 are interchanged, i.e.,[17]

$$\bar{t}^0_0 = \overline{t^2_2}, \qquad \bar{t}^0_1 = \overline{t^3_2}, \qquad \bar{t}^0_2 = \overline{t^0_2}, \qquad \bar{t}^0_3 = \overline{t^1_2},$$
$$\bar{t}^1_0 = \overline{t^2_3}, \cdots, \bar{t}^3_2 = \overline{t^0_1}, \qquad \bar{t}^3_3 = \overline{t^1_1}.$$

Condition (7.4) states that the matrix (t^α_β) is pseudo-unitary in the sense that the form (6.1) is left invariant, that is, in matrix notation,

$$(t^\alpha_\beta)^* \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} (t^\gamma_\delta) = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{7.5}$$

where the asterisk denotes Hermitian conjugate in the usual sense.[29]

Equation (7.4) ensures that the operation of complex conjugation, according to the rules just stated, is a twistor operation. That is, if $A^{\alpha\beta\cdots\delta}_{\rho\cdots\tau}$ is a twistor of valence $\left[\begin{smallmatrix}\cdot\\\cdot\end{smallmatrix}\right]$, then $\bar{A}^{\rho\cdots\tau}_{\alpha\beta\cdots\delta}$ is a twistor of valence $\left[\begin{smallmatrix}\cdot\\\cdot\end{smallmatrix}\right]$ (where,[17] for example, $\bar{A}^{3\cdots0}_{21\cdots2} = \overline{A^{03\cdots1}_{1\cdots2}}$, etc.). Clearly, the operations of addition, outer multiplication, contraction (between upper and lower indices), and index permutation (not mixing upper with lower indices) are also twistor operations. [That is to say, they commute with (7.1).] We might also consider complexified twistor transformations (7.1) for which (7.4) does not hold. These correspond to *complex* conformal transformations of M^*, not preserving M, and the relation between $A^{\alpha\cdots\delta}_{\rho\cdots\tau}$ and $\bar{A}^{\rho\cdots\tau}_{\alpha\cdots\delta}$ would not be preserved. Such transformations will not be discussed here.

It will be appropriate to impose one final condition on the matrix (t^α_β), namely that it should be *unimodular*:

$$|t^\alpha_\beta| = 1. \tag{7.6}$$

This amounts to requiring that the Levi-Civita symbols $\epsilon_{\alpha\beta\gamma\delta}$, $\epsilon^{\alpha\beta\gamma\delta}$ [with fixed components $+1$, -1, 0 according as $(\alpha, \beta, \gamma, \delta)$ is an even permutation, an odd permutation, or no permutation of (0, 1, 2, 3)] shall be *twistors* rather than just "twistor densities." In fact, this condition is *necessary* if a strict interpretation of twistors according to the scheme of Sec. V is to be adhered to. The effect of (7.6) on the twistor algebra is only to enrich it slightly in that the operation of forming *duals* of skew-symmetric twistors is now allowable. Note also that

$$\epsilon_{\alpha\beta\gamma\delta} = \bar{\epsilon}_{\alpha\beta\gamma\delta}, \qquad \epsilon^{\alpha\beta\gamma\delta} = \bar{\epsilon}^{\alpha\beta\gamma\delta}, \tag{7.7}$$

since (2, 3, 0, 1) is an even permutation of (0, 1, 2, 3)

The transformations of the form (7.1) are not the only allowable twistor transformations since, as we have seen, the operations $\mathcal{C}$, $\mathcal{P}$, $\mathcal{T}$, $\mathcal{CP}$, $\mathcal{CT}$, and $\mathcal{PCT}$, give conformal transformations of M but they do not correspond to projective point transformations of C. On the other hand, $\mathcal{PT}$ *does* have the form (7.1), with (7.2) and (7.6) holding, but where (7.4) is replaced by

$$T^\alpha_\beta = -\bar{t}^\alpha_\beta; \qquad t^\alpha_\beta \bar{t}^\beta_\gamma = -\delta^\alpha_\gamma. \tag{7.8}$$

Thus, for $\mathcal{PT}$, we have $\bar{\tilde{A}}^{\cdots}_{\cdots} = -\bar{\tilde{A}}^{\cdots}_{\cdots}$ if the total valence of $A^{\cdots}_{\cdots}$ is odd. (This, again, is assuming that we impose (7.2). Had we chosen $t^\alpha_\beta T^\beta_\gamma = -\delta^\alpha_\gamma$ instead, then $\bar{\tilde{A}}^{\cdots}_{\cdots} = \bar{\tilde{A}}^{\cdots}_{\cdots}$, *but* $\tilde{L}^\alpha \tilde{R}_\alpha = -L^\alpha R_\alpha$, etc.) The operation $\mathcal{C}$ is given by

$$A^{\alpha\beta\cdots\delta}_{\rho\cdots\tau} \to \bar{A}^{\rho\cdots\tau}_{\alpha\beta\cdots\delta}, \tag{7.9}$$

while $\mathcal{PCT}$ is a combination of (7.9) with a transformation of the form (7.1) satisfying (7.8). The operations $\mathcal{P}$ and $\mathcal{T}$ belong to a class of twistor transformations given by

$$\tilde{A}^{\rho\cdots\gamma}_{\alpha\beta\cdots\delta} = (\mp 1)^s A^{\varkappa\lambda\cdots\nu}_{\varphi\cdots\psi} u_{\alpha\varkappa} u_{\beta\lambda} \cdots u_{\delta\nu} \bar{u}^{\rho\varphi} \cdots \bar{u}^{\gamma\psi}, \tag{7.10}$$

where $(u_{\alpha\beta})$ and the transpose of $\mp(\bar{u}^{\alpha\beta})$ are unimodular inverse matrices:

$$u_{\beta\alpha}\bar{u}^{\beta\gamma} = \mp\delta^\gamma_\alpha; \qquad |u_{\alpha\beta}| = 1. \tag{7.11}$$

[29] The representation of twistors in terms of components given in this paper is perhaps the most convenient, but is by no means the only one possible. It amounts to insisting that the coordinate basis twistors $E^\alpha{}_{(0)}$, $E^\alpha{}_{(1)}$, $E^\alpha{}_{(2)}$, $E^\alpha{}_{(3)}$ are null and satisfy $E^\alpha{}_{(0)} \bar{E}_{(2)\alpha} = 1 = E^\alpha{}_{(1)} \bar{E}_{(3)\alpha}$, with all the other scalar products vanishing. Another possible choice would be to require $E^\alpha{}_{(a)} \bar{E}_{(b)\alpha} = (-1)^b \delta_{ab}$ (b not summed!) in which case twistor complex conjugation would take the form $\bar{L}_0 = \overline{L^0}$, $\bar{L}_1 = -\overline{L^1}$, $\bar{L}_2 = \overline{L^2}$, $\bar{L}_3 = -\overline{L^3}$, instead of (3.5), and the fixed matrix in (7.5) would become diag $(1, -1, 1, -1)$. (The simple connection with spinors (2.13) would be lost, however.) The only *essential* restriction on the way in which the twistors are represented is that the signature of the form $X^\alpha \bar{X}_\alpha$ must be $(+\,+\,-\,-)$ [cf., (6.1)].

The minus sign applies when a reversal of space orientation is involved and the plus sign when it is the time orientation that is reversed. Equation (7.11) implies that the matrix $(u_{\alpha\beta})$ is also pseudo-unitary in the sense of (7.4) and (7.5), or (7.8). These transformations give the projective correlations of C. The antiprojective point transformations of C such as $\mathcal{CP}$ and $\mathcal{CJ}$, are given by the combination of (7.10) with (7.9).

Let us now examine some twistor transformations in a little more detail. The *translation* given by

$$x^i = x^i + a^i, \qquad \tilde{\xi}^A = \xi^A \tag{7.12}$$

[using the notation of (3.1), (3.2)] results in

$$(\tilde{X}^\alpha) = (\xi^A, \eta_{A'} - i\xi^B a_{BA'}),$$

whence[30]

$$(t^\alpha_\beta) = \begin{bmatrix} -\epsilon^A{}_B & 0 \\ -ia_{A'B} & \epsilon_{A'}{}^{B'} \end{bmatrix}. \tag{7.13}$$

The *Lorentz rotation* given, with $(b^A{}_B)$ unimodular, by

$$x^{JJ'} = x^{KK'} b^J{}_K \bar{b}^{J'}{}_{K'}, \qquad \tilde{\xi}^A = \xi^B b^A{}_B, \tag{7.14}$$

results in

$$(\tilde{X}^\alpha) = (b^A{}_B \xi^B, -\bar{b}_{A'}{}^{B'} \eta_{B'}),$$

whence

$$(t^\alpha_\beta) = \begin{bmatrix} b^A{}_B & 0 \\ 0 & -\bar{b}_{A'}{}^{B'} \end{bmatrix}. \tag{7.15}$$

Combining (7.14) with (7.12) and (7.15) with (7.13), we get the general restricted Poincaré transformation

$$x^{JJ'} = x^{KK'} b^J{}_K \bar{b}^{J'}{}_{K'} + a^{JJ'}, \qquad \tilde{\xi}^A = \xi^B b^A{}_B \tag{7.16}$$

[actually, "inhomogeneous $SL(2, C)$ transformation"[10]] represented by

$$(t^\alpha_\beta) = \begin{bmatrix} b^A{}_B & 0 \\ -ia_{CA'} b^C{}_B & -\bar{b}_{A'}{}^{B'} \end{bmatrix} \tag{7.17}$$

so that matrices of the form (7.17) give a (two-valued) representation of the Poincaré group.

The *inversion*

$$\tilde{x}^i = 2a^2 x^i (x^k x_k)^{-1} \tag{7.18}$$

$(a^4 > 0)$ is given if we put $\tilde{\xi}^{A'} = a^{-1}\eta^{A'}$, $\tilde{\eta}_A = -a\xi_A$, i.e.,

$$\tilde{X}_\alpha = (-a\xi_A, a^{-1}\eta^{A'})$$

since this agrees with (3.2) [and hence with (3.10)]. Thus, in this case

$$(u_{\alpha\beta}) = \begin{bmatrix} a\epsilon_{AB} & 0 \\ 0 & a^{-1}\epsilon^{A'B'} \end{bmatrix}. \tag{7.19}$$

Here a reversal in time orientation ($a^2 > 0$) *or* space orientation ($a^2 < 0$) is involved, so we get (7.10) rather than (7.1). The space–time reflection

$$x^i = -x^i, \qquad \tilde{\xi}^A = -\xi^A \tag{7.20}$$

is given if

$$\tilde{X}^\alpha = (-\xi^A, \eta_{A'})$$

so that[30]

$$(t^\alpha_\beta) = \begin{bmatrix} \epsilon^A{}_B & 0 \\ 0 & \epsilon_{A'}{}^{B'} \end{bmatrix}. \tag{7.21}$$

For a space reflection or a time reflection, we must introduce a timelike vector v^i at the origin. For convenience we normalize v^i to have length $\sqrt{2}$, so that

$$\tilde{x}^i = \mp x^i \pm v^i v_k x^k; \qquad v_k v^k = 2. \tag{7.22}$$

The upper sign refers to space reflection and the lower sign to time reflection. In spinor terms we have

$$\tilde{x}^{JJ'} = \pm x^{KK'} v^{J'}_K v^J_{K'}; \quad \tilde{\xi}^{A'} = \xi^B v^{A'}_B, \quad \tilde{\eta}_A = \pm \eta_{B'} v^{B'}_A$$

with $v^{B'}_A v^C_{B'} = -\delta^C_A$. Thus,

$$\tilde{X}_\alpha = (\pm v^{B'}_A \eta_{B'}, v^{A'}_B \xi^B),$$

so that

$$(u_{\alpha\beta}) = \begin{bmatrix} 0 & \pm v_A{}^{B'} \\ v^{A'}{}_B & 0 \end{bmatrix}. \tag{7.23}$$

Note that $u_{\alpha\beta}$ is *symmetric* for a space reflection and *skew-symmetric* for a time reflection. Recall, also, that $u_{\alpha\beta}$ was skew-symmetric for the inversion (7.18) [cf., (7.19)]. The significance of this lies in the fact that there are two distinct kinds of projective correlation in C which are involutory (i.e., whose squares are the identity), namely polarity with respect to a quadric (corresponding to $u_{\alpha\beta}$ symmetric) and null-polarity with respect to a linear complex[26] ($u_{\alpha\beta}$ skew-symmetric). The null-polarity is distinguished by the fact that any point in C lies *on* the plane into which it is transformed, whence any null line in M meets the null line into which it is transformed. Thus, in addition to the cases of time reflection or inversion just considered, $u_{\alpha\beta}$ is skew-

[30] The staggering of the spinor indices is to indicate that, in each case, the left-hand index labels rows and right-hand index labels columns. Also, for notational consistency, $-\epsilon^A{}_B$, is used here instead of $\delta^A{}_B$, and $\epsilon_{A'}{}^{B'}$ instead of $\delta_{A'}{}^{B'}$, although they are all numerically equal!

symmetric in the case of a space reflection in a plane (that is, in a timelike hyperplane). The linear complex involved is, in each case, the system of lines in C which are left invariant by the correlation. Some of these lines lie in N corresponding to the points of M which are invariant under the transformation. These points constitute a *hypersphere* in M which is spacelike [as in (7.18) with $a^2 > 0$, or (7.22) with the lower sign] if a time reversal is involved, or timelike [e.g., (7.18) with $a^2 < 0$] if it is a change of spatial orientation that is involved. (A hyperplane is to be regarded as a case of a hypersphere in M.) Transformations for which $u_{\alpha\beta}$ is *symmetric* include, in addition to the space reflection in the origin (or, more correctly, in a timelike line), a reflection in a spacelike line which is accompanied by a time reflection. In these cases, the points of M left invariant by the transformation are represented by the generators which lie in N of the quadric (equation: $X^\alpha X^\beta u_{\alpha\beta} = 0$) defining the polarity in C. These invariant points in M constitute either a timelike circle (a timelike straight line being one case) or a pair of spacelike circles (a spacelike straight line together with a "circle" at infinity being one case).

The involutory projective *point* transformations of C are called harmonic perspectives.[26] They fall into two main classes, depending on whether the invariant points of C constitute two skew lines (when $t^\alpha_\alpha = 0$) or a point and a plane ($t^\alpha_\alpha = \pm 2, \pm 2i$). The *first* class is further subdivided according as the two skew lines both lie in N, both cross N, or lie one in C^+ and one in C^-. (No other cases are possible.) If the lines both lie on N, they correspond to two special invariant points of M. The spacelike 2-sphere in M of intersection of the null cones of these two special points consists also of invarient points. Examples of this type of transformation are the space–time reflection (7.20) in the origin O (whence O and I are the special invariant points and the invariant 2-sphere is at infinity) or a reflection in a plane together with a time reversal (in which case the plane is the invariant "2-sphere" and the special invariant points are both at infinity). If the two skew lines both cross N, we get a timelike 2-sphere of invariant points in M (e.g., a timelike 2-plane, for the case of an ordinary reflection in a line). If neither skew line meets N, there are *no* invariant points in M [e.g., the inversion (7.18) with $a^2 > 0$, followed by the time reflection of (7.22)]. The transformations of the *second* class are also interesting in that no points of M are left invariant. The plane R of invariant points of C meets N in a set of points representing a Robinson congruence. The lines of the Robinson congruence are all invariant under the transformation but not pointwise invariant. Noninvolutory transformations with this property also exist. These are intimately related to twistors and their *four-valuedness* and so will be described briefly next.

Choose a right-handed twistor R_α normalized so that

$$R_\alpha \bar{R}^\alpha = 1.$$

Define

$$t(\theta)^\alpha_\beta = e^{i\theta}(\delta^\alpha_\beta - \bar{R}^\alpha R_\beta) + e^{-3i\theta}\bar{R}^\alpha R_\beta \qquad (7.24)$$

for each real θ. Then $t(\theta)^\alpha_\beta$ satisfies (7.4) and (7.6), so it represents an allowable twistor transformation. Also

$$t(\theta)^\alpha_\beta t(\varphi)^\beta_\gamma = t(\theta + \varphi)^\alpha_\gamma$$

so that these transformations form a one-parameter subgroup of twistor transformations. We can represent any line L of the *Robinson congruence* R, in M, by a twistor L^α satisfying $L^\alpha R_\alpha = 0$ (and $L^\alpha \bar{L}_\alpha = 0$). Thus,

$$t(\theta)^\alpha_\beta L^\beta = e^{i\theta} L^\alpha$$

whence each line of R must be left invariant under the transformation. Furthermore, except for those values

$$\theta = \tfrac{1}{2}n\pi \qquad (n = \cdots, -2, -1, 0, 1, 2, \cdots)$$

for which $t(\theta)^\alpha_\beta$ is proportional to δ^α_β, there are *no points* of M invariant under the transformation. (This follows because, in the C picture, the lines which are left invariant are those which either pass through the point $\bar{R}$ or lie in the plane R. In neither case can these lines lie on N, since $\bar{R}$ is not on N and R does not touch N.)

Finally, observe that for $\theta = \frac{1}{2}\pi$ we have

$$t\left(\frac{\pi}{2}\right)^\alpha_\beta = i\,\delta^\alpha_\beta. \qquad (7.25)$$

This gives the identity transformation on C and therefore also the identity transformation on M. But it *multiplies every twistor of valence* $[^1_0]$ *by* i. Also (7.25) is continuous with the identity twistor transformation via (7.24) with $0 \leq \theta \leq \frac{1}{2}\pi$. Hence *twistors are essentially four-valued under conformal transformations of* M.

VIII. THE KERR THEOREM

Consider the problem of finding all null solutions of Maxwell's equations (i.e., with $F_{ij}F^{ij} = 0 =$

$F_{ij}F_{kl}\epsilon^{ijkl}$) in free space. Robinson[11] showed that the problem could be reduced to that of finding all the shear-free null geodetic congruences in the space–time [cf., (5.8)], since associated with any null field was a congruence of this type, while conversely, given such a congruence, all the corresponding solutions of Maxwell's equations could then be defined in terms of arbitrary complex functions. The result applies to curved space–times as well as flat. We are concerned here only with space–times which are (conformally) flat. With this restriction, Robinson's result applies also to null zero rest-mass fields of arbitrary spin ($>\frac{1}{2}$). (The result for spin 2 has a bearing on the construction of null solutions of Einstein's nonlinear field equations.[31])

The solution of the remainder of the problem for flat space–time, namely the construction of all the shear-free null congruences, is given by a remarkable theorem due to Kerr.[32] The theorem takes the following very natural form when stated in twistor terms.

A congruence of null lines in M is shear-free if and only if it is representable in C as the intersection of N with a complex analytic surface S in C (or as a limiting case of such an intersection). (8.1)

"Complex analytic" means here analytic in the sense of the complex structure of C, where S is a "surface" in the complex sense (i.e., it has four real dimensions). Thus, (8.1) tells us that the shear-free condition [cf., (5.8)] is, effectively, a Cauchy–Riemann type of relation in the C picture. [The reason for the final parenthetic remark in (8.1) is to enable certain exceptional cases to be incorporated, for which the congruence does not form an analytic system even in the real sense. These exceptional cases appear to occur only when the rotation of the congruence also vanishes, cf., (5.7).]

Suppose, first, that a congruence of null lines in M is defined by an equation

$$\varphi(X^\alpha) \equiv \varphi(\xi^A, \eta_{A'}) = 0, \tag{8.2}$$

where φ is analytic and homogeneous in the four complex variables X^α so that (8.2) defines a complex analytic surface S in C. We see that the congruence must be shear free. Putting $\eta_{A'} = -i\xi^A x_{AA'}$, as in (3.2), Eq. (8.2) can be used to solve for ξ^A, up to proportionality, as a function of $x_{AA'}$, i.e., to give the direction of the line X as a function of a point on X with position vector x^i. Euler's condition on φ, for it to be homogeneous (degree n), gives

$$X^\alpha \frac{\partial\varphi}{\partial X^\alpha} \equiv \xi^A \frac{\partial\varphi}{\partial\xi^A} + \eta_{A'} \frac{\partial\varphi}{\partial\eta_{A'}} = n\varphi.$$

Hence, by (3.2) and (8.2),

$$\xi^A\left(\frac{\partial\varphi}{\partial\xi^A} - ix_{AA'}\frac{\partial\varphi}{\partial\eta_{A'}}\right) = 0,$$

whence

$$\frac{\partial\varphi}{\partial\xi^A} - ix_{AA'}\frac{\partial\varphi}{\partial\eta_{A'}} = \kappa\xi_A \tag{8.3}$$

for some κ. The derivative of (8.2) with respect to $x_{AA'}$ [cf., (5.3)] must also vanish

$$\begin{aligned} 0 = \nabla^{BB'}\varphi &= \frac{\partial\varphi}{\partial\xi^A}\nabla^{BB'}\xi^A + \frac{\partial\varphi}{\partial\eta_{A'}}\nabla^{BB'}\eta_{A'} \\ &= \frac{\partial\varphi}{\partial\xi^A}\nabla^{BB'}\xi^A \\ &\quad + \frac{\partial\varphi}{\partial\eta_{A'}}(-ix_{AA'}\nabla^{BB'}\xi^A - i\xi^A\nabla^{BB'}x_{AA'}) \\ &= \kappa\xi_A\nabla^{BB'}\xi^A - i\frac{\partial\varphi}{\partial\eta_{B'}}\xi^B, \end{aligned}$$

by (3.2) and (8.3). Assuming $\kappa \neq 0$, contraction with ξ_B gives $\xi_B\xi_A\nabla^{BB'}\xi^A = 0$, i.e., the condition (5.8) for the null directions defined by ξ^A to be tangent to *shear-free* null straight lines, as required. Finally, κ *cannot* vanish if (8.2) represents a genuine condition ξ^A, since the left-hand side of (8.3) is simply the derivative, with respect to ξ^A, of φ considered as a function of ξ^A and $x_{AA'}$.

The converse result that (8.2) represents essentially the most *general* shear-free congruence in M is somewhat less straightforward because of the existence of exceptional cases. However, if we assume that the congruence is *analytic in the real sense*—and any nonanalytic congruence (presumably necessarily rotation-free) can be approximated arbitrarily closely by analytic ones—then we can see from the form of (5.8) that the congruence is *determined* once the directions of the lines (i.e., of ξ^A) are known on any *spacelike 2-sphere* F in M. [If $\xi^0/\xi^1 = \zeta$, then (5.8) becomes $\partial\zeta/\partial x^{00'} = -\zeta\partial\zeta/\partial x^{10'}$, $\partial\zeta/\partial x^{11'} = -\zeta^{-1}\partial\zeta/\partial x^{01'}$. This defines the propagation of ζ

[31] I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962).

[32] R. P. Kerr, private communication; cf. also R. P. Kerr, Phys. Rev. Letters **11**, 238 (1963); R. P. Kerr and A. Schild, in Proceedings of the American Mathematical Society Symposium, April 1964. According to Kerr's original construction, a general shear-free null geodetic congruence is defined by an analytic relation $\phi(\eta, x^{00'} + \eta x^{10'}, x^{01'} + \eta x^{11'}) = 0$, where $dx^{00'} + \eta dx^{10'} = 0 = dx^{01'} + \eta dx^{11'}$.

off the 2-plane $x^{00'} = x^{11'} = 0$. On this 2-plane—which from the point of view of M is a 2-sphere—ζ can be given arbitrarily as a function of the complex variable $x^{01'}$ and its complex conjugate $x^{10'}$.] The 2-sphere F is the intersection of two null cones (vertices A, B) in M and is therefore represented, in the C picture by the lines lying in N which meet two skew lines A, B in N. The *complexification* F^* of F is, in the C picture, the system of *all* lines in C meeting both A and B. Now (in the M picture), F is a *real environment*[10] for F^*, so that any complex analytic function on F extends uniquely to a complex analytic function on F^*. The null directions of the congruence, defined at points of F, are represented in the C picture as an ∞^2 system of points lying on the lines of the F system (describing a *real* analytic surface). These extend uniquely to an ∞^4 system of points S on the lines of the F^* system (describing a *complex* analytic surface). The intersection of S with N then defines the given shear-free congruence as required (since the congruence defined by S agrees with the given one on F).

The Kerr theorem provides a very convenient means of studying the structure of shear-free null congruences in M, in general. Only a few results are briefly indicated here. For example, the lines of a shear-free null congruence along which the *rotation vanishes* are represented in C by the points where lines of N touch S. We can also generally form the *reciprocal*[26] $\tilde{S}$ of S with respect to the Hermitian correlation defined by N (i.e., the envelope of polar planes with respect to N of points of S). Then $\tilde{S}$ defines a shear-free null congruence in M which is, in a sense, "reciprocal" to the original one. The congruences which are *everywhere rotation-free* are the ones which are *self-reciprocal* (although the individual lines of the congruence do not generally reciprocate to themselves). In this case, S is a *ruled surface*[26] in C, with ∞^1 of its generators lying in N. This ∞^1 system corresponds, in M, to a *curve* and the lines of the congruence are just the null lines meeting this curve.[33] This curve in M is *null* if the ruled surface in C is *developable*.

Of special interest are the *algebraic* shear-free null congruences.[34] In this case S is algebraic variety[26] and φ can be given as a homogeneous polynomial

$$\varphi(X^\alpha) \equiv S_{\alpha\beta\ldots\delta}X^\alpha X^\beta \cdots X^\delta = 0, \qquad (8.4)$$

where $S_{\alpha\beta\ldots\delta}$ is symmetric, of valence $[{}^0_n]$. In terms of ξ^A and $\eta_{A'} = -i\xi^A x_{AA'}$, (8.4) becomes

$$\xi^A\xi^B \cdots \xi^D\Phi_{AB\ldots D} = 0, \qquad (8.5)$$

where the spinor field $\Phi_{AB\ldots D}$ is defined by

$$\Phi_{AB\ldots D} = {}^0\sigma_{(AB\cdots D)} + {}^1\sigma_{A'(B\cdots D}x^{A'}_{A)} + \cdots + {}^n\sigma_{A'B'\ldots D'}x^{A'}_{(A}x^{B'}_{B}\cdots x^{D'}_{D)}, \qquad (8.6)$$

the ${}^r\sigma\ldots$'s being constants which are essentially the coefficients $S_{\alpha\beta\ldots\delta}$. (The round brackets denote symmetrization.) The $\Phi_{AB\ldots D}$ is symmetric, satisfies the "field equation"

$$\nabla^{E'}_{(E}\Phi_{AB\ldots D)} = 0 \qquad (8.7)$$

by virtue of (8.6) and has a "canonical decomposition"[7]

$$\Phi_{AB\ldots D} = {}^1\xi_{(A}{}^2\xi_B \cdots {}^n\xi_{D)}, \qquad (8.8)$$

where each ξ of (8.8) satisfies (8.5). Conversely, *every* solution of (8.7) has the form (8.6) and each resulting ξ of (8.8) [or (8.5)] defines (one branch of) the corresponding algebraic shear-free null congruence. This generalizes the results of Sec. V (for which $n = 1$) and gives us a representation of an arbitrary symmetric twistor of valence $[{}^0_n]$, in terms of a symmetric spinor field satisfying (8.7).

Robinson's construction of the general, null, zero rest-mass field from its associated shear-free congruence can also conveniently be represented in twistor terms: the field can be defined in terms of a complex function on S. However, such matters are not entered into here. The twister description of physical fields, generally, is left to a later paper.

IX. GEOMETRICAL APPLICATIONS OF TWISTORS

Many interesting geometrical properties arise from the interplay between the geometric structure of M^* and that of C. Some of these result simply from the fact that the former is the Klein representation of the latter and are, therefore, essentially classical results of algebraic geometry.[26] Others, however, take into account the reality structure of M and so have a more direct relevance to the structure of the physical world. The natural algebra of the C picture—namely twistor algebra—can be used to derive certain geometrical properties of M. A small selection is given here.

A linear space of dimension r in C can be represented by a (*simple*) *skew-symmetric twistor* of valence $[{}^{r+1}_0]$ or by its *dual* of valence $[{}^0_{3-r}]$, (r =

[33] The rotation-free, shear-free null congruences which are *not* analytic in the real sense emerge here simply as the system of null lines meeting a *nonanalytic* curve in M.

[34] Explicit shear-free null congruences of this type have been used to generate explicit solutions of Einstein's equations. For example, in Kerr's construction of the field of a rotating body, $\varphi(x^\alpha)$ is quadratic. For details, see Ref. 32.

0, 1, 2). The relation between general skew-symmetric twistors and their duals is

$$A_{\alpha\beta\gamma} = A^{\delta}\epsilon_{\alpha\beta\gamma\delta}, \qquad B_{\alpha\beta} = \tfrac{1}{2}B^{\gamma\delta}\epsilon_{\alpha\beta\gamma\delta}, \qquad C_{\alpha} = -\tfrac{1}{6}C^{\beta\gamma\delta}\epsilon_{\alpha\beta\gamma\delta}, \tag{9.1}$$

where conversely

$$A^{\alpha} = -\tfrac{1}{6}A_{\beta\gamma\delta}\epsilon^{\alpha\beta\gamma\delta}, \qquad B^{\alpha\beta} = \tfrac{1}{2}B_{\gamma\delta}\epsilon^{\alpha\beta\gamma\delta}, \qquad C^{\alpha\beta\gamma} = C_{\delta}\epsilon^{\alpha\beta\gamma\delta}. \tag{9.2}$$

Because of (7.7) and the symmetry between (9.1) and (9.2), the operations of complex conjugation and of forming the dual commute. The meaning of, for example, $\bar{B}^{\alpha\beta}$, given $B^{\alpha\beta}$, is therefore unambiguous. We call $B^{\alpha\beta}$ *real* if $B^{\alpha\beta} = \bar{B}^{\alpha\beta}$ (whence also $B_{\alpha\beta} = \bar{B}_{\alpha\beta}$).

A skew-symmetric twistor $P^{\alpha\beta}$ is *simple* if it is of the form

$$P^{\alpha\beta} = X^{\alpha}Y^{\beta} - Y^{\alpha}X^{\beta}, \tag{9.3}$$

whence $P_{\alpha\beta}$, $\bar{P}^{\alpha\beta}$, $\bar{P}_{\alpha\beta}$ are all also simple. ($C^{\alpha\beta\gamma}$ and $A_{\alpha\beta\gamma}$ are necessarily simple.) Equivalent alternative conditions for $P^{\alpha\beta}$ to be simple are

$$P^{\alpha\beta}P_{\alpha\gamma} = 0 \quad \text{or} \quad P^{\alpha\beta}P_{\alpha\beta} = 0. \tag{9.4}$$

As given by (9.3), $P^{\alpha\beta}$ represents, in the C picture, the *line* joining two points X and Y. In the M picture, $P^{\alpha\beta}$ represents a complex *point* P (i.e., point of M^*). Thus, we can use *simple skew-symmetric twistors* $P^{\alpha\beta}$ for *projective coordinates in* M^*.

Now, the Hermitian correlation defined by N effects the correspondence $P^{\alpha\beta} \leftrightarrow \bar{P}_{\alpha\beta}$. This is the complex conjugation operation in M^* [cf., (7.9)] so, in terms of our coordinates $P^{\alpha\beta}$, complex conjugation in M^* is defined by $P^{\alpha\beta} \leftrightarrow \bar{P}^{\alpha\beta}$, the points of M being given when $P^{\alpha\beta}$ is *real* (in the above sense). Corresponding to (6.5)–(6.10) we can also express, in terms of $P^{\alpha\beta}$, the space–time nature of the imaginary part q^i of the position vector of a complex point P of M^*:

$$P^{\alpha\beta}\bar{P}_{\alpha\beta} \gtreqless 0 \text{ according as } q^iq_i \gtreqless 0, \tag{9.5}$$

and if $q^iq_i \geq 0$, then for all Z_{α}

$$Z_{\alpha}\bar{Z}^{\beta}P^{\alpha\gamma}\bar{P}_{\beta\gamma}\begin{cases}\geq 0 & \text{if } q^i \text{ future-pointing,}\\ \leq 0 & \text{if } q^i \text{ past-pointing.}\end{cases} \tag{9.6}$$

[The point of intersection of the plane Z with the line P in C is represented by $Z_{\alpha}P^{\alpha\gamma}$ and, cf., (6.5), (6.6), (6.8), and (6.9).]

A complex point of M^* can be realized in terms of a *real* structure in M in various ways; for example, in terms of the involutory transformation (see Sec. VII) which is represented in C by the harmonic perspective whose axes are the lines P and $\bar{P}$ in C (assuming P and $\bar{P}$ are skew, i.e., $P_{\alpha\beta}\bar{P}^{\alpha\beta} \neq 0$); or in terms of linear systems of Robinson congruences. However, if q^i is *spacelike*, we have a more easily visualizable representation.[35] In this case, the lines P, $\bar{P}$ in C each meets N in an ∞^1 system of points. In the M picture, these become two ∞^1 systems of null lines. No two null lines belonging to the same system can intersect, but every null line of the P system meets every null line of the $\bar{P}$ system (since in the C picture, all the points of the line P are conjugate under the Hermitian correlation to all the points of the line $\bar{P}$.) We can refer to the P system of lines in M as a *null regulus* and the $\bar{P}$ system as its *complementary* null regulus.[26] Any two nonintersecting null lines in M belong to a unique null regulus; the null transversals of these two null lines generate the complementary null regulus. The null reguli in M thus geometrically represent the points in M^* whose position vectors have spacelike imaginary parts.

A *real* skew-symmetric twistor $B^{\alpha\beta}(=\bar{B}^{\alpha\beta})$, which is *not* simple, also has a direct interpretation in M. We can, in fact, normalize $B^{\alpha\beta}$ so that

$$B^{\alpha\beta}B_{\alpha\gamma} = \pm\delta^{\beta}_{\gamma}. \tag{9.7}$$

Then, if we put $B_{\alpha\beta} = u_{\alpha\beta}$ in (7.10) we get an involutory projective correlation, in C, of the type for which the invariant points in M constitute a hypersphere (cf., Sec. VII). Thus, $B^{\alpha\beta}$ represents a *hypersphere* in M which is spacelike or timelike according as the upper or lower sign occurs in (9.7). (In the limiting case when $B^{\alpha\beta}$ becomes simple, the hypersphere becomes a null cone with vertex B.)

We have seen that A^{α} represents a point A in C and that $\bar{A}_{\alpha}$ represents its "polar plane" with respect to N (cf., Sec. VI). A lies on N if and only if $A^{\alpha}\bar{A}_{\alpha} = 0$, which is also the condition for $\bar{A}$ to touch N. In terms of the dual twistor $A_{\alpha\beta\gamma}$, this condition is

$$\bar{A}^{\alpha\beta\gamma}A_{\alpha\beta\gamma} = 0. \tag{9.8}$$

Let X^{α}, Y^{α}, Z^{α} be three (nonzero) null twistors so

[35] When q^i is timelike we may represent P as a parallelism on M (with torsion; left-handed if $P \subset C^+$) which is closely related to Clifford parallelism on S^3. The sets of *null* directions which are to be regarded as parallel are those of the Robinson congruences represented by the points of the line P in C (i.e., by planes through $\bar{P}$). A transitive four-parameter group of (conformal) motions of M preserves this parallelism, namely that given by twistor transformations (7.1) for which the line P is left pointwise invariant. This group is readily seen to be the group of unitary (2×2) matices and leads to Uhlmann's representation (see Ref. 18) of the points of M in terms of such matrices.

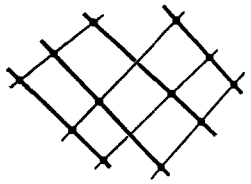

FIG. 6. A geometrical theorem concerning null lines in M.

that, in the M picture, X, Y, and Z are three null lines. Put

$$\bar{A}^{\alpha\beta\gamma} = X^{[\alpha} Y^{\beta} Z^{\gamma]}. \tag{9.9}$$

(Square brackets denote skew-symmetrization.) In the M picture, A is generally a "complexified null line". But if (9.8) is satisfied, then A is a real null line in M which meets each of X, Y, and Z. ($\bar{A}_{\alpha} X^{\alpha} = 0$ follows from $\bar{A}^{[\alpha\beta\gamma} X^{\delta]} = 0$, etc.) Thus, *the condition for the three null lines X, Y, Z to have a common null transversal* is simply (9.9) substituted into (9.8), i.e.,

$$\{XY\}\{YZ\}\{ZX\} = 1, \tag{9.10}$$

where

$$\{XY\} = -(X^{\alpha}\bar{Y}_{\alpha})(Y^{\beta}\bar{X}_{\beta})^{-1} = \{YX\}^{-1}$$

etc. (since X^{α}, Y^{α}, and Z^{α} are null). We assume no two of X, Y, Z meet so that (9.10) is well defined. If X, Y, and Z have *two* common transversals, then (9.9) vanishes and X, Y, and Z belong to a null regulus.

It is possible to derive a host of geometrical theorems concerning incidence of null lines in M, from the condition (9.10). For example, if four null lines in M, of which no two meet, are such that there is a null transversal to each of three different selections of triplets of the lines, then there is also a null transversal to the remaining triplet. (The configuration is that depicted in Fig. 6—except for the case when there is a single common transversal to all four lines.) Some theorems of this general type can be generated by means of a diagrammatic notation: a graph made up of triangles whose *vertices* represent null lines in M, can be used. The triangles represent triples of lines with a common null transversal. Now, if any one of these triangles represents a circuit in the graph linearly dependent on the circuits given by other triangles, then, because of the form of (9.10), we have a geometrical theorem. For the case just considered (cf., Fig. 6), the graph would be that of a tetrahedron.[36] Any other polyhedron of triangles would also give rise to a theorem on null lines.

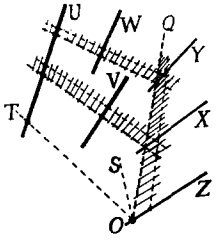

FIG. 7. Construction of the Robinson congruence containing three given null lines U, V, and W.

Finally, let us consider the *construction of a Robinson congruence* R containing three given null lines U, V, W in general position in M. (In the C picture, R is the plane of the three points U, V, W.) We wish to construct the line of R which passes through a general point O in M. Now, a Robinson congruence is characterized by the fact that, if any two null lines X, Y belong to the congruence, then so does every line of the null regulus containing X and Y. (In the C picture, the join of any two points in a plane also lies on the plane.) We have seen that the null regulus containing X and Y is simply the set of null transversals of any pair of null transversals of X and Y. Thus, if we can find X and Y belonging, respectively, to the null regulus containing U, V and to that containing U, W such that the null regulus containing X and Y also contains a line Z through O (see Fig. 7), the construction is complete. What is required, in fact, is to show that there is a null line Q through O which meets a line (namely X) of the U, V regulus and a distinct line (namely Y) of the U, W regulus; i.e., the triplets U, V, Q and U, W, Q must each have a null transversal. In fact, there is, in general, a unique such line Q through O. For, putting $Q^{\alpha} = S^{\alpha} + \zeta T^{\alpha}$, where T and S are null lines through O, with T meeting U, and substituting in (9.10), we get two simultaneous linear equations in ζ and $\bar{\zeta}$. The condition for these equations to have a unique solution reduces to $\{UV\}\{VT\}\{TW\}\{WU\} \neq 1$ which, for general positions of O, is indeed satisfied.[37]

X. THE METRIC TWISTOR

If we wish to use twistors to describe the Minkowski metric structure of M (i.e., $M\{I\}$), rather than just its conformal structure, we may do this by introducing a *metric twistor* $I^{\alpha\beta}$ which represents the point I of M, according to the scheme of Sec. IX. Thus, $I^{\alpha\beta}$ is simple, skew-symmetric, and real (in the sense of Sec. IX):

$$I^{\alpha\beta} I_{\alpha\gamma} = 0, \tag{10.1}$$

$$I^{\alpha\beta} = \bar{I}^{\alpha\beta} \tag{10.2}$$

[36] Since a real null line in M defines both a point in C and a plane in C through this point, the configuration of Fig. 6 is represented in the C picture as a pair of mutually inscribed and circumscribed tetrahedra—a configuration familiar to geometers. We may note that the *full* complexification of a null line in M leads, in the C picture, strictly to a point in C *together* with a plane through it. This gives a five-complex-dimensional system as we would expect.

[37] It only fails if there is a circle through O meeting U, V, and W which lies on a null cone through U.

(with $I^{\alpha\beta} = -I^{\beta\alpha}$). The twistor transformations which leave $I^{\alpha\beta}$ invariant give the *Poincaré group*. (The transformations which leave $I^{\alpha\beta}$ invariant up to *proportionality* give the Poincaré group plus the *dilatations*. These would be just the conformal transformations of M which leave the point I—and its null cone—invariant.)

We may also treat *de Sitter space-time* in a very similar way. In this case there is a hypersphere at infinity rather than a null cone at infinity. As we have seen in Sec. IX, this can be described by a real, skew-symmetric twistor which is *not simple*, i.e., we just drop condition (10.1). Other conformally flat space-times can also be treated. For example, if we retain (10.1) but drop (10.2) [and replace (10.3) by its modulus], we can describe the Einstein static universe. The matter is not pursued further here, however.

We can generate Poincaré covariant operations simply by employing twistor algebra and admitting, as basic elements of the algebra (in addition to $\epsilon_{\alpha\beta\rho\sigma}$, $\epsilon^{\alpha\beta\rho\sigma}$, δ^{α}_{β}), the twistors $I^{\alpha\beta}$, $I_{\alpha\beta}$, satisfying (10.1), (10.2). We can represent points P, of $M^*\{I\}$ by simple skew-symmetric twistors $P^{\alpha\beta}$ normalized, for convenience, so that

$$P^{\alpha\beta}I_{\alpha\beta} = 2. \tag{10.3}$$

($P \in M\{I\}$ if $P^{\alpha\beta} = \bar{P}^{\alpha\beta}$.) Then the correspondence between P and $P^{\alpha\beta}$ is unique.[38] If $Q^{\alpha\beta}$ and $R^{\alpha\beta}$ similarly represent points of $M^*\{I\}$, then an example of a Poincaré covariant operation is

$$aP^{\alpha\beta} + bQ^{\alpha\beta} + cR^{\alpha\beta} - \tfrac{1}{2}\{bcQ^{\gamma\delta}R_{\gamma\delta} + caR^{\gamma\delta}P_{\gamma\delta} + abP^{\gamma\delta}Q_{\gamma\delta}\}I^{\alpha\beta}. \tag{10.4}$$

The significance of this particular expression is that if

$$a + b + c = 1 \tag{10.5}$$

then (10.4) represents the point in $M^*\{I\}$ whose position vector is $ap^i + bq^i + cr^i$, where p^i, q^i, r^i are the respective position vectors of P, Q, R. [This vector operation is clearly Poincaré covariant if (10.5) holds and represents a weighted mean.] Expression (10.4) generalizes to any number of points in an obvious way. Other less involved expressions than (10.4) can, of course, also be given and define Poincaré covariant operations in $M\{I\}$. For example, $P^{\alpha\beta} + aI^{\alpha\beta}$ and $P^{\alpha\beta} - Q^{\alpha\beta}$ represent, respectively, a hypersphere center P and the hyperplane bisecting PQ orthogonally. (These twistors are not simple, so they represent "hyperspheres" rather than points.) We can check the coefficients in (10.4) using

$$P^{\alpha\beta}Q_{\alpha\beta} = -(p_i - q_i)(p^i - q^i), \tag{10.6}$$

which is a familiar Poincaré invariant expression.

The derivation of expressions such as (10.6) is facilitated if we specialize our twistor structure still further by the introduction of an *origin twistor* $O^{\alpha\beta}$ which represents a particular point O (the "origin") in M. Here, $O^{\alpha\beta}$ is to be simple, skew-symmetric, real and normalized as in (10.3)

$$O^{\alpha\beta}O_{\alpha\gamma} = 0, \qquad \bar{O}^{\alpha\beta} = O^{\alpha\beta}, \qquad O^{\alpha\beta}I_{\alpha\beta} = 2. \tag{10.7}$$

By putting $Q^{\alpha\beta} = O^{\alpha\beta}$ in (10.4) and (10.6), we obtain the twistor expressions for the Lorentz covariant vector operations of linear combination and squared magnitude.

Since the transformation group leaving both $I^{\alpha\beta}$ and $O^{\alpha\beta}$ invariant is the Lorentz group, we can expect to be able to express *spinors*, referred to the origin O, in *terms* of twistors. In effect, we carry out the construction given in Secs. II and III for twistors in terms of spinors, but in reverse. This gives us a correspondence between spinor indices and "reduced" twistor indices for which

$$\begin{aligned} O^{\alpha\beta} &\leftrightarrow \epsilon^{AB}, & I_{\alpha\beta} &\leftrightarrow \epsilon_{AB}, \\ O_{\alpha\beta} &\leftrightarrow \epsilon^{A'B'}, & I^{\alpha\beta} &\leftrightarrow \epsilon_{A'B'}. \end{aligned} \tag{10.8}$$

The *orthogonal idempotents* which reduce the twistor space—to a direct sum of two spinor spaces (one unprimed and one primed)—are

$$\begin{aligned} J^{\alpha}_{\beta} &= O^{\alpha\gamma}I_{\beta\gamma} \leftrightarrow \delta^{A}_{B} = \epsilon^{AG}\epsilon_{BG}, \\ \bar{J}^{\alpha}_{\beta} &= O_{\beta\gamma}I^{\alpha\gamma} \leftrightarrow \delta^{B'}_{A'} = \epsilon^{B'G'}\epsilon_{A'G'}. \end{aligned} \tag{10.9}$$

We have [from (10.1), (10.2), (10.7), etc.]

$$\begin{aligned} J^{\alpha}_{\beta}J^{\beta}_{\gamma} &= J^{\alpha}_{\gamma}, & J^{\alpha}_{\beta}\bar{J}^{\beta}_{\gamma} &= 0 = \bar{J}^{\alpha}_{\beta}J^{\beta}_{\gamma}, \\ \bar{J}^{\alpha}_{\beta}\bar{J}^{\beta}_{\gamma} &= \bar{J}^{\alpha}_{\gamma}, & J^{\alpha}_{\beta} + \bar{J}^{\alpha}_{\beta} &= \delta^{\alpha}_{\gamma}. \end{aligned} \tag{10.10}$$

A general twistor $P^{\alpha\beta\cdots\delta}_{\rho\cdots\tau}$, of valence $[{}^{r}_{s}]$ corresponds to a set of 2^{r+s} spinors. To obtain such a spinor, each index of $P^{\alpha\beta\cdots\delta}_{\rho\cdots\tau}$ is transvected with either a J^{α}_{β} or a $\bar{J}^{\alpha}_{\beta}$. The resulting twistor then represents a spinor with an unprimed index corresponding to each free J^{α}_{β} index and a primed index, in the *reverse position*, corresponding to each free $\bar{J}^{\alpha}_{\beta}$ index. For example,

$$P^{\kappa\lambda\mu}{}_{\varphi\chi}J^{\alpha}_{\kappa}\bar{J}^{\beta}_{\lambda}J^{\gamma}_{\mu}\bar{J}^{\varphi}_{\rho}J^{\chi}_{\sigma} \leftrightarrow \Pi^{A}{}_{B'}{}^{GR'}{}_{S}.$$

[38] In practice, the X^{α} and Y^{α} of the decomposition (9.3) (which may be specilized if desired) often turn out to be more convenient coordinates than $P^{\alpha\beta}$. This is of value in connection with physical fields and will be discussed elsewhere.

With the different possibilities for J^α_β and $\bar{J}^\alpha_\beta$, these give 2^{r+s} different spinors, which [by (10.10)] together determine $P^{\alpha\beta\cdots\delta}_{\rho\cdots\tau}$. Any spinor operation can thus be mirrored in twistor terms, using J^α_β and $\bar{J}^\alpha_{\dot\beta}$. The correspondences of (10.8) and (10.9) are consistent with this.

The basic relations (2.13) and (3.1) are expressed as

$$X^\beta J^\alpha_\beta \leftrightarrow \xi^A, \qquad X^\beta \bar{J}^\alpha_\beta \leftrightarrow \eta_{A'},$$
$$L^\beta J^\alpha_\beta \leftrightarrow \lambda^A, \qquad L^\beta \bar{J}^\alpha_\beta \leftrightarrow \mu_{A'}.$$

If we put

$$\nu P^{\alpha\beta} = X^\alpha L^\beta - L^\alpha X^\beta$$

[cf., (9.3)] with ν chosen so that (10.3) holds, we get $\nu = \xi_A \lambda^A$. Then,

$$\left.\begin{aligned} P^{\rho\sigma} J^\alpha_\rho J^\beta_\sigma &\leftrightarrow \epsilon^{AB}, \\ P^{\rho\sigma} J^\alpha_\rho \bar{J}^\beta_\sigma &\leftrightarrow -i p^A_{B'}, \\ P^{\rho\sigma} \bar{J}^\alpha_\rho \bar{J}^\beta_\sigma &\leftrightarrow -\tfrac{1}{2} p_i p^i \epsilon_{A'B'}, \end{aligned}\right\} \qquad (10.11)$$

where $p_{AB'}$ is given as in (3.10). In the C picture P is the line joining two points X and L; in M^*, P is the (complex) point of intersection of two ("complexified") null lines. Thus, according to (3.10), p^i represents the *position vector*, with respect to O, of the point P in M^*. We can express (10.11) in matrix form as

$$(P^{\alpha\beta}) = \begin{bmatrix} \epsilon^{AB} & -ip^A{}_{B'} \\ ip_{A'}{}^B & -\tfrac{1}{2} p_i p^i \epsilon_{A'B'} \end{bmatrix}.$$

Expressions such as (10.6) and (10.4) then follow at once.

ACKNOWLEDGMENTS

The author is greatly indebted particularly to I. Robinson, but also to E. L. Schücking and to R. P. Kerr, for stimulating ideas in the early stages of this work.

This work was done partly at the University of Texas, where it was supported, in part, by the U. S. Air Force under Contract, AF-AFOSR-454-63.