JOURNAL OF
MATHEMATICAL
PHYSICS

VOLUME 7, NUMBER 7 JULY 1966

Space Groups and Selection Rules

C. J. BRADLEY
Department of Metallurgy, University of Oxford, Oxford, England
(Received 15 November 1965)

A rigorous method is presented for obtaining the Clebsch–Gordan coefficients for the Kronecker products of induced representations of finite groups. The result obtained is applied to the theory of space groups, and the final formula for the coefficients has the advantage, as with other subgroup techniques, of being expressed in terms of the characters of small representations, thereby providing an additional proof independent of Lax, yet supporting his point of view the subgroup and full-group procedures are equivalent, and at the same time the result clarifies subgroup treatments so far given. An example using $P23$ is given.

1. INTRODUCTION

A UNITARY irreducible representation $\Gamma^{\mathbf{k}}_{p}$ of a space group $\mathbf{G}$ is uniquely distinguished by giving it two labels, a vector $\mathbf{k}$ out of a symmetrically chosen basic domain of the Brillouin zone and an index p to denote the small representation of the little group $\mathbf{G}^{\mathbf{k}}$ to which it is associated. To be more precise, if $D^{\mathbf{k}}_{p}$ is a small representation of $\mathbf{G}^{\mathbf{k}}$ then $\Gamma^{\mathbf{k}}_{p}$ is the representation induced from $D^{\mathbf{k}}_{p}$ in $\mathbf{G}$. Electron and phonon states in crystals belong to such representations and are therefore similarly labeled.

A transition between initial states belonging to $\Gamma^{\mathbf{m}}_{q}$ and final states belonging to $\Gamma^{\mathbf{h}}_{r}$ caused by a potential belonging to $\Gamma^{\mathbf{k}}_{p}$ is forbidden by the symmetry of $\mathbf{G}$ if the inner Kronecker product $(\Gamma^{\mathbf{k}}_{p} \otimes \Gamma^{\mathbf{m}}_{q})$ does not contain $\Gamma^{\mathbf{h}}_{r}$ or, equivalently, if the triple inner product $(\Gamma^{\mathbf{h}}_{r}{}^{*} \otimes \Gamma^{\mathbf{k}}_{p} \otimes \Gamma^{\mathbf{m}}_{q})$ does not contain the totally symmetric representation of $\mathbf{G}$.

Elliott and Loudon[1] investigate the analogous problem in which the states and the potential belong to three little groups and for which the answer depends on whether or not the triple product $(D^{\mathbf{h}}_{r}{}^{*} \otimes D^{\mathbf{k}}_{p} \otimes D^{\mathbf{m}}_{q})$ subduced onto $\mathbf{N} = \mathbf{G}^{\mathbf{h}} \wedge \mathbf{G}^{\mathbf{k}} \wedge \mathbf{G}^{\mathbf{m}}$ contains the totally symmetric representation of $\mathbf{N}$.

In practice it is often the first problem rather than the second which is of physical significance, that is, when the selection rule depends on contributions obtained from integrals in which a sum is taken over all initial and final states belonging to the appropriate representations of the whole space group. Of course, Elliott and Loudon appreciated this and took care in applying their rule to physical problems to allow the vectors $\mathbf{m}$, $\mathbf{k}$, and $\mathbf{h}$ to vary over their respective stars. That it would be sufficient to apply Elliott and Loudon's result repeatedly to solve the first problem was implied but not proved by the authors. In Sec. 4 we obtain a formula which in fact shows this procedure to be more than sufficient. The next work on the subject, by Lax and Hopfield,[2] was simply a development in technique for dealing with Elliott and Loudon's result. Then Zak[3] set about to prove Elliott and Loudon's formula from first principles, using the characters of induced representations. Although it is not clear whether Zak meant this to

[1] R. J. Elliott and R. Loudon, J. Phys. Chem. Solids **15**, 146 (1960).

[2] M. Lax and J. Hopfield, Phys. Rev. **124**, 115 (1961).
[3] J. Zak, J. Math. Phys. **3**, 1278 (1961).

be so, his paper gave the impression that, if a single pair of vectors $\mathbf{m}'$ and $\mathbf{k}'$ could be found from the stars of $\mathbf{m}$ and $\mathbf{k}$ such that $\mathbf{m}' + \mathbf{k}' \sim \mathbf{h}$ (we use the symbol $\sim$ to denote equivalence), then the problem of the whole space group is solved by investigation of the inner Kronecker product of the small representations of the single triple $\mathbf{m}'$, $\mathbf{k}'$, and $\mathbf{h}$. Confusion may have arisen because this is true for the vast majority of cases, but as we show in Secs. 4 and 5 this is not true in all cases. Zak's proof is adapted to such cases in which only one triple of vectors needs to be considered.

At about this time Birman,[4] trying to avoid subgroup techniques, possibly because he noticed the deficiency just mentioned, derived the Clebsch–Gordan coefficients for the single-valued representations of the diamond and zinc-blende structures, using a rather laborious method on the whole space group. The tables given are useful, but the method is extremely involved and makes no use of the simplifications that can be obtained by using the theory of induced representations.

Subsequent papers by Lax[5,6] are partly concerned with extra selection rules that can be obtained when time reversal is taken into account. In these papers it is realized that, if the vectors $\mathbf{m}$, $\mathbf{k}$, and $\mathbf{h}$ are related, as well as the basis functions of their little groups, then time reversal can yield additional selection rules. Also, in the latter paper, Lax[6] establishes the equivalence of subgroup and full-group procedures which, as will be seen in Sec. 4, we support and clarify. What does not seem, however, to have been made adequately clear so far is that particular relationships between $\mathbf{m}$, $\mathbf{k}$, and $\mathbf{h}$ can affect the value of the Clebsch–Gordan coefficients. The formula we derive in Sec. 4 shows in a completely unambiguous way which relationships, if any, produce such additional information. In other words, in using our method, no *ad hoc* counting technique has to be superimposed on the method of Elliott and Loudon to get the right results. An example to illustrate this is given in Sec. 5.

The purpose of this article, therefore, is to resolve ambiguities and uncertainties in how to obtain the correct value of the Clebsch–Gordan coefficients for inner Kronecker products of space-group representations within the framework of the little group intersection method of Elliott and Loudon. To proceed with this program of work we prove in Sec. 2 certain theorems on induced representations. Some of these theorems were proved or stated first by Mackey,[7,8] but not in a context which makes understandable reading for any but the expert algebraist. From the point of view of the mathematical physicist, an explanation of Mackey's work by Coleman[9] is a better reference but unfortunately this exists only as an unpublished reprint; the theorems of Sec. 2 are therefore not yet widely accessible to mathematical physicists. It is because of this and also because we have simplified some of the proofs that these theorems are given in some detail. We claim no credit except for the work following Theorem 2.8 and embodied by Eq. (2.14), which is the necessary new result for the applications we make later. In Sec. 3 we give a short review of the theory of space groups. Then in Sec. 4 the results of Secs. 2 and 3 are combined to produce the required results, which are simplified and discussed pictorially. The space group $P23$ is used as an example in Sec. 5.

2. INDUCED REPRESENTATIONS

In this section $\mathbf{G}$ is any finite group, and if in a given theorem $\mathbf{H}$, $\mathbf{K}$, or $\mathbf{M}$ appear they are subgroups of $\mathbf{G}$. C, D, and B are respectively representations of $\mathbf{H}$, $\mathbf{K}$, and $\mathbf{M}$ and the character of D in $\mathbf{K}$ is denoted by ψ: All representations are unitary, but unless otherwise stated they are not necessarily irreducible. The order of a group $\mathbf{G}$ will be denoted by $|\mathbf{G}|$, its elements by g_i, $i = 1$ to $|\mathbf{G}|$, and its totally symmetric representation by $A(\mathbf{G})$. An asterisk always denotes complex conjugation. Other notations introduced from time to time also retain their meanings throughout the whole section. We should like to mention in particular the arrow notation used by Coleman, the upward arrow for induction and the downward arrow for subduction. (The precise meaning of these arrows is defined in the appropriate place in the text.) We mention this because we believe that this notation used by Coleman has done more than anything else to aid progress in this field.

Certain trivial or well-known theorems will be stated without proof, for example:

Theorem (2.1): Let Γ^i, $i = 1$ to r, be the irreducible representations of $\mathbf{G}$ and suppose that Γ is a representation equivalent to the direct sum

[4] J. L. Birman, Phys. Rev. **127**, 1093 (1962).
[5] M. Lax, *International Conference on Physical Semiconductors, Exeter* (Institute of Physics and the Physical Society, London, 1962), p. 395.
[6] M. Lax, Phys. Rev. **138**, A793 (1965).
[7] G. W. Mackey, Ann. Math. **55**, 101 (1952).
[8] G. W. Mackey, Ann. Math. **58**, 193 (1953).
[9] A. J. Coleman, Report No. 102, Quantum Chemistry Group, Uppsala University, Uppsala, Sweden.

$\sum_{i=1}^{r} C_i\Gamma^i$, then C_i is the frequency of $A(\mathbf{G})$ in $\Gamma^* \otimes \Gamma^i$ (or $\Gamma \otimes \Gamma^{i*}$).

For proof of this theorem see, for example, Hammermesh.[10]

We now define what is meant by the representation $D \uparrow \mathbf{G}$, the representation D induces from $\mathbf{K}$ into $\mathbf{G}$. Let us factorize $\mathbf{G}$ into left cosets with respect to $\mathbf{K}$ so that

$$\mathbf{G} = \sum_{\sigma} p_\sigma \mathbf{K}, \tag{2.1}$$

where p_σ are left coset representatives, which when chosen we shall suppose fixed once for all. Equation (2.1) means that each element $g \in \mathbf{G}$ is uniquely expressible in the form $p_\lambda k_s$, where p_λ is one of the chosen coset representatives and where $k_s \in \mathbf{K}$ is uniquely determined by g. Let ϕ_r, $r = 1$ to d, be a basis for D so that for all $k_i \in \mathbf{K}$

$$k_i\phi_r = \sum_{t=1}^{d} \phi_t\, D(k_i)_{tr}. \tag{2.2}$$

Define for all σ and r the $d\,|\mathbf{G}|/|\mathbf{K}|$ functions

$$\phi_{\sigma r} = p_\sigma \phi_r. \tag{2.3}$$

Equation (2.3) is a functional identity, so that, for equal values of their argument, $\phi_{\sigma r}$ and $p_\sigma\phi_r$ have the same value. Suppose that $gp_r \in p_\gamma \mathbf{K}$ so that, for some m, $p_\gamma^{-1} g p_r = k_m$. Then

$$\begin{aligned} g\phi_{rr} &= gp_r\phi_r \\ &= p_\gamma k_m \phi_r \\ &= \sum_{t=1}^{d} \phi_{\gamma t}\, D(k_m)_{tr}. \end{aligned} \tag{2.4}$$

Equation (2.4) follows immediately from Eqs. (2.2) and (2.3), and since g is any element of $\mathbf{G}$, it follows that the vector space V spanned by the functions $\phi_{\sigma r}$ is invariant under $\mathbf{G}$. If we denote by $\langle\phi_r|$ the row vector comprised of the d elements $\phi_{r1}, \phi_{r2}, \cdots, \phi_{rd}$, then Eq. (2.4) can be written in matrix notation as

$$g\langle\phi_r| = \langle\phi_\gamma|\, D(p_\gamma^{-1} g p_r).$$

Note that γ is determined *uniquely* by g and r. We now define the symbol $\delta_{\lambda,\sigma\sigma}$ to be unity if $gp_\sigma \in p_\lambda \mathbf{K}$, and to be zero otherwise. The $d\,|\mathbf{G}|/|\mathbf{K}|$ dimensional representation of $\mathbf{G}$ in the vector space V defined by the functions $\phi_{\sigma r}$, $\sigma = 1$ to $|\mathbf{G}|/|\mathbf{K}|$, and $r = 1$ to d, is called the *induced representation* $D \uparrow \mathbf{G}$. From Eq. (2.4) we see that it is comprised of block matrices labeled with Greek letters γ, τ, etc. such that the (γ, τ) block matrix of $D \uparrow \mathbf{G}$ is given by

$$(D \uparrow \mathbf{G})_{\gamma\tau}(g) = D(p_\gamma^{-1} g p_\tau)\, \delta_{\gamma,g\tau}. \tag{2.5}$$

In the representation $D \uparrow \mathbf{G}$ there are only $|\mathbf{G}|/|\mathbf{K}|$ nonvanishing block matrices, one somewhere in each row and one somewhere in each column, just where depending on the array of numbers $\delta_{\gamma,g\tau}$. Within any given block the rows and columns are as in Eq. (2.4) labeled with Latin letters r, t, etc. The character χ of $D \uparrow \mathbf{G}$ is easily determined in terms of ψ. Thus,

$$\begin{aligned} \chi(g) &= \sum_{\sigma} \mathrm{Tr}\,[(D \uparrow \mathbf{G})_{\sigma\sigma}(g)] \\ &= \sum_{\sigma} \mathrm{Tr}\,[D(p_\sigma^{-1} g p_\sigma)\, \delta_{\sigma,g\sigma}] \\ &= \sum_{\sigma} \psi(p_\sigma^{-1} g p_\sigma)\, \delta_{\sigma,g\sigma}. \end{aligned} \tag{2.6}$$

In Eq. (2.6), $\delta_{\sigma,g\sigma}$ implies that the sum over σ is restricted to those σ for which $gp_\sigma \in p_\sigma\mathbf{K}$, that is, those σ for which $p_\sigma \mathbf{K} p_\sigma^{-1} = \mathbf{K}_\sigma$ contains g.

Theorem (2.2): The character χ of $D \uparrow \mathbf{G}$ is independent of the particular choice of coset representatives p_σ.

This follows immediately from the invariant form of Eq. (2.6) under change of p_σ. This theorem implies that, two representations induced by using two different choices of coset representatives, though not identical, are equivalent.

Theorem (2.3): The frequency of $A(\mathbf{G})$ in $D \uparrow \mathbf{G}$ is equal to the frequency of $A(\mathbf{K})$ in D.

Proof: We write $f(\Gamma^i)/\Gamma$ for the frequency of Γ^i in Γ.

$$\begin{aligned} f[A(\mathbf{G})]/(D \uparrow \mathbf{G}) &= \frac{1}{|\mathbf{G}|} \sum_{g\in\mathbf{G}} \chi(g) \\ &= \frac{1}{|\mathbf{G}|} \sum_{g\in\mathbf{G}} \sum_{\sigma} \psi(p_\sigma^{-1} g p_\sigma)\, \delta_{\sigma,g\sigma} \qquad (2.7) \\ &= \frac{1}{|\mathbf{G}|} \sum_{\sigma} \sum_{g\in\mathbf{K}_\sigma} \psi(p_\sigma^{-1} g p_\sigma) \qquad (2.8) \\ &= \frac{1}{|\mathbf{G}|} \sum_{\sigma} \sum_{k\in\mathbf{K}} \psi(k) \\ &= \frac{1}{|\mathbf{K}|} \sum_{k\in\mathbf{K}} \psi(k) \qquad (2.9) \\ &= f[A(\mathbf{K})]/D. \end{aligned}$$

In the proof, Eq. (2.7) follows from Eq. (2.6), Eq. (2.8) by reversal of the order of summation, and Eq. (2.9) because there are $|\mathbf{G}|/|\mathbf{K}|$ cosets σ.

[10] M. Hammermesh, *Group Theory and its Application to Physical Problems* (Pergamon Press, Inc., New York, 1962), Chap. 5.

We now define what is meant by a *subduced representation.* If Γ is a representation of $\mathbf{G}$ then $\Gamma \downarrow \mathbf{K}$ denotes the representation Γ subduces in $\mathbf{K}$: it consists of the restriction of Γ to elements of $\mathbf{K}$ (and is therefore of the same dimension as Γ); it is a well-defined representation since $\mathbf{K}$ is a subgroup of $\mathbf{G}$.

Theorem (2.4): Let Γ be a representation of $\mathbf{G}$; then (using $\sim$ to denote equivalence)

$$\Gamma \otimes (D \uparrow \mathbf{G}) \sim [(\Gamma \downarrow \mathbf{K}) \otimes D] \uparrow \mathbf{G}.$$

Proof: We establish the equivalence by showing that the characters of the two representations coincide. Denote the character of Γ by θ. Then the character of g in $\Gamma \otimes (D \uparrow \mathbf{G})$ is

$$\theta(g)\chi(g) = \theta(g) \sum_{\sigma} \psi(p_\sigma^{-1} g p_\sigma)\ \delta_{\sigma, g\sigma}.$$

Since the character of k in $(\Gamma \downarrow \mathbf{K}) \otimes D$ is $\theta(k)\psi(k)$, it follows from Eq. (2.6) that the character of g in $[(\Gamma \downarrow \mathbf{K}) \otimes D] \uparrow \mathbf{G}$ is

$$\sum_{\sigma} \theta(p_\sigma^{-1} g p_\sigma)\psi(p_\sigma^{-1} g p_\sigma)\ \delta_{\sigma, g\sigma}.$$

But g and $p_\sigma^{-1} g p_\sigma$ are in the same class in $\mathbf{G}$ and θ is a representation of $\mathbf{G}$ so $\theta(p_\sigma^{-1} g p_\sigma) = \theta(g)$ for all σ. The result follows.

Theorem (2.5): Transitivity of Induction. Let $\mathbf{K}$ be a subgroup of $\mathbf{L}$ and $\mathbf{L}$ a subgroup of $\mathbf{G}$. Then $(D \uparrow \mathbf{L}) \uparrow \mathbf{G} \sim D \uparrow \mathbf{G}$.

Proof: We establish the equivalence by showing that if Γ is an irreducible representation of $\mathbf{G}$ then its frequency in $(D \uparrow \mathbf{L}) \uparrow \mathbf{G}$ is equal to its frequency in $D \uparrow \mathbf{G}$.

$$\begin{aligned}
&f(\Gamma)/D \uparrow \mathbf{G} \\
&\quad = f[A(\mathbf{G})]/\Gamma^* \otimes D \uparrow \mathbf{G} && \text{by Theorem (2.1)} \\
&\quad = f[A(\mathbf{G})]/[(\Gamma^* \downarrow \mathbf{K}) \otimes D] \uparrow \mathbf{G} && \text{by Theorem (2.4)} \\
&\quad = f[A(\mathbf{K})]/(\Gamma^* \downarrow \mathbf{K}) \otimes D && \text{by Theorem (2.3).}
\end{aligned}$$

But

$$\begin{aligned}
&f(\Gamma)/(D \uparrow \mathbf{L}) \uparrow \mathbf{G} \\
&\quad = f[A(\mathbf{G})]/\Gamma^* \otimes [(D \uparrow \mathbf{L}) \uparrow \mathbf{G}] && \text{by Theorem (2.1)} \\
&\quad = f[A(\mathbf{G})]/[(\Gamma^* \downarrow \mathbf{L}) \otimes (D \uparrow \mathbf{L})] \uparrow \mathbf{G} \\
&\qquad && \text{by Theorem (2.4)} \\
&\quad = f[A(\mathbf{L})]/(\Gamma^* \downarrow \mathbf{L}) \otimes (D \uparrow \mathbf{L}) && \text{by Theorem (2.3)} \\
&\quad = f[A(\mathbf{L})]/\{[(\Gamma^* \downarrow \mathbf{L}) \downarrow \mathbf{K}] \otimes D\} \uparrow \mathbf{L} \\
&\qquad && \text{by Theorem (2.4)} \\
&\quad = f[A(\mathbf{K})]/(\Gamma^* \downarrow \mathbf{K}) \otimes D && \text{by Theorem (2.3),}
\end{aligned}$$

and because $(\Gamma^* \downarrow \mathbf{L}) \downarrow \mathbf{K} = (\Gamma^* \downarrow \mathbf{K})$ (both sides consisting of identical matrices).

Let us now factorize $\mathbf{G}$ into double cosets with respect to $\mathbf{H}$ and $\mathbf{K}$:

$$\mathbf{G} = \sum_{\alpha} \mathbf{H}\, d_\alpha \mathbf{K} \tag{2.10}$$

in which the d_α are called the double coset representatives, and in the complex $\mathbf{H}\, d_\alpha \mathbf{K}$ we count any element that appears once only. The expansion (2.10) is unique in much the same way as the decomposition (2.1) into left cosets: the double coset representatives are not unique, any element of a double coset serving equally well as its representative; furthermore each element of $\mathbf{G}$ appears in one and only one double coset. Again we suppose that the d_α once chosen remain fixed.

Theorem (2.6): The double coset $\mathbf{H}\, d_\alpha \mathbf{K}$ contains $|\mathbf{H}|/|\mathbf{L}_\alpha|$ left cosets of $\mathbf{K}$, where $\mathbf{L}_\alpha = \mathbf{H} \wedge \mathbf{K}_\alpha$ and $\mathbf{K}_\alpha = d_\alpha \mathbf{K}\, d_\alpha^{-1}$.

Outline proof: If we express $\mathbf{H}$ in terms of left cosets with respect to $\mathbf{L}_\alpha$:

$$\mathbf{H} = \sum_{\gamma} q_{\alpha\gamma} \mathbf{L}_\alpha, \tag{2.11}$$

then what happens is that $h\, d_\alpha \mathbf{K} \in q_{\alpha\gamma}\, d_\alpha \mathbf{K}$ if and only if $h \in q_{\alpha\gamma} \mathbf{L}_\alpha$. Thus the left cosets of $\mathbf{K}$ in $\mathbf{H}\, d_\alpha \mathbf{K}$ are in $(1-1)$ correspondence with the left cosets of $\mathbf{L}_\alpha$ in $\mathbf{H}$, and are therefore $|\mathbf{H}|/|\mathbf{L}_\alpha|$ in number. Note that because of this theorem it is possible to choose in Eq. (2.1) $p_\sigma = q_{\alpha\gamma}\, d_\alpha$ and that as α and γ run over all possible values so does σ.

We define the representation D_α of $\mathbf{K}_\alpha$ by the equation

$$D_\alpha(d_\alpha\, k\, d_\alpha^{-1}) = D(k), \quad \text{for all} \quad k \in \mathbf{K}. \tag{2.12}$$

Theorem (2.7): Mackey's Subgroup Theorem.

$$(D \uparrow \mathbf{G}) \downarrow \mathbf{H} \sim \sum_{\alpha} (D_\alpha \downarrow \mathbf{L}_\alpha) \uparrow \mathbf{H}.$$

It should be emphasized here that the sum over α is over all α appropriate to the double coset decomposition (2.10) even though there may exist distinct α and β in the sum for which $\mathbf{L}_\alpha = \mathbf{L}_\beta$ and $D_\alpha \sim D_\beta$.

Proof: We show that the characters of the two representations coincide. The character of h in $(D \uparrow \mathbf{G}) \downarrow \mathbf{H}$ from Eq. (2.6) is

$$\chi(h) = \sum_{\sigma} \psi(p_\sigma^{-1} h p_\sigma)\ \delta_{\sigma, h\sigma}.$$

Now,

$$\mathbf{K}_\sigma = p_\sigma \mathbf{K} p_\sigma^{-1} = q_{\alpha\gamma}\, d_\alpha \mathbf{K}\, d_\alpha^{-1} q_{\alpha\gamma}^{-1} = q_{\alpha\gamma} \mathbf{K}_\alpha q_{\alpha\gamma}^{-1}.$$

Hence

$$\chi(h) = \sum_{\alpha,\gamma} \psi(d_\alpha^{-1} q_{\alpha\gamma}^{-1} h q_{\alpha\gamma}\, d_\alpha),$$

where for a given α the sum over γ is restricted to those γ for which $q_{\alpha\gamma}^{-1} h q_{\alpha\gamma} \in \mathbf{K}_\alpha$.

Now from Eq. (2.12)

$$D(d_\alpha^{-1} q_{\alpha\gamma}^{-1} h q_{\alpha\gamma}\, d_\alpha) = D_\alpha(q_{\alpha\gamma}^{-1} h q_{\alpha\gamma}).$$

Therefore,

$$\chi(h) = \sum_\alpha [\sum_\gamma \mathrm{Tr}\, D_\alpha(q_{\alpha\gamma}^{-1} h q_{\alpha\gamma})]. \qquad (2.13)$$

But $q_{\alpha\gamma} \in \mathbf{H}$ so $q_{\alpha\gamma}^{-1} h q_{\alpha\gamma} \in \mathbf{K}_\alpha \wedge \mathbf{H}$ and therefore in Eq. (2.13) the restriction on the sum over γ for fixed α is that $q_{\alpha\gamma}^{-1} h q_{\alpha\gamma} \in \mathbf{L}_\alpha$. From Eq. (2.6), with h replacing g, $q_{\alpha\gamma}$ replacing p_σ, and $\mathbf{L}_\alpha$ replacing $\mathbf{K}$, we see that the right-hand side of Eq. (2.13) is just the character of h in $\sum_\alpha (D_\alpha \downarrow \mathbf{L}_\alpha) \uparrow \mathbf{H}$.

Theorem (2.8):

$$(D \uparrow \mathbf{G}) \otimes (C \uparrow \mathbf{G}) \sim \sum_\alpha \{(D_\alpha \otimes C) \downarrow \mathbf{L}_\alpha\} \uparrow \mathbf{G}.$$

Proof: From Theorem (2.4) the left-hand side is equivalent to $\{[(D \uparrow \mathbf{G}) \downarrow \mathbf{H}] \otimes C\} \uparrow \mathbf{G}$, which from Theorem (2.7) we have just proved is equivalent to

$$\sum_\alpha \{[(D_\alpha \downarrow \mathbf{L}_\alpha) \uparrow \mathbf{H}] \otimes C\} \uparrow \mathbf{G}.$$

From Theorem (2.4) again this is in turn equivalent to

$$\sum_\alpha \{[(C \downarrow \mathbf{L}_\alpha) \otimes (D_\alpha \downarrow \mathbf{L}_\alpha)] \uparrow \mathbf{H}\} \uparrow \mathbf{G}.$$

The result follows from the distributivity of subduction, which is obvious, and the transitivity of induction, Theorem (2.5).

We write $E_\alpha = (D_\alpha \otimes C) \downarrow \mathbf{L}_\alpha$ so that E_α is a representation of $\mathbf{L}_\alpha = \mathbf{H} \wedge \mathbf{K}_\alpha$.

If we now write $\mathbf{G} = \sum_\beta \mathbf{L}_\alpha b_\beta \mathbf{M}$ and let $\mathbf{N}_{\alpha\beta} = \mathbf{L}_\alpha \wedge \mathbf{M}_\beta$, where $\mathbf{M}_\beta = b_\beta \mathbf{M} b_\beta^{-1}$ and if we define a representation B_β on $\mathbf{M}_\beta$ by the equation $B_\beta(b_\beta m b_\beta^{-1}) = B(m)$, for all $m \in \mathbf{M}$, then as a direct application of Theorem (2.8)

$$(B \uparrow \mathbf{G}) \otimes (E_\alpha \uparrow \mathbf{G}) \sim \sum_\beta \{(B_\beta \otimes E_\alpha) \downarrow \mathbf{N}_{\alpha\beta}\} \uparrow \mathbf{G},$$

from which we conclude that

$$(B \uparrow \mathbf{G}) \otimes (D \uparrow \mathbf{G}) \otimes (C \uparrow \mathbf{G}) \sim \sum_\beta \sum_\alpha \{(B_\beta \otimes D_\alpha \otimes C) \downarrow \mathbf{M}_\beta \wedge \mathbf{K}_\alpha \wedge \mathbf{H}\} \uparrow \mathbf{G}. \qquad (2.14)$$

Equation (2.14) can obviously be generalized by repeated application of Theorem (2.8) to express an inner Kronecker product of as many induced representations as we want. However, we are interested in the triple product and in particular the result we require in Sec. 4 is the frequency of $A(\mathbf{G})$ in $(B \uparrow \mathbf{G}) \otimes (D \uparrow \mathbf{G}) \otimes (C \uparrow \mathbf{G})$. From Eq. (2.14) and Theorem (2.3), this is the double sum over α and β of the frequency of $A(\mathbf{N}_{\alpha\beta})$ in $(B_\beta \otimes D_\alpha \otimes C)$ restricted to $\mathbf{N}_{\alpha\beta} = \mathbf{M}_\beta \wedge \mathbf{K}_\alpha \wedge \mathbf{H}$. Here α is in correspondence with the double coset representatives of $\mathbf{G}$ with respect to $\mathbf{H}$ and $\mathbf{K}$, and β to those of $\mathbf{L}_\alpha$ and $\mathbf{M}$.

3. SPACE GROUPS

In this section we review, mainly without proof, some well-known facts about space groups. If proofs are required the reader should refer to the standard works, a good bibliography of which is contained in the review article by Koster.[11]

In dealing with the case in which $\mathbf{G}$ is a space group we use the notation of Seitz.[12] Thus, if

$\{R \mid \mathbf{v}\} \in \mathbf{G}$,

$$\{R \mid \mathbf{v}\}\mathbf{r} = R\mathbf{r} + \mathbf{v}. \qquad (3.1)$$

Here R is a point group operation and $\mathbf{v}$ may or may not be a lattice translation $\mathbf{t}$. We denote the identity by $\{E \mid \mathbf{0}\}$. The elements $\{E \mid \mathbf{t}\}$, where $\mathbf{t}$ is any lattice translation, form a subgroup $\mathbf{T}$ which is an invariant Abelian subgroup of $\mathbf{G}$. The irreducible representations of $\mathbf{T}$, $D^{\mathbf{k}}$, are therefore one-dimensional and may be written in the form

$$D^{\mathbf{k}}(\{E \mid \mathbf{t}\}) = \exp(-i\mathbf{k}\cdot\mathbf{t}). \qquad (3.2)$$

In Eq. (3.2) $\mathbf{k}$ is a wave vector in reciprocal space. The basis function for the representation $D^{\mathbf{k}}$ is a Bloch function $\phi_{\mathbf{k}}(\mathbf{r})$ for which

$$\{E \mid \mathbf{t}\}\phi_{\mathbf{k}}(\mathbf{r}) = \phi_{\mathbf{k}}(\mathbf{r} - \mathbf{t}) = \exp(-i\mathbf{k}\cdot\mathbf{t})\phi_{\mathbf{k}}(\mathbf{r}). \qquad (3.3)$$

Two vectors $\mathbf{k}$ and $\mathbf{c}$ are said to be equivalent ($\mathbf{k} \sim \mathbf{c}$) if they differ by a reciprocal lattice vector $\mathbf{g}$ such that $\exp(i\mathbf{g}\cdot\mathbf{t}) = 1$. From Eq. (3.2) $D^{\mathbf{k}}$ and $D^{\mathbf{c}}$ are then identical and the representations of $\mathbf{T}$ are therefore in 1–1 correspondence with a complete set of nonequivalent $\mathbf{k}$ vectors. These can be chosen to form the interior and half the surface of the first Brillouin zone. If $\phi_{\mathbf{k}}(\mathbf{r})$ is a function satisfying Eq. (3.3) then $\{R|\mathbf{v}\}\phi_{\mathbf{k}}(\mathbf{r})$ satisfies an analogous equation with $\exp(-iR\mathbf{k}\cdot\mathbf{t})$

[11] G. F. Koster, *Solid State Physics*, F. Seitz and Turnbull, Eds. (Academic Press Inc., New York 1957), Vol. 5, p. 173.
[12] F. Seitz, Ann. Math. 37, 17 (1936).

replacing $\exp(-i\mathbf{k}\cdot\mathbf{t})$. That is to say $\{R \mid \mathbf{v}\}\phi_{\mathbf{k}}(\mathbf{r})$ is a Bloch function with wave vector $R\mathbf{k}$.

This fact motivates the idea of the group of $\mathbf{k}$. This consists of all the elements of $\mathbf{G}$ which transform a Bloch function with wave vector equivalent to $\mathbf{k}$ into a Bloch function also with wave vector equivalent to $\mathbf{k}$. We denote this group by $\mathbf{G}^{\mathbf{k}}$. If we decompose $\mathbf{G}$ into left cosets with respect to $\mathbf{G}^{\mathbf{k}}$:

$$\mathbf{G} = \sum_{i=1}^{d} \{R_i \mid \mathbf{v}_i\}\mathbf{G}^{\mathbf{k}}, \tag{3.4}$$

then the set of d vectors $R_i\mathbf{k}$, $i = 1$ to d, forms the *star* of $\mathbf{k}$. It should be noticed that the group of $R_i\mathbf{k}$ is a subgroup of $\mathbf{G}$ conjugate to $\mathbf{G}^{\mathbf{k}}$, namely $\{R_i \mid \mathbf{v}_i\}\mathbf{G}^{\mathbf{k}}\{R_i \mid \mathbf{v}_i\}^{-1}$.

The irreducible representations of $\mathbf{G}$ are classified within the framework of these concepts. First we distribute all the mutually nonequivalent $\mathbf{k}$ vectors into stars and choose one $\mathbf{k}$ vector from each star. If h is the macroscopic order of symmetry of $\mathbf{G}$ then $h = |\mathbf{G}|/|\mathbf{T}|$ and the Brillouin zone has the symmetry of a point group of order h. What happens is that the single $\mathbf{k}$ vectors from each star just described can be chosen so that they fill a symmetrically chosen $1/h$ part of the Brillouin zone. This we call the *basic zone*. The scheme for finding all the irreducible representations of $\mathbf{G}$ is then as follows:

(i) Choose a $\mathbf{k}$ vector from the basic zone.

(ii) If we write

$$D^{\mathbf{k}} \uparrow \mathbf{G}^{\mathbf{k}} \sim \sum_{p} \lambda_p^{\mathbf{k}} D_p^{\mathbf{k}}, \tag{3.5}$$

where the $D_p^{\mathbf{k}}$ are irreducible representations of $\mathbf{G}^{\mathbf{k}}$, then $D_p^{\mathbf{k}}$ has the property that

$$D_p^{\mathbf{k}}(\{E \mid \mathbf{t}\}) = \exp(-i\mathbf{k}\cdot\mathbf{t})\mathbf{1}, \tag{3.6}$$

where, in Eq. (3.6), $\mathbf{1}$ is the unit matrix of dimension $\lambda_p^{\mathbf{k}}$. The $D_p^{\mathbf{k}}$ appearing in the sum (3.5) are called the *small representations* of $\mathbf{G}^{\mathbf{k}}$. The second step is to determine all such representations $D_p^{\mathbf{k}}$.

(iii) Select any small representation $D_p^{\mathbf{k}}$ and form the induced representation $\Gamma_p^{\mathbf{k}} = D_p^{\mathbf{k}} \uparrow \mathbf{G}$. The result is that $\Gamma_p^{\mathbf{k}}$ is irreducible and that if we perform the process described above for all $\mathbf{k}$ in the basic zone, for each $\mathbf{k}$ for all p, then we obtain all irreducible representations of $\mathbf{G}$ once and once only. In practice the difficult part of this process is to determine the small representations. Standard methods are available for doing this. It can be shown (see, for example, Lyubarskii[13]) that they are related very simply to the projective representations of the little cogroup, $\mathbf{G}^{\mathbf{k}}/\mathbf{T}$. It is the small representations which are tabulated in the literature so their derivation need not concern us; for the purpose of this article we may assume that they are known.

The above theory has been set purposely within the framework of the theory of induced representations so that we may go on to apply the theory of Sec. 2.

4. INNER KRONECKER PRODUCTS OF SPACE-GROUP REPRESENTATIONS

The problem that we now consider is that of the evaluation of the coefficients $C_{pq,r}^{\mathbf{k}\mathbf{m},\mathbf{h}}$ in the expansion

$$\Gamma_p^{\mathbf{k}} \otimes \Gamma_q^{\mathbf{m}} = \sum_{\mathbf{h},r} C_{pq,r}^{\mathbf{k}\mathbf{m},\mathbf{h}} \Gamma_r^{\mathbf{h}}. \tag{4.1}$$

In Eq. (4.1) the sum on the right-hand side over $\mathbf{h}$ is restricted, by virtue of the scheme outlined in Sec. 3, to vectors within the basic zone; $\mathbf{k}$ and $\mathbf{m}$ also belong to the basic zone. From Theorem (2.1) $C_{pq,r}^{\mathbf{k}\mathbf{m},\mathbf{h}}$ is the frequency of $A(\mathbf{G})$ in $\Gamma_p^{\mathbf{k}} \otimes \Gamma_q^{\mathbf{m}} \otimes \Gamma_r^{\mathbf{h}*}$. Now $\Gamma_r^{\mathbf{h}} = D_r^{\mathbf{h}} \uparrow \mathbf{G}$ is induced from $\mathbf{G}^{\mathbf{h}}$ and, since $D_r^{\mathbf{h}*}$ is a representation of $\mathbf{G}^{-\mathbf{h}} = \mathbf{G}^{\mathbf{h}}$, it follows that $\Gamma_r^{\mathbf{h}*}$ is also induced from $\mathbf{G}^{\mathbf{h}}$ and, from Eq. (2.5), that $\Gamma_r^{\mathbf{h}*} = D_r^{\mathbf{h}*} \uparrow \mathbf{G}$. Also $\Gamma_p^{\mathbf{k}}$ is induced from $\mathbf{G}^{\mathbf{k}}$ and $\Gamma_q^{\mathbf{m}}$ from $\mathbf{G}^{\mathbf{m}}$. Therefore, in order to apply Theorem (2.8) we expand $\mathbf{G}$ into double cosets first with respect to $\mathbf{G}^{\mathbf{h}}$ and $\mathbf{G}^{\mathbf{k}}$:

$$\mathbf{G} = \sum_{\alpha} \mathbf{G}^{\mathbf{h}}\{R_\alpha \mid \mathbf{v}_\alpha\}\mathbf{G}^{\mathbf{k}}. \tag{4.2}$$

We call the set of distinct $\mathbf{k}$ vectors $\mathbf{k}_\alpha = R_\alpha\mathbf{k}$, as α ranges over its values in Eq. (4.2) the *costar* of $\mathbf{k}$ with respect to $\mathbf{G}^{\mathbf{h}}$. The group

$$G_\alpha^{\mathbf{k}} = \{R_\alpha \mid \mathbf{v}_\alpha\}\mathbf{G}^{\mathbf{k}}\{R_\alpha \mid \mathbf{v}_\alpha\}^{-1}$$

is the group of $\mathbf{k}_\alpha$. Notice that, within the framework of this definition, the star of $\mathbf{k}$ is the costar of $\mathbf{k}$ with respect to $\mathbf{T}$. In general, however, a costar of $\mathbf{k}$ will be a subset of the elements in the star of $\mathbf{k}$. Following the procedure of Sec. 2, we then form the group $\mathbf{G}^{\mathbf{h}} \wedge \mathbf{G}_\alpha^{\mathbf{k}} = \mathbf{L}_\alpha$ and make the decomposition

$$\mathbf{G} = \sum_{\beta} \mathbf{L}_\alpha\{R_\beta \mid \mathbf{v}_\beta\}\mathbf{G}^{\mathbf{m}}. \tag{4.3}$$

In Eq. (4.3) the β indices are in correspondence with the elements of the costar of $\mathbf{m}$ with respect to $\mathbf{L}_\alpha$. As before we write $\mathbf{N}_{\alpha\beta} = \mathbf{G}_\alpha^{\mathbf{k}} \wedge \mathbf{G}_\beta^{\mathbf{m}} \wedge \mathbf{G}^{\mathbf{h}}$. In keeping with Eq. (2.12), we write $D_{\alpha p}^{\mathbf{k}}$ for the small representation of $G_\alpha^{\mathbf{k}}$ for which, if $\{R \mid \mathbf{v}\} \in \mathbf{G}^{\mathbf{k}}$,

$$D_{\alpha p}^{\mathbf{k}}(\{R_\alpha \mid \mathbf{v}_\alpha\}\{R \mid \mathbf{v}\}\{R_\alpha \mid \mathbf{v}_\alpha\}^{-1}) = D_p^{\mathbf{k}}(\{R \mid \mathbf{v}\}). \tag{4.4}$$

The representation $D_{\beta q}^{\mathbf{m}}$ is similarly defined.

[13] G. Ya. Lyubarskii, *The Application of Group Theory in Physics* (Pergamon Press, Inc., New York, 1960), Sec. 30.

Application of the result at the end of Sec. 2 leads now to the expression

$$C^{\mathbf{k}\mathbf{m},\mathbf{h}}_{pq,r} = \sum_{\alpha}\sum_{\beta} f[A(\mathbf{N}_{\alpha\beta})]/(D^{\mathbf{k}}_{\alpha p} \otimes D^{\mathbf{m}}_{\beta q} \otimes D^{\mathbf{h}}_{r}{}^{*}) \downarrow \mathbf{N}_{\alpha\beta}. \quad (4.5)$$

Consider now any particular term from the double sum in Eq. (4.5). Its value is

$$(1/|\mathbf{N}_{\alpha\beta}|) \sum_{\{R|\mathbf{v}\}\in\mathbf{N}_{\alpha\beta}} \chi^{\mathbf{k}}_{\alpha p}(\{R \mid \mathbf{v}\}) \times \chi^{\mathbf{m}}_{\beta q}(\{R \mid \mathbf{v}\})\chi^{\mathbf{h}}_{r}{}^{*}(\{R \mid \mathbf{v}\})$$

in which the sum is over all $\{R \mid \mathbf{v}\} \in \mathbf{N}_{\alpha\beta}$ and $\chi^{\mathbf{k}}_{\alpha p}$ is the character of $D^{\mathbf{k}}_{\alpha p}$, etc.

Performing first a sum over translations we see that this sum vanishes unless

$$\mathbf{k}_{\alpha} + \mathbf{m}_{\beta} \sim \mathbf{h}. \quad (4.6)$$

On the other hand, if Eq. (4.6) holds then the sum falls into $|\mathbf{T}|$ equal portions and, denoting the coset representatives of $\mathbf{T}$ in $\mathbf{N}_{\alpha\beta}$ by $\{S \mid \mathbf{w}\}$ and on using Eq. (4.4), we find its value to be

$$(|\mathbf{T}|/|\mathbf{N}_{\alpha\beta}|) \sum_{\{S|\mathbf{w}\}} \chi^{\mathbf{k}}_{p}(\{R_{\alpha} \mid \mathbf{v}_{\alpha}\}^{-1}\{S \mid \mathbf{w}\}\{R_{\alpha} \mid \mathbf{v}_{\alpha}\}) \times \chi^{\mathbf{m}}_{q}(\{R_{\beta} \mid \mathbf{v}_{\beta}\}^{-1}\{S \mid \mathbf{w}\}\{R_{\beta} \mid \mathbf{v}_{\beta}\})\chi^{\mathbf{h}}_{r}{}^{*}(\{S \mid \mathbf{w}\}). \quad (4.7)$$

In order to evaluate $C^{\mathbf{k}\mathbf{m},\mathbf{h}}_{pq,r}$ *we must therefore determine the costar of* $\mathbf{k}$ *with respect to* $\mathbf{G}^{\mathbf{h}}$ *and the costar of* $\mathbf{m}$ *with respect to* $\mathbf{G}^{\mathbf{h}} \wedge \mathbf{G}^{\mathbf{k}}_{\alpha}$, *and find all triples of vectors, one vector from each costar, which together with* $\mathbf{h}$ *satisfy Eq.* (4.6). Since $\mathbf{k}$, $\mathbf{m}$, and $\mathbf{h}$ must belong to the basic zone this is very easy to do pictorially.

Having found all such triples of $\mathbf{k}$ vectors, we must evaluate for each of them a sum of the form (4.7). The grand total of such sums is the Clebsch–Gordan coefficient $C^{\mathbf{k}\mathbf{m},\mathbf{h}}_{pq,r}$.

We can see now how this fits into the scheme of Elliott and Loudon, and subsequent authors of subgroup techniques. First, if one merely requires the selection rule over the whole group and, by considering all pairs of vectors one from the star of $\mathbf{k}$ and one from the star of $\mathbf{m}$ which have a sum equivalent to $\mathbf{h}$ and evaluate, for each triple of vectors so formed, a sum of the form (4.7), then certainly one cannot avoid finding a nonzero contribution if one exists (thereby answering the question whether or not the transition is forbidden). On the other hand, if one considers only one such triple of vectors, one gets the selection rule for that triple only and not necessarily the selection rule over the whole group. Secondly, if one requires the precise value of the Clebsch–Gordan coefficient, it is no value either running through all triples of vectors from the various stars or restricting one's consideration to a single trio. Put in this light it becomes abundantly clear that one needs a rule for determining which triples of vectors have to be considered and this rule is the one italicized in this section and given in terms of the costars of $\mathbf{k}$ and $\mathbf{m}$.

5. EXAMPLE

To illustrate the results of Sec. 4, we give here an example using the space group $P23$. This example is chosen to produce cases in which more than one term survives in the sum (4.5). $P23$ is based on the simple cubic lattice, and, to prevent us having to introduce details of the lattice and the Brillouin zone and definitions of the operators, we use for them exactly the notation of Altmann and Cracknell.[14] The point group of $P23$ is the tetrahedral group which contains the 12 elements E, C_{2m}, $C_{3j}^{\pm}(m = x, y, z; j = 1, 2, 3, 4)$. We consider only two points in the basic zone: $\Gamma = (0, 0, 0)$ and $M = (1/2, 1/2, 0)$. Coordinates are given here in units of reciprocal lattice vectors in the x, y, and z directions. We define the points $M^{+} = C_{31}^{+}M = (0, 1/2, 1/2)$ and $M^{-} = C_{31}^{-}M = (1/2, 0, 1/2)$. The star of Γ is just the point Γ by itself, and the star of M consists of the three points M, M^{+}, and M^{-}.

In Table I we list the characters of the small representations of $\mathbf{G}^{\Gamma}$ and $\mathbf{G}^{M}$. The group of Γ is the whole space group, but the group of M contains

TABLE I. *The small representations of* Γ *and* M. *Note.* The symmetry operations should be identified from Fig. 4 of Altmann and Cracknell. The suffices m, j take the following values with reference to this figure: $m = x, y,$ and z; $j = 1, 2, 3,$ and 4.

Γ	E	$3C_{2m}$	$4C_{3j}^{+}$	$4C_{3j}^{-}$
A	1	1	1	1
^{1}E	1	1	ω^{a}	ω
^{2}E	1	1	ω	ω^{a}
T	3	−1	0	0
M	E	C_{2x}	C_{2y}	C_{2z}
A_1	1	1	1	1
B_1	1	−1	−1	1
B_2	1	−1	1	−1
B_3	1	1	−1	−1

[a] $\omega = \exp(2\pi i/3)$. All translations are represented by the identity in Γ. But in M, $\{E \mid \mathbf{t}_1\}$ and $\{E \mid \mathbf{t}_2\}$ are represented by −1 and $\{E \mid \mathbf{t}_3\}$ by 1.

[14] S. L. Altmann and A. P. Cracknell, Rev. Mod. Phys. **37**, 19 (1965).

multiples of the translations with just four rotation operators: E, C_{2x}, C_{2y}, and C_{2z}.

For our example we consider the inner Kronecker product of pairs of representations of M. We write MA_1 for $A_1 \uparrow \mathbf{G}$, etc. In the notation of Sec. 4 $\mathbf{k} = M = (1/2, 1/2, 0)$ and $\mathbf{m} = M = (1/2, 1/2, 0)$. It is soom verified that, with this choice of $\mathbf{k}$ and $\mathbf{m}$, the only possible values of $\mathbf{h}$ that can appear in the right-hand side of Eq. (4.1) are $\mathbf{h} = \Gamma = (0, 0, 0)$ and $\mathbf{h} = M = (1/2, 1/2, 0)$.

Take first the case $\mathbf{h} = \Gamma$. From Eq. (4.2) the costar of $\mathbf{k} = M$ with respect to $\mathbf{G}^\Gamma$ is just the point $M = (1/2, 1/2, 0)$, and the one value of α is in correspondence with the identity $\{E \mid 0\}$. The group intersection $\mathbf{L}_\alpha = \mathbf{G}^\Gamma \wedge \mathbf{G}^M = \mathbf{G}^M$. From Eq. (4.3) the costar of $\mathbf{m} = M$ with respect to $\mathbf{G}^M$ consists of M, M^+, and M^- and the values of β are in correspondence with $\{E \mid 0\}$, $\{C_{31}^+ \mid 0\}$, and $\{C_{31}^- \mid 0\}$. Since $M + M \sim \Gamma$, $M + M^+ \sim M^-$, $M + M^- \sim M^+$, the only pair of α, β values that are compatible with Eq. (4.6) are when we choose $\mathbf{k}_\alpha = M$ and $\mathbf{m}_\beta = M$. The triple intersection group $\mathbf{N}_{\alpha\beta} = \mathbf{G}^M$.

Take now the case $\mathbf{h} = M$. The costar of $\mathbf{k} = M$ with respect to $\mathbf{G}^M$ consists of M, M^+, and M^- and the corresponding α are $\{E \mid 0\}$, $\{C_{31}^+ \mid 0\}$, and $\{C_{31}^- \mid 0\}$. Since $\mathbf{G}^M$ is invariant under these operations, the groups of M, M^+, and M^- coincide so that in each case $\mathbf{L}_\alpha = \mathbf{G}^M$. Again the costar of $\mathbf{m} = M$ with respect to $\mathbf{L}_\alpha$ consists (for each α) of M, M^+, and M^-, and the values of β are in correspondence with $\{E \mid 0\}$, $\{C_{31}^+ \mid 0\}$, and $\{C_{31}^- \mid 0\}$. Of the nine possible pairs of α, β values we now have two pairs which are compatible with Eq. (4.6): when $\mathbf{k}_\alpha = M^+$ and $\mathbf{m}_\beta = M^-$, and when $\mathbf{k}_\alpha = M^-$ and $\mathbf{m}_\beta = M^+$. In both cases $\mathbf{N}_{\alpha\beta} = \mathbf{G}^M$.

Consider now in detail the product $MB_2 \otimes MB_3$. There are three cases to consider:

(i) $\mathbf{k}_\alpha = M$, $\mathbf{m}_\beta = M$, $\mathbf{h} = \Gamma$.

From Table I the sum (4.7) becomes

$$\tfrac{1}{4}[\chi_r^\Gamma(E) - \chi_r^\Gamma(C_{2x}) - \chi_r^\Gamma(C_{2y}) + \chi_r^\Gamma(C_{2z})],$$

which is zero when $r = A$, ^1E or ^2E and unity when $r = T$.

(ii) $\mathbf{k}_\alpha = M^+$, $\mathbf{m}_\beta = M^-$, $\mathbf{h} = M$.

Now

$$C_{31}^- C_{2z} C_{31}^+ = C_{2z}, \quad C_{31}^- C_{2y} C_{31}^+ = C_{2x}, \quad C_{31}^- C_{2z} C_{31}^+ = C_{2y},$$

and, on using these relations and Table I, the sum (4.7) is now

TABLE II. The inner Kronecker products of representations belonging to M for $P23$.

$$\begin{aligned}
A_1 \otimes A_1 &= A + {}^1E + {}^2E + 2A_1 \\
A_1 \otimes B_1 &= T + B_2 + B_3 \\
A_1 \otimes B_2 &= T + B_3 + B_1 \\
A_1 \otimes B_3 &= T + B_1 + B_2 \\
B_1 \otimes B_1 &= A + {}^1E + {}^2E + 2B_1 \\
B_1 \otimes B_2 &= T + B_3 + A_1 \\
B_1 \otimes B_3 &= T + B_2 + A_1 \\
B_2 \otimes B_2 &= A + {}^1E + {}^2E + 2B_2 \\
B_2 \otimes B_3 &= T + A_1 + B_1 \\
B_3 \otimes B_3 &= A + {}^1E + {}^2E + 2B_3
\end{aligned}$$

Notes. (i) Each line of the table is an abbreviated form of an equation like (5.1). Since no confusion can arise, the prefix Γ or M has constantly been omitted.

(ii) To obtain a Clebsch–Gordan coefficient $C_{pq,r}^{\mathbf{k}\,\mathbf{m},\,\mathbf{h}}$, pickup the number multiplying the symbol corresponding to r on the right-hand side of the equation whose left-hand side contains the symbols p and q. For example $C_{A_1A_1,A_1}^{MM,M} = 2$.

$$\tfrac{1}{4}[\chi_r^M(E) + \chi_r^M(C_{2x}) + \chi_r^M(C_{2y}) + \chi_r^M(C_{2z})],$$

which is zero when $r = B_1$, B_2, or B_3 and unity when $r = A_1$.

(iii) $\mathbf{k}_\alpha = M^-$, $\mathbf{m}_\beta = M^+$, $\mathbf{h} = M$.

This time the sum (4.7) is

$$\tfrac{1}{4}[\chi_r^M(E) - \chi_r^M(C_{2x}) - \chi_r^M(C_{2y}) + \chi_r^M(C_{2z})],$$

which is zero when $r = A_1$, B_2, and B_3, and unity when $r = B_1$.

Collecting these results together, the equation corresponding to (4.1) is

$$MB_2 \otimes MB_3 = \Gamma T + MA_1 + MB_1. \qquad (5.1)$$

As a check on Eq. (5.1), note that the dimension of each side is 9 (representations of M in $P23$ being of dimension 3).

In Table II we list all 10 such products for representations of M.

To conclude we make just two remarks. First, although we chose $\mathbf{k} = \mathbf{m}$ in the above example, this is neither necessary nor sufficient for more than one pair of α, β values to survive. Secondly, provided the sum (4.7) is evaluated as indicated in the example, it is not necessary to have at one's disposal a character table of $\mathbf{N}_{\alpha\beta}$.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Professor A. J. Coleman for introducing him to the theory of induced representations and for many of the ideas in Sec. 2, and also to Dr. S. L. Altmann for valuable discussions on all aspects of group theory.

Kinetic Equation for the Electron–Phonon Gas

WILLARD R. CHAPPELL
Smithsonian Astrophysical Observatory and Harvard University, Cambridge, Massachusetts
(Received 16 December 1965; final manuscript received 21 February 1966)

The random phase approximation equations of motion for the electron–phonon system are solved to obtain long-time expressions for the basic operators that appear in the calculation of correlation functions. These operator expressions are used to obtain a kinetic equation for the electrons and an expression for the phonon spectral function. The kinetic equation has the same form as the Balescu–Lenard equation. Expressions for the density autocorrelation function and the two-particle correlation function are also obtained. The latter is shown to have the same form as the test-particle result for a classical plasma.

INTRODUCTION

THE approach to equilibrium in the electron–phonon system has been studied by several authors. Pines and Schrieffer[1] used the Bohm–Pines collective-coordinate approach and first-order perturbation theory to obtain coupled kinetic equations for electrons, phonons, and plasmons. Wyld and Pines[2] have shown that the Balescu–Lenard quantum collision integral yields the coupled kinetic equations for the electrons and plasmons. Ron[3] used a hierarchy approach and the Bogoliubov (adiabatic) assumption to obtain coupled kinetic equations that included the Balescu–Lenard collision integral. Michel and Van Leeuwen[4] have used a Green's function approach to obtain kinetic equations for the electron–phonon system excluding the coulomb interaction.

In this paper we use an approach originally developed by Wyld and Fried for the electron gas.[5] This method consists of solving the random phase approximation (RPA) equations for the long-time behavior of certain operators. These long-time operators are then used to calculate asymptotic expressions for correlation functions. The resulting kinetic equation has the form of the Balescu–Lenard equation for the quantum electron gas. The wavenumber- and frequency-dependent dielectric function that appears in this equation is the dielectric function for the electron–phonon gas. The spectral function for the phonon Green's function is also calculated. When the explicit contributions of the optical and acoustical resonances are considered, the results of Pines and Schrieffer[1] are obtained. In equilibrium, the acoustic phonon distribution is given by the Bose–Einstein distribution with the renormalized phonon frequencies. An expression for the density autocorrelation function is derived that agrees with the result of Lee and Tzoar.[6] Finally, the two-particle correlation function is expressed in a form similar to the test-particle result of Rostoker.[7]

EQUATIONS OF MOTION

The electron–phonon Hamiltonian is assumed to be[8]

$$H = \sum_{\mathbf{p},s} E_{\mathbf{p}} C^{\dagger}_{\mathbf{p}s} C_{\mathbf{p}s} + \tfrac{1}{2} \sum_{\mathbf{k}} (p^{\dagger}_{\mathbf{k}} p_{\mathbf{k}} + \Omega^2_{\mathbf{k}} q^{\dagger}_{\mathbf{k}} q_{\mathbf{k}}) + \sum_{\mathbf{k}} v_{\mathbf{k}} \rho_{-\mathbf{k}} q_{\mathbf{k}} + \tfrac{1}{2} \sum_{\mathbf{k}} \phi(\mathbf{k}) \rho_{\mathbf{k}} \rho_{-\mathbf{k}}, \quad (1)$$

where $C_{\mathbf{p}s}$ and $C^{\dagger}_{\mathbf{p}s}$ are the annihilation and creation operators, respectively, for Bloch electrons with momentum $\mathbf{p}$ and spin s, and obey the usual anticommutation relations

$$\begin{aligned} &[C_{\mathbf{p}s}, C_{\mathbf{p}'s'}]_+ = [C^{\dagger}_{\mathbf{p}s}, C^{\dagger}_{\mathbf{p}'s'}]_+ = 0, \\ \text{and} \quad &[C_{\mathbf{p}s}, C^{\dagger}_{\mathbf{p}'s'}]_+ = \delta_{\mathbf{p},\mathbf{p}'} \delta_{s,s'}. \end{aligned} \quad (2)$$

The quantity $\rho_{\mathbf{k}}$ is given by

$$\rho_{\mathbf{k}} = \sum_{\mathbf{p},s} C^{\dagger}_{\mathbf{p}-\frac{1}{2}\mathbf{k},s} C_{\mathbf{p}+\frac{1}{2}\mathbf{k},s}. \quad (3)$$

The operators $q_{\mathbf{k}}$ and $p_{\mathbf{k}}$ are the phonon canonical operators and obey the relations

$$\begin{aligned} &q^{\dagger}_{\mathbf{k}} = q_{-\mathbf{k}}, \qquad p^{\dagger}_{\mathbf{k}} = p_{-\mathbf{k}}, \\ &[q_{\mathbf{k}}, q_{\mathbf{k}'}]_- = [p_{\mathbf{k}}, p_{\mathbf{k}'}]_- = 0, \\ \text{and} \quad &[q_{\mathbf{k}}, p_{\mathbf{k}'}]_- = i\, \delta_{\mathbf{k},\mathbf{k}'}. \end{aligned} \quad (4)$$

The quantity $\phi(k) = 4\pi e^2/k^2$ is the coulomb interaction; Ω_k, the bare phonon frequency; and v_k, the

[1] D. Pines and J. R. Schrieffer, Phys. Rev. 125, 804 (1962).
[2] H. W. Wyld, Jr. and D. Pines, Phys. Rev. 127, 1851 (1962).
[3] A. Ron, J. Math. Phys. 4, 1182 (1963).
[4] K. H. Michel and J. M. J. Van Leeuwen, Physica 30, 410 (1964), and K. H. Michel, *ibid.* 30, 2194 (1964).
[5] H. W. Wyld, Jr. and B. D. Fried, Ann. Phys. (N. Y.) 23, 374 (1963).
[6] Y. C. Lee and N. Tzoar, Phys. Rev. 140, A396 (1965).
[7] N. Rostoker, Phys. Fluids 7, 491 (1964).
[8] J. Bardeen and D. Pines, Phys. Rev. 99, 1140 (1955).

coupling of bare phonons to the electrons. These latter quantities are given by[8]

$$\Omega_k^2 = 4\pi Z^2 e^2 N/M$$

and

$$(v_{-k})^* = v_k = -ik(N/M)^{\frac{1}{2}}(4\pi Ze^2/k^2), \quad (5)$$

where Ze, M, and N are the charge, mass, and the total numbers of ionic cores. We have taken $\hbar = 1$ and chosen for convenience to work in a unit volume.

To obtain a kinetic equation for the electrons, we must find an asymptotic equation of motion for the electron distribution function $F_s(\mathbf{p}) = \langle C^\dagger_{\mathbf{p}s} C_{\mathbf{p}s}\rangle$, where $\langle A\rangle = T_r \rho A$ and ρ is the density matrix. The exact equation of motion for this quantity is given by

$$\partial F_s(\mathbf{p})/\partial t = [\partial F_s(\mathbf{p})/\partial t]_{\rm ph} + [\partial F_s(\mathbf{p})/\partial t]_{\rm el}, \quad (6)$$

where

$$[\partial F_s(\mathbf{p})/\partial t]_{\rm ph} = -2 \sum_{\mathbf{k}} \mathrm{Im}\,[v_k \langle b_s(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k}) q_{\mathbf{k}}\rangle] \quad (7)$$

and

$$[\partial F_s(\mathbf{p})/\partial t]_{\rm el} = 2 \sum_{\mathbf{k}} \phi(k)\ \mathrm{Im}\,\langle b_s(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k}) \rho_{-\mathbf{k}}\rangle. \quad (8)$$

The operator $b_s(\mathbf{k}, \mathbf{p})$ is defined by

$$b_s(\mathbf{k}, \mathbf{p}) = C^\dagger_{\mathbf{p}-\frac{1}{2}\mathbf{k},s} C_{\mathbf{p}+\frac{1}{2}\mathbf{k},s}. \quad (10)$$

Thus, we need to obtain asymptotic expressions for the correlation functions $\langle bs(\mathbf{k}, \mathbf{p})\rho_{-k}\rangle$ and $\langle b_s(\mathbf{k}, \mathbf{p}) q_{\mathbf{k}}\rangle$.

Wyld and Fried used the RPA equation of motion for $b_s(\mathbf{k}, \mathbf{p})$ and the Bogoliubov (adiabatic) assumption to obtain asymptotic expressions for the operators $b_s(\mathbf{k}, \mathbf{p})$ and $\rho_{\mathbf{k}}$ for the electron gas.[5] They then used these expressions to obtain an asymptotic expression for the correlation function. We use the same method to obtain asymptotic expressions for the correlation functions that enter this problem.

The RPA or linearized equations of motion are given by[9,10]

$$\partial q_{\mathbf{k}}/\partial t = p_{-\mathbf{k}}, \quad (11)$$

$$\partial p_{\mathbf{k}}/\partial t = -\Omega_k^2 q_{-\mathbf{k}} - v_k \rho_{-\mathbf{k}}, \quad (12)$$

and

$$\partial b_s(\mathbf{k}, \mathbf{p})/\partial t = -i\Omega(\mathbf{k}, \mathbf{p}) b_s(\mathbf{k}, \mathbf{p}) + i v_k \Delta_s(\mathbf{k}, \mathbf{p}) q_{\mathbf{k}} + i\phi(k)\Delta_s(\mathbf{k}, \mathbf{p}) \sum_{\mathbf{p}'s'} b_{s'}(\mathbf{k}, \mathbf{p}'), \quad (13)$$

where

$$\Omega(\mathbf{k}, \mathbf{p}) = E_{\mathbf{p}+\frac{1}{2}\mathbf{k}} - E_{\mathbf{p}-\frac{1}{2}\mathbf{k}}$$

and

$$\Delta_s(\mathbf{k}, \mathbf{p}) = F_s(\mathbf{p} + \tfrac{1}{2}\mathbf{k}) - F_s(\mathbf{p} - \tfrac{1}{2}\mathbf{k}). \quad (14)$$

We will use the Bogoliubov assumption that the quantity $F_s(\mathbf{p})$ can be considered independent of time in the calculation of the correlation functions. We then introduce the one-sided Fourier transformation:

$$\tilde{O}(\omega) = \int_0^\infty e^{i\omega t} O(t)\, dt,$$
$$O(t) = \frac{1}{2\pi} \int e^{-i\omega t} \tilde{O}(\omega)\, d\omega, \quad (15)$$

where $\mathrm{Im}\ \omega > 0$. With the use of the above transformation Eqs. (11)–(13) become

$$-i\omega \tilde{q}_{\mathbf{k}}(\omega) = q_{\mathbf{k}}(0) + \tilde{p}_{-\mathbf{k}}(\omega), \quad (16)$$

$$-i\omega \tilde{p}_{\mathbf{k}}(\omega) = p_{\mathbf{k}}(0) - \Omega_k^2 \tilde{q}_{-\mathbf{k}}(\omega) - v_k \tilde{\rho}_{-\mathbf{k}}(\omega), \quad (17)$$

and

$$-i[\omega - \Omega(\mathbf{k}, \mathbf{p})]\tilde{b}_s(\mathbf{k}, \mathbf{p}, \omega) = b_s(\mathbf{k}, \mathbf{p}, 0) + i v_k \Delta_s(\mathbf{k}, \mathbf{p}) \tilde{q}_{\mathbf{k}}(\omega) + i\phi(k)\Delta_s(\mathbf{k}, \mathbf{p}) \sum_{\mathbf{p}'s'} \tilde{b}_{s'}(\mathbf{k}, \mathbf{p}', \omega). \quad (18)$$

These equations are easily solved for the operators $\tilde{q}_{\mathbf{k}}(\omega)$, $\tilde{p}_{\mathbf{k}}(\omega)$, and $\tilde{\rho}_{\mathbf{k}}(\omega)$. The results are given by

$$\tilde{q}_{\mathbf{k}}(\omega) = i[\omega^2 \epsilon(\mathbf{k}, \omega)]^{-1} \times \{K(\mathbf{k}, \omega)[\omega q_{\mathbf{k}}(0) + i p_{-\mathbf{k}}(0)] + v_{-k} \hat{F}(\mathbf{k}, \omega)\}, \quad (19)$$

$$\tilde{p}_{\mathbf{k}}(\omega) = [\omega^2 \epsilon(-\mathbf{k}, \omega)]^{-1} \times \{K(-\mathbf{k}, \omega)[i\omega p_{\mathbf{k}}(0) + \Omega_k^2 q_{-\mathbf{k}}(0)] + v_k \omega \hat{F}(-\mathbf{k}, \omega)\}, \quad (20)$$

and

$$\tilde{\rho}_{\mathbf{k}}(\omega) = i[\omega^2 \epsilon(\mathbf{k}, \omega)]^{-1} \times \{(\omega^2 - \Omega_k^2)\hat{F}(\mathbf{k}, \omega) - v_k Q(\mathbf{k}, \omega)[\omega q_{\mathbf{k}}(0) + i p_{-\mathbf{k}}(0)]\}, \quad (21)$$

where

$$\hat{F}(\mathbf{k}, \omega) = \sum_{\mathbf{p},s} \frac{b_s(\mathbf{k}, \mathbf{p}, 0)}{\omega - \Omega(\mathbf{k}, \mathbf{p})}, \quad (22)$$

$$Q(\mathbf{k}, \omega) = \sum_{\mathbf{p},s} \frac{\Delta_s(\mathbf{k}, \mathbf{p})}{\omega - \Omega(\mathbf{k}, \mathbf{p})}, \quad (23)$$

$$K(\mathbf{k}, \omega) = 1 + \phi(k) Q(\mathbf{k}, \omega), \quad (24)$$

and

$$\epsilon(\mathbf{k}, \omega) = K(\mathbf{k}, \omega) - \Omega_k^2/\omega^2. \quad (25)$$

[9] D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin Company, Inc., New York, 1963), p. 246.
[10] R. K. Nesbet, J. Math. Phys. 6, 621 (1965).

The quantities $K(\mathbf{k}, \omega)$ and $\epsilon(\mathbf{k}, \omega)$ are the RPA wavenumber- and frequency-dependent dielectric constants for the electron gas and electron–phonon gas (neglecting periodic effects), respectively.

ASYMPTOTIC BEHAVIOR

To obtain the asymptotic expressions as $t \to \infty$ for the operators $q_{\mathbf{k}}(t)$, $p_{\mathbf{k}}(t)$, $\rho_{\mathbf{k}}(t)$, and $b_s(k, p, t)$, we must obtain the inverse transformation. For a stable system, the only poles arise from the quantities $[\omega^2\epsilon(\mathbf{k}, \omega)]^{-1}$ and $\hat{F}(\mathbf{k}, \omega)$. The zeros of $\epsilon(\mathbf{k}, \omega)$ correspond to the acoustic and optical phonon excitations, and the location of these poles is given by the solution of the dispersion relation

$$\omega^2 = \Omega_k^2/K(\mathbf{k}, \omega). \tag{26}$$

Since these poles are in the lower half of the ω-plane, we assume that the corresponding contributions represent transients that can be ignored. Thus the only pole we need to consider is that arising from $\hat{F}(\mathbf{k}, \omega)$. We thus obtain

$$q_{\mathbf{k}}(t) \sim v_{-k} \sum_{p,s} \frac{b_s(\mathbf{k}, \mathbf{p}, 0)e^{-i\Omega(\mathbf{k},\mathbf{p})t}}{\Omega^2(\mathbf{k}, \mathbf{p})\epsilon_+(\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p}))}, \tag{27}$$

$$p_{\mathbf{k}}(t) \sim iv_k \sum_{p,s} \frac{\Omega(\mathbf{k}, \mathbf{p})b_s(-\mathbf{k}, \mathbf{p}, 0)e^{i\Omega(\mathbf{k},\mathbf{p})t}}{\Omega^2(\mathbf{k}, \mathbf{p})\epsilon_+[\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p})]^*}, \tag{28}$$

and

$$\rho_{\mathbf{k}}(t) \sim \sum_{p,s} \frac{[\Omega^2(\mathbf{k}, p) - \Omega_k^2]b_s(\mathbf{k}, \mathbf{p}, 0)e^{-i\Omega(\mathbf{k},\mathbf{p})t}}{\Omega^2(\mathbf{k}, \mathbf{p})\epsilon_+[\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p})]}, \tag{29}$$

where

$$\epsilon_+(\mathbf{k}, \omega) = \epsilon(\mathbf{k}, \omega + i\eta). \tag{30}$$

The quantity η is the positive infinitesimal. It is present because of the analytic continuation of $\epsilon(\mathbf{k}, \omega)$, which is necessary for the integration. Henceforth, we drop the subscript $+$.

We obtain the asymptotic formula for $b_s(\mathbf{k}, \mathbf{p}, t)$ by substituting the expressions given by Eqs. (27) and (29) into (13) and by solving the resulting equation subject to the condition

$$\rho_{\mathbf{k}}(t) = \sum_{p,s} b_s(\mathbf{k}, \mathbf{p}, t). \tag{31}$$

The result is

$$b_s(\mathbf{k}, \mathbf{p}, t) = b_s(\mathbf{k}, \mathbf{p}, 0)e^{-i\Omega(\mathbf{k},\mathbf{p})t} - \phi(k)\Delta_s(\mathbf{k}, \mathbf{p}) \times \sum_{p',s'} \frac{b_{s'}(\mathbf{k}, \mathbf{p}', 0)e^{-i\Omega(\mathbf{k},\mathbf{p}')t}}{\epsilon[\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p}')][\Omega(\mathbf{k}, \mathbf{p}') - \Omega(\mathbf{k}, \mathbf{p}) + i\eta]}. \tag{32}$$

We now have what constitutes a full description of the asymptotic behavior of the system (within the RPA). In the process of calculating correlation functions, we will encounter the quantity $\langle b_s(\mathbf{k}, \mathbf{p}, 0)b_{s'}(-\mathbf{k}, \mathbf{p}', 0)\rangle$. This quantity can be written as

$$\delta_{s,s'}\delta_{p,p'}F_s(\mathbf{p} - \tfrac{1}{2}\mathbf{k})[1 - F_s(\mathbf{p} + \mathbf{k}/2)] + g_{ss'}(\mathbf{k}, \mathbf{p}, \mathbf{p}', 0),$$

where $g_{ss'}(\mathbf{k}, \mathbf{p}, \mathbf{p}', 0)$ is the usual two-particle correlation function at $t = 0$. We have used the Bogoliubov assumption to replace $F_s(\mathbf{p}, 0)$ by $F_s(\mathbf{p}, t)$. The initial two-particle correlation function will always be multiplied by a factor of the form $\exp[if(\mathbf{k}, \mathbf{p})t]$. The Riemann–Lebesgue theorem can then be used to show that such contributions will vanish as $t \to \infty$ (phase mixing).

Using Eqs. (7), (8), (27), (29), and (32) we obtain

$$\frac{\partial F_s(\mathbf{p})}{\partial t} = 2\pi \sum_{\mathbf{k},\mathbf{p}',s'} \frac{|\phi(k)|^2}{|\epsilon[\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p} + \frac{1}{2}\mathbf{k})]|^2} \times [G_s^+(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k})G_{s'}^-(\mathbf{k}, \mathbf{p}' - \mathbf{k}/2) - G_s^-(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k})G_{s'}^+(\mathbf{k}, \mathbf{p}' - \tfrac{1}{2}\mathbf{k})] \times \delta[\Omega(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k}) - \Omega(\mathbf{k}, \mathbf{p}' - \tfrac{1}{2}\mathbf{k})], \tag{33}$$

where

$$G_s^{\pm}(\mathbf{k}, \mathbf{p}) = F_s(\mathbf{p} \pm \tfrac{1}{2}\mathbf{k})[1 - F_s(\mathbf{p} \mp \tfrac{1}{2}\mathbf{k})]. \tag{34}$$

The above kinetic equation has the form of the Balescu–Lenard equation. The ordinary Balescu–Lenard equation is obtained if we let the ion mass become infinite. It is readily demonstrated that Eq. (33) satisfies the usual conservation laws (total momentum and spin) and that there is an H theorem. The general time-independent solution is the Fermi–Dirac distribution.

The form of Eq. (33) is somewhat surprising at first glance. We would expect to obtain terms that correspond to the emission and absorption of acoustical phonons by the electrons, as well as terms that describe the coulomb interactions between the electrons.[1] In the next section we show that such emission-absorption terms are obtained from Eq. (33) when appropriate approximations are made.

GREEN'S FUNCTION AND ACOUSTIC PHONONS

We can also obtain asymptotic expressions for quantities such as

$$D^>(\mathbf{k}, t - t') = \langle q_{\mathbf{k}}(t)q_{-\mathbf{k}}(t')\rangle \tag{35}$$

and

$$D^<(\mathbf{k}, t - t') = \langle q_{-\mathbf{k}}(t')q_{\mathbf{k}}(t)\rangle. \tag{36}$$

These quantities are closely related to the phonon Green's function.[11] With the use of Eq. (27) $D^>$ and $D^<$ are found to be given by

$$D^>(\mathbf{k}, t-t') = |v_k|^2 \sum_{p,s} \int d\omega e^{-i\omega(t-t')}\, \delta[\omega - \Omega(\mathbf{k}, \mathbf{p})] \times \frac{F_s(\mathbf{p} - \tfrac{1}{2}\mathbf{k})[1 - F_s(\mathbf{p} + \tfrac{1}{2}\mathbf{k})]}{\omega^4\, |\epsilon(\mathbf{k}, \omega)|^2} \quad (37)$$

and

$$D^<(\mathbf{k}, t-t') = |v_k|^2 \sum_{p,s} \int d\omega e^{-i\omega(t-t')}\, \delta(\omega - \Omega(\mathbf{k}, \mathbf{p})) \times \frac{F_s(\mathbf{p} + \tfrac{1}{2}\mathbf{k})[1 - F_s(\mathbf{p} - \tfrac{1}{2}\mathbf{k})]}{\omega^4\, |\epsilon(\mathbf{k}, \omega)|^2}, \quad (38)$$

where we have neglected the initial two-particle correlations. The spectral function

$$B(\mathbf{k}, \omega) = D^>(\mathbf{k}, \omega) - D^<(\mathbf{k}, \omega), \quad (39)$$

where

$$D^>(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t} D^>(\mathbf{k}, t), \quad (40)$$

is given by

$$B(\mathbf{k}, \omega) = \frac{2\pi\, |v_k|^2}{\omega^4\, |\epsilon(\mathbf{k}, \omega)|^2} \sum_{p,s} \Delta_s(\mathbf{k}, p)\, \delta(\omega - \Omega(\mathbf{k}, \mathbf{p})). \quad (41)$$

The contributions of the acoustic phonons can be isolated if we take only the contributions from acoustic resonances in $1/|\epsilon(\mathbf{k}, \omega)|^2$. Near these resonances we can write

$$\epsilon(\mathbf{k}, \omega) = [2K(\mathbf{k}, 0)/\omega_k](\omega - \omega_k + i\gamma_k), \qquad \omega \approx \omega_k,$$
$$\epsilon(\mathbf{k}, \omega) = -[2K(\mathbf{k}, 0)/\omega_{-k}] \times (\omega + \omega_{-k} + i\gamma_{-k}), \qquad \omega \approx -\omega_{-k}, \quad (42)$$

where

$$\omega_k^2 = \Omega_k^2/K(\mathbf{k}, 0) \quad (43)$$

and

$$\gamma_k = -\frac{\pi\phi(k)\omega_k}{2K(\mathbf{k}, 0)} \sum_{p,s} \Delta_s(\mathbf{k}, \mathbf{p})\, \delta(\omega_k - \Omega(\mathbf{k}, \mathbf{p})). \quad (44)$$

If we use an analysis similar to that of Wyld and Pines,[2] the acoustic contribution to $1/|\epsilon(\mathbf{k}, \omega)|^2$ is found to be

$$\frac{1}{|\epsilon(\mathbf{k}, \omega)|^2} = \frac{\pi\omega_k^2}{4\gamma_k K(\mathbf{k}, 0)^2}\, \delta(\omega - \omega_k) + \frac{\pi\omega_{-k}^2}{4\gamma_{-k} K(\mathbf{k}, 0)^2}\, \delta(\omega + \omega_{-k}). \quad (45)$$

[11] G. Baym, Ann. Phys. (N. Y.) 14, 1 (1961).

We then find that

$$B(\mathbf{k}, \omega) = -\pi[\omega_k^{-1}\, \delta(\omega - \omega_k) - \omega_{-k}^{-1}\, \delta(\omega + \omega_{-k})], \quad (46)$$

which agrees with the results of Baym.[11]

Equation (45) can also be used to obtain the acoustic contribution to $D^<(\mathbf{k}, \omega)$. The result is

$$D^<(\mathbf{k}, \omega) = \frac{2\pi^2\phi(k)}{4\gamma_k K(\mathbf{k}, 0)}\, \delta(\omega - \omega_k) \times \sum_{p,s} G_s^+(\mathbf{k}, \mathbf{p})\, \delta(\omega_k - \Omega(\mathbf{k}, \mathbf{p})) + \frac{2\pi^2\phi(k)}{4\gamma_{-k} K(\mathbf{k}, 0)}\, \delta(\omega + \omega_{-k}) \times \sum_{p,s} G_s^+(\mathbf{k}, \mathbf{p})\, \delta(\omega_{-k} + \Omega(\mathbf{k}, \mathbf{p})). \quad (47)$$

For a system of noninteracting quasi-phonons with frequencies ω_k, the result is

$$D^<(\mathbf{k}, \omega) = 2\pi\left[\frac{N_k}{2\omega_k}\, \delta(\omega - \omega_k) + \frac{(1 + N_{-k})}{2\omega_{-k}}\, \delta(\omega + \omega_{-k})\right]. \quad (48)$$

In this approximation we can take the identification

$$N_k(t) = \frac{\pi\phi(k)\omega_k}{2\gamma_k K(\mathbf{k}, 0)} \sum_{p,s} G_s^+(\mathbf{k}, \mathbf{p})\, \delta(\omega_k - \Omega(\mathbf{k}, p)). \quad (49)$$

The above equation agrees with the results of Pines and Schrieffer.[1] if the Fermi distribution for $F_s(\mathbf{p})$ is used in the above equation, we obtain the Bose–Einstein distribution:

$$N_k = (e^{\beta\omega_k} - 1)^{-1}. \quad (50)$$

The acoustic contribution to the kinetic equation [Eq. (33)] is obtained by using the approximation for $1/\epsilon(\mathbf{k}, \omega)$ given by Eq. (45). If Eq. (49) and a similar equation for $1 + N_k$ are then used, we obtain

$$\left[\frac{\partial F_s(\mathbf{p})}{\partial t}\right]_{sc} = \sum_{\mathbf{k}} \frac{\pi\phi(k)\omega_k}{K(\mathbf{k}, 0)} \{\delta(\omega_k - \Omega(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k})) \times [G_s^+(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k})(1 + N_k) - G_s^-(\mathbf{k}, \mathbf{p} + \tfrac{1}{2}\mathbf{k})N_k] - \delta(\omega_k - \Omega(\mathbf{k}, \mathbf{p} - \tfrac{1}{2}\mathbf{k}))[G_s^+(\mathbf{k}, \mathbf{p} - \tfrac{1}{2}\mathbf{k}) \times (1 + N_k) - G_s^-(\mathbf{k}, \mathbf{p} - \tfrac{1}{2}\mathbf{k})N_k]\}, \quad (51)$$

which agrees with the results of Pine and Schrieffer for the contribution of the acoustic phonons to the kinetic equation.[1] The contribution of the optical phonons (plasmons) to Eq. (33) can be obtained by approximating $1/\epsilon(\mathbf{k}, \omega)$ by its behavior in the neighborhood of the plasmon poles.[2]

Thus, in the approximation where $1/\epsilon(\mathbf{k}, \omega)$ is approximated by its behavior near its poles (acoustic

and plasmon), and when $\gamma_k/\omega_k \ll 1$, the kinetic equation given by Eq. (33) can be written approximately as the sum of two parts. One part is given by Eq. (51) and gives the contribution of the acoustical phonons as a difference of gain and loss terms arising from the Cherenkov-like emission and absorption of acoustical phonons. The other term is a plasmon contribution having the same general form.

DENSITY AUTOCORRELATION FUNCTION AND TEST-PARTICLE RESULT

The expressions obtained for the operators $b_s(\mathbf{k}, \mathbf{p}, t)$ and $\rho(\mathbf{k}, t)$ can also be used to obtain the density autocorrelation function and the two-particle correlation function. The density autocorrelation function is given by

$$\langle \rho(-\mathbf{k}, t)\rho(\mathbf{k}, t')\rangle = \sum_{\mathbf{p},s} \frac{[\Omega^2(\mathbf{k}, \mathbf{p}) - \Omega_k^2]^2 G_s^+(\mathbf{k}, \mathbf{p}) e^{i\Omega(\mathbf{k},\mathbf{p})(t-t')}}{\Omega^4(\mathbf{k}, \mathbf{p})\,|\epsilon(\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p}))|^2}, \quad (52)$$

and the spectral function by

$$S(\mathbf{k}, \omega) = \frac{2\pi(\omega^2 - \Omega_k^2)^2}{\omega^4\,|\epsilon(\mathbf{k}, \omega)|^2} \sum_{\mathbf{p},s} G_s^+(\mathbf{k}, \mathbf{p})\ \delta(\omega - \Omega(\mathbf{k}, \mathbf{p})), \quad (53)$$

where

$$\langle \rho(\mathbf{x}, t)\rho(\mathbf{x}', t')\rangle = \sum_{\mathbf{k}} \int \frac{d\omega}{2\pi} S(\mathbf{k}, \omega) e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')} e^{i\omega(t-t')}. \quad (54)$$

The above results agree with those obtained by Lee and Tzoar[6] [with a suitable redefinition of $\epsilon(\mathbf{k}, \omega)$].

It is possible to write the result for the two-particle correlation function in a manner that resembles the test-particle result obtained by Rostoker for a classical plasma.[7] If we define the quantity

$$P_{s,s'}(\mathbf{k}, \mathbf{p} \mid \mathbf{p}') = -\frac{\phi(k)\Delta_{s,s'}(\mathbf{k}, \mathbf{p}')}{\epsilon[\mathbf{k}, \Omega(\mathbf{k}, \mathbf{p})]^*[\Omega(\mathbf{k}, \mathbf{p}) - \Omega(\mathbf{k}, \mathbf{p}') - i\eta]} \quad (55)$$

and assume that $g_{s,s'}(\mathbf{k}, \mathbf{p}, \mathbf{p}', 0) = 0$, we obtain

$$g_{s,s'}(\mathbf{k}, \mathbf{p}, \mathbf{p}', t) = P_s^*(\mathbf{k}, \mathbf{p}' \mid \mathbf{p}) G_{s'}^-(\mathbf{k}, \mathbf{p}') + P_{s,s'}(\mathbf{k}, \mathbf{p} \mid \mathbf{p}') G_s^-(\mathbf{k}, \mathbf{p}) + \sum_{\mathbf{p}'',s''} P_s^*(\mathbf{k}, \mathbf{p}'' \mid \mathbf{p}) P_{s,s'}(\mathbf{k}, \mathbf{p}'' \mid \mathbf{p}') G_{s'',s'}^-(\mathbf{k}, \mathbf{p}''), \quad (56)$$

where the identity[4]

$$\langle b_s(\mathbf{k}, \mathbf{p}, t) b_{s'}(-\mathbf{k}, \mathbf{p}', t)\rangle = \delta_{\mathbf{p},\mathbf{p}'}\,\delta_{s,s'} G_s^-(\mathbf{k}, \mathbf{p}) + \delta_{\mathbf{k},0} F_s(\mathbf{p}) F_{s'}(\mathbf{p}') + g_{s,s'}(\mathbf{k}, \mathbf{p}, \mathbf{p}', t) \quad (57)$$

has been used. Equation (56) closely resembles the test-particle result for the two-particle correlation function.[7]

SUMMARY

We have derived asymptotic expressions for certain electron and phonon operators by using a method developed by Wyld and Fried for the electron gas. We obtained these expressions by solving the set of RPA equations of motion as an initial value problem and then considering the poles in the complex ω plane. The use of these operator expressions and the Bogoliubov (adiabatic) assumption for the particles enabled us to obtain a kinetic equation and the expression for various correlation functions. The kinetic equation that was obtained has the form of the Balescu–Lenard equation for an electron gas. The difference is that the dielectric constant for the electron gas is replaced by the dielectric constant for the electron–phonon gas. Certain phonon–phonon correlation functions were also obtained, as well as the density autocorrelation function and the two-particle correlation function. The latter result was shown to have the appearance of the test-particle result of Rostoker for the classical electron gas.[7]

The general picture obtained here of the approach to equilibrium is that the phonons are dressed by the particles in such a manner that the initial bare phonons decay in a time given by the damping time for a quasiphonon. The equal-time phonon-correlation functions at long times become functions of time only through the functional dependence on the one-electron distribution function. Meanwhile, the contribution of the initial two-particle correlation function decays through phase mixing. A similar picture has been obtained by the test-particle method for a classical plasma interacting with an electromagnetic field.[12]

ACKNOWLEDGMENTS

The author acknowledges helpful discussions with Professor P. Argyres of Massachusetts Institute of Technology, Professor H. Ehrenreich of Harvard University, Professor G. Carmi of Yeshiva University, and V. O'Donnell of Harvard University.

[12] R. E. Aamodt, O. C. Eldridge, and N. Rostoker, Phys. Fluids 7, 1952 (1964).

Majorization of Feynman Graphs*

YUTZE CHOW
Department of Physics, University of Wisconsin at Milwaukee, Milwaukee, Wisconsin
(Received 28 September 1965)

The different methods of majorization of Feynman graphs are reviewed and their respective domains of application and relative merits are discussed. The initial steps of Wu's majorization method are discussed in a variational formalism. Effort is made to present the rather complex arguments involved in Nakanishi's "path method" in a systematic manner.

I. INTRODUCTION

IN order to investigate analytic properties of Feynman amplitudes it is convenient to study the so-called V-function[1] appearing in a parametrized Feynman amplitude. Since a general physical process involves usually an infinite set of Feynman graphs, it is therefore more interesting to know a lower bound of the domain in which *all* the graphs of the set are analytic. "Majorization" is just the criterion sorted out toward this end. To state exactly, *majorization* means that, for two Feynman graphs G and G' with the same set of external momenta $\{p_i\}$, but arbitrarily different otherwise, the graph G *majorizes* the graph G' if

$$\forall \{p_i\} : \text{Min } V_G \leq \text{Min } V_{G'}, \tag{1}$$

where V_G denotes the V-function of the graph G.

Before defining the V-function, we introduce some notations and definitions. Let the total numbers of internal lines (external lines and vertices) of, a Feynman graph G be denoted respectively by J, ${}^eJ$, and N. An *internal* line is usually denoted by a subscript i to the left of the symbol, like ${}_ij$, and an *external* line is denoted by a superscript e to the left of the symbol, like ${}^ej$. Similar convention is used for vertices; an *external vertex* is a vertex attached to some external lines besides internal lines. An *internal vertex* attaches only to internal lines. When we say "shrinking a line ${}_ij$" it means ${}_ij$ is first deleted from the given graph G, then the two terminal vertices of ${}_ij$ are identified. This, of course, yields a new graph called a "reduced graph of G". The operation "shrinking" is denoted by the symbol $\downarrow$. Thus $G\,(\downarrow_i i_i j)$ means to shrink internal lines ${}_ii$ and ${}_ij$ from the graph G. The operation "deleting" is denoted by the symbol $\uparrow$. The graphs resulted by deleting some of the internal lines of a graph G are called the subgraphs of G.

The following definitions are also useful here:

(a) The *star*, $\mathcal{S}_v$, of a vertex v, is defined as the set containing all internal and external lines attached to the vertex v, and the vertex itself. A star is called an *external star*, $\mathcal{S}^*$, if the vertex concerned is an external vertex. An *internal star*, $\mathcal{S}^0$, contains an internal vertex. The subset of the totality of *external* lines of an external star $\mathcal{S}^*$ is denoted ${}^e\mathcal{S}$. The subset of the totality of *internal* lines of an external star $\mathcal{S}^*$ is denoted by ${}_i\mathcal{S}^*$.

(b) A *path* $\mathcal{P}$ is a connected graph which contains only some vertices and internal lines such that each of the two terminating vertices attaches to one line only, while each of the rest of the vertices attaches to exactly two lines. $\mathcal{P}(v_iv_j)$ denotes a path with terminating vertices v_i and v_j.

(c) A graph G is *connected* if

$$\forall v_i, v_j \in G \Rightarrow \exists\, \mathcal{P}(v_iv_j) \subset G,$$

where $\subset$ means "is contained in". Also, we shall use the symbols $\cup$ and $\cap$ for unions and intersections.

(d) A *loop*, $\mathcal{C}$, is a connected graph such that each vertex attaches to exactly two internal lines. A direction is usually assigned to a loop.

(e) An *intermediate-state* set, $\mathcal{I}$, of a given Feynman graph G is a set of internal lines such that G is separated into two disjoint connected subgraphs if $\mathcal{I}$ is deleted from G, yet G will remain connected if only a proper subset of $\mathcal{I}$ is deleted.

The 4-momentum assigned to an internal line j will be referred to as an internal *momentum*, q_j. The internal momenta may be decomposed into two parts:

$$q_j = K_j + P_j \tag{2}$$

with

$$K_j \equiv \sum_{l=1}^{L} \varepsilon_{jl} k_l \tag{3}$$

and

* This research was supported in part by the Graduate School of University of Wisconsin at Milwaukee, Wisconsin.
[1] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) Suppl. 18, (1961).

$$P_i \equiv \sum_{i=1}^{{}^{\circ}N} \eta_{ii} p_i, \tag{4}$$

where $\{p_1, \cdots, p^{{}^{\circ}N}\}$ is the set of external momenta and ${}^{\circ}N$ the total number of external vertices in G. K_j and P_j will be referred to, respectively, as the *integration momentum* and *constant momentum* of the internal line j. $\{K_1, \cdots, K_L\}$ is the set of *basic momenta* each of which flows along a loop, and such that *any* internal momentum can be expressed as a linear combination of the basic momenta. $\{K_1, \cdots K_L\}$ is usually called a *base* and the loops associated with it are called *basic loops*, to be denoted by $\{\mathfrak{C}_1^0, \cdots, \mathfrak{C}_L^0\}$. The *base* is clearly not unique for a given G.

We also define here that a set of momenta $\{p_1, \cdots, p_N\}$ is *Euclidean* if, for any set of real coefficients $\{a_1, \cdots, a_N\}$, the following condition is satisfied:

$$\left(\sum_{i=1}^{N} a_i p_i\right)^2 \leq 0, \tag{5}$$

or symbolically,

$$p_i \in \mathbf{E}.$$

Next, the coefficients appearing in (3) are defined by:

$$\varepsilon_{ji} = 0, \quad \text{if } j \notin \mathfrak{C}_i^0,$$

$$\varepsilon_{ji} = \pm, \text{ if } j \in \mathfrak{C}_i^0 \text{ and } \pm j \,||\, \mathfrak{C}_i^0.$$

The coefficients η_{ji} in (4) may take any constant values provided conservation of internal momenta at each vertex is satisfied.

The V-function of a graph G is then defined as

$$V = \sum_{j=1}^{J} \alpha_j (m_j^2 + P_j^2) - \frac{1}{U} \sum_{l,l'=1}^{L} \Delta_{ll'} b_l b_{l'}, \tag{6}$$

where $\{\alpha_1, \cdots, \alpha_J\} \equiv$ the set of Feynman parameters;

$U = \det \ddot{a}$,

$\Delta_{ll'} = ll'$th cofactor of $\ddot{a}$,

$\ddot{a} =$ the $L \times L$ matrix of elements $a_{ll'}$,

$$a_{ll'} = \sum_{j=1}^{J} \alpha_j \varepsilon_{jl} \varepsilon_{jl'}. \tag{7}$$

Using the U and V functions, the parametrized Feynman amplitude of a graph G can be expressed in the form

$$F_G = (i\pi^2)^L (J - 2L - 1)! \int_0^1 \cdots \int_0^1 \prod_{j=1}^{J} d\alpha_j \times \frac{\delta\left(1 - \sum_{j=1}^{J} j\right)}{U^2 (V - i\epsilon)^{J-2L}} \tag{8}$$

for $J > 2L$, ϵ is an infintesimal, positive number. The U-function can never be negative; the zeros correspond to the ultraviolet divergence which is not of present concern since we are interested in analytic properties only. The V-function is the object of present investigation.

If a graph G majorizes another graph G', and if one can show that

$$\text{Min } V_{G'} > 0$$

is established for certain value (or certain region of values) of $\{p_i\}$, then $V_G > 0$ is true for the same value (or region of values) of $\{p_i\}$. In this way, we can get a lower bound of the domain of analyticity by means of the few Feynman graphs which majorize many (or infinitely many) other much more complicated Feynman graphs. The presently available methods of majorization are: Wu's majorization method derived from electric circuit analogy[2] and its extension by Boyling,[3] Nakanishi's majorization by the so-called "Path method"[1,4] and Symanzik[5] and Logunov *et al.* majorization by purely algebraic manipulation.[6]

II. WU'S METHOD OF MAJORIZATION

The essential advantage of Wu's majorization method lies in the fact that it imposes no requirement of the set of external momenta's being Euclidean. Originally, the derivation of Wu's majorization was based upon electric circuit analogy. We carry out here some initial steps of the formalism in variational relation without going through the development parallel to the details given in Wu's original paper.

We first introduce double subscripts for the external momenta, e.g., p_{ni} denotes the ith external momentum (labeled independently of the vertex upon which it is incident) incident upon the nth vertex, we have, therefore,

$$\forall {}^{\circ}v_n : p_n = \sum_{\forall {}^{\circ}i \in \mathcal{S}_n} p_{ni}, \tag{9}$$

where p_n is the effective external momentum at the vertex ${}^{\circ}v_n$. The conservation rules can be written as

$$\forall \mathcal{S}_n^* : \sum_{\forall {}^{\circ}i \in G} p_{ni} = 0 \tag{10}$$

[2] T. T. Wu, Phys. Rev. **123**, 678 (1961).
[3] J. B. Boyling, Cambridge preprint.
[4] N. Nakanishi, Institute for Advanced Study Preprints (1963).
[5] K. Symanzik, Progr. Theoret. Phys. (Kyoto) 20, 690 (1958).
[6] A. A. Logunov, I. T. Todorov, and N. A. Chernikov, Zh. Eksperim. i Teor. Fiz. **42**, 1285 (1962). [English transl.: Soviet Phys.—JETP 15 891 (1962)]. I. T. Todorov, Doctoral dissertation, Joint Inst. of Nuclear Research, Report P–1205, Dubna (1963), and literature cited there.

and

$$\forall \mathcal{S}_n : \sum_{\forall {}_1 j \in \mathcal{S}_n} q_j = p_n \tag{11}$$

with the understanding that $p_n = 0$ if $\mathcal{S}_n$ is an *internal star*.

Next, let us use the notation $q_{nn'\mu_j}$ to denote the internal momentum which initiates from the vertex n and terminates at the vertex n' with μ_j denoting the multiplicity (that is when there are several lines incident upon the *same* pair of vertices n and n'). Thus (11) can be written as:

$$\forall n : \sum_{n',\mu_j} q_{nn'\mu_j} = p_n \tag{12}$$

and also, obviously,

$$q_{nn'\mu_j} = -q_{n'n\mu_j}. \tag{13}$$

In order to take into account of (12) by the method of Lagrangian multipliers, we define

$$R_n \equiv \sum_{n',\mu_j} q_{nn'\mu_j} - p_n. \tag{14}$$

Therefore,

$$\partial R_n / \partial q_{mm'\mu_j} = \delta_{nm} - \delta_{nm'}, \tag{15}$$

where ${}_1 j \in \mathcal{S}_m \cap \mathcal{S}_{m'}$ is assumed.

Next, let us introduce a set of Lagrangian multipliers $\{\lambda\}$. Thus, the requirement of

$$\frac{\partial}{\partial q_{mm'\mu_j}} \Big(\sum_j \alpha_j q_j^2 + \sum_n \lambda_n R_n \Big) = 0 \tag{16}$$

is equivalent to the condition

$$\forall {}_1 j : \partial V / \partial q_j = 0 \tag{17}$$

with (12) satisfied.

Using (15), condition (16) becomes

$$\forall n, n' : 2\alpha_j \bar{q}_{nn'\mu_j} + \lambda_n - \lambda_{n'} = 0, \tag{18}$$

which is the important variational relation we are looking for.

Next, (12) can be written into

$$\forall n : \sum_{n',\mu_j} \bar{q}_{nn'\mu_j} - p_n = 0, \tag{19}$$

where the $\bar{q}$, determined by (18) and (19) together, gives:

$$V = \psi(\{q\})|_{\{q\}=\{\bar{q}\}}. \tag{20}$$

We note that $\bar{q}_j$ is a function of α; thus V is also.

From (18) we get

$$\sum_{n',\mu_j} \bar{q}_{nn'\mu_j} + \sum_{n'} \sum_{{}_1 j \in \mathcal{S}_n \cap \mathcal{S}_{n'}} (\lambda_n - \lambda_{n'})/2\alpha_j = 0 \tag{21}$$

or

$${}_1 j \in \mathcal{S}_n \cap \mathcal{S}_{n'} : \sum_{j,n'} (\lambda_n - \lambda_{n'})/2\alpha_j = p_n. \tag{22}$$

From (18), we conclude immediately

$$\forall {}_1 j, {}_1 j' \in \mathcal{S}_n \cap \mathcal{S}_{n'} : \alpha_j \bar{q}_{nn'\mu_j} = \alpha_{j'} \bar{q}_{nn'\mu_{j'}}. \tag{23}$$

There appears to be a difficulty in generalizing (23) further if there is no internal line connecting *directly* two vertices v_n and $v_{n'}$, i.e., if

$$\mathcal{S}_n \cap \mathcal{S}_{n'} = \mathbb{O}, \tag{24}$$

where $\mathbb{O}$ denotes the empty set of lines (may contain vertices!).

However, this difficulty can be easily solved by defining $\lambda_n - \lambda_{n'}$ in the following *formal* way:

$$\lambda_n - \lambda_{n'} = (\lambda_n - \lambda_{n_1}) + (\lambda_{n_1} - \lambda_{n_2}) + \cdots + (\lambda_{n_r} - \lambda_{n'}), \tag{25}$$

where

$$\begin{aligned} &\mathcal{S}_n \cap \mathcal{S}_{n_1} \neq \mathbb{O}, \\ &\mathcal{S}_{n_1} \cap \mathcal{S}_{n_2} \neq \mathbb{O}, \\ &\cdots \\ \text{and} \quad &\mathcal{S}_{n_r} \cap \mathcal{S}_{n'} \neq \mathbb{O}. \end{aligned} \tag{26}$$

Since the above technique is good even for

$$\mathcal{S}_n \cap \mathcal{S}_{n'} = \mathbb{O}, \tag{27}$$

(18) and (25) lead to the conclusion

$$\alpha_j \bar{q}_{nn'\mu_j} = \alpha_{j_a} \bar{q}_{nn_1\mu_a} + \alpha_{j_b} \bar{q}_{n_1 n_2 \mu_b} + \cdots + \alpha_{j_s} \bar{q}_{n_r n'\mu_s}. \tag{28}$$

We can also write (28) in the form

$$\forall \mathcal{C} \in G : \sum_{\forall j \in \mathcal{C}} \alpha_j \bar{q}_j = 0 \tag{29}$$

or more precisely,

$$\forall \mathcal{C} \in G : \sum_{\forall j \in \mathcal{C}} \alpha_j \mathcal{E}_{j\mathcal{C}} \bar{q}_j = 0, \tag{30}$$

which is just one of two Landau–Bjorken–Nakanishi conditions, namely

$$\forall \mathcal{C} \in G^{\downarrow} : \sum_{j \in \mathcal{C}} \alpha_j^0 \mathcal{E}_{j\mathcal{C}} q_j = 0, \tag{31}$$

where $G^{\downarrow}$ denotes a reduced graph of G.

The only difference between (30) and (31) lies in the fact that in (30) we are dealing with the *original* given Feynman graph, while in (31) we are dealing with *all the possible reduced graphs* as well as the original graph. Therefore, it is clear that, when the original graph is considered for (31) (i.e., when one studies the so-called "principal

singularities"), we can say in this case that (31) has actually nothing to do with the singularity condition in view of (30). The singularity condition is rather given by the other equation of the Landau–Bjorken–Nakanishi conditions,[7-9] namely

$$\forall j \in {}_{c}[J] : m_j^2 + q_j^2 = 0, \tag{32}$$

where $[J]$ denotes a subset of internal lines $\{{}_1 1, \cdots, {}_1 J\}$ and the left subscript c means its complement. We note that the α_j's in (31) are fixed due to the presence of condition (32) and the requirement $\tilde{\alpha} = 1$. But, in (30), the α is of course entirely unspecified.

From the mathematical point of view, (30) or (31) is merely a trick to do integration with respect to $\{k\}$, and thus, strictly speaking, they alone have nothing to do with singularities. The singularity conditions are essentially governed by (32) and (31) *together*; it is therefore clear that the way to state the singularity conditions is by no means unique in this manner.

Due to the property shown in (30), it is convenient to introduce $E_{nn'}$, E_n, and $E_{n'}$ by

$$E_{nn'} \equiv E_n - E_{n'} = \alpha_j \bar{q}_{nn'\mu j}, \quad {}_1 j \in \mathcal{S}_n \cap \mathcal{S}_{n'}, \tag{33}$$

and we note, that, except for its difference, E_n alone is not well defined.

For two vertices whose stars have no common line, we can define $E_{nn'}$ by means of the method used in (25), e.g.,

$$\begin{aligned} E_{nn'} &\equiv E_n - E_{n'} \\ &= (E_n - E_{n_1}) \\ &\quad + (E_{n_1} - E_{n_2}) + \cdots + (E_{n_r} - E_{n'}). \end{aligned} \tag{34}$$

From (20),

$$\begin{aligned} V &= \psi(\{q\})|_{(q)=(\bar{q})} \\ &= \sum_{j=1}^{J} \alpha_j(\bar{q}_j^2 + m_j^2) \\ &= \tfrac{1}{2} \sum_{n,n',\mu j} \alpha_j \bar{q}_{nn'\mu j}^2 + \sum_j \alpha_j m_j^2, \quad {}_1 j \in \mathcal{S}_n \cap \mathcal{S}_{n'} \\ &= \tfrac{1}{2} \sum_{n,n',\mu j} (E_n - E_{n'}) \bar{q}_{nn'\mu j} + \sum_j \alpha_j m_j^2 \end{aligned}$$

or

$$V = \sum_{n,n',\mu j} E_n \bar{q}_{nn'\mu j} + \sum_j \alpha_j m_j^2.$$

Using (12), we obtain

$$V = \sum_n E_n p_n + \sum_j \alpha_j m_j^2. \tag{35}$$

In order for (35) to be meaningful, we can set

$$E_N \equiv 0, \tag{36}$$

where N denotes the Nth vertex. Thus,

$$\forall n : E_{nN} = E_n. \tag{37}$$

Therefore, (35) can be written as

$$V = \sum_n E_{nN} p_n + \sum_j \alpha_j m_j^2. \tag{38}$$

Now, let us consider the modified graph, $G^{(rs)}$, obtained from the original Feynman graph by deleting *all* the external lines *except* those at the external vertices r and s to which we now assign the new external momenta $+1$ and -1 respectively. The quantity $E_{mn}^{(rs)}$ is then defined as

$$E_{mn}^{(rs)} = E_{mn}|_{G \to G^{(rs)}} \tag{39}$$

and we also introduce the notation

$$E^{(rs)} \equiv E_{rs}^{(rs)} \equiv E_{rs}|_{G \to G^{(rs)}}. \tag{40}$$

The explicit meaning of $E^{(rs)}$ is

$$E^{(rs)} = \alpha_j [\bar{q}_{rs\mu j}]_{G \to G^{(rs)}}. \tag{41}$$

The following properties are either self-evident or are easily shown.

(a) *Symmetry:*

$$E^{(rs)} = E^{(sr)}, \tag{42}$$

$$E_{mn}^{(rs)} = E_{nm}^{(sr)}; \tag{43}$$

(b) *Antisymmetry:*

$$E_{mn}^{(rs)} = -E_{nm}^{(rs)}, \tag{44}$$

$$E_{mn}^{(rs)} = -E_{mn}^{(sr)}; \tag{45}$$

(c) *Reciprocity:*

$$E_{mn}^{(rs)} = E_{rs}^{(mn)}; \tag{46}$$

(d) *Contraction:*

$$E_{mn}^{(rs)} + E_{nk}^{(rs)} = E_{mk}^{(rs)}, \tag{47}$$

$$E_{mn}^{(rs)} + E_{mn}^{(st)} = E_{mn}^{(rt)}. \tag{48}$$

Further, there are following properties.

(e) *Proposition:*

$$E_{nn'} = \sum_{m=1}^{{}_N N} E_{nn'}^{(mn')} p_m, \tag{49}$$

[7] Landau, Nucl. Phys. **13**, 181 (1959).
[8] Bjorken, Stanford University preprint (1959).
[9] N. Nakanishi, Progr. Theoret. Phys. (Kyoto) **22**, 128 (1959).

where $^{\circ}N$ denotes the total number of *external* vertices.

Proof: This is obvious from (33) and the definition of $E^{(mn')}_{nn'}$.

(f) *Proposition:*

$$E^{(mn)} = E^{(mr)} + E^{(rn)} - 2E^{(mr)}_{nr}. \tag{50}$$

Proof:

$$\begin{aligned}\text{Right-hand side} &= E^{(mr)}_{mr} + E^{(rn)}_{rn} + 2E^{(mr)}_{rn} \\ &= (E^{(mr)}_{mr} + E^{(mr)}_{rn}) + (E^{(rn)}_{rn} + E^{(mr)}_{rn}) \\ &= E^{(mr)}_{mn} + E^{(mn)}_{rn} \\ &= E^{(mr)}_{mn} + E^{(rn)}_{mn} \\ &= E^{(mn)}_{mn} = E^{(mn)} = \text{left-hand side},\end{aligned}$$

where we have used antisymmetry, reciprocity and contraction properties. Q.E.D.

In the introduction, the V-function is expressed in the form of (6). However, for our present purpose, it is more convenient to write (6) in terms of external momenta explicitly using incidence matrices[1,5,6,10]:

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 - \sum_{n,n'=1}^{N-1} (\vec{h}^{-1})_{nn'} p_n p_{n'} \tag{51}$$

with

$$\vec{h} = \begin{bmatrix} h_{11} & \cdots & h_{1,N-1} \\ \vdots & & \vdots \\ h_{N-1,1} & \cdots & h_{N-1,N-1} \end{bmatrix}, \tag{52}$$

$$h_{nn'} \equiv \sum_{j=1}^{J} \epsilon_{jn} \cdot \epsilon_{jn'} / \alpha_j, \tag{53}$$

where

$\epsilon_{jn} = 0$, if the internal line j does not initiate or terminate at the vertex n,

$\epsilon_{jn} = 1$, if the internal line j initiates from the vertex n,

and

$\epsilon_{jn} = -1$, if the internal line j terminates at the vertex n.

We would like to write (51) further into the form

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 - \sum_{\substack{n,n'=1 \\ (n<n')}}^{N} w_{nn'} p_n p_{n'} \tag{54}$$

with $w_{n'n} \equiv w_{nn'}$.

To find $w_{nn'}$ function, we use (51):

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 + \frac{1}{h} \sum_{n,n'=1}^{{}^{\circ}N-1} \Delta^h_{nn'} p_n p_{n'} = \sum_{j=1}^{J} \alpha_j m_j^2 + \frac{1}{h} \left[\sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N-1} 2\Delta^h_{nn'} p_n p_{n'} - \sum_{n=1}^{{}^{\circ}N-1} \Delta^h_{nn} p_n^2 \right]. \tag{55}$$

Using momentum conservation,

$$p_n = - \sum_{\substack{n'=1 \\ (n' \neq n)}}^{{}^{\circ}N} p_{n'}; \tag{56}$$

thus (55) becomes

$$V = \sum_{j=1}^{J} \alpha_j m_j^2 - \frac{1}{h} \left[\sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N-1} (\Delta^h_{nn} + \Delta^h_{n'n'} - 2\Delta^h_{nn'}) p_n p_{n'} + \sum_{n=1}^{{}^{\circ}N-1} \Delta^h_{nn} p_n p_{{}^{\circ}N} \right]. \tag{57}$$

Comparison between (54) and (57) gives immediately

$$w_{n'n} = w_{nn'}$$

$$= (\Delta^h_{nn} + \Delta^h_{n'n'} - 2\Delta^h_{nn'})/h, \text{ for } n, n' < {}^{\circ}N; \tag{58}$$

$$w_{Nn} = w_{nN} = \Delta^h_{nn}/h, \text{ for } n < {}^{\circ}N. \tag{59}$$

Similarly, through some arrangement of indices one can express $w_{nn'}$ in terms of ϵ_{jl} and η_{jn}. The derivation is given in the Appendix.

Now, we are in a position to find the relationship between the $w_{nn'}$ functions and the $E^{(nn')}$ functions. From (49), we have

$$E_n - E_{n''} = \sum_{n'=1}^{{}^{\circ}N} E^{(n'n'')}_{nn'} p_{n'}, \tag{60}$$

that is,

$$\sum_{n=1}^{{}^{\circ}N} (E_n - E_{n''}) \cdot p_n = \sum_{n,n'=1}^{{}^{\circ}N} p_n p_{n'}. \tag{61}$$

But, since

$$\sum_{n=1}^{{}^{\circ}N} p_n E_{n''} = \left(\sum_{n=1}^{{}^{\circ}N} p_n \right) E_{n''} = 0, \tag{62}$$

thus, (61) becomes

$$\sum_{n=1}^{{}^{\circ}N} E_n p_n = \sum_{n,n'=1}^{{}^{\circ}N} E^{(n'n'')}_{nn'} p_n p_{n'}.$$

By virtue of (35), (60) gives

$$V = \sum_j \alpha_j m_j^2 + \sum_{n,n'=1}^{{}^{\circ}N} E^{(n'n'')}_{nn'} p_n p_{n'}. \tag{63}$$

For convenience, let us put $n'' \equiv {}^{\circ}N$ (there is no loss of generality); thus

$$V = \sum_j \alpha_j m_j^2 + \sum_{n,n'=1}^{{}^{\circ}N-1} E^{(n'{}^{\circ}N)}_{n{}^{\circ}N} p_n p_{n'}. \tag{64}$$

[10] Y. Chow, J. Math Phys. 5, 1255 (1964).

Comparison of (64) with the first equation of (55) gives the most interesting feature:

$$E_{n\,{}^{\circ}N}^{(n'\,{}^{\circ}N)} = (\vec{h})_{nn'}^{-1}. \tag{65}$$

Due to this idenitification, we write down the following expression in complete analogy with (57):

$$V = \sum_i \alpha_i m_i^2 - \sum_{\substack{n,n'=1\\(n'>n)}}^{{}^{\circ}N-1} (E^{(n\,{}^{\circ}N)} + E^{(n'\,{}^{\circ}N)} - 2E_{n\,{}^{\circ}N}^{(n'\,{}^{\circ}N)}) p_n p_{n'} - \sum_{n=1}^{{}^{\circ}N-1} E^{(n\,{}^{\circ}N)} p_n p_{{}^{\circ}N}. \tag{66}$$

Comparison between (54) and (66) yields

$$w_{n'n} = w_{nn'} = E^{(n\,{}^{\circ}N)} + E^{(n'\,{}^{\circ}N)} - 2E_{n\,{}^{\circ}N}^{(n'\,{}^{\circ}N)}, \text{ for } n, n' < {}^{\circ}N \tag{67}$$

and

$$w_{{}^{\circ}Nn} = w_{n\,{}^{\circ}N} = E^{(n\,{}^{\circ}N)}, \text{ for } n < {}^{\circ}N, \tag{68}$$

which shows $E^{(nm')}$ nonnegative since $w_{nn'}$ is.

Further, by using the property (50), we can simplify (67) and (68) into

$$\forall n, n' : w_{n'n} = w_{nn'} = E^{(nn')}, \tag{69}$$

which also defines $w_{nn} = 0$, since $E^{(nn)}$ is obviously zero. Thus, the V-function can now be written simply as

$$V = \sum_{i=1}^{J} \alpha_i m_i^2 - \sum_{\substack{n,n'=1\\(n'>n)}}^{{}^{\circ}N} E^{(nn')} p_n p_{n'}. \tag{70}$$

Since we have no intention of going into further development parallel to Wu's work,[2] we shall merely summarize here some of his majorization results with illustrative graphs.

Notations

(1) *When* a Feynman graph G is majorized by the graphs G', G'', and G''' *together*, we write symbolically:

$$G \multimap G' \dot{+} G'' \dot{+} G''',$$

where $\multimap$ denotes "being majorized by" and a circle on the plus sign means it is not the ordinary sum but

$$\mathfrak{D}_G \supset \mathfrak{D}_{G'} \cap \mathfrak{D}_{G''} \cap \mathfrak{D}_{G'''},$$

where $\mathfrak{D}_G$ means the domain in which $\min V_G > 0$.

(2) (a) A *subgraph* obtained from G by deleting ${}_ij$ is denoted by $G(\uparrow_i j)$.

(b) A *reduced* graph obtained from G by shrinking ${}_ii$ is denoted by $G(\downarrow_i i)$.

(c) A *reduced subgraph* obtained from G by deleting ${}_ij$ and shrinking ${}_ii$ is denoted by $G(\uparrow_i j \downarrow_i i)$.

Lemmas

In the following lemmas all the graphs, on both sides of the majorization signs ($\multimap$), are required to be *strongly connected* (i.e., there is no one-particle intermediate state).

(I) $$\forall\, {}_ii : G \multimap G(\downarrow_i i) \dot{+} G(\uparrow_i i). \tag{71}$$

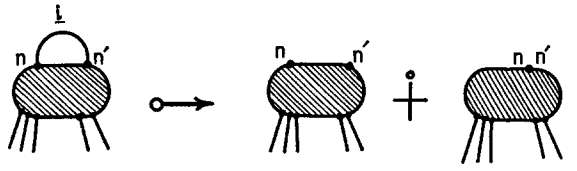

FIG. 1. Majorization corresponding to expression (71). In all figures, the following change of notation is made respective of *all* letters: ${}^{\circ}J$ in text is shown as J with an overbar, ${}^{\circ}j$ is shown as j with an overbar, ${}_iJ$ is shown as J with an underbar, and ${}_ij$ is shown as j with an underbar.

(II) Let $G = \bigcup_{i=1}^{r} G_i$ and

$$\exists v_n : G_i \cap G_j = v_n \text{ for any } i \neq j,$$

then

$$G \multimap G(\downarrow_i G_1) \dot{+} G(\downarrow_i G_2) \dot{+} \cdots \dot{+} G(\downarrow_i G_r). \tag{72}$$

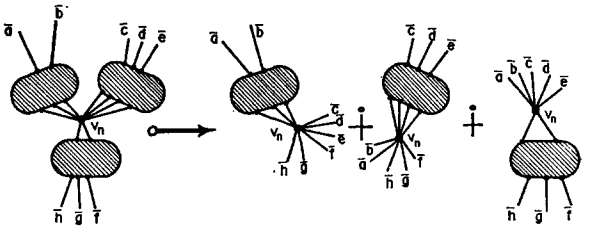

FIG. 2. Majorization corresponding to expression (72).

(III) If $\exists\, {}_i\mathcal{S}_v = {}_ii \cup {}_ij$ in G, then

$$G \multimap G(\downarrow {}_ii \uparrow {}_ij) \dot{+} G(\uparrow {}_ii \downarrow {}_ij) \dot{+} G(\downarrow {}_ii, {}_ij). \tag{73}$$

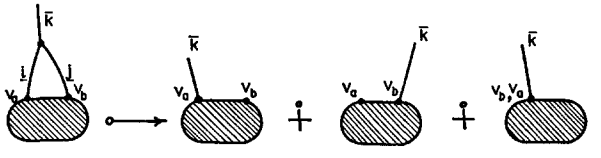

FIG. 3. Majorization corresponding to expression (73).

(IV) If, for $v_n \neq v_{n'}$,

$$\exists\, \mathcal{S}_n \text{ and } \mathcal{S}_{n'} : {}_i\mathcal{S}_n = {}_ih \cup {}_ii, \quad {}_i\mathcal{S}_{n'} = {}_ih \cup {}_ij,$$

then

$$G \multimap G(\downarrow {}_ii, {}_ij) \dot{+} G(\downarrow {}_ii\, {}_ih, {}_ij) \dot{+} G(\uparrow {}_ii \downarrow {}_ij, {}_ih) \dot{+} G(\uparrow {}_ij \downarrow {}_ii, {}_ih) \dot{+} G(\uparrow {}_ih \downarrow {}_ii, {}_ij). \tag{74}$$

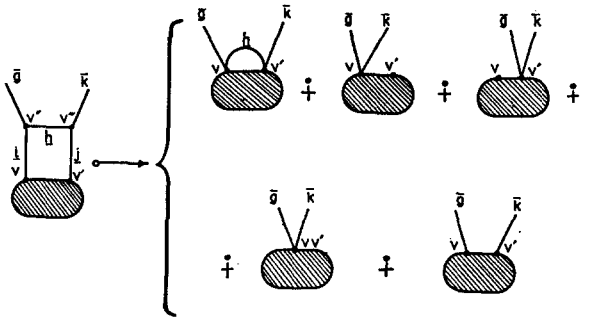

FIG. 4. Majorization corresponding to expression (74).

The above lemmas are derived by Wu under the assumption of the equal-mass case. Certain generalization can be made at the cost of some imposed restriction on mass value. Some limited generalization is carried out by Boyling[3] by taking into consideration of baryon number conservation, and we shall not discuss it here since the method follows closely that of Wu.

III. NAKANISHI'S PATH METHOD OF MAJORIZATION

Before proceeding with our discussion of several important theorems of Nakanishi in connection with his Path method of majorization, we first introduce the concept of "equal-mass graph."

Let m_j and M be the masses of internal lines and external lines ($^{\circ}N \equiv 2$), respectively. Further, we assume

$$\forall j : m_j/M = \text{a } \textit{rational} \text{ number, by approximation.}$$

Therefore we can write

$$\forall j : m_j/M \equiv c_j/d_j, \text{ with } c_j \text{ and } d_j \text{ as } \textit{integers}.$$

Next, let us find the least-common multiplier, n, of the set $\{d_j\}$ and define a mass

$$m^* \equiv M/n. \tag{75}$$

Further, we define

$$n_j \equiv nc_j/d_j, \tag{76}$$

which has the obvious property of n_j being integers and

$$\forall_j : m_j = m^* n_j.$$

This provides a means to replace each ${}_ij \in G$ by a bundle of n_j new internal lines; each new line now has a mass m^*. This modified graph, denoted by G^*, will be referred to as the *equal-mass graph* of G.

For convenience of discussion, let us introduce the notation $P^*_{j\mu}$ for a *constant momentum* of the equal-mass graph. Here the subscript j corresponds to the original line ${}_ij$ of the *original* graph and μ stands for the multiplicity resulted from splitting ${}_ij$ into many lines in the equal-mass graph. Using this notation, we state the following.

Proposition:

$$\text{If } \exists \{P^*_{j\mu}\} : m^{*2} + P^{*2}_{j\mu} \geq 0, \quad \forall j, \mu \text{ for } G^*, \text{ then}$$
$$\exists \{P_j\} : m_j^2 + P_j^2 \geq 0, \quad \forall j \in G. \tag{77}$$

Proof: By assumption we have

$$n_j m^{*2} + n_j P^{*2}_{j\mu} \geq 0, \quad \forall j, \mu \text{ for } G^*,$$

that is,

$$m_j^2 + n_j P^{*2}_{j\mu} \geq 0, \quad \forall j, \mu \text{ for } G, \tag{78}$$

Thus, if we simply choose

$$\forall {}_ij \in G : P_j \equiv n_j P^*_{j\mu},$$

then (78) becomes

$$\forall {}_ij \in G : m_j^2 + P_j^2 \geq 0; \qquad \text{Q.E.D.}$$

A. The Self-Energy Graphs

The case of $^{\circ}N \equiv 2$ corresponds to the self-energy graphs. Due to the property of (77) we shall study G^* instead of G, for the self-energy graphs. First, let us convert the condition of a stable particle into the language of G^*. Stability of the particle of a self-energy graph G means

$$\forall \mathcal{g} \in G : \sum_{j \in \mathcal{g}} m_j \geq M$$

that is,

$$\forall \mathcal{g} \in G : \sum_{j \in \mathcal{g}} m_0 n_j \geq m_0 n$$

or

$$\forall \mathcal{g} \in G^* : \tau(\mathcal{g}) \geq n, \tag{79}$$

where $\tau(\mathcal{g})$ denotes the *total number of internal lines* contained in the intermediate-state set $\mathcal{g}$.

The following *Path theorem* for a self-energy graph is due to Nakanishi.

Theorem N.1:

If the particle of a self-energy graph, G, is stable, then

$$\exists \{\mathcal{P}_1, \cdots, \mathcal{P}_n\} \in G^* : {}_i\mathcal{P} \cap {}_i\mathcal{P}_j = \mathbb{O}, \forall i, j, \tag{80}$$

where all paths are assumed to initiate and terminate at the two external vertices of the graph, and n is defined by expression (75). The set of n disjoint paths will be denoted simply by $\{\mathcal{P}\}$.

Proof: The proof will be given by means of mathematical induction (with respect to J). First, for the case of

$$\forall g \in G^* : \tau(g) = n, \tag{81}$$

we have exactly n paths; thus (80) is true.

Next, we *assume* the theorem to be true for $\tau(g) < J$ and proceed to show that the theorem is true for $\tau(g) = J$ to complete the proof.

To proceed the proof we consider systematically the different possibilities.

If the given graph has the property

$$\forall g \in G^* : \exists\, {}_ij \notin g, \tag{82}$$

then obviously this line ${}_ij$ can be deleted without effecting the stability condition (79). After deleting such lines, we obtain the condition

$$\forall\, {}_ij \in G^* : \exists\, g \ni {}_ij. \tag{83}$$

A graph with the property (82) may be further classified into different cases.

(1) Case of $G^* \equiv S_{\bullet_v} \cup S_{\bullet_{v'}}$: A typical example is given in Fig. 5.

Thus, we can easily see

$$\forall\, {}_ij \in G^* : {}_ij \in S_{\bullet_v} \cup S_{\bullet_{v'}} \tag{84}$$

and

$$\forall n, n' : {}_iS_n^0 \cap {}_iS_{n'}^0 = \textcircled{0}. \tag{85}$$

Let there be totally r internal stars, then we can write

$$g = ({}_iS_{\bullet_v} \cap {}_iS_{\bullet_{v'}}) \cup \left(\bigcup_{i=1}^{r} {}_i\tilde{S}_i^0\right), \tag{86}$$

where

$$\tilde{S}_i^0 = \text{either } S_i^0 \cap S_{\bullet_v} \text{ or } S_i^0 \cap S_{\bullet_{v'}} \tag{87}$$

The stability condition (79) now becomes

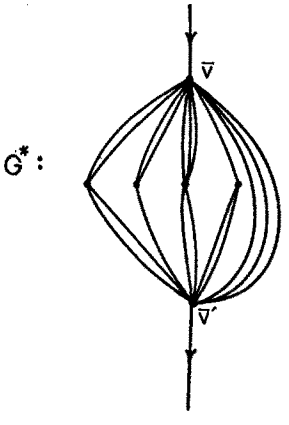

FIG. 5. An example of G^*.

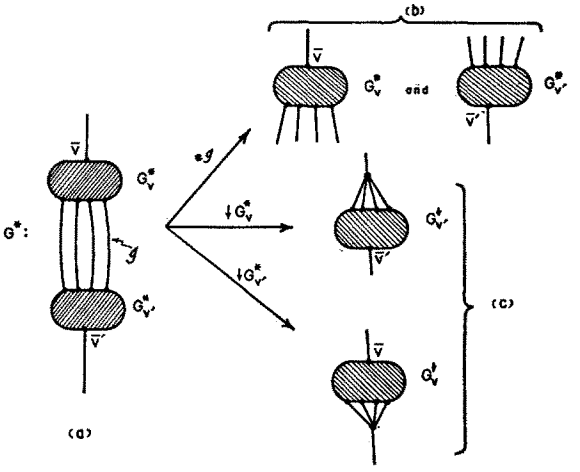

FIG. 6. G^* of Case (a).

$$\tau(S_{\bullet_v} \cap S_{\bullet_{v'}}) + \sum_{i=1}^{r} \min \tau(\tilde{S}_i^0) \geq n \tag{88}$$

with

$$\min \tau(\tilde{S}_i^0) \equiv \min \{\tau(S_i^0 \cap S_{\bullet_v}), \tau(S_i^0 \cap S_{\bullet_{v'}})\}. \tag{89}$$

Condition (88) gives the number of disjoint paths existing in G^*; it is greater than n, thus (80) is proved in this case.

(2) Case of $\exists\, {}_ij \in G^* : {}_ij \notin {}_i(S_{\bullet_v}) \cup {}_i(S_{\bullet_{v'}})$.

In this case, there are two possibilities:

(a) $$\forall\, {}_ij \in G^* : \exists\, g \ni {}_ij \text{ with } \tau(g) = n. \tag{90}$$

(b) $$ {}_ij : \forall g \ni {}_ij \text{ with } \tau(g) > n. \tag{91}$$

However, case (b) is reducible to case (a) if we simply delete all those ${}_ij$ satisfying (91). Thus, only case (a) shall be analyzed below.

Before further discussion we introduce here the concept of "opening" an internal line. To open an internal line, we artificially insert a vertex into this line and then delete this vertex so that this *internal* line now becomes two new *external* lines. The operation "opening" will be denoted by the symbol #.

In case (a), the graph G^* can be drawn in the form as shown in Fig. 6 (a) where we have

$$G^* = G_v^* \cup g \cup G_{v'}^*, \tag{92}$$

with

$$G_v^* \cap g = \textcircled{0} \text{ and } G_{v'}^* \cap g = \textcircled{0}, \tag{93}$$

here g satisfies (90). By *opening* g, we get two separate connected graphs $G_v^\#$ and $G_{v'}^\#$ of Fig. 6 (b). On the other hand, by shrinking $G_v^\#$ or $G_{v'}^\#$ we get $G_v^\downarrow$ and $G_{v'}^\downarrow$, of Fig. 6 (c).

We now have

$$\tau[\mathcal{g} \in G^{\downarrow}_{\bullet}] < \tau[\mathcal{g} \in G^{\#}_{\bullet}] \tag{94}$$

and

$$\tau[\mathcal{g} \in G^{\downarrow}_{\bullet\cdot}] < \tau[\mathcal{g} \in G^{\#}_{\bullet\cdot}]. \tag{95}$$

To use mathematical induction, we first assume

$$\exists \{\mathcal{P}\}_{\mathcal{D}} \in G^{\downarrow}_{\bullet} \tag{96}$$

and

$$\exists \{\mathcal{P}\}'_{\mathcal{D}} \in G^{\downarrow}_{\bullet\cdot}. \tag{97}$$

By virtue of (94)–(97), we thus conclude

$$\exists \{\mathcal{P}\}_{\mathcal{D}} \in G^*; \qquad \text{Q.E.D.}$$

Theorem N.2: For a self-energy graph with stable particle,

$$\exists \{P_i\} : m_i^2 + P_i^2 \geq 0, \quad \forall_i j \in G. \tag{98}$$

Proof: Since Theorem N.1 guarantees n disjoint paths in G^*, thus we can choose the constant momentum in G^* to be (p is the external momenta, and single subscript is used for P for simplicity):

$$\forall_i i \in \{\mathcal{P}\}_{\mathcal{D}} : P_i^* \equiv p/n \tag{99}$$

and

$$\forall_i i \notin \{\mathcal{P}\}_{\mathcal{D}} : P_i^* \equiv 0, \tag{100}$$

which lead to

$$\forall_i i \in \{\mathcal{P}\}_{\mathcal{D}} : m^{*2} + P_i^2 = 0, \tag{101}$$

$$\forall_i i \notin \{\mathcal{P}\}_{\mathcal{D}} : m^{*2} + P_i^2 = m^{*2} > 0; \tag{102}$$

thus (98) is established for G^*. But we show in Sec. IIIB that if (98) is true for G^* then it is also true for G; Q.E.D.

B. Vertex Graphs

The case of ${}^{\circ}N \equiv 3$ corresponds to the so-called vertex graphs (Fig. 7). In studying the self-energy graph, there are two very distinguishing features: first of all, the two external momenta are equal due to conservation, secondly, the particles of the external lines must be the same. These features do not exist in the vertex graph.

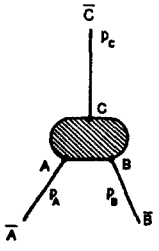

FIG. 7. A vertex graph.

In order to use the method of self-energy graphs to vertex graphs, we have to impose the restriction of *parallel* external momenta for a vertex graph, i.e., we can write:

$$p_v \equiv c_v p, \; v = A, B, C, \text{ (three vertex labels)}, \tag{103}$$

where c_v are the proportionality constants.

By choosing the following 4-momenta assignment:

$$p_v \equiv (M_v, 0, 0, 0), \qquad v = A, B, C, \tag{104}$$

then we can define n_A, n_B, and n_C by first finding a mass m^* such that any element of the set $\{M_A, M_B, M_C, m_1, \cdots, m_J\}$

$$\begin{aligned} M_v &\equiv n_v m^*, \qquad v = A, B, C, \\ m_j &\equiv n_j m^*, \qquad j = 1, \cdots, J, \end{aligned} \tag{105}$$

where n_v and n_j are integers.

Then an equal-mass graph G^* can be constructed by simply changing each *internal* line ${}_i j$ into n_j equal-mass lines, and changing external lines ${}^{\bullet}A$ and ${}^{\bullet}B$ into n_A and n_B equal-mass lines, respectively. Each new internal line has a mass m^* and each new external line has a mass M. Then, by virtue of (77), it is sufficient for us to study only G^*.

Next, we define $\mathcal{g}_v$ ($v = A, B, C$) in G^* as an intermediate state set that, if deleted, separates the external vertex v from the other two external vertices.

The following theorem can be easily proved.

Theorem N.3: For a vertex graph G, with external vertices A, B, and C, and parallel external momenta, if

$$\begin{aligned} &\forall \mathcal{g}_A \in G^* : \tau(\mathcal{g}_A) \geq n_A, \\ &\forall \mathcal{g}_B \in G^* : \tau(\mathcal{g}_B) \geq n_B, \\ &\forall \mathcal{g}_C \in G^* : \tau(\mathcal{g}_C) \geq n_C, \end{aligned}$$

then (1) $\exists\, n_A$ disjoint paths from A to C; (2) $\exists\, n_B$ disjoint paths from B to C; where n_A, n_B, n_C, are defined by (105).

Proof: First, we note that the conservation of 4-momenta gives

$$n_C = n_A + n_B,$$

since our choice of momenta is (104). Now, by identifying together the ends of the external lines of vertices A and B and by introducing a new external line ${}^{\circ}C'$, we get an artificial *self-energy graph*, $G^{*\prime}$, as shown in Fig. 8. The n_C is, therefore, just the n of this artificial self-energy graph. Then, in $G^{*\prime}$, there are four possible ways an intermediate-state

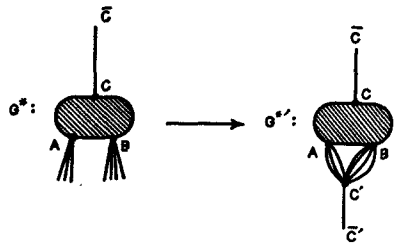

FIG. 8. Forming of an artificial self-energy graph from a vertex graph.

set, $\mathcal{G}'$, can separate $G^{*\prime}$ into two parts, as shown in Fig. 9.

We have the following cases, as implied by Figure 9:

Case (a): $\mathcal{G}'_C = \mathcal{G}_C$; thus $\tau(\mathcal{G}_C) \geq n_C \equiv n$, by assumption;

Case (b): $\mathcal{G}'_A$ has $\tau(\mathcal{G}'_A) = (\mathcal{G}_A) + n_B \geq n_A + n_B = n_C \equiv n$ by assumption;

Case (c): $\mathcal{G}'_B$ has $\tau(\mathcal{G}'_B) = \tau(\mathcal{G}_B) + n_A > n_A + n_B = n_C \equiv n$, by assumption;

Case (d): $\mathcal{G}'_{AB}$ has $\tau(\mathcal{G}_{AB}) = n_A + n_B = n_C \equiv n$, from graph.

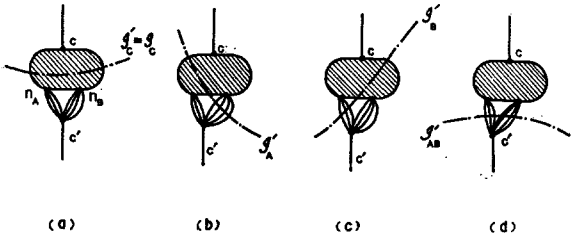

FIG. 9. Different possible intermediate states.

thus in all cases we have

$$\tau(\mathcal{G}') \geq n, \qquad \mathcal{G}' \equiv \text{either } \mathcal{G}'_A, \mathcal{G}'_B, \mathcal{G}'_C, \mathcal{G}'_{AB},$$

which leads, by Theorem N.2, to

$$\exists \{\mathcal{P}\}_{\mathcal{D}} \text{ from } A \text{ to } B.$$

This implies that there are totally n_A and n_B disjoint paths from A to C and from B to C respectively. Q.E.D.

C. The General Case of any $\bar{N}$

For the general graph with ${}^{\circ}N$ external lines, we classify the external lines as

$$\text{the incoming external lines} : \{a_1, \cdots, a_r\} \equiv {}^{\circ}a,$$

$$\text{the outgoing external lines} : \{b_1, \cdots, b_{r'}\} \equiv {}^{\circ}b.$$

Further, let an intermediate-state set $\mathcal{G}_k$, if deleted, separate the G into two sets:

$$\{a_1, \cdots, a_k, b_1, \cdots, b_{k'}\} \equiv {}^{\circ}h$$

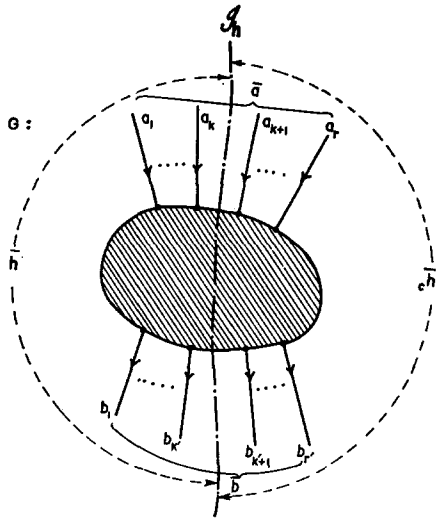

FIG. 10. A general graph.

and

$$\{a_{k+1}, \cdots, a_{r'}, b_{k'+1}, \cdots, b_{r'}\} \equiv {}^{\circ}_{c}h,$$

which is depicted in Fig. 10.

Next, for the case of parallel external momenta we can again choose the momenta as (104) and also define n_i similar to (105). Further, we introduce the concept of *capacity of in-flow and out-flow of external momenta*, n_{in} and n_{out}, as defined by

$$n_{\text{in}} \equiv \sum_{\mathcal{G}_i \leftrightarrow ({}^{\circ}h \cap {}^{\circ}a)} n_i$$

and

$$n_{\text{out}} \equiv \sum_{\mathcal{G}_i \leftrightarrow ({}^{\circ}h \cap {}^{\circ}b)} n_i$$

and the total capacity of flow, n, which is a generalization of the n of the last sections:

$$n = \sum_{\mathcal{G}_i \leftrightarrow {}^{\circ}a} n_i = \sum_{\mathcal{G}_i \leftrightarrow {}^{\circ}b} n_i; \tag{106}$$

then we have the following theorem.

Theorem N.4: For a general Feynman graph with parallel external momenta, if

$$\forall \mathcal{G} \in G^* : \tau(\mathcal{G}) \geq |n_{\text{in}} - n_{\text{out}}|, \tag{107}$$

then

$$\exists \{\mathcal{P}_1, \cdots, \mathcal{P}_n\}_{\mathcal{D}} \in G^*.$$

Proof: We shall make use of mathematical induction with respect to the number of effective external lines.

First, let us identify together all the equal-mass *incoming* external lines incident upon any two vertices, say a_1 and a_2, and introduce a new external line, a_{12}, to the new vertex, as shown in Fig. 11.

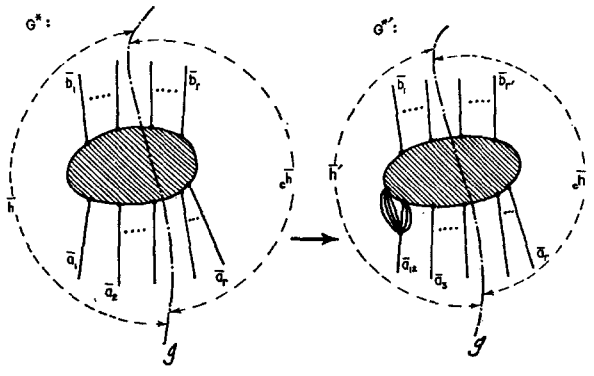

FIG. 11. Introduction of a new external line a_{12}.

The different possible intermediate-state sets for $G^{*\prime}$, as shown in Fig. 12, are

(1) $\mathcal{G}_1' = \mathcal{G}_1$, thus by assumption $\tau(\mathcal{G}') \geq |n_{\text{in}} - n_{\text{out}}|$,

(2) $\mathcal{G}_2'$: we have $\tau(\mathcal{G}_2') = n_{a_1} + n_{a_2}$. (108)

But,

$$|n_{\text{in}} - n_{\text{out}}| = \sum_{v_i \leftrightarrow ({}^{\bullet}h' \cap {}^{\bullet}a')} n_i - \sum_{v_i \leftrightarrow ({}^{\bullet}h' \cap {}^{\bullet}b)} n_i = |(n_{a_1} + n_{a_2}) - 0|; \quad (109)$$

thus (108) and (109) satisfy

$$\tau(\mathcal{G}_2) \geq |n_{\text{in}} - n_{\text{out}}|.$$

(3) $\mathcal{G}_3'$: from the graph of Fig. 12, we have

$$\begin{aligned} \tau(\mathcal{G}_3') &= n_{a_1} + \tau(\mathcal{G}_1') \\ &= n_{a_1} + \tau(\mathcal{G}_1) \\ &\geq n_{a_1} + |n_{\text{in}} - n_{\text{out}} + n_{a_1}| \\ &\geq |n_{\text{in}} - n_{\text{out}}|, \end{aligned}$$

where, in this case,

$$n_{\text{in}} = \sum_{v_i \leftrightarrow ({}^{\bullet}h - {}^{\bullet}a_1) \cap {}^{\bullet}a} n_i$$

and

$$n_{\text{out}} = \sum_{v_i \leftrightarrow ({}^{\bullet}h - {}^{\bullet}a_1) \cap {}^{\bullet}b} n_i.$$

(4) $\mathcal{G}_4'$: Similar to the above case, thus

$$\tau(\mathcal{G}_4') \geq |n_{\text{in}} - n_{\text{out}}|.$$

Therefore we conclude

$$\forall \mathcal{G}' \in G^{*\prime} : \tau(\mathcal{G}') \geq |n_{\text{in}} - n_{\text{out}}|.$$

Using mathematical induction, by assuming (107) is true for ${}^{\bullet}N - 1$ external lines, i.e.,

$$\exists \{\mathcal{P}_1, \cdots, \mathcal{P}_n\}_{\mathcal{D}} \in G^{*\prime}.$$

By opening the set of all internal lines of the star $\mathcal{S}_{a_{12}}$ we get back G^* from $G^{*\prime}$; thus

$\exists\, n_{a_1}$ disjoint paths initiating from the vertex a_1,
$\exists\, n_{a_2}$ disjoint paths initiating from the vertex a_2,
...
and $\exists\, n_{a_r}$ disjoint paths initiating from the vertex a_r.

Therefore, the total number of disjoint paths is

$$\sum_{i=1}^{r} n_{a_i} \quad \text{i.e.,} \quad \sum_{v_i \leftrightarrow {}^{\bullet}a} n_i,$$

which is equal to n, as according to the definition (106). This proves that the statement is also true for ${}^{\bullet}N$ external lines. Thus it completes the mathematical induction. Q.E.D.

IV. THE MAJORIZATION METHOD OF SYMANZIK AND LOGUNOV ET AL.

For the convenience of discussion, let us write expression (51) into

$$V = X + H,$$

where we have introduced the notation

$$X \equiv \sum_{j=1}^{J} \alpha_j m_j^2$$

and

$$H \equiv \sum_{n,n'=1}^{N-1} (\overleftrightarrow{h}^{-1})_{nn'} p_n p_{n'}.$$

The region of analyticity $\mathfrak{D}$ is given by

$$p \in \mathfrak{D} \text{ if } \forall \alpha_j \geq 0 : V(\alpha, p) \geq 0. \quad (110)$$

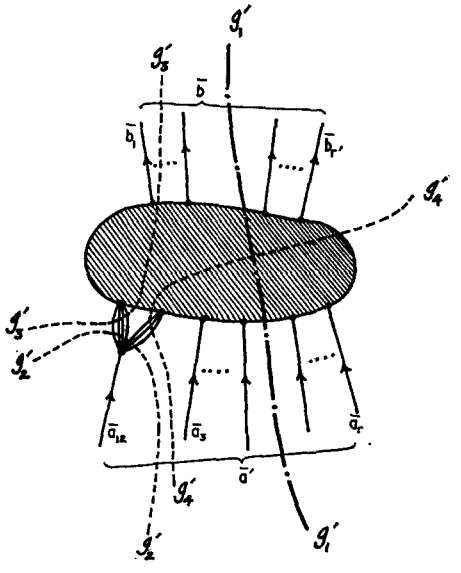

FIG. 12. Different possible intermediate states for a general graph.

By introducing

$$\mathfrak{K}^2(\alpha, p) \equiv -H/X$$

and

$$\mathfrak{K}^2(p) = \operatorname*{Sup}_{\alpha_i \geq 0} \{\mathfrak{K}^2(\alpha, p)\},$$

(110) becomes

$$p \in \mathfrak{D} \text{ if } \mathfrak{K}^2(p) < 1.$$

Majorization then means, in this notion,

$$G_1 \hookrightarrow G_2 \text{ if } \mathfrak{K}_1(p) \geq \mathfrak{K}_2(p), \quad \forall p \in \mathbf{E},$$

where G_1 and G_2 are assumed to have the same set of external momenta, p.

We note $\mathfrak{K}(p)$ is a homogeneous function of degree one, and further

$$\mathfrak{K}(p) \quad \text{is a } \textit{norm} \quad \text{if} \quad p \in \mathbf{E}.$$

One can show the *conjugate norm* to $\mathfrak{K}(p)$ is

$$\overline{\mathfrak{K}}(x) = \operatorname*{Inf}_{\alpha_i \geq 0} \{[X(\alpha)H(\alpha, p)]^{\frac{1}{2}}\}.$$

Now, we are in a position to give a summary of results due to Symanzik[5] and Logunov *et al.*[6]:

Theorem A:

$$\forall_i i \in G : G \hookrightarrow G(\uparrow_i i) \text{ if } p \in \mathbf{E}. \tag{111}$$

Remark: Comparing (111) with (71) we can see that in relaxing the Euclidean condition, an extra graph (reduced graph) is required to majorize G.

Theorem B: Let $\mathcal{C} \in G$ be a loop consiting of $r + 1$ internal lines, r of the lines have mass m and one line has a mass m', with $m \geq m'$. Let G' be a modified graph obtained from G by exchanging $m \rightleftarrows m'$ for those lines belonging to $\mathcal{C}$. Then

$$G \hookrightarrow G'.$$

Theorem C: Let G_1 and G_2 have the same set of external momenta, p. Let their norms be $\mathfrak{K}_1(p)$ and $\mathfrak{K}_2(p)$. Then

$$G_1 \hookrightarrow G_2 \quad \text{for} \quad \forall p \in \mathbf{E}$$

if and only if

$$\overline{\mathfrak{K}}_1(x) \leq \overline{\mathfrak{K}}_2(x).$$

Theorem D: A graph G is majorized by the set of graphs $\{G_1, \cdots, G_r\}$ if

$$\min_{i=1,\cdots,r} \{\overline{\mathfrak{K}}_i(x)\} \leq \overline{\mathfrak{K}}(x).$$

V. CONCLUSION

For the three different majorization methods discussed in the previous sections, the method of T. T. Wu has the advantage of imposing no Euclidean requirement for the external momenta. However, there are restrictions like values of masses involved, etc. On the other hand, the path method of Nakanishi and the algebraic method of Symanzik and Logunov *et al.* are applicable only if the set of external momenta is Euclidean. However, the path method has one distinguished feature; that is, it is a method of very general nature and further, this method can meet easily certain conservation requirements, like Baryon number or charge conservation, due to the existence of some paths as required by the method. The Symanzik method and its extension by the Russian workers are essentially algebraic; majorizations can be established if one can show that the required inequality in terms of norms are fulfilled. Thus, they are not as general as the path method. It is our hope that especially since the path method and the Symanzik method are both restricted to the Euclidean set of external momenta, one should be able to find out whether there is a region of overlapping of these two methods in a quantitative manner.

ACKNOWLEDGMENTS

I am greatly indebted to Professor N. Nakanishi for explaining the subtle points of his path method and to Professor S. S. Schweber for his interest and encouragement. I am also grateful to Dr. Y. Shimamoto for his hospitality during my stay at the Applied Mathematics Department, Brookhaven National Laboratory in the summer of 1964; part of the work was carried out there.

APPENDIX

$$V = \sum_{i=1}^{J} \alpha_i m_i^2 + \sum_{i=1}^{J} \alpha_i P_i^2 - \frac{1}{U} \sum_{n,n'-1}^{{}^{*}N} \sum_{l,l'-1}^{L} \sum_{i,i'-1}^{J} p_n p_{n'} \alpha_i \alpha_{i'} \varepsilon_{il} \varepsilon_{i'l'} \eta_{in} \eta_{i'n'} \Delta_{ll'}. \tag{A1}$$

The last sum of the right-hand side of (A1) can be split into three parts:

$$\sum_{n=n'} \{\quad\} \equiv A_1, \qquad \sum_{n>n'} \{\quad\} \equiv A_2,$$

$$\sum_{n<n'} \{\quad\} \equiv A_3.$$

Writing A_1 into the form

$$A_1 = \sum_{n=1}^{{}^{*}N} c_n p_n^2,$$

and due to momentum conservation, we have

$$\sum_{n=1}^{{}^{\circ}N} c_n p_n^2 = -\sum_{n=1}^{{}^{\circ}N} \sum_{\substack{n'=1 \\ (n'\neq n)}}^{{}^{\circ}N} c_n p_n p_{n'}$$

$$= -\sum_{n=1}^{{}^{\circ}N} \sum_{n'=1}^{n-1} c_n p_n p_{n'} - \sum_{n=1}^{{}^{\circ}N} \sum_{n'=n+1}^{{}^{\circ}N} c_n p_n p_{n'}$$

$$= -\sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} (c_n + c_{n'}) p_n p_{n'};$$

thus

$$A_1 = -\sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} \sum_{l,l',j,j'} \Delta_{ll'} \alpha_i \alpha_{i'} \varepsilon_{jl} \varepsilon_{j'l'} \times (\eta_{jn}\eta_{j'n} + \eta_{jn'}\eta_{j'n'}) p_n p_{n'}.$$

Therefore, the three sums A_1, A_2, and A_3 together give

$$-\frac{1}{U} \sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} \sum_{l,l'=1}^{L} \sum_{j,j'=1}^{J} \Delta_{ll'} \alpha_i \alpha_{i'} \varepsilon_{jl} \varepsilon_{j'l'} (2\eta_{jn}\eta_{j'n'} - \eta_{jn}\eta_{j'n} - \eta_{jn'}\eta_{j'n'}) p_n p_{n'}. \quad (A2)$$

As to the term $\sum_j \alpha_j P_j^2$, it contributes

$$\sum_{n,n'=1}^{{}^{\circ}N} \sum_{j=1}^{J} \alpha_j \eta_{jn}\eta_{jn'} p_n p_{n'}$$

$$= \sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} 2\alpha_j \eta_{jn}\eta_{jn'} p_n p_{n'} - \sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} \sum_{j=1}^{J} \alpha_j (\eta_{jn}^2 + \eta_{jn'}^2) p_n p_{n'}$$

$$= -\sum_{\substack{n,n'=1 \\ (n'>n)}}^{{}^{\circ}N} \sum_{j=1}^{J} \alpha_j (\eta_{jn} - \eta_{jn'})^2 p_n p_{n'}. \quad (A3)$$

Summing up (A2) and (A3), we obtain

$$w_{nn'} = \sum_{j=1}^{J} \alpha_j (\eta_{jn} - \eta_{jn'})^2 + \frac{1}{U} \sum_{j,j'=1}^{J} \sum_{l,l'=1}^{L} \Delta_{ll'} \alpha_i \alpha_{i'} \varepsilon_{jl} \varepsilon_{j'l'} \times (2\eta_{jn}\eta_{j'n'} - \eta_{jn}\eta_{j'n} - \eta_{jn'}\eta_{j'n'}),$$

$$n, n' = 1, \cdots, {}^{\circ}N, (n \neq n'),$$

which, therefore, also furnishes an expression of $E^{(nn')}$ in terms of ε_{jl} and η_{jn}.

Some Collapsing Cylinders and their Exterior Vacuum Metrics in General Relativity*

W. J. Cocke†
Institut Henri Poincaré, Paris and Aerospace Corporation, El Segundo, California
(Received 4 June 1965)

A cylindrical form of the Friedman metric is used to obtain nonstatic infinite cylinders of incoherent fluid. The vacuum metrics exterior to the cylinders are determined from the hypothesis that the metric tensor and its first partial derivatives be continuous. This hypothesis is applied by requiring the continuity of the first and second fundamental forms of the boundaries of the cylinders. The exterior metrics are nonstatic and may be expressed in the Einstein–Rosen form, and equations governing their behavior are derived. It is found that the exterior metric carries a flux of gravitational "C energy," the direction of which in the Einstein–Rosen frame is the same as that of the surface of the cylinders.

1. INTRODUCTION

THIS paper is concerned with finding the gravitational field surrounding certain nonstatic infinite cylinders of a zero-pressure perfect fluid. We use the general theory of relativity throughout, together with the usual differentiability and the continuity hypotheses.

We do not treat the general problem of the cylindrically symmetric zero-pressure fluid, but restrict ourselves to a special class of fluid motions. This special class is obtained by cutting infinitely long cylinders of fluid out of the well-known Friedman universe, for whose metric we have obtained a cylindrically symmetric form.

The Lichnerowicz junction conditions are used to ensure the existence of an exterior vacuum field surrounding the column of fluid, such that the continuity conditions across the hypersurface S separating the fluid from the vacuum are fulfilled. These continuity requirements then determine the vacuum field up to a transformation of coordinates.

We do not use the continuity conditions as such, but rather an invariant form of them; viz., the continuity of the first and second fundamental forms intrinsic to the hypersurface S. This enables us to avoid the difficult problem of actually finding a system of coordinates in which the continuity conditions are fulfilled explicitly.

The exterior vacuum metric is expressible in terms of a diagonal metric form found by Einstein and Rosen, and equations completely determining the metric tensor in this form are derived. However, the complexity of these equations prevents us from actually solving them.

The vacuum metrics all turn out to be nonstatic, and it is shown that there is a flux of gravitational radiation in the form of "C energy" associated with these fluid columns. We demonstrate that, if the direction of time is chosen so that the columns are collapsing toward the symmetry axis, then, in the Einstein–Rosen exterior frame, the flux of "C energy" is also inward.

In Sec. II we present the cylindrical form of the Friedman metric, and the transformation reducing it to the usual spherical form is given. In Sec. III the general problem of finding the exterior vacuum metric is discussed and the results of the application of the invariant continuity conditions presented. The actual calculations are quite complicated and are given in the Appendix, where it is demonstrated that the Einstein–Rosen metric form may be used in the vacuum, and a set of equations governing its behavior is derived. In Sec. IV this set of equations is further investigated, and we state how, in the Einstein–Rosen form, they show the exterior metric is exactly determined. Section V is devoted to a study of the radiative properties of the cylinders, carried out as mentioned above. A few remarks and suggestions for further work then follow.

We use the following conventions: We take the space–time metric to have signature +2 and use Latin tensor indices for the range 1, 2, 3, 4. Greek indices indicate 1, 2, 3. We use gravitational units in which $G = c = 1$, and write the field equations as $G_{ij} = R_{ij} - \frac{1}{2}Rg_{ij} = -8\pi T_{ij}$, where R_{ij} is the Ricci tensor and T_{ij} the energy-momentum tensor. Covariant differentiation is denoted by a semicolon, and partial differentiation by a comma.

* This work was supported by the National Science Foundation and by Air Force contract number AF 04(695)–669.

† National Science Foundation Postdoctoral Fellow 1964–65.

II. FRIEDMAN UNIVERSE IN CYLINDRICAL FORM

The cylindrical form of the Friedman universe may be conveniently written as

$$-ds^2 = g_{ij}\,dx^i\,dx^j = e^{g(t)}\left[\frac{dr^2}{1-kr^2} + (1-kr^2)\,dz^2 + r^2\,d\phi^2\right] - dt^2, \tag{1}$$

where $g(t)$ is an arbitrary real function of the time t, and k is any real constant, positive, negative, or zero. If $k > 0$, we must have $0 \leq r < k^{-\frac{1}{2}}$; if $k \leq 0$, $0 \leq r < \infty$. Always, $-\infty < z < \infty$, $0 \leq \phi < 2\pi$. The allowable range of t must be such that $g(t)$ is real and finite.

We now show that the metric (1) is the same as the Friedman metric, which may be written as[1]

$$-ds^2 = e^{g(t)}\left[\frac{d\bar{r}^2}{1-k\bar{r}^2} + \bar{r}^2(d\theta^2 + \sin^2\theta\,d\phi^2)\right] - dt^2. \tag{2}$$

The reader may easily verify that the following transformation reduces (1) to (2):

$$r = \bar{r}\sin\theta, \qquad z = f(\bar{r}, \theta),$$

where

$$\partial f/\partial\bar{r} = \cos\theta(1-k\bar{r}^2\sin^2\theta)^{-1}(1-k\bar{r}^2)^{-\frac{1}{2}}$$

and

$$\partial f/\partial\theta = -\bar{r}\sin\theta(1-k\bar{r}^2)^{\frac{1}{2}}(1-k\bar{r}^2\sin^2\theta)^{-1}.$$

One may show that $\partial^2 f/\partial\theta\partial\bar{r} = \partial^2 f/\partial\bar{r}\partial\theta$, so that the above differential equations are integrable. The fact that a cylindrically symmetric transform of (2) exists should surprise no one, since it is well known that (2) is translationally invariant.[2]

A more abstract study of metrics of the type (1) has also been made by Kompaneets and Chernov;[3] however, the connection with the Friedman universe was not mentioned.

Calculation of the conservative Einstein tensor G_{ij} for both (1) and (2) gives the field equations[4]

$$\begin{aligned} G^1_1 = G^2_2 = G^3_3 &= (d^2g/dt^2) + \tfrac{3}{4}(dg/dt)^2 + ke^{-g} = -8\pi p, \\ G^4_4 &= \tfrac{3}{4}(dg/dt)^2 + 3ke^{-g} = 8\pi\epsilon, \\ G^i_j &= 0 \quad \text{for} \quad i \neq j, \end{aligned} \tag{3}$$

where p is the pressure, and ϵ is the total proper energy density per unit volume. Equations (3) still leave room for assigning an equation of state $p = p(\epsilon)$, which then implies a single differential equation for $g(t)$.

III. GENERAL MATCHING PROBLEM

We now consider the problem of cutting out an infinitely long cylinder of material, whose motion we take to be governed by the cylindrical Friedman metric (1), and surrounding it with the appropriate vacuum field. In order for this to be possible, we must satisfy certain well-known continuity conditions[5] at the hypersurface S separating the material from the vacuum: These conditions require that there exist an admissible coordinate system in some neighborhood of S with respect to which the metric tensor g_{ij} and its first partial derivatives $g_{ij,k}$ are continuous across S. These conditions, in turn, imply the Lichnerowicz junction conditions,[6] which state that for a perfect fluid the pressure must vanish on S and that, if n_i is the unit vector normal to S, the fluid 4-velocity u^i must satisfy $u^i n_i = 0$ on S. It has been further shown[7] that, given an interior solution of the field equations for which there exists a hypersurface S on which $p = 0$ and $u^i n_i = 0$, then there exists a physically unique vacuum metric on the other side of S for which g_{ij} and $g_{ij,k}$ are continuous across S.

Let us mention that there are two equivalent ways of representing the hypersurface S. The first is to find a function $f(x^i)$ such that S is the locus of all points x^i satisfying $f(x^i) = 0$. The unit vector normal to S (for spacelike S) is then $n_i = f_{,i}(g^{ik}f_{,i}f_{,k})^{-\frac{1}{2}}$. The second way is to find a parametrization of S in terms of three coordinates u^α on S: $x^i = x^i(u^\alpha)$. The equation $f[x^i(u^\alpha)] = 0$ then becomes an identity in the u^α. Of course, the choice of the function $f(x^i)$ is not unique, and one may change the direction in which n_i points by changing the sign of $f(x^i)$. We use these two representations interchangeably in the discussions that follow.

What is the situation for our interior metric (1)? We consider a cylindrical hypersurface represented by the equation $f(x^i) = r - r_0 = 0$, such that $kr_0^2 < 1$, so that the outward-pointing normal is $n_i = (g_{11})^{\frac{1}{2}}\delta^1_i$.

Now, since we have everywhere $u^i = \delta^i_4 u^4$, the

[1] R. C. Tolman, *Relativity, Thermodynamics, and Cosmology* (Oxford University Press, Oxford, 1962), p. 370.
[2] Ref. 1, pp. 372–375.
[3] A. S. Kompaneets and A. S. Chernov, Zh. Eksperim. i Teor. Fiz. **47**, 1939 (1964) [English translation: Soviet Phys.-JETP **20**, 1303 (1965)].
[4] Ref. 1, p. 377.
[5] A. Lichnerowicz, *Théories relativistes de la gravitation* (Masson et Cie., Paris, 1955), pp. 3–5. J. L. Synge, *Relativity: The General Theory* (North-Holland Publishing Company, Amsterdam, 1960), pp. 39–41.
[6] Lichnerowicz, Ref. 5, pp. 63–64. Synge, Ref. 5., p. 187.
[7] Lichnerowicz, Ref. 5, pp. 27–33, 63–64.

condition $u^i n_i = 0$ is obviously satisfied on S. What about $p = 0$? Since the pressure is a function of the time only, this can only be true for $r = r_0$ if $p = 0$ everywhere. This is the case of incoherent matter, or "dust." Reference to Eqs. (3) then gives the condition

$$-(d^2g/dt^2) - \tfrac{3}{4}(dg/dt)^2 - ke^{-g} = 8\pi p = 0, \qquad (4)$$

which determines g as a function of time. The reader is referred to standard texts[8] for a discussion of the solutions of Eq. (4).

Thus if we choose $g(t)$ to satisfy (4) and pick $r = r_0$ as the hypersurface S, we may be assured that there exists a physically unique solution of the vacuum field equations in the region $r > r_0$ for which g_{ij} and $g_{ij,k}$ are continuous across S. We demonstrate in the next section that the desired vacuum solution is one of the class of cylindrical metrics known as "Einstein–Rosen waves," which may be written in the form[9]

$$-ds^2 = e^{2\gamma-2\psi}(d\rho^2 - d\tau^2) + e^{2\psi}\,dz^2 + \rho^2 e^{-2\psi}\,d\phi^2, \qquad (5)$$

where γ and ψ are functions of $x^1 = \rho$ and $x^4 = \tau$. The vacuum field equations $R_{ij} = 0$ then give the following equations[9] which must be satisfied by ψ and γ:

$$\gamma' = \rho[(\psi')^2 + (\dot{\psi})^2], \qquad \dot{\gamma} = 2\rho\psi'\dot{\psi}, \qquad (6)$$

where we use a prime for $\partial/\partial\rho$ and a dot for $\partial/\partial\tau$. The integrability condition $\partial^2\gamma/\partial\tau\partial\rho = \partial^2\gamma/\partial\rho\partial\tau$ then implies the equation

$$\psi'' + \frac{1}{\rho}\psi' - \ddot{\psi} = 0, \qquad (7)$$

which is itself, by the way, also contained in $R_{ij} = 0$. This is the cylindrical wave equation.

Note that, since it is not evident that the metric (5) comprises all cylindrically symmetric vacuum solutions, we cannot say *a priori* that the vacuum solutions which we are seeking may be written in this form. We show, however, that such is the case by considering the first and second fundamental forms intrinsic to S. Both fundamental forms are defined by using a parametrization of S $x^i = x^i(u^1, u^2, u^3)$ in terms of three coordinates u^α on S.

The first fundamental form is the metric which S inherits from the space-time in which it is imbedded and may be written in terms of the u^α as

$$\gamma_{\mu\nu}(u^\alpha) \equiv g_{ij}(\partial x^i/\partial u^\mu)(\partial x^j/\partial u^\nu). \qquad (8)$$

One defines the second fundamental form[10] via the unit vector n_i normal to S as

$$\Omega_{\mu\nu}(u^\alpha) \equiv -n_{i;j}(\partial x^i/\partial u^\mu)(\partial x^j/\partial u^\nu). \qquad (9)$$

From now on, we use the symbols $(-)$ and $(+)$ to designate indexed quantities defined, respectively, in terms of the interior and exterior coordinates. Note that since we are dealing with two different regions of space-time having only S in common (the one covered by the interior coordinates $x^{(-)i}$, and the other by the exterior coordinates $x^{(+)i}$), we obtain two different sets of fundamental forms, depending on which coordinate system we use to evaluate them, and, of couse, depending on different parametrizations of S with respect to the two systems. We denote these two sets $\gamma_{\mu\nu}^{(-)}$, $\Omega_{\mu\nu}^{(-)}$, $\gamma_{\mu\nu}^{(+)}$, $\Omega_{\mu\nu}^{(+)}$, in obvious notation.

We now introduce two hypotheses which we use in connection with a necessary and sufficient condition for the metrics to be compatible across S.

First, we suppose that the $g_{ij}^{(\pm)}$ and $g_{ij,k}^{(\pm)}$ are *separately* continuous up to and on S in their respective domains of definition. Secondly, we suppose that we are given two parametric representations of S, $x^{(-)i} = h^{(-)i}(u^1, u^2, u^3)$ and $x^{(+)i} = h^{(+)i}(u^1, u^2, u^3)$, such that all of S is covered by the same domains of the u^α in both cases. The $h^{(\pm)i}(u^\alpha)$ are supposed to have continuous second partial derivatives.

Then if the above hypotheses are true, the following proposition is known from the theory of continuity conditions[11]: A necessary and sufficient condition that the two metrics $g_{ij}^{(-)}$ and $g_{ij}^{(+)}$ be compatible across S is the equality of the two sets of fundamental forms as functions of the u^α:

$$\gamma_{\mu\nu}^{(-)}(u^\alpha) = \gamma_{\mu\nu}^{(+)}(u^\alpha), \qquad \Omega_{\mu\nu}^{(-)}(u^\alpha) = \Omega_{\mu\nu}^{(+)}(u^\alpha).$$

These equations have certain obvious invariance properties and establish the *existence* of an admissible coordinate system in some neighborhood of S (for which the g_{ij} and $g_{ij,k}$ are continuous across S) and permit us to skirt the tedious problem of actually having to find this admissible coordinate system.

We now present the results of the above matching conditions for our problem, the calculations for which are rather tedious and are relegated to the Appendix. They show that the metric (5) can be matched to the collapsing cylinder whose boundary

[8] Ref. 1, pp. 413, 416. See also L. Landau and E. Lifshitz, *Théorie du champ* (Éditions de la Paix, Moscow), pp. 445–447.

[9] J. Weber, *General Relativity and Gravitational Waves* (Interscience Publishers, Inc., New York, 1961), p. 99.

[10] L. P. Eisenhart, *Riemannian Geometry* (Princeton University Press, Princeton, 1949), pp. 146–150.

[11] G. Darmois, "Les equations de la gravitation einsteinienne," *Memorial des sciences mathématiques* **XXV** (Paris, 1927), p. 30.

S is $r = r_0$ in the frame (1), where we must have $\alpha = kr_0^2 < 1$. As stated above we always take $g(t)$ to satisfy the zero-pressure equation (4) and parametrize S in the frame (1) via $x^{(-)i} = h^{(-)i}(u^\alpha)$, where $x^{(-)1} = r = r_0$, $x^{(-)2} = z = u^2$, $x^{(-)3} = \phi = u^3$, $x^{(-)4} = t = u^1 = u$. Note that $t = u$ is then the proper time of a particle "at rest" on S and that the particle's 4-velocity is then $V^{(-)i} = \partial x^{(-)i}/\partial u$. Hence $V^{(-)i}V_i^{(-)} = -1$.

In the frame (5) we again use $u = u^1$ and write S similarly as $x^{(+)1} = \rho = \rho(u)$, $x^{(+)2} = z = u^2$, $x^{(+)3} = \phi = u^3$, and $x^{(+)4} = \tau = \tau(u)$, where $\rho(u)$ and $\tau(u)$ are functions yet to be determined. We still take u to be the proper time of a particle on S, so that again we have

$$V^{(+)i}V_i^{(+)} = e^{2\gamma(u)-2\psi(u)} \times [(d\rho/du)^2 - (d\tau/du)^2] = -1, \quad (10)$$

where $\gamma(u) = \gamma[\rho(u), \tau(u)]$, etc.

We use the above parametrizations in the Appendix to show that $\gamma_{\mu\nu}^{(+)} = \gamma_{\mu\nu}^{(-)}$ implies that, with $\alpha = kr_0^2$ and $c_0 = r_0(1-\alpha)^{\frac{1}{2}}$,

$$\rho(u) = c_0 e^{\sigma(u)} \quad (A1)$$

and

$$e^{\psi(u)} = (1-\alpha)^{\frac{1}{2}} e^{\frac{1}{2}\sigma(u)}. \quad (A2)$$

Equating the two second fundamental forms gives the information

$$d\tau/du = \pm e^{\frac{1}{2}\sigma}(1-2\alpha), \quad (A9)$$

$$\psi'(u) = (2\alpha-1)^{-1}e^{-\sigma}\left[\frac{\alpha}{c_0} + e^{2(\gamma-\psi+\sigma)} \times \frac{c_0}{2}\left(\frac{dg}{du}\right)^2\right] \equiv \theta(u), \quad (A11)$$

$$\psi(u) = \pm\tfrac{1}{2}e^{2\gamma-2\psi+\frac{1}{2}\sigma}(dg/du) \equiv \omega(u), \quad (A12)$$

where Eq. (10) yields the relation

$$e^{2\psi(u)-2\gamma(u)} = e^{\sigma}(1-2\alpha)^2 - c_0^2 e^{2\sigma}(dg/du)^2. \quad (11)$$

The signs in Eqs. (A9) and (A12) may always be chosen so that $d\tau/du > 0$, with positive signs for $\alpha < \frac{1}{2}$ and negative ones for $\frac{1}{2} < \alpha < 1$.

In the Appendix we also need the equation $d^2x^{(+)i}/du^2 + \Gamma_{km}^{(+)i}[dx^{(+)k}/du][dx^{(+)m}/du] = 0$, which is satisfied if $\rho(u)$, $\tau(u)$ is a geodesic. This must be so since S is generated by the streamlines of an incoherent fluid. Further, one can substitute the expressions given by the above equations to show that this geodesic equation implies the zero-pressure condition Eq. (4).

Thus we are assured that the necessary and sufficient conditions $\gamma_{\mu\nu}^{(+)} = \gamma_{\mu\nu}^{(-)}$ and $\Omega_{\mu\nu}^{(+)} = \Omega_{\mu\nu}^{(-)}$ may be fulfilled by the use of the above parametrizations.

IV. DETERMINATION OF THE VACUUM METRIC

Having proved that the junction conditions may be satisfied by the exterior metric form (5), we need only show how the full dependence of ψ and γ is determined. Note that actually we have only shown that the form (5) holds on S itself and not in the exterior away from S. However, if we can show that ψ and γ are not overdetermined in the exterior, we are assured that the conditions of the theorem on compatibility given in Sec. III are fulfilled. Thus no additional terms vanishing on S would be necessary in Eq. (5).

Further, from Ref. 7 we also know that this exterior solution is physically unique; i.e., determined up to a transformation of coordinates. However, we also state how one may prove that the functions ψ and γ are completely determined via Eqs. (A2), (A11), and (A12) in some neighborbood of S. This is in agreement with the fact that the only transformation that conserves the form of (5) and leaves $\rho \geq 0$ is a trivial change of scale.

We begin by discussing the cylindrical wave equation

$$\psi'' + (1/\rho)\psi' - \ddot{\psi} = 0, \quad (7)$$

the general real solution for which is

$$\psi(\rho,\tau) = \sum_\lambda\{[a(\lambda)\sin\lambda\tau + b(\lambda)\cos\lambda\tau]J_0(\lambda\rho) + [c(\lambda)\sin\lambda\tau + d(\lambda)\cos\lambda\tau]N_0(\lambda\rho)\} + c_1\log\rho + c_2\tau + c_3. \quad (12)$$

$J_0(x)$ and $N_0(x)$ are, respectively, the ordinary Bessel function and the Neumann function, both of order zero, and the $a(\lambda)$, $b(\lambda)$, $c(\lambda)$, and $d(\lambda)$ are arbitrary real functions of λ. The c_n are constants, and the symbol $\sum_\lambda$ indicates summation over any discrete series of positive values of $\lambda \neq 0$ plus integration over any positive intervals of λ. We assume that all integrals and infinite series are "termwise" differentiable as many times as we like with respect to both ρ and τ for whatever intervals of ρ and τ are under consideration. We examine this assumption more carefully later on.

Let us now consider the relations (A2), (A11), and (A12), which we write using Eq. (12) as

$$\psi(u) = \tfrac{1}{2}g(u) + \log(c_0/r_0) = \sum_\lambda\{[a(\lambda)\sin\lambda\tau(u) + b(\lambda)\cos\lambda\tau(u)]J_0[\lambda\rho(u)] + [c(\lambda)\sin\lambda\tau(u) + d(\lambda)\cos\lambda\tau(u)]N_0[\lambda\rho(u)]\} + c_1\log\rho(u) + c_2\tau(u) + c_3, \quad (13)$$

$$\psi'[\rho(u), \tau(u)] = \theta(u), \tag{14}$$

$$\psi[\rho(u), \tau(u)] = \omega(u), \tag{15}$$

where we do not bother to write out the full expressions for the first derivatives of Eq. (12).

We see that we have three equations to satisfy. Actually, however, there is an identity which connects them; namely, Eq. (A10). Thus we may work with any pair of the above equations.

What about the assumption of infinite differentiability of (12)? Examination of explicit expressions for $g(u)$, $\theta(u)$, and $\omega(u)$ would show that they are all analytic except for those values of u where $g(u)$ goes to infinity, which values correspond to the singular states of the Friedman universe. Thus if we wish a solution valid only in a certain closed interval of τ such that for the corresponding closed interval I of u the functions g, θ, and ω are analytic, then we may evidently restrict ourselves to an ordinary "Fourier series" for $\sum_\lambda$. In this case both Bessel functions will also be analytic in some open interval of ρ containing the corresponding closed interval $\{\rho(u): u \in I\}$, and since $J_0(x)$ and $N_0(x)$ for large positive x behave as sines and cosines divided by $x^{\frac{1}{2}}$, their presence in the series should not interfere with reworking the usual proofs of the infinite termwise differentiability of Fourier series representing analytic functions.[12] It does not seem profitable to pursue this topic any further here.

Let us also note that any pair of the Eqs. (13), (14), (15) certainly have linearly independent "kernels," so that essentially we have two independent "Fourier series" to solve for two independent functions. Thus the problem is not overdetermined.

We now state how to prove that, given the assumption of infinite termwise differentiability for the chosen interval I of u, the solution is also unique in the following sense: Given two different sets of constants $\hat{a}(\lambda)$, $\hat{b}(\lambda)$, $\hat{c}(\lambda)$, $\hat{d}(\lambda)$, $\hat{c}_n$ and $\bar{a}(\lambda)$, $\bar{b}(\lambda)$, $\bar{c}(\lambda)$, $\bar{d}(\lambda)$, $\bar{c}_n$, each of which separately solves Eqs. (13)–(15) on S for the same $\psi(u)$, $\theta(u)$, $\omega(u)$, then the corresponding vacuum solutions $\hat{\psi}(\rho, \tau)$ and $\bar{\psi}(\rho, \tau)$ thus obtained are identical in some neighborhood of S.

We proceed as follows: By virtue of the linearity of Eqs. (12)–(15) in the unknown parameters, we may form a new solution ϕ of Eq. (7) by writing

$$\phi(\rho, \tau) \equiv \hat{\psi}(\rho, \tau) - \bar{\psi}(\rho, \tau).$$

Obviously, $\phi(u) \equiv \hat{\psi}(u) - \bar{\psi}(u) = \phi'(u) = \dot{\phi}(u) = 0$.

Using Eq. (10) and the fact that $\phi(\rho, \tau)$ is a solution of Eq. (7) in some neighborhood of S, one may prove that $\phi''(u) = \dot{\phi}'(u) = \ddot{\phi}(u) = 0$ on S. One may easily extend the process to show that the partial derivatives up to any given order vanish on S. Thus, since $\phi(\rho, \tau)$ is by construction analytic in some neighborhood of S, it vanishes indentically in this neighborhood. Hence $\hat{\psi}(\rho, \tau) = \bar{\psi}(\rho, \tau)$, and the solutions of Eqs. (13)–(15) are unique.

Having determined $\psi(\rho, \tau)$, we now inquire as to how to find $\gamma(\rho, \tau)$. Since Eq. (7) has been satisfied, the field equations (6) are integrable, and $\gamma(\rho, \tau)$ is thus determined up to an arbitrary constant. However, we may also use Eq. (10) to find $\gamma(u)$, which raises the suspicion that $\gamma(\rho, \tau)$ may be over determined. But if one uses Eqs. (6) to find $\gamma'(u)$ and $\dot{\gamma}(u)$, the equality $d\gamma(u)/du = \gamma'(u)\, d\rho/du + \dot{\gamma}(u)\, d\tau/du$ is easily seen to be satisfied by virtue of Eq. (4). Thus $\gamma(\rho, \tau)$ is also exactly determined.

V. RADIATION IN THE EXTERIOR METRIC

We now examine the interesting question of whether or not our cylinders emit (or absorb) gravitational radiation. Is there in some sense a flux of gravitational energy in the exterior space whose metric is given by the solution of Eqs. (13), (14), and (15)?

Such questions are ordinarily very difficult to deal with, since the total energy of a gravitating system is satisfactorily defined only when the system is bounded and when the metric is asymptotically Minkowskian. Note that in the present infinite cylinder problem these conditions are obviously not fulfilled. However, for the special case of cylindrical systems a very useful type of energy has been devised by Thorne.[13] This is the so-called "C energy," and the reader is referred to the original paper for the general treatment.

For our purposes it is enough to note that in the exterior vacuum field given by the metric (5), the total "C energy" per unit length in the z direction contained inside the surface ρ = const on the hypersurface τ = const is given by

$$E(\rho, \tau) = \tfrac{1}{4}\gamma(\rho, \tau). \tag{16}$$

This energy per unit length contains not only a gravitational energy inside ρ = const but also the mass energy of the incoherent fluid in the cylinder. Thus if we can determine the sign of $\dot{E}(\rho, \tau) = \partial E/\partial\tau$ in the exterior we will know whether the "C energy" inside ρ = const is increasing or decreasing.

But would this tell us whether the cylinder under

[12] G. P. Tolstow, *Fourierreihen* (Deutscher Verlag der Wissenschaften, Berlin, 1955), p. 119.

[13] K. S. Thorne, Phys. Rev. **138**, B251 (1965).

scrutiny is "really" radiating or absorbing? After all, for nonstatic systems there is nothing special about the Einstein–Rosen frame, except for the fact that it admits simple expressions for the field equations.

Thus it appears that the question of the absolute sign of the energy flux in the exterior cannot be invariantly stated.[14] This is shown by noting that the sign of the radial component of the "C energy" flux vector P^i given by Thorne's Eq. (17) can be changed by a transformation of coordinates satisfying the following criteria: (1) The cylindrical "standard coordinate system" form $-ds^2 = e^{2(\gamma-\psi)}(dr^2-dt^2) + e^{2\psi}dz^2 + \alpha^2 e^{-2\psi}d\phi^2$, where $\alpha = \alpha(r, t)$, is preserved; (2) $\partial r'/\partial r > 0$, $\partial t'/\partial t > 0$; and (3) the Jacobian of the transformation is positive.

However, it is still instructive to examine the sign of $\dot{E}(\rho, \tau)$ in the exterior. We have from Eq. (6) $4\dot{E}(\rho, \tau) = \dot{\gamma}(\rho, \tau) = 2\rho\psi'\dot{\psi}$, which we may evaluate very near the surface of the cylinder by means Eqs. (A11), (A12), and (11). This will tell us the rate of change of the total "C energy" inside $\rho = \text{const}$ at a time when this particular ρ is located just outside $\rho(u)$.

Let us always choose the signs to conform to $d\tau/du > 0$. Then

$$\begin{aligned}\dot{\gamma}(u) &= 2\rho(u)\psi'(u)\dot{\psi}(u) \\ &= -e^{-\frac{1}{2}\sigma}(dg/du)\,|2\alpha - 1|^{-1} \\ &\quad \times [\alpha + \tfrac{1}{2}c_0^2 e^{2(\gamma-\psi+\sigma)}(dg/du)^2].\end{aligned} \tag{17}$$

Since c_0 is positive we see that if α is nonnegative, then $\dot{\gamma}(u)$ has the opposite sign from $d\rho/du = c_0 e^{\sigma} dg/du$. Thus for collapse, when $d\rho/du < 0$, it follows that $\dot{E}(u) > 0$ for $\alpha \geq 0$; and we can say that the gravitational "C energy" flux is directed inward for observers at constant ρ situated just outside the cylinder.

For $\alpha < 0$ the answer is not immediately apparent. However, one can substitute the actual solution of Eq. (4) for $g(u)$ in parametrized form, which for $\alpha < 0$ may be written as (see Ref. 8) $g(v) = 2 \ln (\cosh v - 1)$, $u(v) = (-k)^{-\frac{1}{2}}(\sinh v - v)$. Detailed examination with the help of Eq. (11) would show that the quantity in square brackets in Eq. (17) is again always positive as long as $e^{2\psi-2\gamma} > 0$.

Thus from the standpoint of the Einstein–Rosen coordinate system the collapsing cylinders are all associated with an incoming flux of gravitational "C energy". Of course we may reverse our point of view and consider the cylinders to be "exploding" from an initially singular state. If $k > 0$, this is so anyway during half of the finite "lifetime" of the cylinder, since there is a nonsingular moment of time symmetry for $g(u)$ in this case. In an explosive phase, then, the flux of "C energy" just outside the surface S is always outward.

It would be interesting to compare these results with those obtained from the use of the various energy-momentum pseudotensors, but we do not go into that here.

CONCLUSIONS AND OBSERVATIONS

Having theoretically solved the problem of finding the vacuum metric exterior to the collapsing cylinder, one should next try to solve the equations for the expansion coefficients. However, this is seen to be an extremely complicated problem, even for the simple case $k = 0$, and the author has made no progress toward resolving it. Evidently, a great many questions remain to be answered regarding the vacuum metric, particularly about its behavior "in the large."

Let us also remark that a complete study of its properties might shed some light on the collapse of an elongated "cigar-shaped" object, albeit of a very special sort. One might approximate the fields close to the central portions of such an object by using the metric studied here. For example, it might be possible to get an estimate of the amount of gravitational energy absorbed during a collapse.

ACKNOWLEDGMENTS

The author would like to thank the relativity group at the Institut Henri Poincaré for their kind hospitality, especially Mmes. M. A. Tonnelat and Y. Choquet-Bruhat. He would also like to thank Dr. L. Bel for enlightening discussions.

APPENDIX

We now proceed to show how the invariant continuity conditions stated above in Sec. III may be satisfied. Let us first investigate the form $\gamma_{\mu\nu}$. From Eq. (8) together with the parametrization $h^{(-)i}(u^\alpha)$ given in Sec. III, one easily obtains, with $\alpha = kr_0^2$,

$$\gamma_{11}^{(-)} = -1, \quad \gamma_{22}^{(-)} = e^{\sigma(u)}(1-\alpha), \quad \gamma_{33}^{(-)} = r_0^2 e^{\sigma(u)}$$
$$\gamma_{\mu\nu}^{(-)} = 0 \quad \text{for} \quad \mu \neq \nu.$$

Evaluating $\gamma_{\mu\nu}$ now from the outside via Eq. (5) and using Eq. (10) one finds

[14] One might also ask whether or not the total integrated proper fluid energy per unit length changes. However, for an incoherent fluid, the total proper energy contained in any spacelike 3-volume across the surface of which there is no fluid flow is a constant of the motion. Thus the total integrated fluid energy per unit column length would be a constant in any case.

$$\gamma_{11}^{(+)} = -1, \quad \gamma_{22}^{(+)} = e^{2\psi(u)}, \quad \gamma_{33}^{(+)} = \rho(u)^2 e^{-2\psi(u)},$$
$$\gamma_{\mu\nu}^{(+)} = 0 \quad \text{for} \quad \mu \neq \nu.$$

The condition $\gamma_{\mu\nu}^{(+)} = \gamma_{\mu\nu}^{(-)}$ then yields the relations [we abbreviate $c_0 \equiv r_0(1-\alpha)^{\frac{1}{2}}$]

$$\rho(u) = c_0 e^{\sigma(u)}, \tag{A1}$$

$$e^{\psi(u)} = c_0 e^{\frac{1}{2}\sigma(u)}/r_0. \tag{A2}$$

We now calculate the second fundamental form $\Omega_{\mu\nu}$. From Eq. (9) we have

$$\Omega_{\mu\nu} = -n_{i;j}\frac{\partial x^i}{\partial u^\mu}\frac{\partial x^j}{\partial u^\nu} = (\Gamma_{ij}^k n_k - n_{i,j})\frac{\partial x^i}{\partial u^\mu}\frac{\partial x^j}{\partial u^\nu}.$$

In the frame (1), again using $n_i = (g_{11})^{\frac{1}{2}}\delta_i^1$ and the preceding parametrization $h^{(-)i}(u^\alpha)$,

$$\Omega_{\mu\nu}^{(-)} = \left[\Gamma_{ij}^1(g_{11})^{\frac{1}{2}} - (g_{11})_{,j}^{\frac{1}{2}}\delta_i^1\right]\frac{\partial x^i}{\partial u^\mu}\frac{\partial x^j}{\partial u^\nu}$$
$$= (g_{11})^{\frac{1}{2}}\Gamma_{ij}^1\frac{\partial x^i}{\partial u^\mu}\frac{\partial x^j}{\partial u^\nu}.$$

Straightforward calculation yields

$$\Omega_{11}^{(-)} = 0, \quad \Omega_{22}^{(-)} = kc_0 e^{\frac{1}{2}\sigma(u)},$$
$$\Omega_{33}^{(-)} = -c_0 e^{\frac{1}{2}\sigma(u)}, \quad \Omega_{\mu\nu}^{(-)} = 0, \quad \text{for} \quad \mu \neq \nu. \tag{A3}$$

For the exterior metric (5), the situation is more complicated, as we do not know the form of the function $f[x^{(+)i}]$, such that S is the locus of all points satisfying $f[x^{(+)i}] = 0$, except for the fact that it does not depend on z and ϕ. However, there are two identities which $n_i^{(+)}$ must satisfy. These are

$$n_i^{(+)}n^{(+)i} = e^{2\psi(u)-2\gamma(u)}\{[n_1^{(+)}]^2 - [n_4^{(+)}]^2\} = 1 \tag{A4}$$

and, since $f[x^{(+)i}(u^\alpha)] = 0$ is an identity in u^α,

$$f_{,i}^{(+)}\partial x^{(+)i}/\partial u^\alpha = 0 = n_i^{(+)}\partial x^{(+)i}/\partial u^\alpha,$$

which is identically satisfied for all α except $\alpha = 1$, which gives

$$n_1^{(+)}(d\rho/du) + n_4^{(+)}(d\tau/du) = 0. \tag{A5}$$

We may use Eq. (10) to solve Eqs. (A4) and (A5) assuming $d\rho/du \neq 0$, finding

$$n_1^{(+)} = \pm e^{2\gamma(u)-2\psi(u)}(d\tau/du),$$
$$n_4^{(+)} = \mp e^{2\gamma(u)-2\psi(u)}(d\rho/du), \tag{A6}$$

where we leave the signs arbitrary.

In order to evaluate $\Omega_{\mu\nu}^{(+)}$, we must consider the derivative $n_{i,j}^{(+)}$, whereas Eqs. (A6) present $n_i^{(+)}$ as functions of u^α and not of $x^{(+)i}$. However, we may differentiate twice the identity $f[x^{(+)i}(u^\alpha)] = 0$, obtaining

$$n_{i,j}^{(+)}[\partial x^{(+)i}/\partial u^\mu][\partial x^{(+)j}/\partial u^\nu] = -n_i^{(+)}\,\partial^2 x^{(+)i}/\partial u^\nu\,\partial u^\mu.$$

Equation (9) then becomes

$$\Omega_{\mu\nu}^{(+)} = n_k^{(+)}\left[\frac{\partial^2 x^{(+)k}}{\partial u^\mu\,\partial u^\nu} + \Gamma_{ij}^k\frac{\partial x^{(+)i}}{\partial u^\mu}\frac{\partial x^{(+)j}}{\partial u^\nu}\right].$$

Or, explicitly, with the use of Eqs. (A6)

$$\Omega_{11}^{(+)} = n_k^{(+)}\left[\frac{d^2x^{(+)k}}{du^2} + \Gamma_{ij}^k\frac{dx^{(+)i}}{du}\frac{dx^{(+)j}}{du}\right],$$

$$\Omega_{22}^{(+)} = \mp e^{2\psi(u)}\left[\psi'(u)\frac{d\tau}{du} + \psi(u)\frac{d\rho}{du}\right],$$

$$\Omega_{33}^{(+)} = \pm e^{-2\psi(u)}\left\{\rho(u)^2\left[\psi'(u)\frac{d\tau}{du} + \dot{\psi}(u)\frac{d\rho}{du}\right] - \rho(u)\frac{d\tau}{du}\right\},$$

$$\Omega_{\mu\nu}^{(+)} = 0, \quad \text{for} \quad \mu \neq \nu.$$

The condition $\Omega_{\mu\nu}^{(-)} = \Omega_{\mu\nu}^{(+)}$ then yields, from Eqs. (A3), the three relations

$$n_k^{(+)}\left[\frac{d^2x^{(+)k}}{du^2} + \Gamma_{ij}^k\frac{dx^{(+)i}}{du}\frac{dx^{(+)j}}{du}\right] = 0, \tag{A7}$$

$$e^{2\psi(u)}\left[\psi'(u)\frac{d\tau}{du} + \psi(u)\frac{d\rho}{du}\right] = \mp kc_0 e^{\frac{1}{2}\sigma(u)},$$

$$e^{-2\psi(u)}\left\{\rho(u)\frac{d\tau}{du} - \rho(u)^2\left[\psi'(u)\frac{d\tau}{du} + \psi(u)\frac{d\rho}{du}\right]\right\} = \pm c_0 e^{\frac{1}{2}\sigma(u)}.$$

Using Eqs. (A1) and (A2), we may solve the last two of the above equations to get

$$\psi'(u)\frac{d\tau}{du} + \psi(u)\frac{d\rho}{du} = \mp e^{-\frac{1}{2}\sigma}kr_0^2/c_0. \tag{A8}$$

$$\frac{d\tau}{du} = \pm e^{\frac{1}{2}\sigma(u)}(1-2\alpha). \tag{A9}$$

We may also differentiate Eq. (A2) to obtain

$$d\psi(u)/du = \psi'(u)(d\rho/du) + \psi(u)(d\tau/du) = dg/2\,du. \tag{A10}$$

Finally, with the help of Eqs. (A1), (A2), and (A9), one may solve Eqs. (A8) and (A10) to obtain

$$\psi'(u) = (2\alpha-1)^{-1}e^{-\sigma}\left[\frac{\alpha}{c_0} + e^{2(\gamma-\psi+\sigma)}\frac{c_0}{2}\left(\frac{dg}{du}\right)^2\right], \tag{A11}$$

$$\psi(u) = \pm\tfrac{1}{2}e^{2\gamma-2\psi+\frac{1}{2}\sigma}\frac{dg}{du}. \tag{A12}$$

Note from Eq. (A9) that for $\alpha = \frac{1}{2}$ we have

$d\tau/du = 0$; and therefore from Eq. (10) $e^{2\psi(u)-2\gamma(u)} \leq 0$, which means that in this case ρ becomes a timelike coordinate, and the formalism fails.

What about $\frac{1}{2} < \alpha < 1$? If we desire $d\tau/du > 0$, we simply choose the lower signs in Eqs. (A6), (A8), and (A9). In principle this causes no difficulty. However, we then note that by Eq. (A6) we have $n_1^{(+)} < 0$, no matter what sign is chosen for $d\tau/du$.

This fact raises the question of how to decide whether or not a displacement δx^i originating on S projects into the vacuum or into the fluid. If δx^i is expressed in terms of the interior coordinates (1), we already know the answer: $\delta x^{(-)i} = \delta r \delta_1^i + \delta t \delta_4^i$ projects into the vacuum from S if and only if $\delta r > 0$, and into the fluid if and only if $\delta r < 0$. Since $n_i^{(-)} = (g_{11})^{\frac{1}{2}}\delta_i^1$, we may put the matter invariantly by saying that $\delta x^{(-)i}$ projects into the vacuum (fluid) if and only if $n_i^{(-)}\delta x^{(-)i} > 0 (< 0)$. We must also interpret $n_i^{(+)}$ in the same manner; otherwise we would have to posit $\Omega_{\mu\nu}^{(+)} = -\Omega_{\mu\nu}^{(-)}$. Thus we likewise conclude that a displacement $\delta x^{(+)i} = \delta\rho\delta_1^i + \delta\tau\delta_4^i$ projects into the vacuum (fluid) if and only if $n_i^{(+)}\delta x^{(+)i} > 0 (< 0)$. Now consider the special displacement $\delta y^{(+)i} = \delta\rho\delta_1^i$. We have $n_i^{(+)}\delta y^{(+)i} = n_1^{(+)}\delta\rho$, and thus if $n_1^{(+)} < 0$, $\delta y^{(+)i}$ will project into the *vacuum* if and only if $\delta\rho < 0$. Hence it appears that the *exterior* part of our universe might also turn out to be a cylinder of finite spacelike radius, covered by $0 \leq \rho < \rho(u)$. The final answer to this problem, however, would depend on the behavior of $g_{33} = \rho^2 e^{-2\psi(\rho,\tau)}$ as $\rho \to 0$ away from $\rho(u)$. As stated in the Introduction, the complete form of $\psi(\rho, \tau)$ and $\gamma(\rho, \tau)$ has not yet been discovered.

We now turn to consider the condition (A7), which will be satisfied if $\rho(u)$, $\tau(u)$ represent a geodesic; and since we are dealing with a surface S generated by steamlines of an incoherent fluid, this must be the case. In the frame (1), the trajectory $r(u) = r_0$, $t(u) = u$ is certainly a geodesic.

Further, if one were to use Eqs. (10), (A1), (A2), (A9), (A11), (A12), and (6) to evaluate the Christoffel symbols Γ_{ij}^k and the derivatives of $\rho(u)$ and $\tau(u)$ in terms of g and dg/du, a tedious but elementary calculation would show that

$$d^2x^{(+)m}/du^2 + \Gamma_{ij}^m[dx^{(+)i}/du][dx^{(+)j}/du] = 0$$

is identically satisfied for $m = 4$, and that $m = 1$ gives back the differential Eq. (4) with t replaced by u.

In this section we have thus proved that the necessary and sufficient conditions $\gamma_{\mu\nu}^{(+)} = \gamma_{\mu\nu}^{(-)}$, $\Omega_{\mu\nu}^{(+)} = \Omega_{\mu\nu}^{(-)}$ may be fulfilled by the use of the interior metric (1), for which S is expressed via $f[x^{(-)i}] = r - r_0 = 0$, and the exterior metric (5) with S represented via the above expressions for $\rho(u)$ and $\tau(u)$.

Covariant Conservation Laws for General Relativistic Exploding Matter

PHILIPPE H. DROZ-VINCENT
Centre National de la Recherche Scientifique, Paris and University of North Carolina, Chapel Hill, N. C.*
(Received 23 October 1965)

In this paper we attempt a general relativistic extension of some simple notions, e.g., total angular momentum, so far correctly defined only in the frame of special relativity. The mathematical aspect of the problem leads us to ask, given a tensor of vanishing divergence, how to construct an integral object, conserved in time and covariant.

A formalism based upon bitensors ensures the covariance, while the assumption of an exploding (or imploding) schema of matter seems to be the only available means to preserve conservation when the space–time is curved. This can be formulated in a general theorem then applied to different physical situations. The total linear momentum occurs as a vector fixed at the point of explosion (or implosion). Its length turns out (at least when a pure matter schema is concerned) to be superior to the total mass of the fluid. The excess appears as the energy carried by the explosion. The case of a so called "uncompressible" holonomic fluid gives a quite analogous result. In order to find a nontrivial angular momentum we also consider, in the last section, the case of a fluid possessing an intrinsic spin density. Both linear and angular momenta are conserved and covariantly defined. Moreover, they reduce to the conventional ones when the curvature vanishes.

INTRODUCTION

THE attempt of this paper is to point out how bilocal techniques, and especially geodesic parallel transport, permit one in certain cases to define integral conserved convariant quantities in the presence of an arbitrary Riemannian curvature. As an example of the difficulty of finding such quantities, consider the definition of the total linear momentum of a continuous distribution of matter. Insofar as we stay in a Minkowski space no problem occurs, consider the definition of the total linear momentum of a continuous distribution of matter. Insofar as we stay in a Minkowski space no problem occurs since: (a) The integral of a linear momentum distribution yields a covariant vector. (b) The vanishing of the divergence of the matter tensor ensures that the result of integration is a conserved quantity. (No external forces are involved.) As soon as curvature is taken into account, the definition of total linear momentum faces a two-step problem: 1. How can one add elementary vectors fixed at different points so as to obtain a simple covariant object under general coordinate transformation? 2. Even if the first step is passed, will the result be a conserved quantity?

The first problem can be easily solved by performing geodesic parallel transport to an arbitrary given point before adding the elementary vectors. According to the properties of parallelism, the result of this calculation will reduce to the usual Lorentz covariant vector in case of flat space. The second problem appears to have no solution in the general case. Nevertheless, we shall see that we can ensure conservation for a certain special type of matter distribution, viz. the exploding (or imploding) schema. The conserved integral vector will be a local vector fixed at the explosion point. For the sake of generality the mathematical formalism which yields these results is given in a form applicable not only to vectors, but also to tensors of any rank. This generalized formalism is also applied to the case of angular momentum. The orbital momentum turns out to vanish with respect to the imploding point when the usual pure matter schema is assumed. Therefore, in order to apply the general formalism to a nontrivial object, a fluid possessing an intrinsic spin density is examined. The Weyssenhoff–Raabe description of such a fluid is translated into Riemannian language, always maintaining the explosion assumption. In this way, a satisfactory definition of integral conserved momentum can be found. However, the Riemannian background is in this case given *a priori*; the fluid *is not* the source of curvature (Sec. IV). The physical meaning is that the fluid must be interpretated as a "test droplet" in the gravitational fluid produced by a certain external source. The droplet differs from the usual test body by two features: explosion and intrinsic spin.

Notation: In this paper ∇ denotes the covariant differentiation and the speed of light is equal to unity.

* On leave from Institut Poincaré, Paris, France.

I. MATHEMATICAL FORMALISM

Let us consider a differentiable manifold V_4, where an affinity and a volume element density $\eta_{\alpha\beta\gamma\delta}\, dx^\alpha \wedge dx^\beta \wedge dx^\gamma \wedge dx^\delta$ are supposedly defined.

The labels $\alpha, \beta, \gamma \cdots$ or $A, B, C \cdots$ refer to x while $\alpha', \beta', \gamma' \cdots$ or $A', B', C' \cdots$ refer to another point, say x'. As usual $\alpha, \beta, \gamma, \cdots$ run from 1 to 4 and $A, B, C \cdots$ are short for a certain set of Greek labels.

We shall use bitensors which are tensors at both x and x'.

It is well known that an integrand of the type

$$\omega^A = T^{\nu A}\, d\Sigma_\nu \tag{I.1}$$

(where $T^{\nu A}$ is a tensor and $d\Sigma_\nu$ the 3-dimensional element

$$d\Sigma_\nu = \eta_{\nu\alpha\beta\gamma}\, dx^\alpha \wedge dx^\beta \wedge dx^\gamma) \tag{I.2}$$

cannot be integrated over any three-dimensional surface in a covariant manner because of the presence of the label A. The geometrical reason for this is the impossibility of adding tensors taken at different points of a curved space. However, if $T^{\nu A}$ is merely a stress-energy tensor defined in a Riemannian manifold, B. S. DeWitt and R. Brehme have pointed out, that it is natural to consider, instead of $T^{\nu\mu}$, the transported quantities

$$\hat{T}^{\nu\mu'}_{(x,x')} = T^{\nu\rho} g^{\mu'}{}_{\rho}, \tag{I.3}$$

where $g^{\mu'}{}_{\rho}$ is the bitensor of geodesic parallel displacement.[1] Since $\hat{T}^{\nu\mu'}$ is a vector at the point x, it provides us with the covariant integral

$$\dot{I}^{\mu'}_{(x',\Sigma)} = \int_\Sigma \hat{T}^{\nu\mu'}\, d\Sigma_\nu, \tag{I.4}$$

which, if convergent, is a vector at the point x'. But when the space is curved, the vanishing divergence equation

$$\nabla_\nu T^{\nu\mu} = 0 \tag{I.5}$$

generally does not imply that $\nabla_\nu \hat{T}^{\nu\mu'}$ also vanishes.

Returning now to the general case of any tensor $T^{\nu A}$ given in a variety V_4 where a volume element is defined, it will be possible to build covariant integrals as follows:

Let $\theta = (\theta^{\alpha'}{}_{\beta})$ be any bitensor which reduces to $\delta^\alpha{}_\beta$ when x' tends to x, and $\theta^{-1} = (\theta^\rho{}_{\sigma'})$ the reciprocal of θ, i.e.,

$$\theta^{\alpha'}{}_{\beta}\theta^{\beta}{}_{\sigma'} = \delta^\alpha{}_\sigma, \tag{I.6}$$

$$\theta^\rho{}_{\sigma'}\theta^{\sigma'}{}_{\nu} = \delta^\rho{}_\nu. \tag{I.7}$$

If A involves p contravariant labels $\alpha_1 \cdots \alpha_p$ and q covariant labels $\beta_1 \cdots \beta_q$, let us take

$$\theta^{(p)} = (\theta^{\alpha'}{}_{\sigma_1}\theta^{\alpha'}{}_{\sigma_2} \cdots \theta^{\alpha'}{}_{\sigma_p}) \tag{I.8}$$

and

$$\theta^{(-q)} = (\theta^{\tau}{}_{\beta'_1}\theta^{\tau}{}_{\beta'_2} \cdots \theta^{\tau}{}_{\beta'_q}). \tag{I.9}$$

So, by the tensorial product, we get

$$\theta^{(p)} \otimes \theta^{(-q)} = (\theta^{A'}{}_{B}) \tag{I.10}$$

$$\theta^{\alpha_1'\cdots\alpha_p',\tau_1\cdots\tau_q}_{\sigma_1\cdots\sigma_p\beta_1'\cdots\beta_q'} = \theta^{\alpha_1'}_{\sigma_1} \cdots \theta^{\alpha_p'}_{\sigma_p}\theta^{\tau_1}_{\beta_1'} \cdots \theta^{\tau_q}_{\beta_q'}, \tag{I.11}$$

which reduces to $\delta^A{}_B$ when x' tends to x. Instead of $T^{\nu A}$ and ω^A we shall use

$$\hat{T}^{\nu A'}_{(x,x')} = T^{\nu B}\theta^{A'}{}_{B} \tag{I.12}$$

and

$$\hat{\omega}^{A'} = \omega^B \theta^{A'}{}_{B}, \tag{I.13}$$

which, being a scalar at the point x, yields the covariant integral

$$\dot{I}^{A'}_{(x',\Sigma)} = \int_\Sigma \hat{\omega}^{A'}{}_{(x,x')}. \tag{I.14}$$

Let us now assume that an affinity Γ is defined on V_4 and require that θ is the bitensor of parallel displacement along a set of given (not necessarily geodesic) curves joining each point x to the fixed x'.

Even remaining strongly undetermined, the choice of θ is restricted in a convenient manner, since in the case of a flat space, $\theta^{A'}{}_{B}$ reduces to the Kronecker $\delta^A{}_B$ within a special frame both at the points x and x'.

So $\dot{I}^{A'}$ appears as a natural generalization of the Lorentz covariant integrals we can build in Minkowskian space, and the indetermination in θ cancels out with a vanishing curvature. In all that follows let us suppose that V_4 is merely a Riemannian variety. Even if the null divergence condition

$$\nabla_\nu T^{\nu A} = 0 \tag{I.15}$$

is satisfied, with the most general Riemannian curvature, $\nabla_\nu \hat{T}^{\nu A'}$ is nonvanishing and $d\hat{\omega}^{A'} \neq 0$. In fact we have

$$\nabla_\nu \hat{T}^{\nu A'} = (\nabla_\nu T^{\nu B})\theta^{A'}{}_{B} + T^{\nu B}\nabla_\nu \theta^{A'}{}_{B}. \tag{I.16}$$

A sufficient condition to have a null divergence $\nabla_\nu \hat{T}^{\nu A'}$ if $\nabla_\nu T^{\nu A}$ vanishes,

$$T^{\nu B}\nabla_\nu \theta^{A'}{}_{C} = 0. \tag{I.17}$$

In Eq. (I.17) $\nabla_\nu \theta^{A'}{}_{C}$ involves linearly the covariant derivatives of $\theta^{\alpha'}{}_{\beta}$ and $\theta^\rho{}_{\sigma'}$. But, θ^{-1} being the inverse

[1] B. S. DeWitt and R. W. Brehme, Ann. of Phys. (N. Y.) 9, 220 (1960).

of θ, we have

$$\nabla_\nu \theta^\rho{}_{\sigma'} = -\theta^\rho{}_\alpha \cdot \theta^\beta{}_{\sigma'} \nabla_\nu \theta^{\alpha'}{}_\beta. \tag{I.18}$$

So, $\nabla_\nu \theta^{A'}{}_C$ is linear and homogeneous in the covariant derivatives $\nabla_\nu \theta^{\alpha'}{}_\beta$. Hence, a sufficient condition to have (I.17) is

$$T^{\nu\beta} \nabla_\nu \theta^{\alpha'}{}_\beta = 0. \tag{I.19}$$

Our purpose is to exhibit some particular choices of $T^{\nu A}$ which allow us to find θ such that (I.19) holds without any condition on the curvature.

An interesting type of $T^{\nu A}$

We consider now the case where $T^{\nu A}$ can be written as

$$T^{\nu A} = v^\nu T^A, \tag{I.20}$$

where the trajectories of v^ν (congruence of curves everywhere tangent to v^ν) are outgoing from (or incoming to) a common point. Let us take this common point as x', and θ as the bitensor of parallel displacement along the trajectories of v^ν. This means that, applying θ to any vector V^σ at the point x, we get, at the point x', the vector $V^{\sigma'}$ generated from V^σ by parallel displacement along the trajectory joining x to x' (insofar as the path is unique).

In other words, θ can be defined by integrating the differential system

$$v^\nu \nabla_\nu \theta^{\mu'}{}_\sigma = 0 \tag{I.21}$$

on each path, with the initial condition

$$\theta^\mu{}_{\sigma'(x,x'=x)} = \delta^\mu{}_\sigma \tag{I.22}$$

(cf. DeWitt, Brehme,[1] Lichnerowicz,[2] if geodesics). Eqs. (I.21) and (I.22) have been introduced by B. S. DeWitt[1] and used by Lichnerowicz[2] with geodesics but are now extended to any kind of path.

From (I.20) and (I.21) we obtain immediately

$$T^{\nu A} \nabla_\nu \theta^{\alpha'}{}_\beta = 0, \tag{I.23}$$

i.e., condition (I.19).

Theorem: If $\nabla_\nu T^{\nu A} = 0$, $\nabla_\nu \hat{T}^{\nu A'}$ also vanishes, the Gauss formula can be applied to (I.14) with any convenient hypersurface.

II. TOTAL LINEAR MOMENTUMS FOR OUTPOURING PURE MATTER

Let V_4 be the space–time of general relativity. The matter is described, as usual, by the stress-energy tensor

$$T^{\alpha\beta} = \rho u^\alpha u^\beta, \tag{II.1}$$

but we assume that the current lines (spacelike geodesics trajectories of u^α) are all outgoing from a singular point x'. (Such a pattern can be found, for instance, with an expanding universe reducing to a point at the initial time.) $T^{\alpha\beta}$ is of the type (I.20), and the trajectories of u^α being geodesics, the corresponding choice of $\theta^{\alpha'}{}_\beta$ is merely

$$\theta^{\alpha'}{}_\beta = g^{\alpha'}{}_\beta, \tag{II.2}$$

the geodesic parallel displacement tensor.

Since the divergence of $T^{\alpha\beta}$ is zero, the divergence of

$$\hat{T}^{\alpha\nu'} = T^{\alpha\beta} \theta^{\nu'}{}_\beta \tag{II.3}$$

also vanishes. The matter distribution is necessarily contained in the future of x'. We can assume that it is bounded by a 3-conoid K generated by timelike geodesics issuing from x'. The inside points of K situated between two spacelike hypersurfaces Σ_1 and Σ_2 intersecting K define a four-dimensional domain Ω to which the Gauss formula can be applied.

Since

$$u^\alpha \, d\Sigma_\alpha = 0 \tag{II.4}$$

over K, the timelike part of $\partial\Omega$ is irrelevant in

$$\int_{\partial\Omega} \hat{\omega}^{\nu'} = \int_{\partial\Omega} \hat{T}^{\alpha\nu'} \, d\Sigma_\alpha, \tag{II.5}$$

hence

$$\int_{\Sigma_1} \hat{T}^{\alpha\nu'} \, d\Sigma_\alpha = \int_{\Sigma_2} T^{\alpha\nu'} \, d\Sigma_\alpha \tag{II.6}$$

and

$$P^{\nu'}_{(x')} = \int_\Sigma \hat{T}^{\alpha\nu'} \, d\Sigma_\alpha \tag{II.7}$$

has the same value for all spacelike Σ intersecting the future of x'. The parallel displacement, being performed continuously along orientable timelike trajectories, maps any timelike vector future oriented at the point x onto a timelike vector future oriented at the point x'. Choosing u^α everywhere future oriented we get

$$\hat{T}^{\alpha\nu'} \, d\Sigma_\alpha \equiv \hat{\omega}^{\nu'} = \rho u^\alpha g^{\nu'}{}_\beta u^\beta \, d\Sigma_\alpha \tag{II.8}$$

as a timelike vector future oriented at the point x', and defined for each x (zero valued when $\overset{\frown}{x'x}$ is spacelike). The future timelike property is conserved by addition. Hence the integral (II.7) is a future timelike vector the length of which defines a positive proper mass P generally different from the inertial mass

$$M = \int_\Sigma \rho u^\alpha \, d\Sigma_\alpha. \tag{II.9}$$

[2] A. Lichnerowicz, *Propagateurs et commutateurs en Relativité Generale* (Institut des Hautes Etudes Scientifiques, Paris, 1961), p. 9.

It is noteworthy that the length of $P^{\nu'}$ cannot be inferior to M. In effect we must remember that, if timelike paths are compared in Minkowski space, the straight line joining two points is always *longer* than any other timelike line. This provides a pseudo-Euclidian triangular inequality: When two timelike vectors of the same orientation (i.e., either both future-oriented or both past-oriented) are added, we have

$$||\mathbf{A} + \mathbf{B}|| \geq ||\mathbf{A}|| + ||\mathbf{B}||. \tag{II.10}$$

Naturally, this property holds when the vectors are fixed at the point x' of space–time, since they lie in the same tangent Minkowski space. Of course, (II.10) remains valid for any number of timelike vectors, provided they have the same orientation, and hence convergent series or integrals can be considered.

Application of (II.10) to (II.7) yields

$$||\mathbf{P}|| = \left|\left|\int \hat{\omega}\right|\right| \geq \int ||\hat{\omega}||, \tag{II.11}$$

if $\hat{\omega}$ denotes the vector having the components $\hat{\omega}^{\nu'}$. But since $g^{\nu'}{}_{\beta}u^{\beta}$ is a unit vector

$$||\hat{\omega}|| = |\rho u^{\alpha}\, d\Sigma_{\alpha}|, \tag{II.12}$$

and hence $\int ||\hat{\omega}||$ is simply M. This yields

$$P \geq M. \tag{II.13}$$

Generally, $P - M$ is rigorously positive and represents the energy involved in the explosion. The excess of P over M is, of course, a feature of the explosion rather than a curvature effect, since $P - M$ differs from zero even in flat space.

III. TOTAL CURRENT RESULTANT FOR AN INCOMPRESSIBLE HOLONOMIC FLUID.

Let us first recall that a holonomic fluid is described by a stress-energy tensor

$$T_{\alpha\beta} = ru_{\alpha}u_{\beta} - t_{\alpha\beta}, \tag{III.1}$$

where r is a positive scalar, u_{α} is a unit vector, and $t_{\alpha\beta}$ is assumed such that we can write

$$(1/r)\nabla_{\alpha}t^{\alpha}{}_{\beta} = \partial_{\beta} \log F; \tag{III.2}$$

F is called the index of the fluid. From these assumptions it follows[3–5] that the current lines, trajectories of u^{α}, are geodesics of the conformal metric

$$\bar{g}_{\mu\nu}\, dx^{\mu}\, dx^{\nu} = d\bar{s}^2 = F^2\, ds^2. \tag{III.3}$$

[3] Eisenhart, Trans. Am. Math. Soc. **26**, 205 (1924).
[4] J. L. Synge, Proc. London Math. Soc. **43**, 376 (1937).
[5] A. Lichnerowicz, *Theories Relativistes de la Gravitation et de l'Electromagnetisme* (Masson, Paris, 1955), Chap. IV, pp. 71–75; Chap. V.

It is convenient[5] to introduce the current vectors $\mathbf{C}$ and $\bar{\mathbf{C}}$ colinear to $\mathbf{u}$.

$$C^{\alpha} = Fu^{\alpha}, \tag{III.4}$$

$$C_{\alpha} = \bar{C}_{\alpha} = Fu_{\alpha}, \tag{III.5}$$

$$\bar{C}^{\alpha} = \bar{g}^{\alpha\beta}\bar{C}_{\beta} = (1/F)u^{\alpha}. \tag{III.6}$$

Note that

$$\bar{C}^{\alpha}\bar{C}_{\alpha} = 1. \tag{III.7}$$

The trajectories of C being geodesics of $d\bar{s}$,

$$\bar{C}^{\alpha}\bar{\nabla}_{\alpha}\bar{C}^{\beta} = 0, \tag{III.8}$$

where $\bar{\nabla}$ is the covariant differentiation operator with respect to the Riemannian space defined by $d\bar{s}^2$.

We shall now assume that the fluid we deal with is incompressible, i.e.,[5]

$$\nabla_{\rho}C^{\rho} = 0. \tag{III.9}$$

This equation is equivalent to

$$\partial_{\rho}[(|g|)^{\frac{1}{2}}C^{\rho}] = 0. \tag{III.10}$$

So let us introduce J^{α} by

$$(|g|)^{\frac{1}{2}}C^{\alpha} = (|\bar{g}|)^{\frac{1}{2}}J^{\alpha}. \tag{III.11}$$

We have

$$\partial_{\rho}[(|\bar{g}|)^{\frac{1}{2}}J_{\rho}] = 0, \tag{III.12}$$

hence

$$\bar{\nabla}_{\rho}\, J^{\rho} = 0. \tag{III.13}$$

This formula is merely a means of putting (III.9) in terms of the affinity $\bar{\Gamma}$ resulting from $d\bar{s}^2$. Let us now consider the tensor $J^{\alpha}\bar{C}^{\beta}$ which presents a certain formal analogy with a pure-matter tensor. If we calculate its divergence by $\bar{\nabla}$ differentiation, we get

$$\bar{\nabla}_{\alpha}(J^{\alpha}\bar{C}^{\beta}) = \bar{\nabla}_{\alpha}J^{\alpha}\cdot C^{\beta} + J^{\alpha}\bar{\nabla}_{\alpha}\bar{C}^{\beta}. \tag{III.14}$$

The first term in the right-hand side cancels out from (III.13). The second term is proportional to $\bar{C}^{\alpha}\bar{\nabla}_{\alpha}\bar{C}^{\beta}$, since J^{α} is proportional to $\bar{C}^{\alpha}$. Thus, from (III.8) this term also vanishes and, for any incompressible holonomic fluid, the tensor $J^{\alpha}\bar{C}^{\beta}$ satisfies

$$\bar{\nabla}_{\alpha}(J^{\alpha}\bar{C}^{\beta}) = 0. \tag{III.15}$$

The situation is, mathematically, quite the same as within the pure-matter picutre. Respectively, $\bar{\nabla}_{\alpha}$, $\bar{g}_{\mu\nu}$, J^{μ}, and $\bar{C}^{\beta}$ play the role of ∇_{β}, $g_{\mu\nu}$, ρu^{μ}, and u^{β}. If we now introduce the topological assumption that all the current lines are outpouring from a point x', we can perform the geodesic parallel displacement of $\bar{C}$ with respect to the affinity of

$d\bar{s}^2$ and along the different paths $\overset{\frown}{xx'}$. So we obtain, for

$$j^{\alpha\mu'} = j^\alpha \bar{C}^\beta \bar{g}^{\mu'}{}_\beta \tag{III.16}$$

a vanishing $\overline{\nabla}$ divergence at each point x. As above, the matter distribution is contained inside of a three-dimensional conoid generated by timelike geodesics going from x' to the future. (Both ds^2 and $d\bar{s}^2$ define the same timelike directions, since $d\bar{s}^2$ is conformal to ds^2.) We have

$$j^\alpha \, d\Sigma_\alpha = 0 \tag{III.17}$$

over the conoid and the usual argument with Gauss theorem leads to the invariance of

$$I^{\mu'}_{(x')} = \int_\Sigma j^\alpha \bar{C}^\beta \bar{g}^{\mu'}{}_\beta \, d\Sigma_\alpha \tag{III.18}$$

with respect to the spacelike Σ intersecting the future of x'. For the same reason as $P^{\nu'}$ in the preceding section, $I^{\mu'}$ is a timelike vector at the point x'.

Remark: In the relativistic sense, the incompressibility of a fluid is not inconsistent with the existence of a point such as x', the only condition required being that (III.9) holds.

IV. ANGULAR MOMENTA

Momentum of a vector with respect to a point

The usual Lorentz covariant expression

$$m^{\alpha\beta} = v^\alpha (x^\beta - x'^\beta) - v^\beta (x^\alpha - x'^\alpha) \tag{IV.1}$$

giving the moment of a vector v^α with respect to any point x' can be formally generalized in a Riemannian space, since we have a natural generalization of $(x^\beta - x'^\beta)$. From the biscalar geodesic interval

$$s_{(x,x')} = \int_{x'}^{x} ds, \tag{IV.2}$$

calculated along the geodesic path joining x' to x, it is convenient to put

$$\sigma_{(x,x')} = \tfrac{1}{2} s^2 \tag{IV.3}$$

which provides the gradient

$$\sigma_\nu = \partial_\nu \sigma_{(x,x')}, \tag{IV.4}$$

studied in detail by B. S. DeWitt and R. Brehme.[1] This manifestly covariant object is a scalar at the point x' and a vector at the point x. As a vector it is tangent at x to the geodesic path $\overset{\frown}{x'x}$. When the space is merely Minkowskian, we get, in a special frame, after the labels have been raised,

$$\sigma^\nu = x^\nu - x'^\nu. \tag{IV.5}$$

So the looked-for generalization will be taken as

$$m^{\alpha\beta} = v^\alpha \sigma^\beta_{(x,x')} - v^\beta \sigma^\alpha_{(x,x')}. \tag{IV.6}$$

But if we apply this definition to the matter picture stated in Sec. II and x' being the point from which the world lines emanate, we obtain a vanishing result for the moment of the vector ρu^α. In effect u^α is, for each x, colinear to $\sigma^\alpha_{(xx')}$, since it is tangent to the geodesic world line.

An analogous trivial result is obtained for the incompressible fluid of Sec. III if we deal with the vector J^α and the $\dot{\sigma}^\beta$ deduced from $d\bar{s}^2$. Nevertheless, we can get from definition (IV.6) nontrivial results if we consider a fluid pattern possessing intrinsic spin. In view of describing such a fluid we can extend the Weyssenhoff's model[6] to the Riemannian universe as follows.

The fluid with intrinsic spin in a Riemannian universe

A momentum vector g^ν, noncolinear to the velocity u^μ, is assumed. The energy-momentum tensor

$$T^{\mu\nu} = u^\mu g^\nu \tag{IV.7}$$

is supposed to satisfy the four equations

$$\nabla_\mu T^{\mu\nu} = 0, \tag{IV.8}$$

that is to say

$$g^\nu \nabla_\mu u^\mu + u^\mu \nabla_\mu g^\nu = 0, \tag{IV.9}$$

which shows that the momentum vector g remains parallel to itself along any current lines (the trajectories of u^ν). With respect to any point ω the moment of energy–momentum tensor will be, according to definition (IV.6),

$$L^{\mu\nu\rho}_{(x,\omega)} = T^{\mu\nu}\sigma^\rho_{(x,\omega)} - T^{\mu\rho}\sigma^\nu_{(x,\omega)}, \tag{IV.10}$$

in other words

$$L^{\mu\nu\rho} = u^\mu (g^\nu \sigma^\rho_{(x,\omega)} - g^\rho \sigma^\nu_{(x,\omega)}). \tag{IV.11}$$

From (IV.8) the divergence of this is merely

$$\nabla_\mu L^{\mu\nu\rho} = u^\mu (g^\nu \tilde{u}^\rho - g^\rho \tilde{u}^\nu), \tag{IV.12}$$

where

$$\tilde{u}^\rho_{(x,\omega)} = u^\mu \nabla_\mu \sigma^\rho_{(x,\omega)} \tag{IV.13}$$

reduces to u^ρ in the flat case. As in the flat space of Weyssenhoff's theory, the nonvanishing divergence of L lends us to construct a total angular momentum

$$J^{\mu\nu\rho} = L^{\mu\nu\rho} + u^\mu S^{\nu\rho} \tag{IV.14}$$

[6] J. V. Wyssenhoff and Raabe, Acta Phys. Polon. 9, 8 (1947).

involving the skew-symmetric quantities $S^{\nu\rho}$ supposed to describe the intrinsic spin properties. But in the curved case, the divergence of L, containing $\tilde{u}$, is ω dependent. Hence, before assuming anything about the divergence of $\dot{J}$ we need to fix a particular choice of ω. Only when ω has been fixed can we introduce $S^{\nu\rho}(x, \omega)$ as a skew-metric tensor at the point x, and postulate

$$\nabla_\mu \dot{J}^{\mu\nu\rho} = 0, \tag{IV.15}$$

which comprises 10 independent equations. Having 4 equations in (IV.8) we need three more equations and shall take

$$S^{\nu\rho}_{(x,\omega)} u_\rho = 0. \tag{IV.16}$$

This formalism seems merely to be a translation of the Weyssenhoff theory into the Riemannian language, but here arises the necessity of specifying a particular point of the space-time. This situation becomes more meaningful if we now assume that all the current lines have a common point x' and that ω has been fixed identical to x'.

From (IV.11) and (IV.14) it is easy to see that

$$\dot{J}^{\mu\nu\rho}_{(x,x')} = u^\mu (g^\nu \sigma^\rho_{(x,x')} - g^\rho \sigma^\nu_{(x,x')} + S^{\nu\rho}_{(xx')}). \tag{IV.17}$$

Hence $J^{\mu\nu\rho}_{(xx')}$ appears as a particular case of the $T^{\nu A}$ tensor studied in Sec. I [(I.20) *et seq.*]. Since the divergence of $\dot{J}$ is null, the theorem given in Sec. I yields

$$\nabla_\mu \hat{\dot{J}}^{\mu\alpha'\beta'} = 0, \tag{IV.18}$$

where

$$\hat{\dot{J}}^{\mu\alpha'\beta'} = \dot{J}^{\mu\nu\rho} \theta^{\alpha'}_\nu \theta^{\beta'}_\rho \tag{IV.19}$$

and $\theta^{\alpha'}{}_\rho$ is the parallel displacement bitensor along the current lines (u-trajectories). On the other hand, Eqs. (IV.7) and (IV.8) exhibit $T^{\mu\nu}$ as another example of the $T^{\nu A}$ tensor to which the theorem of Sec. I can be applied. Then we have

$$\nabla_\mu \hat{T}^{\mu\alpha'} = 0 \tag{IV.20}$$

with

$$\hat{T}^{\mu\alpha'} = T^{\mu\nu} \theta^{\alpha'}{}_\nu. \tag{IV.21}$$

Using hypersurfaces intersecting in the future of x', the conoid which bounds the matter, we get from Gauss's theorem

$$\dot{J}^{\alpha'\beta'}_{(x')} = \int_\Sigma \hat{\dot{J}}^{\mu\alpha'\beta'} \, d\Sigma_\mu \tag{IV.22}$$

and

$$P^{\alpha'}_{(x')} = \int_\Sigma \hat{T}^{\mu\alpha'} \, d\Sigma_\mu \tag{IV.23}$$

as Σ invariant integrals.

In (IV.23), $T^{\mu\alpha'} d\Sigma_\mu$ is proportional to $g^\nu \theta^{\alpha'}{}_\nu$. Since g^ν is timelike and future oriented, this property also belongs to $g^\nu \theta^{\alpha'}{}_\nu$ and $P^{\alpha'}$. As in Sec. II, $P^{\alpha'}$ represents the total momentum of the fluid, but, here we have a nonzero total angular momentum $\dot{J}^{\alpha'\beta'}$.

CONCLUDING REMARKS

The basic assumption of the existance of a singular point such as x' may seem very restrictive. So it is natural to investigate whether one could take x' at infinity in the past. Unfortunately, the above results, so far, seem certain to fail with this modification. Hence, at this stage, the requirement of having integral conservation laws in curved space leads to the consideration of an instant, where all the matter of rhe universe is concentrated at a single point. For simplicity this point has been here chosen in the past (exploding schema) but the above calculations also hold with matter collapsing to a point in the future.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor B. S. DeWitt and Professor C. DeWitt for both their stimulating discussions and their hospitality at the Institute of Field Physics, University of North Carolina.

Electromagnetic Fields of Moving Dipoles and Multipoles

J. R. Ellis
*Bedford College, London, England**
(Received 23 September 1965)

A method of invariant Green's functions which is frequently used to find the fields of a moving charge is applied to evaluate the fields of dipoles and multipoles. Concise expressions are obtained from the integrals by successively integrating by parts. Two well-known methods for finding the radiation field on the world line of a charge are repeated for a dipole. It is found that a nonrotating (Fermi–Walker propagated) electric dipole has vanishing radiation field on the world line when it moves hyperbolically, as in the corresponding case for a charge. Radiation on the world lines of multipoles is also discussed (with particular reference to quadrupoles and octupoles), and the problem of evaluating the radiation reaction for a dipole is described within the context of the methods here given. It is further shown that a classical description of mass renormalization is possible to within the approximation of first-order terms, but not beyond.

INTRODUCTION

We devote our attention to a familiar part of classical electromagnetic theory dealing with moving point charges, dipoles, and multipoles, and in particular to a standard treatment for moving charges involving the use of invariant Green's functions. The latter method is frequently given in textbooks as a means of evaluating the potentials, fields, radiation, radiation reaction, and mass renormalization, for moving charges. The purpose of this article is to show that the same techniques can be used for the calculation of similar quantities for dipoles and multipoles.

Invariant Green's functions have an advantage over other methods in that they avoid altogether the ugly dependence of the retarded (or advanced) proper time on the field point in the differentiation of the potentials to find the fields. Another advantage which they clearly possess is demonstrated in the rapidity with which the value of the radiation field on the world line of a moving charge can be obtained by their use. The method is to be compared with the lengthier, explicit, and original method of Dirac which he gave some thirty years ago. The radiation on the world line has a close correspondence with the value of the force of reaction (we should not expect this to be entirely the case for a dipole) and the associated principle of (classical) mass renormalization is almost always described with the aid of Green's functions.

As far as the calculation of the potentials, fields, and radiation is concerned, the method possesses interesting possibilities of generalization to dipoles and multipoles which may not be widely known. They are the sort of quantities which should have been calculated; and, so far as the author is aware, only the potential of a dipole has been calculated by the method of Green's functions. The (concise) expression for the fields of a dipole which the Green's function method gives rise to is related to a similar form given by one other method (which, none the less, has a wide range of applications), while the "world line radiation field" (for want of a better word) does not seem to have appeared at all in the literature either by the original method of Dirac or by the method of Green's functions.

Although the connection between this radiation field and the force of reaction would not be so straightforward for a dipole (or multipole) as it is for a charge, it would seem a desirable quantity to calculate (especially for an electric dipole) either as a counterpart to the corresponding expression for a charge, or from the point of view of establishing whether it vanishes for hyperbolic motion as might seem plausible in certain circumstances in view of the fact that an electric dipole can be constructed from rigidly connected charges.

Before we can calculate the force and the torque of reaction for a moving dipole, we must possess adequate knowledge of the equations of motion. Different forms of such equations have been given in the literature but they seem to lack an interaction of the type $\dot{\mathbf{q}} \times \mathbf{H}$ (for an electric dipole of moment $\mathbf{q}$) which one might expect would exist (by analogy with $e\mathbf{V} \times \mathbf{H}$) together with the usual expression for the force $(\mathbf{q} \cdot \nabla)\mathbf{E}$ in the rest system. Such equations of motion can be obtained by assuming a first-order homogeneous Lagrangian in velocities through the moment tensor, so that the generalized momentum contains dipole moment terms in addition to the usual eA^{α} when a charge is present.

* Present address: University of Sussex, Falmer, Brighton, England.

It then appears that while the torque of reaction depends directly on the radiation field along the world line, the force of reaction depends upon this as well as its derivative, and the proceedure for finding the latter part can be given following the method of calculating the radiation field of a quadrupole.

Mass renormalization for the infinite part of the self-force of a dipole can be carried out without necessarily specifying the spin terms in the Lagrangian, but this is no longer possible for the torque of reaction and so we have only been able to discuss the situation as it appears for the force.

The first section deals very briefly with the Green's function (methods) for a moving charge; though the subject matter is very familiar, it is a convenient starting point. We describe the similar situations for dipoles and multipoles (including multipole expansions) in the later sections, leaving the problem of radiation reaction for a dipole to the end.

THE MOVING CHARGE

We begin[1] with Maxwell's equations for the electromagnetic field.[2] $F^{\mu\nu}(x)$ arising from the current density vector $j^\mu(x)$:

$$F^{\mu\nu}{}_{,\nu} = j^\mu, \qquad F^{*\mu\nu}{}_{,\nu} = 0,$$

where

$$F^{\mu\nu} = A^{\mu,\nu} - A^{\nu,\mu}.$$

Imposing the Lorentz gauge, $A^\mu{}_{,\mu} = 0$, the equations are equivalent to

$$\Box A^\mu = j^\mu \qquad (\Box \equiv \partial_\nu \partial^\nu). \tag{1}$$

If there is no incoming field, we have the solution[3]

$$\left.\begin{aligned} A^\mu(x) &= A^\mu_{\rm ret}(x) = \int D_{\rm ret}(x - x') j^\mu(x')\, d^4x' \\ \text{with}^4 \quad D_{\rm ret}(x) &= \frac{1}{2\pi}\,\theta(x^0)\,\delta(x^2) \end{aligned}\right\}. \tag{2}$$

For a point charge e moving along a given world line L, $z^\alpha = z^\alpha(\tau)$, we may substitute

$$j^\mu(x') = 4\pi e \int_{-\infty}^{\infty} \dot z^\mu(\tau)\,\delta^4(x' - z(\tau))\, d\tau \tag{3}$$

and integrate over x' to obtain

$$A^\mu_{\rm ret}(x) = 4\pi e \int_{-\infty}^{\infty} \dot z^\mu(\tau) D_{\rm ret}(x - z(\tau))\, d\tau; \tag{4}$$

and, on further differentiating and integrating by parts

writing $\quad x^\mu - z^\mu(\tau) \equiv l^\mu, \quad \dot z^\alpha(\tau)(x_\alpha - z_\alpha(\tau)) \equiv v\,,$

$$\begin{aligned} F^{\mu\nu}_{\rm ret}(x) &= A^{\mu,\nu}_{\rm ret}(x) - A^{\nu,\mu}_{\rm ret}(x), \\ &= 4\pi e \int_{-\infty}^{\infty} \dot z^\mu \frac{(d/d\tau) D_{\rm ret}(l)}{(d/d\tau) l^2} \cdot (l^2)^\nu\, d\tau \\ &\qquad - (\mu \leftrightarrow \nu) \\ &= 4\pi e \int_{-\infty}^{\infty} D_{\rm ret}(l) \frac{d}{d\tau} \left\{ \frac{\dot z^\mu l^\nu - \dot z^\nu l^\mu}{v} \right\} d\tau, \end{aligned} \tag{5}$$

where we have used

$$(d/d\tau) l^2 = -2v, \qquad (l^2)^\nu = 2l^\nu.$$

Evaluating (4) and (5) gives,[5] for $x^0 > z^0$,

$$\left.\begin{aligned} A^\mu_{\rm ret}(x) &= [e\dot z^\mu / v]_{\tau = \tau_r}, \\ F^{\mu\nu}_{\rm ret}(x) &= \left[\frac{e}{v} \frac{d}{d\tau} \left\{ \frac{\dot z^\mu l^\nu - \dot z^\nu l^\mu}{v} \right\} \right]_{\tau = \tau_r}, \end{aligned}\right\} \tag{6}$$

the potentials and the fields of an arbitrarily moving charge.

The radiation field $F^{\mu\nu}_{\rm rad} = F^{\mu\nu}_{\rm ret} - F^{\mu\nu}_{\rm adv}$ measured along the world line $x^\alpha = z^\alpha(\tau)$ has the value

$$F^{\mu\nu}_{\rm rad}(z(\tau)) = \tfrac{4}{3} e [\dot z^\mu (d^3 z^\nu / d\tau^3) - \dot z^\nu (d^3 z^\mu / d\tau^3)] \tag{7}$$

and was first obtained by Dirac.[6] This original method by Dirac was to evaluate $F^{\mu\nu}_{\rm ret}(x)$ given in (6)

---

[1] Gaussian units are used and the metric of special relativity is taken in the form $d\tau^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu$ with $g_{\mu\nu} = \text{diag}\,(1, -1, -1, -1)$ and $x^0 \equiv ct$. Greek suffixes take the values 0, 1, 2, 3. A comma or ∂ denotes differentiation with respect to the event x^α; differentiation with respect to τ is denoted by a dot. Duality is defined according to the definition

$$iF^{*\mu\nu} = \tfrac{1}{2} g^{\frac{1}{2}} \xi^{\mu\nu\alpha\beta} F^{\alpha\beta} \qquad (\text{or } iF^{*\mu\nu} = \tfrac{1}{2} g^{-\frac{1}{2}} \xi^{\mu\nu\alpha\beta} F_{\alpha\beta})$$

where $g = \det(g_{\mu\nu}) = -1$, and $\xi_{\mu\nu\alpha\beta}$, $\xi^{\mu\nu\alpha\beta}$ are alternating tensor densities having the numerical value of $\delta^{0123}_{\mu\nu\alpha\beta}$. The following notations are used:

$$x^\mu \equiv (ct, \mathbf{x}), \quad A^\mu \equiv (\phi, \mathbf{A}), \quad j^\mu \equiv (4\pi\rho, 4\pi \mathbf{j}/c),$$
$$(F^{01}, F^{02}, F^{03}) \equiv \mathbf{E}, \qquad (F^{23}, F^{31}, F^{12}) \equiv \mathbf{H}.$$

[2] Suffixes will be dropped in functional notation and also in products, thus $x^2 = x^\mu x_\mu$, $(xy) = x^\mu y_\mu$.

[3] See, for example, W. E. Thirring, *Principles of Quantum Electrodynamics* (Academic Press, Inc., New York, 1958), p. 19.

[4] $D_{\rm ret}(x)$ is the four-dimensional retarded Green's function; the θ function is defined by $\theta(x^0) = 1$, $x^0 > 0$, and $\theta(x^0) = 0$, $x^0 < 0$.

[5] Using

$$2\delta(l^2) = \delta(\tau - \tau_r)/|v|_{\tau=\tau_r} + \delta(\tau - \tau_a)/|v|_{\tau=\tau_a},$$

where τ_r, τ_a are retarded and advanced proper times where the world line L intersects the null cone drawn from the event x^α into the past and future, respectively.

[6] P. A. M. Dirac, Proc. Roy. Soc. (London) **A167**, 148 (1938).

at the point $x^\mu = z^\mu(\tau_0) + \gamma^\mu$, just "off" the world line, where the infinitesimal vector γ^μ was orthogonal to $\dot{z}^\mu(\tau_0)$ and $\gamma^2 = -\epsilon^2$. The result[7]

$$\begin{aligned} F_{\rm ret}(z(\tau_0)+\gamma) &= -e[1-\ddot{z}\gamma]^{-\frac{1}{2}}\{-\epsilon^{-3}\dot{z}^\mu\gamma^\nu, -\tfrac{1}{2}\epsilon^{-1}[1-\ddot{z}\gamma]^{-1}\ddot{z}^\mu\dot{z}^\nu, \\ &\quad + \tfrac{1}{2}\epsilon^{-1}(d^3z^\mu/d\tau^3)\gamma^\nu + \tfrac{1}{8}\epsilon^{-1}\ddot{z}^2\dot{z}^\mu\gamma^\nu \\ &\quad + \tfrac{2}{3}(d^3z^\mu/d\tau^3)\dot{z}^\nu\} - (\mu\leftrightarrow\nu) \end{aligned} \tag{8}$$

contains no even powers of ϵ. The value of $-F^{\mu\nu}_{adv}$ is obtained by changing ϵ to $-\epsilon$. Hence the value of $F^{\mu\nu}_{\rm rad}(z(\tau))$ is *finite* along the world line and equal to (7).

The other, shorter, method of calculating the radiation field along the world line of the moving charge,[8] which we referred to, is by replacing the retarded Green's function $D_{\rm ret}$ by the antisymmetric one[9] $D \equiv D_{\rm ret} - D_{\rm adv}$ in the integral (5) for the fields, with $x^\mu = z^\mu(\tau)$, and expanding of the integrand.

THE MOVING DIPOLE

Dipole fields are often described by Hertz vectors, and in fact we use the relation

$$A^\mu = \Pi^{\mu\nu}{}_{,\nu} \qquad (\Pi^{\mu\nu} = -\Pi^{\nu\mu}) \tag{9}$$

in the wave equation (1) and solve the equations

$$\Box\Pi^{\mu\nu} = j^{\mu\nu} \qquad (j^{\mu\nu} = -j^{\nu\mu}),$$

where $j^{\mu\nu}{}_{,\nu} = j^\mu$. The solution [cf. Eq. (2)]

$$\Pi^{\mu\nu}(x) = \Pi^{\mu\nu}_{\rm ret}(x) = \int D_{\rm ret}(x-x')j^{\mu\nu}(x')\,d^4x' \tag{10}$$

is applied to the case where $j^{\mu\nu}(x')$ arises solely from a dipole singularity along the world line:

$$j^{\mu\nu}(x') = 4\pi\int_{-\infty}^{\infty} p^{\mu\nu}(\tau)\,\delta^4(x'-z(\tau))\,d\tau. \tag{11}$$

We note that the conserved four-current then has the form

$$j^\mu(x) = 4\pi\int_{-\infty}^{\infty} p^{\mu\nu}(\tau)\,\partial_\nu\delta^4(x-z(\tau))\,d\tau. \tag{12}$$

The derivative of the delta function indicates the dipole character of the singularity. By substituting (11) in (10) and carrying out the integrations over x', we find

$$\Pi^{\mu\nu}_{\rm ret}(x) = 4\pi\int_{-\infty}^{\infty} p^{\mu\nu}(\tau)D_{\rm ret}(x-z(\tau))\,d\tau, \tag{13}$$

and so from (9)

$$\begin{aligned} A^\mu_{\rm ret}(x) &= 4\pi\int_{-\infty}^{\infty} p^{\mu\nu}\,\partial_\nu D_{\rm ret}(l)\,d\tau \\ &= 4\pi\int_{-\infty}^{\infty} p^{\mu\nu}\frac{\frac{d}{d\tau}D_{\rm ret}(l)}{\frac{d}{d\tau}(l^2)}\cdot(l^2)_{,\nu}\,d\tau \qquad (14) \\ &= 4\pi\int_{-\infty}^{\infty} D_{\rm ret}(l)\frac{d}{d\tau}\left\{\frac{p^{\mu\nu}l_\nu}{v}\right\}d\tau. \qquad (15) \end{aligned}$$

Differentiating under the integral sign in (14) we further have

$$\begin{aligned} F^{\mu\nu}_{\rm ret}(x) &= -4\pi\,\partial^\nu\int_{-\infty}^{\infty} p^{\mu\beta}l_\beta\left(\frac{1}{v}\frac{d}{d\tau}D_{\rm ret}(l)\right)d\tau \\ &\quad - (\mu\leftrightarrow\nu) \qquad (16) \\ &= 4\pi\int_{-\infty}^{\infty} p^{\mu\beta}l_\beta l^\nu\frac{1}{v}\frac{d}{d\tau}\left(\frac{1}{v}\frac{d}{d\tau}D_{\rm ret}(l)\right)d\tau \\ &\quad - 4\pi\int_{-\infty}^{\infty} p^{\mu\beta}\,\delta^\nu_\beta\left(\frac{1}{v}\frac{d}{d\tau}D_{\rm ret}(l)\right)d\tau - (\mu\leftrightarrow\nu) \\ &= 4\pi\int_{-\infty}^{\infty} D_{\rm ret}(l)\frac{d}{d\tau}\left\{\frac{1}{v}\frac{d}{d\tau}\left(\frac{p^{\mu\beta}l_\beta l^\nu}{v}\right)+\frac{p^{\mu\nu}}{v}\right\}d\tau \\ &\quad - (\mu\leftrightarrow\nu). \qquad (17) \end{aligned}$$

The latter part of the integrand in (16) which is enclosed within round brackets is a function of l^2, hence differentiation of this part by ∂^ν has been effected in the same way as in (14). Using the expression for $D_{\rm ret}$ in (2) we have for the potentials[10] from (15) assuming $x^0 > z^0$,

$$A^\mu_{\rm ret}(x) = \left[\frac{1}{v}\frac{d}{d\tau}\left\{\frac{p^{\mu\nu}l_\nu}{v}\right\}\right]_{\tau=\tau_r}, \tag{18}$$

---

[7] Commas separate terms of increasing order of infinitesimals.

[8] See, for example, Ref. 3, p. 23 and A. O. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (The Macmillan Company, New York, 1964), p. 188.

[9] $D(x) = (1/2\pi)\,\epsilon(x^0)\,\delta(x^2)$, where $\epsilon(x^0) = 1$, $x^0 > 0$, and $\epsilon(x^0) = -1$, $x^0 < 0$.

[10] A. Bialas, Acta Phys. Polon. 20, 831 (1961).

and for the fields[11] from (17)

$$F^{\mu\nu}_{\rm ret}(x) = \left[\frac{1}{v}\frac{d}{d\tau}\right. \times \left\{\frac{1}{v}\frac{d}{d\tau}\left(\frac{\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}l_\beta l^\gamma}{v}\right) + \frac{2p^{\mu\nu}}{v}\right\}\Bigg]_{\tau=\tau_r}. \qquad (19)$$

The expressions (18) and (19) have the obvious advantages of conciseness, and it is clear that it is with these expressions that one should seek to evaluate the radiation fields. However, the implicit differentiations can be carried out to give the expressions (first obtained by Bhabha and Corben[12] following a direct method of differentiation given by Bhabha[13,14]):

$$A^\mu_{\rm ret}(x) = [\{\dot p^{\mu\nu}k_\nu - p^{\mu\nu}(\ddot z k)k_\nu\}/v + \{p^{\mu\nu}k_\nu - p^{\mu\nu}\dot z_\nu\}/v^2]_{\tau=\tau_r}, \qquad (20)$$

$$F^{\mu\nu}_{\rm ret}(x) = \left[\frac{2P_1^{[\mu\nu]}}{v} + \frac{2Q_1^{[\mu\nu]}}{v^2} + \frac{2R_1^{[\mu\nu]}}{v^3}\right]_{\tau=\tau_r} \qquad (21)$$

with[15]

$$P_1^{\mu\nu} = \{\ddot p^\mu{}_\alpha - 3(\ddot z k)\dot p^\mu{}_\alpha + 3(\ddot z k)^2 p^\mu{}_\alpha - ((d^3z/d\tau^3)k)p^\mu{}_\alpha\}k^\alpha k^\nu,$$

$$Q_1^{\mu\nu} = \dot p^{\mu\nu} + 3\dot p^\mu{}_\alpha k^\alpha k^\nu - 4\dot p^\mu{}_\alpha k^{(\alpha}\dot z^{\nu)} - (\ddot z k)p^{\mu\nu} - 6(\ddot z k)p^\mu{}_\alpha k^\alpha k^\nu + 6(\ddot z k)p^\mu{}_\alpha k^{(\alpha}\dot z^{\nu)} - 2p^\mu{}_\alpha k^{(\alpha}\ddot z^{\nu)},$$

$$R_1^{\mu\nu} = p^{\mu\nu} + 3p^\mu{}_\alpha k^\alpha k^\nu - 6p^\mu{}_\alpha k^{(\alpha}\dot z^{\nu)} + 2p^\mu{}_\alpha \dot z^\alpha \dot z^\nu,$$

where $k^\mu = l^\mu/v$.

If we agree to write for the electric and magnetic four-moments,

$$q^\mu = p^{\mu\nu}\dot z_\nu, \qquad m^\mu = p^{\mu^*\nu}\dot z_\nu,$$

we may use the identity

$$p^{\mu\nu} = \delta^{\mu\nu}_{\alpha\beta}q^\alpha \dot z^\beta + \xi^{\mu\nu\alpha\beta}m_\alpha \dot z_\beta$$

to split $p^{\mu\nu}$ into electric and magnetic components. The corresponding "rest-moments" in the instantaneous Lorentz rest system $(-q^2)^{\frac{1}{2}}$, $(-m^2)^{\frac{1}{2}}$, are related by

$$q^\alpha q_\alpha - m^\alpha m_\alpha = \tfrac{1}{2}p^{\mu\nu}p_{\mu\nu},$$

and the angle between the moment vectors is given by

$$m^\alpha q_\alpha = \tfrac{1}{4}p^{\mu\nu}p_{\mu^*\nu}.$$

For a pure electric dipole we have[16]

$$p^{\mu\nu} = q^\mu \dot z^\nu - q^\nu \dot z^\mu, \qquad (22)$$

and the expressions (20) and (21) reduce to those previously given for an electric dipole constructed from rigidly connected charges.[17]

To gain information about the world line radiation field for a dipole we first use the original (explicit) method of Dirac[6] with our expression (19). The calculation is very lengthy and for this reason we will not reproduce it here. We here give only the result for $F^{\mu\mu}_{\rm ret}$ analogous to (8):

---

[11] See G. N. Ward, Proc. Roy. Soc. (London) **A279**, 562 (1964) for a derivation of a similar form of expression (19) here given, using analytical continuation. The method is an application of a technique of integrating Maxwell's equations based on the use of an integral identity. The equivalence of the two expressions can be seen by using the formula

$$\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}l_\beta l^\gamma = p^{\mu\nu}l_\beta l^\beta - \tfrac{1}{2}\delta^{\mu\nu\beta}_{\lambda\sigma\tau}p^{\lambda\sigma}l_\beta l^\tau$$

together with the identity

$$-\delta^{\mu\nu\beta}_{\lambda\sigma\tau} = \xi^{\mu\nu\beta\alpha}\xi_{\alpha\tau\lambda\sigma}.$$

[12] H. J. Bhabha and H. C. Corben, Proc. Roy. Soc. (London) **A178**, 273 (1941).

[13] H. J. Bhabha, Proc. Roy. Soc. (London) **A172**, 384 (1939).

[14] See also H. J. Bhabha, H. C. Corben, and Harish-Chandra, Proc. Roy. Soc. (London) **A185**, 250 (1946). Other works related to spinning charges with magnetic dipole moment include those of M. Mathisson, Proc. Cambridge Phil. Soc. **38**, 40 (1942); and A. Bialas, Acta Phys. Polon. **22**, 349, 499 (1962).

[15] Square brackets denote complete antisymmetrization: e.g.,

$$A_{[\alpha\beta\gamma]} = 1/3!\,\{A_{\alpha\beta\gamma} + A_{\beta\gamma\alpha} + A_{\gamma\alpha\beta} - A_{\beta\alpha\gamma} - A_{\alpha\gamma\beta} - A_{\gamma\beta\alpha}\};$$

round brackets, complete symmetrization: e.g.,

$$A_{(\alpha\beta\gamma)} = 1/3!\,\{A_{\alpha\beta\gamma} + A_{\beta\gamma\alpha} + A_{\gamma\alpha\beta} + A_{\beta\alpha\gamma} + A_{\alpha\gamma\beta} + A_{\gamma\beta\alpha}\}.$$

[16] G. N. Ward, Proc. Cambridge Phil. Soc. **61**, 547 (1965).

[17] J. R. Ellis, Proc. Cambridge Phil. Soc. **59**, 759 (1963). Equation (22) provides another means of calculating the fields in vector form. We use

$$(p^{01}, p^{02}, p^{03}) \equiv -\beta\mathbf{n}, \qquad (p^{23}, p^{31}, p^{12}) \equiv \beta\mathbf{n}\times\mathbf{V}/c$$

since $\mathbf{n} = \mathbf{q} - q\,\mathbf{V}/c$ is the relative vector moment of the dipole. However, the calculations are still somewhat lengthy. Recent observations by G. N. Ward show that the formulas previously given for the fields **E** and **H** are valid also in the case of time-varying rest-moment $M(t)$; in this case the relative vector moment occurring in these expressions is taken as

$$\mathbf{n} = M(t)\hat{\mathbf{n}}[1 - |\mathbf{V}|^2/c^2]^{\frac{1}{2}}/[1 - |\hat{\mathbf{n}}\times\mathbf{V}|^2/c^2]^{\frac{1}{2}}.$$

$$
\begin{aligned}
F^{\mu\nu}_{\text{adv}}(z(\tau_0)+\gamma) = [1-\ddot{z}\gamma]^{-\frac{1}{2}}\{&3\epsilon^{-5}p^\mu{}_\alpha\gamma^\alpha\gamma^\nu - \epsilon^{-3}[1-\ddot{z}\gamma]^{-1}p^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu \\
&+\epsilon^{-3}p^{\mu\nu}, +2\epsilon^{-3}[1-\ddot{z}\gamma]^{-1}\dot{p}^\mu{}_\alpha\gamma^{(\alpha}\dot{z}^{\nu)} \\
&+\epsilon^{-3}[1-\ddot{z}\gamma]^{-1}p^\mu{}_\alpha\gamma^{(\alpha}\ddot{z}^{\nu)}, -\tfrac{1}{2}\epsilon^{-3\mu}\dot{p}_\alpha\gamma^\alpha\gamma^\nu \\
&-\tfrac{1}{8}\epsilon^{-3}\ddot{z}^2p^\mu{}_\alpha\gamma^\alpha\gamma^\nu + \epsilon^{-3}((d^3z/d\tau^3)\gamma)p^\mu{}_\alpha\gamma^{(\alpha}\dot{z}^{\nu)} + \tfrac{3}{2}\epsilon^{-1}\dot{p}^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu - \tfrac{1}{2}\epsilon^{-1}\ddot{p}^{\mu\nu} \\
&+3\epsilon^{-1}\dot{p}^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} + \tfrac{3}{4}\epsilon^{-1}p^\mu{}_\alpha\ddot{z}^\alpha\ddot{z}^\nu + \epsilon^{-1}p^\mu{}_\alpha\dot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) + \tfrac{5}{8}\epsilon^{-1}\ddot{z}^2p^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu \\
&-\tfrac{1}{8}\epsilon^{-1}\ddot{z}^2p^{\mu\nu}, -\tfrac{1}{2}\epsilon^{-3}(\ddot{z}\gamma)\ddot{p}^\mu{}_\alpha\gamma^\alpha\gamma^\nu - \tfrac{1}{2}\epsilon^{-3}((d^3z/d\tau^3)\gamma)\dot{p}^\mu{}_\alpha\gamma^\alpha\gamma^\nu \\
&-\tfrac{1}{8}\epsilon^{-3}((d^4z/d\tau^4)\gamma)p^\mu{}_\alpha\gamma^\alpha\gamma^\nu - \tfrac{1}{4}\epsilon^{-3}\ddot{z}^2(\ddot{z}\gamma)p^\mu{}_\alpha\gamma^\alpha\gamma^\nu + 2\epsilon^{-3}(\ddot{z}\gamma)((d^3z/d\tau^3)\gamma)p^\mu{}_\alpha\gamma^{(\alpha}\dot{z}^{\nu)} \\
&-\epsilon^{-1}(d^3p^\mu{}_\alpha/d\tau^3)\gamma^{(\alpha}\dot{z}^{\nu)} - \tfrac{3}{2}\epsilon^{-1}\ddot{p}^\mu{}_\alpha\gamma^{(\alpha}\ddot{z}^{\nu)} - \epsilon^{-1}\dot{p}^\mu{}_\alpha\gamma^{(\alpha}(d^3z^{\nu)}/d\tau^3) - \tfrac{1}{4}\epsilon^{-1}p^\mu{}_\alpha\gamma^{(\alpha}(d^4z^{\nu)}/d\tau^4) \\
&+3\epsilon^{-1}(\ddot{z}\gamma)\ddot{p}^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu - \tfrac{1}{2}\epsilon^{-1}(\ddot{z}\gamma)\ddot{p}^{\mu\nu} - \tfrac{5}{4}\epsilon^{-1}\ddot{z}^2\dot{p}^\mu{}_\alpha\gamma^{(\alpha}\dot{z}^{\nu)} \\
&+6\epsilon^{-1}(\ddot{z}\gamma)\dot{p}^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} + \tfrac{5}{2}\epsilon^{-1}((d^3z/d\tau^3)\gamma)\dot{p}^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu - \tfrac{1}{2}\epsilon^{-1}((d^3z/d\tau^3)\gamma)\dot{p}^{\mu\nu} \\
&+\tfrac{15}{8}\epsilon^{-1}\ddot{z}^2(\ddot{z}\gamma)p^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu - \tfrac{1}{4}\epsilon^{-1}\ddot{z}^2(\ddot{z}\gamma)p^{\mu\nu} + \tfrac{5}{8}\epsilon^{-1}((d^4z/d\tau^4)\gamma)p^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu \\
&-\tfrac{1}{8}\epsilon^{-1}((d^4z/d\tau^4)\gamma)p^{\mu\nu} - \tfrac{5}{8}\epsilon^{-1}\ddot{z}^2p^\mu{}_\alpha\gamma^{(\alpha}\ddot{z}^{\nu)} + \tfrac{3}{2}\epsilon^{-1}(\ddot{z}\gamma)p^\mu{}_\alpha\ddot{z}^\alpha\ddot{z}^\nu \\
&+2\epsilon^{-1}(\ddot{z}\gamma)p^\mu{}_\alpha\dot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) + \tfrac{5}{2}\epsilon^{-1}((d^3z/d\tau^3)\gamma)p^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} - \tfrac{5}{4}\epsilon^{-1}(\ddot{z}(d^3z/d\tau^3))p^\mu{}_\alpha\gamma^{(\alpha}\dot{z}^{\nu)} \\
&-\tfrac{4}{3}(d^3p^\mu{}_\alpha/d\tau^3)\dot{z}^\alpha\dot{z}^\nu + \tfrac{1}{3}(d^3p^{\mu\nu}/d\tau^3) - 4\ddot{p}^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} - 2\dot{p}^\mu{}_\alpha\ddot{z}^\alpha\ddot{z}^\nu \\
&-\tfrac{8}{3}\dot{p}^\mu{}_\alpha\dot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) - \tfrac{4}{3}p^\mu{}_\alpha\ddot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) - \tfrac{2}{3}p^\mu{}_\alpha\dot{z}^{(\alpha}(d^4z^{\nu)}/d\tau^4) - 2\ddot{z}^2\dot{p}^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu \\
&+\tfrac{1}{3}\ddot{z}^2\dot{p}^{\mu\nu} - 2\ddot{z}^2p^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} - 2(\ddot{z}(d^3z/d\tau^3))p^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu + \tfrac{1}{3}(\ddot{z}(d^3z/d\tau^3))p^{\mu\nu}\} - (\mu\leftrightarrow\nu). \qquad (23)
\end{aligned}
$$

As before, changing ϵ to $-\epsilon$ everywhere yields $-F^{\mu\nu}_{\text{adv}}$, and because (remarkably perhaps) only odd powers of ϵ occur in the above expression, the radiation field for a dipole is *finite* just as for a moving charge. We have

$$
\begin{aligned}
F^{\mu\nu}_{\text{rad}}(z(\tau)) = &\tfrac{2}{3}(d^3p^{\mu\nu}/d\tau^3) - \tfrac{8}{3}(d^3p^\mu{}_\alpha/d\tau^3)\dot{z}^\alpha\dot{z}^\nu \\
&-8\ddot{p}^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} - 4\dot{p}^\mu{}_\alpha\ddot{z}^\alpha\ddot{z}^\nu \\
&-\tfrac{16}{3}\dot{p}^\mu{}_\alpha\dot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) - \tfrac{8}{3}p^\mu{}_\alpha\ddot{z}^{(\alpha}(d^3z^{\nu)}/d\tau^3) \\
&-\tfrac{4}{3}p^\mu{}_\alpha\dot{z}^{(\alpha}(d^4z^{\nu)}/d\tau^4) \\
&-4\ddot{z}^2\dot{p}^\mu{}_\alpha\dot{z}^\alpha\dot{z}^\nu + \tfrac{2}{3}\ddot{z}^2\dot{p}^{\mu\nu} - 4\ddot{z}^2p^\mu{}_\alpha\dot{z}^{(\alpha}\ddot{z}^{\nu)} \\
&-4(\ddot{z}(d^3z/d\tau^3))p^\mu_\alpha\dot{z}^\alpha\dot{z}^\nu + \tfrac{2}{3}(\ddot{z}(d^3z/d\tau^3))p^{\mu\nu} \\
&-(\mu\leftrightarrow\nu). \qquad (24)
\end{aligned}
$$

The alternative method of calculating the radiation field which we referred to earlier[8] may also be used to obtain the radiation field. It is instructive to give this now in detail since the method admits many possibilities for generalization to multipoles. We start from (17) where D now replaces D_{ret}:

$$
\begin{aligned}
F^{\mu\nu}_{\text{rad}}(x) = 4\pi\int_{-\infty}^{\infty} D(x-z(\tau))\,\frac{d}{d\tau}\Bigg\{&\frac{1}{v(\tau,x)}\,\frac{d}{d\tau} \\
&\times\left(\frac{\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}(\tau)(x_\beta - z_\beta(\tau))(x^\gamma - z^\gamma(\tau))}{v(\tau,x)}\right) \\
&+\frac{2p^{\mu\nu}(\tau)}{v(\tau,x)}\Bigg\}\,d\tau. \qquad (25)
\end{aligned}
$$

We evaluate this for a point on the world line itself: $x^\alpha = z^\alpha(0)$, and we require the following expansions

$$
\begin{aligned}
z(0)-z(\tau) &= -\tau\dot{z}, -\tfrac{1}{2}\tau^2\ddot{z}, -\tfrac{1}{6}\tau^3(d^3z/d\tau^3), \\
&\quad -\tfrac{1}{24}\tau^4(d^4z/d\tau^4), -\cdots, \\
p(\tau) &= p, +\tau\dot{p}, +\tfrac{1}{2}\tau^2\ddot{p}, \\
&\quad +\tfrac{1}{6}\tau^3(d^3p/d\tau^3), +\cdots, \\
\dot{z}(\tau) &= \dot{z}, +\tau\ddot{z}, +\tfrac{1}{2}\tau^2(d^3z/d\tau^3), \\
&\quad +\tfrac{1}{6}\tau^3(d^4z/d\tau^4), +\cdots, \\
1/v(\tau, z(0)) &= -\tau^{-1}\{1,, +\tfrac{1}{6}\tau^2\ddot{z}^2, \\
&\quad +\tfrac{5}{24}\tau^3(\ddot{z}(d^3z/d\tau^3)), +\cdots\}.
\end{aligned} \qquad (26)
$$

The z's and p's (without argument) here refer to $\tau = 0$. Since

$$
(z(0)-z(\tau))^2 = \tau^2,, -\tfrac{1}{12}\tau^4\ddot{z}^2, -\tfrac{1}{12}\tau^5(\ddot{z}(d^3z/d\tau^3)), +\cdots
$$

and

$$
\epsilon(z(0)-z(\tau)) = -\epsilon(\tau),
$$

we have for the Green's function[9]

$$
D(z(0)-z(\tau)) = -\frac{1}{2\pi}\,\epsilon(\tau)\,\delta\{\tau^2(1,, -\tfrac{1}{12}\tau^2\ddot{z}^2, -\tfrac{1}{12}\tau^3(\ddot{z}(d^3z/d\tau^3)), +\cdots)\}. \qquad (27)
$$

An expression of the form $\epsilon(\tau)\delta\{\tau^2 g(\tau)\}$, where g is any function of τ, without zeros, can be reduced quite easily:

$$\begin{aligned}
&\epsilon(\tau)\delta\{\tau^2 g(\tau)\} \\
&\quad = \epsilon(\tau)\lim_{h\to 0}\delta\{(\tau^2 - h^2)g(\tau)\}, \\
&\quad = \epsilon(\tau)\left\{\lim_{h\to 0}\frac{\delta(\tau - h)}{\left|\frac{d}{d\tau}\{(\tau^2 - h^2)g(\tau)\}\right|_{\tau=h}}\right. \\
&\qquad \left. + \lim_{h\to 0}\frac{\delta(\tau + h)}{\left|\frac{d}{d\tau}\{(\tau^2 - h^2)g(\tau)\}\right|_{\tau=-h}}\right\} \\
&\quad = \lim_{h\to 0}\left\{\frac{\epsilon(\tau)\delta(\tau - h)}{|2\tau g(\tau)|} + \frac{\epsilon(\tau)\delta(\tau + h)}{|2\tau g(\tau)|}\right\}, \\
&\quad = \lim_{h\to 0}\left\{\frac{\delta(\tau - h)}{2\tau\,|g(\tau)|} + \frac{\delta(\tau + h)}{2\tau\,|g(\tau)|}\right\}, \\
&\quad = -\frac{\delta'(\tau)}{|g(\tau)|}.
\end{aligned}$$

Hence (27) becomes

$$D(z(0) - z(\tau)) = \frac{1}{2\pi}\,\delta'(\tau)\{1,, +\tfrac{1}{12}\tau^2\ddot{z}^2, \\ +\tfrac{1}{12}\tau^3(\ddot{z}(d^3z/d\tau^3)), + \cdots\}. \qquad (28)$$

The integrand in (25) is now expanded to four orders by means of (26) and (28):

$$\begin{aligned}
F^{\mu\nu}_{\mathrm{rad}}(z(0)) = 2\int_{-\infty}^{\infty}\delta'(\tau) \\
\times \{1,, +\tfrac{1}{12}\tau^2\ddot{z}^2, +\tfrac{1}{12}\tau^3(\ddot{z}(d^3z/d\tau^3))\} \\
\times \{A^{\mu\nu}\tau^{-2},, +B^{\mu\nu}, +C^{\mu\nu}\tau\}\,d\tau, \qquad (29)
\end{aligned}$$

where, $A^{\mu\nu}$, $B^{\mu\nu}$, $C^{\mu\nu}$ are certain constants (there is no term in τ^{-1}). It is then clear that the world line radiation is finite and equal to

$$F^{\mu\nu}_{\mathrm{rad}}(z(0)) = -2C^{\mu\nu} - \tfrac{1}{6}(\ddot{z}(d^3z/d\tau^3))A^{\mu\nu}, \qquad (30)$$

and it remains to evaluate $A^{\mu\nu}$ and $C^{\mu\nu}$.

We have

$$\begin{aligned}
&\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}(\tau)(z_\beta(0) - z_\beta(\tau))(z^\gamma(0) - z^\gamma(\tau)) \\
&\qquad = \tau^2\{a^{[\mu\nu]}, +\tau b^{[\mu\nu]}, +\tau^2 c^{[\mu\nu]}, +\tau^3\,d^{[\mu\nu]}\},
\end{aligned}$$

where

$$\begin{aligned}
\tfrac{1}{2}a^{\mu\nu} &= p^\mu{}_\beta\dot{z}^\beta\dot{z}^\nu, \\
\tfrac{1}{2}b^{\mu\nu} &= \dot{p}^\mu{}_\beta\dot{z}^\beta\dot{z}^\nu + p^\mu{}_\beta\ddot{z}^{(\beta}\dot{z}^{\nu)}, \\
\tfrac{1}{2}c^{\mu\nu} &= \tfrac{1}{2}\ddot{p}^\mu{}_\beta\dot{z}^\beta\dot{z}^\nu + \dot{p}^\mu{}_\beta\ddot{z}^{(\beta}\dot{z}^{\nu)} \\
&\quad + \tfrac{1}{4}p^\mu{}_\beta\ddot{z}^\beta\ddot{z}^\nu + \tfrac{1}{3}p^\mu{}_\beta(d^3z^{(\beta}/d\tau^3)\dot{z}^{\nu)}, \\
\tfrac{1}{2}d^{\mu\nu} &= \tfrac{1}{6}(d^3p^\mu{}_\beta/d\tau^3)\dot{z}^\beta\dot{z}^\nu + \tfrac{1}{2}\ddot{p}^\mu{}_\beta\ddot{z}^{(\beta}\dot{z}^{\nu)} \\
&\quad + \tfrac{1}{4}\dot{p}^\mu{}_\beta\ddot{z}^\beta\ddot{z}^\nu + \tfrac{1}{3}\dot{p}^\mu{}_\beta(d^3z^{(\beta}/d\tau^3)\dot{z}^{\nu)} \\
&\quad + \tfrac{1}{12}p^\mu{}_\beta(d^4z^{(\beta}/d\tau^4)\dot{z}^{\nu)} + \tfrac{1}{6}p^\mu{}_\beta(d^3z^{(\beta}/d\tau^3)\ddot{z}^{\nu)},
\end{aligned}$$

so that

$$\begin{aligned}
&\frac{d}{d\tau}\left(\frac{\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}(\tau)(z_\beta(0) - z_\beta(\tau))(z^\gamma(0) - z^\gamma(\tau))}{v(\tau, z(0))}\right) + 2p^{\mu\nu}(\tau) \\
&\qquad = (2p^{\mu\nu} - a^{[\mu\nu]}) + (2\dot{p}^{\mu\nu} - 2b^{[\mu\nu]})\tau \\
&\qquad\quad + (\ddot{p}^{\mu\nu} - 3c^{[\mu\nu]} - \tfrac{1}{2}\ddot{z}^2 a^{[\mu\nu]})\tau^2 \\
&\qquad\quad + (\tfrac{1}{3}(d^3p^{\mu\nu}/d\tau^3) - 4d^{[\mu\nu]} \\
&\qquad\quad - \tfrac{2}{3}\ddot{z}^2 b^{[\mu\nu]} - \tfrac{5}{6}(\ddot{z}(d^3z/d\tau^3))a^{[\mu\nu]})\tau^3.
\end{aligned}$$

In the product of this expression with $1/v(\tau, z(0))$, the constant $C^{\mu\nu}$ is twice the coefficient of τ^2, and $A^{\mu\nu}$ is minus the coefficient of τ^{-1}. Hence the radiation field (30) at the point $\tau = 0$ on the world line of the dipole is given by

$$\begin{aligned}
F^{\mu\nu}_{\mathrm{rad}}(z(0)) = &-2\{-\tfrac{2}{3}(d^3p^{\mu\nu}/d\tau^3) + 8d^{[\mu\nu]} + \tfrac{4}{3}\ddot{z}^2 b^{[\mu\nu]} \\
&+ \tfrac{5}{3}(\ddot{z}(d^3z/d\tau^3))a^{[\mu\nu]} - \tfrac{1}{3}\ddot{z}^2(2\dot{p}^{\mu\nu} - 2b^{[\mu\nu]}) \\
&- \tfrac{5}{12}(\ddot{z}(d^3z/d\tau^3))(2p^{\mu\nu} - a^{[\mu\nu]})\} \\
&- \tfrac{1}{6}(\ddot{z}(d^3z/d\tau^3))\{2p^{\mu\nu} - a^{[\mu\nu]}\},
\end{aligned}$$

and this again yields the expression (24).

The electric dipole.

In the special case when $p^{\mu\nu} = q^\mu\dot{z}^\nu - q^\nu\dot{z}^\mu$ with q^μ orthogonal to $\dot{z}^\mu$ [Eq. (22)] we have an electric dipole. This may be regarded as having been constructed from rigidly connected charges. The potentials and fields have been calculated[17] and the radiation field (24) for this particular case can be obtained in terms of q^μ and $\dot{z}^\mu$ by substituting the following:

$$\begin{aligned}
\tfrac{1}{2}p^{\mu\beta} &= q^{[\mu}\dot{z}^{\beta]}, \\
\tfrac{1}{2}\dot{p}^{\mu\beta} &= \dot{q}^{[\mu}\dot{z}^{\beta]} + q^{[\mu}\ddot{z}^{\beta]}, \\
\tfrac{1}{2}\ddot{p}^{\mu\beta} &= \ddot{q}^{[\mu}\dot{z}^{\beta]} + 2\dot{q}^{[\mu}\ddot{z}^{\beta]} + q^{[\mu}(d^3z^{\beta]}/d\tau^3), \\
\tfrac{1}{2}(d^3p^{\mu\beta}/d\tau^3) &= (d^3q^{[\mu}/d\tau^3)\dot{z}^{\beta]} + 3\ddot{q}^{[\mu}\ddot{z}^{\beta]} \\
&\quad + 3\dot{q}^{[\mu}(d^3z^{\beta]}/d\tau^3) + q^{[\mu}(d^4z^{\beta]}/d\tau^4), \\
p^{\mu\beta}\dot{z}_\beta &= q^\mu, \\
p^{\mu\beta}\ddot{z}_\beta &= -(q\ddot{z})\dot{z}^\mu, \\
p^{\mu\beta}(d^3z_\beta/d\tau^3) &= -\ddot{z}^2 q^\mu - (q(d^3z/d\tau^3))\dot{z}^\mu, \\
p^{\mu\beta}(d^4z_\beta/d\tau^4) &= -3(\ddot{z}(d^3z/d\tau^3))q^\mu - (q(d^4z/d\tau^4))\dot{z}^\mu, \\
\dot{p}^{\mu\beta}\dot{z}_\beta &= \dot{q}^\mu - (\dot{q}\dot{z})\dot{z}^\mu, \\
\dot{p}^{\mu\beta}\ddot{z}_\beta &= -(\dot{q}\ddot{z})\dot{z}^\mu + \ddot{z}^2 q^\mu - (q\ddot{z})\ddot{z}^\mu,
\end{aligned}$$

$$\dot{p}^{\mu\beta}(d^3z_\beta/d\tau^3) = -\ddot{z}^2\dot{q}^\mu - (\dot{q}(d^3z/d\tau^3))\dot{z}^\mu + (\ddot{z}(d^3z/d\tau^3))q^\mu - (q(d^3z/d\tau^3))\ddot{z}^\mu,$$

$$\ddot{p}^{\mu\beta}\dot{z}_\beta = \ddot{q}^\mu - (\ddot{q}\dot{z})\dot{z}^\mu - 2(\dot{q}\dot{z})\ddot{z}^\mu - \ddot{z}^2q^\mu,$$

$$\ddot{p}^{\mu\beta}\ddot{z}_\beta = -(\ddot{q}\ddot{z})\dot{z}^\mu + 2\ddot{z}^2\dot{q}^\mu - 2(\dot{q}\ddot{z})\ddot{z}^\mu + (\ddot{z}(d^3z/d\tau^3))q^\mu - (q\ddot{z})(d^3z^\mu/d\tau^3),$$

$$(d^3p^{\mu\beta}/d\tau^3)\dot{z}_\beta = (d^3q^\mu/d\tau^3) - ((d^3q/d\tau^3)\dot{z})\dot{z}^\mu - 3(\ddot{q}\ddot{z})\ddot{z}^\mu - 3\ddot{z}^2\dot{q}^\mu - 3(\dot{q}\dot{z})(d^3z^\mu/d\tau^3) - 3(\ddot{z}(d^3z/d\tau^3))q^\mu.$$

After collecting and rearranging terms, we find

$$F^{\mu\nu}_{\text{rad}}(z(0)) = -\tfrac{4}{3}(d^4z^{[\mu}/d\tau^4)q^{\nu]} - \tfrac{4}{3}\ddot{z}^2\dot{z}^{[\mu}q^{\nu]} - \tfrac{4}{3}(\ddot{z}(d^3z/d\tau^3))\dot{z}^{[\mu}q^{\nu]} - \tfrac{8}{3}(d^3z^{[\mu}/d\tau^3)\dot{q}^{\nu]} - \tfrac{8}{3}(d^3q^{[\mu}/d\tau^3)\dot{z}^{\nu]} - \tfrac{8}{3}(\dot{q}\ddot{z})\dot{z}^{[\mu}\ddot{z}^{\nu]} - \tfrac{16}{3}(\ddot{q}\dot{z})\dot{z}^{[\mu}\ddot{z}^{\nu]} - \tfrac{16}{3}(\dot{q}\dot{z})z^{[\mu}(d^3z^{\nu]}/d\tau^3). \quad (31)$$

Ignoring questions of radiation reaction, it would be interesting to investigate whether, from a kinematical point of view, the expression (31) has any close resemblance to the properties of the radiation (7) produced by a moving charge. For instance, if we specify the kinematical condition that the dipole moment q^μ must not rotate about the world line (i.e., q^μ undergoes Fermi–Walker propagation), does the radiation (31) vanish for hyperbolic motion as it does for the case of a moving charge?

This question can be resolved by writing in the equations of Fermi–Walker propagation,

$$-\dot{q}^\nu = (\ddot{z}q)\dot{z}^\nu \quad \text{(and therefore } q^\alpha q_\alpha = \text{constant)},$$

$$-\ddot{q}^\nu = ((d^3z/d\tau^3)q)\dot{z}^\nu + (\ddot{z}\dot{q})\dot{z}^\nu + (\ddot{z}q)\ddot{z}^\nu,$$

$$-(d^3q^\nu/d\tau^3) = ((d^4z/d\tau^4)q)\dot{z}^\nu + 2((d^3z/d\tau^3)\dot{q})\dot{z}^\nu + 2((d^3z/d\tau^3)q)\ddot{z}^\nu + (\ddot{z}\dot{q})\dot{z}^\nu + 2(\ddot{z}\dot{q})\ddot{z}^\nu + (\ddot{z}q)(d^3z^\nu/d\tau^3).$$

We find

$$F^{\mu\nu}_{\text{rad}} = \tfrac{4}{3}q^{[\mu}\{(d^4z^{\nu]}/d\tau^4) + \ddot{z}^2\ddot{z}^{\nu]} + (\ddot{z}(d^3z/d\tau^3))\dot{z}^{\nu]}\}, \quad (32)$$

and we see that the expression does in fact vanish for hyperbolic motion, $(d^3z^\nu/d\tau^3) + \ddot{z}^2\dot{z}^\nu = 0$. [This statement can be made stronger, because we can also show that $(d^4z^\nu/d\tau^4) + \ddot{z}^2\ddot{z}^\nu + (\ddot{z}(d^3z/d\tau^3))\dot{z}^\nu$ will only vanish when the motion is hyperbolic. It is possible from (32) that $F^{\mu\nu}_{\text{rad}}$ may also vanish when q^μ is proportional to $(d^4z^\mu/d\tau^4) + \ddot{z}^2\ddot{z}^\mu + (\ddot{z}(d^3z/d\tau^3))\dot{z}^\mu$; this condition includes the possibility of hyperbolic motion[18].]

The magnetic dipole.

If we need the radiation on the world line of a magnetic dipole, this can be easily found from that of an electric dipole. When we replace q^α by m^α in (31), we obtain $-F^{\text{rad}}_{\mu^*\nu}$ for a magnetic dipole.

THE MOVING MULTIPOLE

We can describe the fields of a 2^m-pole by means of the equations

$$\left.\begin{aligned} A^\mu_{(m)} &= \frac{1}{m!}\Pi^{\mu\nu_1\nu_2\cdots\nu_m}{}_{,\nu_1\nu_2\cdots\nu_m} \\ j^\mu_{(m)} &= \frac{1}{m!}j^{\mu\nu_1\nu_2\cdots\nu_m}{}_{,\nu_1\nu_2\cdots\nu_m} \end{aligned}\right\}, \quad (33)$$

where

$$\Box\Pi^{\mu\nu_1\nu_2\cdots\nu_m} = j^{\mu\nu_1\nu_2\cdots\nu_m} \quad (34)$$

and

$$\Pi^{\mu\nu_1\nu_2\cdots\nu_m} = \Pi^{[\mu\nu_1]\nu_2\cdots\nu_m},$$
$$\Pi^{\mu\nu_1\nu_2\cdots\nu_m} = \Pi^{\mu\nu_1(\nu_2\cdots\nu_m)}, \quad (35)$$
$$\Pi^{[\mu\nu_1\nu_2]\nu_3\cdots\nu_m} = 0.$$

The representation (33) will be a valid representation of Maxwell's equations provided Eq. (34) holds and the first condition of (35) is satisfied (we have $j^\mu_{(m),\mu} = 0$, $A^\mu_{(m),\mu} = 0$). The latter conditions of (35) are sufficient in view of the identities

$$\Pi^{\mu\nu_1\nu_2\cdots\nu_m}{}_{,\nu_1\nu_2\cdots\nu_m} = \Pi^{\mu\nu_1(\nu_2\cdots\nu_m)}{}_{,\nu_1\nu_2\cdots\nu_m}$$
$$\Pi^{[\mu\nu_1\nu_2]\nu_3\cdots\nu_m}{}_{,\nu_1\nu_2\cdots\nu_m} = 0.$$

Solving (34) for a multipole restricted to the world line $z^\alpha = z^\alpha(\tau)$, we obtain [following (10) and (13)]:

$$\Pi^{\mu\nu_1\nu_2\cdots\nu_m}_{\text{ret}}(x) = 4\pi\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau)D_{\text{ret}}[x - z(\tau)]\,d\tau, \quad (36)$$

where the tensor $p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau)$ represents the character of the 2^m-pole in much the same way that $p^\mu(\tau)$ characterizes a dipole. Without imposing further conditions on $p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau)$ than those of (35), we

[18] It can be shown that other solutions to this condition (if they exist) must satisfy

$$\ddot{z}^2 = k, \qquad (d^3z/d\tau^3)^2 - k^2 = l/\{(m\tau + n)^2 - 1\},$$

where k, l, m, n are constants.

can go on to construct concise solutions of Maxwell's equations.

We have the following[10] derived from (33):

$$A^{\mu}_{(m)\,\mathrm{ret}}(x) = \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau) \times \partial_{\nu_m}\partial_{\nu_{m-1}}\cdots\partial_{\nu_1} D_{\mathrm{ret}}(x - z(\tau))\,d\tau, \quad (37)$$

$$j^{\mu}_{(m)}(x) = \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau) \times \partial_{\nu_m}\partial_{\nu_{m-1}}\cdots\partial_{\nu_1}\delta^4(x - z(\tau))\,d\tau, \quad (38)$$

and we evaluate (37) by the same principles used for a dipole. We have, for the first two differentiations,

$$A^{\mu}_{(m)\,\mathrm{ret}}(x) = \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}\partial_{\nu_m}\partial_{\nu_{m-1}}\cdots\partial_{\nu_2} \times\left\{-l_{\nu_1}\left(\frac{1}{v}\frac{d}{d\tau}\right)D_{\mathrm{ret}}(l)\right\}d\tau$$

$$= \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\nu\mu_1\nu_2\cdots\nu_m}\partial_{\nu_m}\partial_{\nu_{m-1}}\cdots\partial_{\nu_2} \times\left\{l_{\nu_1}l_{\nu_2}\left(\frac{1}{v}\frac{d}{d\tau}\right)^2 D_{\mathrm{ret}}(l) - g_{\nu_1\nu_2}\left(\frac{1}{v}\frac{d}{d\tau}\right)D_{\mathrm{ret}}(l)\right\}d\tau.$$

After s differentiations ($s \leq m$) we obtain by induction on s (see Appendix A)

$$A^{\mu}_{(m)\,\mathrm{ret}}(x) = \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}\partial_{\nu_m}\partial_{\nu_{m-1}}\cdots\partial_{\nu_{s+1}} \times\left\{\sum_{r=0}^{\frac{1}{2}(s-1),\frac{1}{2}s}\frac{(-1)^{s-r}s!}{(s-2r)!\,r!\,2^r}\,l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r}} \times g_{\nu_{s-2r+1}\nu_{s-2r+2}}g_{\nu_{s-2r+3}\nu_{s-2r+4}}\cdots g_{\nu_{s-1}\nu_s)} \times\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r}D_{\mathrm{ret}}(l)\right\}d\tau. \quad (39)$$

The summation is extended to $\frac{1}{2}(s-1)$ or $\frac{1}{2}s$ whichever is an integer. After completing m differentiations, we integrate by parts successively until D_{ret} emerges undifferentiated in the integrand,

$$A^{\mu}_{(m)\,\mathrm{ret}}(x) = \int_{-\infty}^{\infty} 4\pi v D_{\mathrm{ret}}(l) \times\left[\sum_{r=0}^{\frac{1}{2}(m-1),\frac{1}{2}m}\frac{1}{(m-2r)!\,r!\,2^r}\left(\frac{1}{v}\frac{d}{d\tau}\right)^{m-r} \times\left\{\frac{1}{v}\,p^{\mu\nu_1\nu_2\cdots\nu_m}l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{m-2r}} \times g_{\nu_{m-2r+1}\nu_{m-2r+2}}\cdots g_{\nu_{m-1}\nu_m)}\right\}\right]d\tau. \quad (40)$$

The retarded fields of the 2^m-pole are obtained by further differentiating (39):

$$F^{\mu\nu}_{(m)\,\mathrm{ret}}(x) = A^{\mu}_{(m)\,\mathrm{ret}}{}^{,\nu}(x) - A^{\nu}_{(m)\,\mathrm{ret}}{}^{,\mu}(x),$$

where

$$A^{\mu}_{(m)\,\mathrm{ret}}{}^{,\nu}(x) = \int_{-\infty}^{\infty} 4\pi v D_{\mathrm{ret}}(l) \times\left[\sum_{r=0}^{\frac{1}{2}m,\frac{1}{2}(m+1)}\frac{m+1}{(m-2r+1)!\,r!\,2^r}\left(\frac{1}{v}\frac{d}{d\tau}\right)^{m-r+1} \times\left\{\frac{1}{v}\,p^{\mu}{}_{\nu_1\nu_2\cdots\nu_m}l^{(\nu_1}l^{\nu_2}\cdots l^{\nu_{m-2r+1}} \times g^{\nu_{m-2r+2}\nu_{m-2r+3}}\cdots g^{\nu_m\nu)}\right\}\right]d\tau. \quad (41)$$

From the expression (2) for $D_{\mathrm{ret}}(l)$, explicit expressions for the potentials (40) and fields (41) can be written down (for $x^0 > z^0$) from the formula

$$\int_{-\infty}^{\infty} 4\pi v D_{\mathrm{ret}}(l)\cdot[\cdots]\,d\tau = [\cdots]_{\tau=\tau_r}.$$

For example, for a moving quadrupole ($m = 2$) we have

$$A^{\mu}_{(2)\,\mathrm{ret}}(x) = \frac{1}{2}\left[\frac{1}{v}\frac{d}{d\tau}\left\{\frac{1}{v}\frac{d}{d\tau}\left(\frac{p^{\mu\alpha\beta}l_\alpha l_\beta}{v}\right) + \frac{p^{\mu\alpha\beta}}{v}\,g_{\alpha\beta}\right\}\right]_{\tau=\tau_r},$$

$$F^{\mu\nu}_{(2)\,\mathrm{ret}}(x) = \frac{1}{2}\left[\frac{1}{v}\frac{d}{d\tau}\left\{\frac{1}{v}\frac{d}{d\tau}\left[\frac{1}{v}\frac{d}{d\tau}\left(\frac{p^{\mu}{}_{\alpha\beta}l^\alpha l^\beta l^\nu}{v}\right) + \frac{p^{\mu}{}_{\alpha\beta}}{v}\,(l^\alpha g^{\beta\nu} + l^\beta g^{\nu\alpha} + l^\nu g^{\alpha\beta})\right]\right\}\right]_{\tau=\tau_r} - (\mu \leftrightarrow \nu). \quad (42)$$

Finally we note that any multipole may be split into electric and magnetic parts by means of the identity:

$$p^{\mu\nu_1\nu_2\cdots\nu_m} = \delta^{\mu\nu_1}_{\alpha\beta}\,q^{\alpha\nu_2\nu_3\cdots\nu_m}\dot{z}^\beta + \xi^{\mu\nu_1\alpha\beta}m_{\alpha\nu_2\nu_3\cdots\nu_m}\dot{z}_\beta,$$

where

$$q^{\alpha\nu_2\nu_3\cdots\nu_m} = p^{\alpha\beta\nu_2\nu_3\cdots\nu_m}\dot{z}_\beta$$

and

$$m^{\alpha\nu_2\nu_3\cdots\nu_m} = p^{\alpha^*\beta\nu_2\nu_3\cdots\nu_m}\dot{z}_\beta$$

are the electric and magnetic multipole moments.

The multipole expansion of a moving source.

The current, potentials, and fields of any moving point source can be described by multipole expansions:

$$\left.\begin{aligned} j^{\mu}(x) &= \underset{(0)}{j^{\mu}}(x) + \sum_{m=1}^{\infty} \underset{(m)}{j^{\mu}}(x) \\ A^{\mu}_{\mathrm{ret}}(x) &= \underset{(0)}{A^{\mu}_{\mathrm{ret}}}(x) + \sum_{m=1}^{\infty} \underset{(m)}{A^{\mu}_{\mathrm{ret}}}(x) \\ F^{\mu\nu}_{\mathrm{ret}}(x) &= \underset{(0)}{F^{\mu\nu}_{\mathrm{ret}}}(x) + \sum_{m=1}^{\infty} \underset{(m)}{F^{\mu\nu}_{\mathrm{ret}}}(x) \end{aligned}\right\}. \tag{43}$$

The first terms $\underset{(0)}{j^{\mu}}(x)$, $\underset{(0)}{A^{\mu}_{\mathrm{ret}}}(x)$, $\underset{(0)}{F^{\mu\nu}_{\mathrm{ret}}}(x)$ are those of (3) and (6). The expansion of the potential is frequently given in the form

$$A^{\mu}_{\mathrm{ret}}(x) = \frac{e\dot{z}^{\mu}}{v} + \sum_{m=1}^{\infty} \frac{1}{m!} \left(\frac{p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau_r)}{v(\tau_r, x)} \right)_{,\nu_1\nu_2\cdots\nu_m}.$$

With regard to these expansions we may easily show that it is sufficient to choose the multipole tensors $p^{\mu\nu_1\nu_2\cdots\nu_m}(\tau)$ to satisfy the following conditions in addition to those identical to (35):

$$\left.\begin{aligned} p^{\mu\nu_1\nu_2\cdots\nu_m}\dot{z}_{\nu_m} &= 0, \\ p^{\mu\nu_1\nu_2\cdots\nu_m}g_{\nu_1\nu_m} &= 0, \\ p^{\mu\nu_1\nu_2\cdots\nu_{m-1}\nu_m}g_{\nu_{m-1}\nu_m} &= 0. \end{aligned}\right\} \tag{44}$$

The first of these conditions will prevent the appearance of terms in the four-current of lower polarity than m when $p^{\mu\nu_1\nu_2\cdots\nu_m}$ is resolved and (38) integrated by parts with the help of $\dot{z}_{\lambda}\partial^{\lambda}/\partial z^{\lambda} = d/d\tau$[10]; while the second two prevent the appearance of all solution terms which contribute lower polarity than m in (40), and likewise could be included in terms of lower order. [The latter two conditions in (44) together with the conditions (35) should also arise in any Lagrangian approach where the same properties would hold for $F^{\mu\nu_1\nu_2\cdots\nu_m}_{(\mathrm{in})}$.] Such arguments may be made rigorous.

Under the circumstances of (44) we find that the expansions (43) can be written [from (40) and (41)]

$$\begin{aligned} A^{\mu}_{\mathrm{ret}}(x) &= \left[\frac{e\dot{z}^{\mu}}{v} + \sum_{m=1}^{\infty} \frac{1}{m!} \left(\frac{1}{v}\frac{d}{d\tau} \right)^m \right. \\ &\quad \left. \times \left\{ \frac{1}{v} l^{\nu_1} l^{\nu_2} \cdots l^{\nu_m} p^{\mu}{}_{\nu_1\nu_2\cdots\nu_m} \right\} \right]_{\tau=\tau_r}, \\ F^{\mu\nu}_{\mathrm{ret}}(x) &= \left[\frac{2e}{v}\frac{d}{d\tau}\left\{ \frac{\dot{z}^{[\mu}l^{\nu]}}{v} \right\} \right. \\ &\quad + \sum_{m=1}^{\infty} \frac{2}{m!} \left(\frac{1}{v}\frac{d}{d\tau} \right)^{m+1} \left\{ \frac{1}{v} l^{\nu_1} l^{\nu_2} \cdots l^{\nu_m} l^{[\nu} p^{\mu]}{}_{\nu_1\nu_2\cdots\nu_m} \right\} \\ &\quad + \sum_{m=1}^{\infty} \frac{2}{m!} \left(\frac{1}{v}\frac{d}{d\tau} \right)^{m} \left\{ \frac{1}{v} l^{\nu_2} l^{\nu_3} \cdots l^{\nu_m} p^{[\mu\nu]}{}_{\nu_2\nu_3\cdots\nu_m} \right\} \\ &\quad + \sum_{m=1}^{\infty} \frac{2(m-1)}{m!} \left(\frac{1}{v}\frac{d}{d\tau} \right)^{m} \\ &\quad \left. \times \left\{ \frac{1}{v} l^{\nu_1} l^{\nu_2} \cdots l^{\nu_{m-1}} p^{[\mu}{}_{\nu_1\nu_2\cdots\nu_{m-1}}{}^{\nu]} \right\} \right]_{\tau=\tau_r}. \end{aligned} \tag{45}$$

We can illustrate the validity of these expressions by finding the far fields of an unaccelerated source, as these are known in connection with general relativity.[19] With $\ddot{z}^{\mu} = 0$ we have for the product of any two functions pq,

$$\begin{aligned} \left(\frac{1}{v}\frac{d}{d\tau} \right)^{m+1} pq &= \frac{1}{v^{m+1}} q \frac{d^{m+1}p}{d\tau^{m+1}} \\ &\quad + \tfrac{1}{2}m(m+1)\frac{1}{v^{m+2}} q \frac{d^m p}{d\tau^m} \\ &\quad + (m+1)\frac{1}{v^{m+1}}\frac{dq}{d\tau}\frac{d^m p}{d\tau^m} + \cdots \end{aligned}$$

so that we find

$$F^{\mu\nu}_{\mathrm{ret}}(x) = \left[\frac{{}_0N^{\mu\nu}}{v} + \frac{{}_0III^{\mu\nu}}{v^2} + O(v^{-3}) \right]_{\tau=\tau_r},$$

where

$$\begin{aligned} {}_0N^{\mu\nu} &= 2\sum_{m=1}^{\infty} \frac{1}{m!} k^{\nu_1} k^{\nu_2} \cdots k^{\nu_m} k^{[\nu} \frac{d^{m+1}}{d\tau^{m+1}} p^{\mu]}{}_{\nu_1\nu_2\cdots\nu_m} \\ {}_0III^{\mu\nu} &= 2e\dot{z}^{[\mu}k^{\nu]} + 2\sum_{m=1}^{\infty} \frac{1}{m!} k^{\nu_2} k^{\nu_3} \cdots k^{\nu_{m-1}} \\ &\quad \times \{ (\tfrac{1}{2}m+1)(m+1) k^{\nu_1} k^{\nu_m} k^{[\nu} \\ &\quad - (m+1)\dot{z}^{\nu_1} k^{\nu_m} k^{[\nu} - (m+1) k^{\nu_1} k^{\nu_m} \dot{z}^{[\nu} \\ &\quad + k^{\nu_m} g^{\nu_1[\nu} + (m-1) k^{\nu_1} g^{\nu_m[\nu} \} \frac{d^m}{d\tau^m} p^{\mu]}{}_{\nu_1\nu_2\cdots\nu_m}. \end{aligned}$$

These agree with the known fields of an unaccelerated source.

Radiation fields

The radiation field of a pure multipole can be calculated by either of the two methods available for charges and dipoles. We ask the question whether the radiation is finite (unique) for any particular multipole. For instance, is the radiation field finite on the world line of a quadrupole?

The first (explicit) method for determining the radiation field is bound to be a very lengthy procedure (the more lengthy the calculation the higher the order of the multipole). Nevertheless, it should be possible to determine at what stage *even* powers of ϵ begin to appear in the expressions for the retarded field. The question of finiteness of the radia-

[19] J. N. Goldberg, Phys. Rev. **131**, 1377 (1963), Eqs. (A8) and (A9).

tion field on the world line seems to be more easily resolvable when viewed within the second method involving delta functions, and it is this method we now refer to. It should be pointed out that complete verification by the long method would be desirable.

We here contend ourselves with resolving the question about the finiteness of the fields; it does not seem worthwhile working them out in detail in view of the extreme lengthy computations. The radiation fields of a moving 2^m-pole are obtained from (41) by replacing D_{ret} by D, and the procedure for evaluating the radiation along the world line follows an identical pattern to that for charges and dipoles as we have already pointed out. Since in any calculation of this type we need to expand the integrand as a power series in τ, let us see when in the sequence dipole, quadrupole, octupole, . . . we first meet a term in τ^{-2} in the integrand. Any term involving an even inverse power of τ would prevent the appearance of a unique and finite value for the fields.

$$F^{\mu\nu}_{\underset{(1)}{\text{rad}}}(z(0)) = 4\pi \int_{-\infty}^{\infty} D(l) \frac{d}{d\tau} \left\{ \frac{1}{v} \frac{d}{d\tau} (O(\tau)) + O\left(\frac{1}{\tau}\right) \right\} d\tau,$$

$$F^{\mu\nu}_{\underset{(2)}{\text{rad}}}(z(0)) = 4\pi \int_{-\infty}^{\infty} D(l) \frac{d}{d\tau} \left\{ \frac{1}{v} \frac{d}{d\tau} \left\{ \frac{1}{v} \frac{d}{d\tau} (O(\tau^2)) + O(1) \right\} \right\} d\tau,$$

$$F^{\mu\nu}_{\underset{(3)}{\text{rad}}}(z(0)) = 4\pi \int_{-\infty}^{\infty} D(l) \frac{d}{d\tau} \left\{ \frac{1}{v} \frac{d}{d\tau} \left[\frac{1}{v} \frac{d}{d\tau} \left\{ \frac{1}{v} \frac{d}{d\tau} (O(\tau^3)) + O(\tau) \right\} + O\left(\frac{1}{\tau}\right) \right] \right\} d\tau.$$

We find (to the appropriate order) for the respective multipoles,

$$F^{\mu\nu}_{\underset{(1)}{\text{rad}}}(z(0)) = 2 \int_{-\infty}^{\infty} \delta'(\tau) \{1,, +\alpha\tau^2, +\beta\tau^3\} \times \{A^{\mu\nu}_{(1)}\tau^{-2},, +B^{\mu\nu}_{(1)}, +C^{\mu\nu}_{(1)}\tau\} \, d\tau,$$

$$F^{\mu\nu}_{\underset{(2)}{\text{rad}}}(z(0)) = 2 \int_{-\infty}^{\infty} \delta'(\tau) \{1,, +\alpha\tau^2, +\beta\tau^3, +\gamma\tau^4\} \times \{A^{\mu\nu}_{(2)}\tau^{-2},, +B^{\mu\nu}_{(2)}, +C^{\mu\nu}_{(2)}\tau, +D^{\mu\nu}_{(2)}\tau^2\} \, d\tau,$$

$$F^{\mu\beta}_{\underset{(3)}{\text{rad}}}(z(0)) = 2 \int_{-\infty}^{\infty} \delta'(\tau) \{1,, +\alpha\tau^2, +\beta\tau^3, +\gamma\tau^4, +\delta\tau^5\} \times \{A^{\mu\nu}_{(3)}\tau^{-4},, +B^{\mu\nu}_{(3)}\tau^{-2},, +C^{\mu\nu}_{(3)}, +D^{\mu\nu}_{(3)}\tau\} \, d\tau,$$

where we have written

$$D(l) = D(z(0) - z(\tau)) \equiv -\frac{1}{2\pi} \delta(\tau)\tau^{-1} \times \{1,, +\alpha\tau^2, +\beta\tau^3, +\gamma\tau^4, +\delta\tau^5, +\cdots\}.$$

Hence

$$F^{\mu\nu}_{\underset{(1)}{\text{rad}}}(z(0)) = -2\{C^{\mu\nu}_{(1)} + \beta A^{\mu\nu}_{(1)}\}, \text{ [Eq. (30)]}$$

$$F^{\mu\nu}_{\underset{(2)}{\text{rad}}}(z(0)) = -2\{C^{\mu\nu}_{(2)} + \beta A^{\mu\nu}_{(2)}\},$$

and $F^{\mu\nu}_{\underset{(3)}{\text{rad}}}(z(0))$ does not possess a unique and finite value because the integrand contains the term $\beta A^{\mu\nu}_{(3)}\tau^{-2}\delta(\tau)$. Hence, we can assume that the radiation field is finite only for charges, dipoles, and quadrupoles, but this is no longer the case for higher-order multipoles. In a similar way, because of the number of differentiations involved, we can say that the potential of the radiation field is finite along the world line for charges, dipoles, quadrupoles, and octupoles (not for any other multipoles in general). It may perhaps be worth noting that these are the multipoles whose ordinary potentials may be constructed from those of rigidly connected monopoles according to the progression rod, parallelogram, parallelopiped, where alternating poles occupy the positions of adjacent vertices. Presumably higher-order multipoles in general lack such a construction (in view of the dimensionality of space). Perhaps, therefore, it is not altogether surprising that the radiation potential is no longer finite for these multipoles.

RADIATION REACTION FOR DIPOLES

The Lagrangian density for a charged particle interacting with a free external field $F^{\mu\nu}_{\text{in}}$ is

$$\mathfrak{L} = \mathfrak{L}_f + \mathfrak{L}_p^{(0)} + \mathfrak{L}_{\text{int}}^{(0)}, \tag{46}$$

where

$$\mathfrak{L}_f = \frac{1}{16\pi} (F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{in}})(F^{\text{ret}}_{\mu\nu} + F^{\text{in}}_{\mu\nu}),$$

$$\mathfrak{L}_p^{(0)} = \int_{-\infty}^{\infty} m_0(\dot{z}^\mu \dot{z}_\mu)^{\frac{1}{2}} c^2 \delta^4(x - z(\tau)) \, d\tau,$$

$$\mathfrak{L}_{\text{int}}^{(0)} = e \int_{-\infty}^{\infty} \dot{z}^\mu (A^{\text{ret}}_\mu + A^{\text{in}}_\mu) \delta^4(x - z(\tau)) \, d\tau.$$

The part of the action arising from the last term of (46), namely,

$$S^{(0)}_{\text{int}} = \int \mathfrak{L}^{(0)}_{\text{int}}\, d^4x \quad \left\{ = \int_{-\infty}^{\infty} L^{(0)}_{\text{int}}\, d\tau \quad \text{with} \right.$$

$$\left. L^{(0)}_{\text{int}} = e\dot{z}^\alpha(\tau)(A^{\text{ret}}_\alpha(z(\tau)) + A^{\text{in}}_\alpha(z(\tau))) \right\}$$

is equivalent to $(1/4\pi)\int j^\mu_{(0)} A_\mu d^4x$ where $j^\mu_{(0)}$ is given by (3). For a dipole, we would expect a similar action term where (12) takes the place of (3), i.e.,

$$S^{(1)}_{\text{int}} = \frac{1}{4\pi}\int j^\mu_{(1)} A_\mu\, d^4x$$

$$= \int d^4x \int_{-\infty}^{\infty} (A^\mu_{\text{ret}} + A^\mu_{\text{in}}) p_{\mu\alpha}\, \partial^\alpha \delta^4(x - z(\tau))\, d\tau,$$

$$= -\int_{-\infty}^{\infty} p_{\mu\alpha} \left\{ \int (A^{\mu,\alpha}_{\text{ret}} + A^{\mu,\alpha}_{\text{in}}) \delta^4(x - z(\tau))\, d^4x \right\} d\tau,$$

$$= \int d^4x \int_{-\infty}^{\infty} -\tfrac{1}{2} p_{\mu\alpha}(F^{\mu\alpha}_{\text{ret}} + F^{\mu\alpha}_{\text{in}}) \delta^4(x - z(\tau))\, d\tau,$$

$$= \int_{-\infty}^{\infty} L^{(1)}_{\text{int}}\, d\tau,$$

where

$$L^{(1)}_{\text{int}} = -\tfrac{1}{2} p^{\alpha\beta}(\tau)[F^{\text{ret}}_{\alpha\beta}(z(\tau)) + F^{\text{in}}_{\alpha\beta}(z(\tau))]. \qquad (47)$$

The complete action for a dipole (with or without added charge) would then also contain terms depending on the mechanical spin. We assume that τ is initially an arbitrary chosen parameter in the variation and we also require to have the Lagrangian L, and in particular $L^{(1)}_{\text{int}}$, homogeneous of the first order in $\dot{z}^\mu$. For $L^{(1)}_{\text{int}}$ this can be prescribed in one of two ways, which we have not as yet indicated. A first method[20] is to multiply (47) by $(\dot{z}^\mu \dot{z}_\mu)^{\frac{1}{2}}$. This will contribute "extra" terms to the total Minkowski force (which would otherwise only be composed of the one given by the canonical expression $\partial L^{(1)}_{\text{int}}/\partial z_\mu$). A more interesting possibility, however, is to regard $p^{\mu\nu}$ as already homogeneous of the first order in $\dot{z}^\mu$, this is by virtue of the relation $p^{\mu\nu} = \delta^{\mu\nu}_{\alpha\beta} q^\alpha \dot{z}^\beta + \xi^{\mu\nu\alpha\beta} m_\alpha \dot{z}_\beta$ and the variation of τ in q^μ and m^μ instead of in $p^{\mu\nu}$. The four force resulting from this interaction Lagrangian can be compared with the ordinary expression for a charge:

$$\frac{\partial L^{(0)}_{\text{int}}}{\partial z_\mu} - \frac{d}{d\tau}\left(\frac{\partial L^{(0)}_{\text{int}}}{\partial \dot{z}_\mu}\right)$$

$$= e\dot{z}_\alpha A^{\alpha,\mu} - \frac{d}{d\tau}(eA^\mu) = -eF^{\mu\nu}\dot{z}_\nu,$$

$$\frac{\partial L^{(1)}_{\text{int}}}{\partial z_\mu} - \frac{d}{d\tau}\left(\frac{\partial L^{(1)}_{\text{int}}}{\partial \dot{z}_\mu}\right)$$

$$= -\tfrac{1}{2} p_{\alpha\beta} F^{\alpha\beta,\mu} - \frac{d}{d\tau}(q_\alpha F^{\mu\alpha} - m_\alpha F^{\mu^*\alpha})$$

$$= -\dot{q}_\alpha F^{\mu\alpha} - q_\alpha F^{\mu\nu,\alpha}\dot{z}_\nu + \dot{m}_\alpha F^{\mu^*\alpha} + m_\alpha F^{\mu^*\nu,\alpha}\dot{z}_\nu.$$

[We have made use of the identity $F^{[\mu^*\nu,\alpha]} = 0$ in the last term, although this will be strictly admissible only for the radiative part of the fields $F^{\mu\nu}(= F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{in}})$ which we consider shortly.] We note that the generalized momentum for a charge and a dipole is

$$p^\mu = G^\mu + eA^\mu + q_\alpha F^{\mu\alpha} - m_\alpha F^{\mu^*\alpha},$$

where G^μ is the mechanical momentum. A general Lagrangian method[21] will then give complete equations of motion for the momentum G^μ and spin $S^{\mu\nu}$:

$$\begin{aligned} \dot{G}^\mu &= -eF^{\mu\nu}\dot{z}_\nu - \dot{q}_\alpha F^{\mu\alpha} - q_\alpha F^{\mu\nu,\alpha}\dot{z}_\nu \\ &\quad + \dot{m}_\alpha F^{\mu^*\alpha} + m_\alpha F^{\mu^*\nu,\alpha}\dot{z}_\nu, \\ \dot{S}^{\mu\nu} + 2G^{[\mu}\dot{z}^{\nu]} &= 2q^{[\mu}F^{\nu]\alpha}\dot{z}_\alpha + \xi^{\mu\nu}\lambda\sigma m^\alpha F_{\alpha\sigma}\dot{z}_\lambda, \end{aligned} \qquad (48)$$

and $G^\mu = m_0 c^2 \dot{z}^\mu - \dot{S}^{\mu\nu}\dot{z}_\nu$. Resolving $-\dot{S}^{\mu\nu}$ into two parts,

$$-\dot{S}^{\mu\nu} = \delta^{\mu\nu}_{\alpha\beta} U^\alpha \dot{z}^\beta + \xi^{\mu\nu\alpha\beta} T_\alpha \dot{z}_\beta$$

we have for the torque T^μ,

$$T^\mu = \xi^{\mu\nu\alpha\beta} q_\nu F_{\alpha\gamma}\dot{z}^\gamma \dot{z}_\beta + 2m_\alpha F^{\alpha[\mu}\dot{z}^{\nu]}\dot{z}_\nu. \qquad (49)$$

It may be verified that in the rest-system where $\dot{z}^\alpha = (1, \mathbf{0})$, $q^\alpha = (0, \mathbf{q})$, $m^\alpha = (0, \mathbf{m})$, expressions (48), (49) reduce to $\dot{G}^\mu = (P, \mathbf{F})$, $T^\mu = (0, \mathbf{T})$, where

$$\mathbf{F} = e\mathbf{E} + \dot{\mathbf{q}} \times \mathbf{H} + (\mathbf{q}\cdot\nabla)\mathbf{E} - \dot{\mathbf{m}} \times \mathbf{E} + (\mathbf{m}\cdot\nabla)\mathbf{H},$$

$$\mathbf{T} = \mathbf{q} \times \mathbf{E} + \mathbf{m} \times \mathbf{H}.$$

These expressions for the force and couple include an interactionary force arising from the circular current elements of the electric dipole and the magnetic field, and a dual term arising from the electric field. Together with the usual formulas for the force and couple, these expressions composing $\mathbf{F}$ and $\mathbf{T}$ would be expected from ordinary electromagnetic theory. (The expression for the power P is $\dot{\mathbf{q}}\cdot\mathbf{E} + \dot{\mathbf{m}}\cdot\mathbf{H}$ and this would also be anticipated in view of the analogy with the expression $e\mathbf{V}\cdot\mathbf{E}$ for a charge.)

We now write

$$F^{\mu\nu} = \tfrac{1}{2} F^{\mu\nu}_{\text{rad}} + \tfrac{1}{2}(F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{adv}}) \qquad (50)$$

[20] See Ref. 8 p. 74.

[21] See Ref. 8, p. 77.

in the expressions (48) and (49) and we have the following values for the force and torque of reaction, for a pure electric dipole (for brevity),

$$F^{\mu}_{\text{react}} = -\tfrac{1}{2}\dot{q}_{\alpha}(\tau)F^{\mu\alpha}_{\text{rad}}(z(\tau)) - \tfrac{1}{2}q_{\alpha}(\tau)F^{\mu\nu;\alpha}_{\text{rad}}(z(\tau))\dot{z}_{\nu}(\tau),$$
$$T^{\mu}_{\text{react}} = \tfrac{1}{2}\xi^{\mu\nu\alpha\beta}q_{\nu}(\tau)F^{\text{rad}}_{\alpha\gamma}(z(\tau))\dot{z}^{\gamma}(\tau)\dot{z}_{\beta}(\tau), \tag{51}$$

provided the infinite part arising from the second term of (50) can be taken to have a purely inertial character (after the fashion of monopoles[22]).

The value of $F^{\mu\nu}_{\text{rad}}(z(\tau))$ occurring in these expressions has been obtained [see (24), (31)] and the value of $F^{\mu\nu;\alpha}_{\text{rad}}(z(\tau))$ can be found from the following:

$$A^{\mu\nu;\alpha}_{\text{rad}}(x) = 4\pi\int_{-\infty}^{\infty} D(l)\frac{d}{d\tau}\left\{\frac{1}{v}\frac{d}{d\tau}\left[\frac{1}{v}\frac{d}{d\tau}\left(\frac{p^{\mu\beta}l_{\beta}l^{\nu}l^{\alpha}}{v}\right) + \frac{1}{v}p^{\mu\beta}(\delta^{\alpha}_{\beta}l^{\nu} + l_{\beta}g^{\alpha\nu} + \delta^{\nu}_{\beta}l^{\alpha})\right]\right\}d\tau,$$

where $x = z(0)$ [cf. Eq. (42) for a quadrupole], the integrand being expanded to the first five orders in τ. Because there are three derivatives involved, the result will be finite. We should not expect the force to be orthogonal to $\dot{z}^{\mu}$.

We may investigate the term of the action which gives rise to the infinite part of the self-force[23]

$$\int d^4x\int_{-\infty}^{\infty} -\tfrac{1}{2}p_{\mu\nu}\cdot\tfrac{1}{2}(F^{\mu\nu}_{\text{ret}} + F^{\mu\nu}_{\text{adv}})\delta^4(x - z(\tau))\,d\tau$$
$$= -2\pi\int d^4x\int_{-\infty}^{\infty} p_{\mu\nu}(\tau)\delta^4(x - z(\tau))\int_{-\infty}^{\infty}\bar{D}(x - z(\tau'))$$
$$\times\frac{d}{d\tau'}\left\{\frac{1}{v(\tau', x)}\frac{d}{d\tau'}\left\{\frac{\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}(\tau')(x_{\beta} - z_{\beta}(\tau'))(x^{\gamma} - z^{\gamma}(\tau'))}{v(\tau', x)}\right\} + \frac{2p^{\mu\nu}(\tau')}{v(\tau', x)}\right\}d\tau'\,d\tau,$$
$$= -2\pi\int_{-\infty}^{\infty} p_{\mu\nu}(\tau)\int_{-\infty}^{\infty}\bar{D}(z(\tau) - z(\tau'))\frac{d}{d\tau'}\left\{\frac{1}{v(\tau', z(\tau))}\frac{d}{d\tau'}\right.$$
$$\left.\times\left\{\frac{\delta^{\mu\nu}_{\alpha\gamma}p^{\alpha\beta}(\tau')(z_{\beta}(\tau) - z_{\beta}(\tau'))(z^{\gamma}(\tau) - z^{\gamma}(\tau'))}{v(\tau', z(\tau))}\right\} + \frac{2p^{\mu\nu}(\tau')}{v(\tau', z(\tau))}\right\}d\tau'\,d\tau,$$
$$= \int_{-\infty}^{\infty}\mu(\tau)\,d\tau$$

with

$$\mu(\tau) = -2\pi p_{\mu\nu}(\tau)\int_{-\infty}^{\infty}\frac{1}{4\pi}\{1,, +\tfrac{1}{12}\lambda^2\ddot{z}^2, + \tfrac{1}{12}\lambda^3(\ddot{z}(d^3z/d\tau^3))\}\frac{\delta(\lambda)}{|\lambda|} \times \{\underset{(1)}{A^{\mu\nu}}\lambda^{-2},, +\underset{(1)}{B^{\mu\nu}}, +\underset{(1)}{C^{\mu\nu}}\lambda\}\,d\lambda,$$

where we have written $\tau' = \tau + \lambda$ in the last integral.

We have

$$\mu(\tau) = (2m^{\mu}m_{\mu} - q^{\mu}q_{\mu})\int_{-\infty}^{\infty}\{1 + O(\lambda^2)\}\frac{\delta(\lambda)}{|\lambda^3|}\,d\lambda.$$

Although the first term of $\int\mu(\tau)\,d\tau$ arising from this expression is merely dependent on the strength of the dipole and may therefore be incorporated in a Lagrangian term

$$\int_{-\infty}^{\infty} m_0(\tau)(\dot{z}^{\mu}\dot{z}_{\mu})^{\frac{1}{2}}c^2\,d\tau,$$

the next nonfinite term (in $1/|\lambda|$) depends on the world line, and so this problem of mass renormalization is not really solved for a dipole. The problem also remains as to how far the mechanical spin terms of the Lagrangian (which we have not specified) may be "renormalized" to account for a similar contribution from the reactional torque given by Eq. (51).

APPENDIX

The validity of the potential (39) is established by induction on s. For a specific s, we have on further differentiation,

[22] See, for example, P. G. Bergmann, *Handbuch der Physik*, S. Flügge, Ed. (Springer-Verlag, Berlin, 1962), Vol. 4, p. 178.

[23] $$\bar{D}(x) = \frac{1}{4\pi}\delta(x^2).$$

$$A^{\mu}_{(m)\,\mathrm{ret}}(x) = \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}\,\partial_{\nu_m}\,\partial_{\nu_{m-1}}\cdots\partial_{\nu_{s+2}}\Bigg\{\sum_{r=0}^{\frac{1}{2}(s-1),\frac{1}{2}s}\frac{(-1)^{s-r}s!}{(s-2r)!\,r!\,2^r}$$

$$\times\Big[(g_{(\nu_1\bar{\nu}_{s+1}}l_{\nu_2}\cdots l_{\nu_{s-2r}} + l_{(\nu_1}g_{\nu_2\bar{\nu}_{s+1}}l_{\nu_3}\cdots l_{\nu_{s-2r}}$$

$$+\cdots + l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r-1}}g_{\nu_{s-2r}\bar{\nu}_{s+1}})g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)}\cdot\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r}D_{\mathrm{ret}}(l)$$

$$-\,l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r}}g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)}l_{\nu_{s+1}}\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r+1}D_{\mathrm{ret}}(l)\Big]\Bigg\}\,d\tau$$

(the bar written above a suffix denotes that it is not included in the symmetrization)

$$= \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}\,\partial_{\nu_m}\,\partial_{\nu_{m-1}}\cdots\partial_{\nu_{s+2}}$$

$$\times\Bigg\{\sum_{r=0}^{\frac{1}{2}(s-1),\frac{1}{2}(s-2)}\frac{(-1)^{s-r}s!}{(s-2r-1)!\,r!\,2^r}(g_{\nu_{s+1}(\nu_1}l_{\nu_2}l_{\nu_3}\cdots l_{\nu_{s-2r}}g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)})\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r}D_{\mathrm{ret}}(l)$$

$$+\sum_{r=0}^{\frac{1}{2}(s+1),\frac{1}{2}s}\frac{(-1)^{s-r+1}(s-2r+1)s!}{(s-2r+1)!\,r!\,2^r}(l_{\nu_{s+1}}l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r}}g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)})\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r+1}D_{\mathrm{ret}}(l)\Bigg\}d\tau$$

$$= \frac{4\pi}{m!}\int_{-\infty}^{\infty} p^{\mu\nu_1\nu_2\cdots\nu_m}\,\partial_{\nu_m}\,\partial_{\nu_{m-1}}\cdots\partial_{\nu_{s+2}}$$

$$\times\Bigg\{\sum_{r=0}^{\frac{1}{2}(s+1),\frac{1}{2}s}\frac{(-1)^{s-r+1}(s+1)!}{(s-2r+1)!\,r!\,2^r}\Big[\frac{2r}{s+1}(g_{\nu_{s+1}(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r+2}}g_{\nu_{s-2r+3}\nu_{s-2r+4}}\cdots g_{\nu_{s-1}\nu_s)})$$

$$+\frac{s-2r+1}{s+1}(l_{\nu_{s+1}}l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r}}g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)})\Big]\cdot\left(\frac{1}{v}\frac{d}{d\tau}\right)^{s-r+1}D_{\mathrm{ret}}(l)\Bigg\}\,d\tau.$$

Result (39) now follows with $s+1$ in place of s on account of the identity

$$2rg_{\nu_{s+1}(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r+2}}g_{\nu_{s-2r+3}\nu_{s-2r+4}}\cdots g_{\nu_{s-1}\nu_s)} + (s-2r+1)l_{\nu_{s+1}}l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r}}$$

$$\times\, g_{\nu_{s-2r+1}\nu_{s-2r+2}}\cdots g_{\nu_{s-1}\nu_s)} = (s+1)l_{(\nu_1}l_{\nu_2}\cdots l_{\nu_{s-2r+1}}\,g_{\nu_{s-2r+2}\nu_{s-2r+3}}\cdots g_{\nu_s\nu_{s+1})}.$$

The factor $2r$ arises because there are two ways of arranging the suffixes ν_{s+1} and ν_1 in $g_{\nu_{s+1}\nu_2}$ and r ways of arranging $g_{\nu_{s+1}\nu_2}$ between the other $r-1$ g's. The factor $(s-2r+1)$ arises because there are this number of ways of arranging $l_{\nu_{s+1}}$ between the $s-2r$ other l's.

Non-Markovian Model for the Approach to Equilibrium*

GÉRARD G. EMCH
University of Maryland, College Park, Maryland
(Received 1 November 1965)

This paper is intended to provide the axiomatic study of nonequilibrium quantum statistical mechanics with some simple and rigorously solvable models. The models considered here are obtained as generalizations of the Ising model. They illustrate and allow a rational discussion of the following concepts relevant to the theory of irreversible phenomena: coarse-graining and time-smoothing, ergodicity, recurrences, impossibility of a Markovian description of the approach to equilibrium for some physical systems, justification of the various random phase assumptions, properties of the interaction responsible for the approach to equilibrium, master equations, etc.

I EXPERIMENTAL BACKGROUND

THE phenomena called free-induction relaxation exhibits an oscillatory approach to equilibrium. From the first time it was observed, it was interpreted[1] as the result of the dipolar interaction between nuclear spins arranged in a rigid lattice.

Just to show that what follows is not only a mathematical game, but also a topic of some definite physical relevance, we shall recall briefly the experimental situation. This is also intended to provide a safer basis for the present theoretical considerations.

A CaF_2 crystal is placed in a magnetic field **B**, the direction of which we shall call z. The system is allowed to reach thermal equilibrium. When this is achieved, a rf pulse is applied to turn the net magnetic moment $\boldsymbol{\mu}$ in the x direction, orthogonal to the z direction of **B**. The time-evolution of the x component μ_x of $\boldsymbol{\mu}$ is then observed. It exhibits an oscillatory decay to zero, starting from a nonzero initial value. This is interpreted as the result of the dipolar interaction between the $(I = \frac{1}{2})$-spin of the fluorine nuclei. No relaxation via lattice vibration is needed to account for this phenomena, so that the spin system may be considered as isolated. It has been shown that the interaction responsible for this approach to equilibrium can be reduced to the following form (we do not want to worry here about units):

$$V = \sum_{i,j} (a_{ij}\sigma_i\sigma_j + b_{ij}\sigma_i^x\sigma_j^x) - B\sum_i \sigma_i^z, \qquad (1)$$

where $i, j \cdots$ denote the position of the fluorine nuclei, and σ^x, σ^y, σ^z are the Pauli matrices. In this form the model has not been solved exactly. Even if the following simplifying assumption does not lead to a quantitative agreement with experiment, it provides a qualitative description of the observed oscillatory approach to equilibrium, and makes the model exactly solvable:

$$a_{ij} = 0 \quad \text{for all} \quad (i, j). \qquad (2)$$

This form of the model is the basis of our considerations. The aim of these is indeed to discuss the *consequences* of the definite approach to equilibrium encountered in this problem. The remarkable fact about this model is that it does not involve any kind of repeated random phase assumption (the quantum analog of Boltzmann's *Stosszahlansatz*), and that no approximations at all are needed. The approach to equilibrium is shown to be a consequence of only the particular choice of a wide class of initial conditions (connected with the partial information obtained from macroscopic measurements) and of the peculiar form of the Hamiltonian. Because of its great simplicity, this model also allows some enlightening of many aspects of the theory of nonequilibrium processes. This is the principal motive for the present investigation.

II THEORETICAL ANALYSIS OF THE MODEL

We express the main features of our model in a quite naive mathematical form. We try to proceed in such a way that:

(i) the simplicity of the model is exhibited;
(ii) the road to the slightly more sophisticated considerations of the next section is prepared.

Consider an infinite linear chain (in several of the considerations to come it is convenient to consider the infinite linear chain as the limit of a finite ring) of fixed spin-$\frac{1}{2}$ identical particles. The quantum mechanical evolution of the system will be described by a continuous one-parameter group of unitary operators $\{U^t\}$ acting on the Hilbert space

* This research was carried out at the Institute for Fluid Dynamics and Applied Mathematics of the University of Maryland under the support of the Office of Naval Research—Contract NONR 595 (22).
[1] I. J. Lowe and R. E. Norberg, Phys. Rev. 107, 46 (1957).

$$\mathfrak{H} = \prod_i \otimes \mathfrak{H}^i, \tag{3}$$

where the index i runs over all the particles and $\mathfrak{H}^i$ are the identical replicas of the two-dimensional Hilbert space used to describe each individual particle.

We furthermore suppose that U^t can be written as

$$U^t = \prod_{n\geq 0} U_n^t, \tag{4}$$

where all the unitary groups $\{U_n^t\}$ are defined on $\mathfrak{H}$, commute among themselves, and are respectively generated by

$$\begin{aligned} H_0 &= -B \sum_i \sigma_i^z, \\ H_n &= \epsilon(n) \sum_i \sigma_i^z \sigma_{i+n}^z, \quad \text{for} \quad n > 0. \end{aligned} \tag{5}$$

We finally suppose that the real-valued function $\epsilon(n)$ satisfies the following conditions:

$$\begin{aligned} &\text{(i)} \quad \epsilon(n) \text{ is monotonically nonincreasing in } n, \\ &\text{(ii)} \quad \lim_{n\to\infty} \epsilon(n) = 0. \end{aligned} \tag{6}$$

These assumptions correspond to physical situations. For the moment we do not specify the exact form of $\epsilon(n)$, nor do we indicate whether $\epsilon(n)$ reaches its asymptotic value for finite n. Let us write for the generator of $\{U^t\}$:

$$H = \sum_{n=0}^{\infty} H_n. \tag{7}$$

Incidently, we note that

$$H^I \equiv H_0 + H_1 \tag{8}$$

is the ordinary one-dimensional Ising model with nearest-neighbor interaction only. We see that it is indeed quite essential for the approach to equilibrium to consider the generalized form of the Ising model, where the interaction extends over all pairs of neighbors.

Usually the observable of interest in the Ising model is

$$S^z = \sum_i \sigma_i^z, \tag{9}$$

which commutes with H. In this sense the ordinary Ising model, even generalized in the above way, is a classical system and is moreover only suitable for describing equilibrium situations.

Here, on the contrary, we want to consider the time-dependence of the expectation value of the observable:

$$S^z = \sum_i \sigma_i^z. \tag{10}$$

Let us now denote by $\langle S^z\rangle(0)$ the initially observed expectation value of S^z. We now have to specify the initial state of the system. As usual, many different density matrices $\rho(0)$ lead to the prescribed expectation value. If we now speak the language used, for instance, in Ref. 2, we say that these various $\rho(0)$ correspond to different microscopic states but are macroscopically equivalent. Each of them would, in principle, lead to a different time-dependence of $\langle S^z\rangle(t)$. We then have to make an assumption on the initial state we want to consider. The most reasonable choice for this is the state which maximizes the microsopic entropy, and is subject to the constraint:

$$\mathrm{Tr}\, S^z \rho(0) = \langle S^z\rangle(0). \tag{11}$$

If one takes the usual expression for the entropy, the solution of this problem is well known (it is just a transposition of one of the most satisfactory ways to derive the canonical distribution[3]):

$$\rho(0) = e^{-\zeta S^z}/\mathrm{Tr}\, e^{-\zeta S^z}, \tag{12}$$

where ζ is determined by the constraint (11). Incidentally, we remark that (12) could also be justified in a more traditional (but approximate) way, following more closely the actual preparation of the system in the laboratory as described in the first section (see again Ref. 1).

The problem now is to calculate

$$\langle S^z\rangle(t) = \mathrm{Tr}\, S^z \rho(t) \tag{13}$$

with

$$\rho(t) = U^t \rho(0) U^{-t}, \tag{14}$$

where $\rho(0)$ and U^t are prescribed by (12), (4), and (5). The expression (13) is more easily calculated if one writes it in the form

$$\langle S^z\rangle(t) = \mathrm{Tr}\, \{U^{-t} S^z U^t \rho(0)\}. \tag{15}$$

Since Tr is independent of the basis in which it is evaluated, we choose as a convenient basis:

$$\Psi_{\{\alpha_i\}} = \prod_i \otimes \psi_{\alpha_i}, \tag{16}$$

---

[2] G. Emch, Lecture notes, 8th Theoretical Physics Institute, University of Colorado, Summer 1965. Preprint *JILA*, University of Colorado and Natl. Bur. Std., Boulder, 1965. For further details see G. Emch, Helv. Phys. Acta **37**, 270 (1964); *ibid.*, **37**, 532 (1964); *ibid.*, **38**, 164 (1965), and references quoted therein. See also: G. Emch and C. Favre, Preprint Geneva 1965.

[3] J. von Neumann, *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955). See also the book published under the same title by G. W. Mackey (W. A. Benjamin Company, Inc., New York, 1963).

where $\{\psi_{\alpha_i}\}$ (with $\alpha_i = \pm 1$) is the basis defined in each $\mathfrak{H}^i$ by

$$\sigma^z \psi_{\alpha_i} = \alpha_i \psi_{\alpha_i}. \tag{17}$$

We next note that $\rho(0)$ is diagonal in the basis $\{\Psi_{\{\alpha_i\}}\}$, so that only the diagonal part of $(U^{-t}S^zU^t)$ is relevant for the evaluation of (15). Now, for any bounded operator A acting on $\mathfrak{H}$, it is convenient to define

$$\mathfrak{U}_n^t A \equiv U_n^{-t} A U_n^t. \tag{18}$$

Since the U_n^t commute among themselves,

$$\begin{aligned} U^{-t}S^zU^t &= \{\prod_{n\geq 0} \mathfrak{U}_n^t\} S^z \\ &= \sum_i \{\prod_{n\geq 0} \mathfrak{U}_n^t\} \sigma_i^z \\ &\equiv \mathfrak{U}^t S^z. \end{aligned}$$

Let us then first calculate

$$\mathfrak{U}_0^t \sigma_i^z = \{\prod_j \mathfrak{U}_{0,j}^t\} \sigma_i^z, \tag{19}$$

where

$$\mathfrak{U}_{0,j}^t A \equiv U_{0,j}^{-t} A U_{0,j}^t \tag{20}$$

with

$$U_{0,j}^t = \exp(+iB\sigma_j^z t). \tag{21}$$

Since

$$\sigma_j^z \text{ commutes with } \sigma_i^z \text{ for } j \neq i, \tag{22}$$

one has

$$\mathfrak{U}_0^t \sigma_i^z = \mathfrak{U}_{0,i}^t \sigma_i^z. \tag{23}$$

One can now use the property

$$(\sigma_i^z)^2 = I$$

to write

$$U_{0,i}^t = I \cos Bt + i\sigma_i^z \sin Bt. \tag{24}$$

Combining now (23) and (24) one gets:

$$\mathfrak{U}_0^t \sigma_i^z = \sigma_i^z \cos 2Bt + \sigma_i^y \sin 2Bt \tag{25}$$

and therefore

$$\begin{aligned} U^{-t}S^zU^t = (\{\prod_{n>0} \mathfrak{U}_n^t\} S^z) \cos 2Bt \\ + (\{\prod_{n>0} \mathfrak{U}_n^t\} S^y) \sin 2Bt. \end{aligned} \tag{26}$$

This takes care of the influence of the magnetic field. Next, we must calculate the influence of the spin–spin interaction. To do so, let us first evaluate

$$\mathfrak{U}_1^t \sigma_i^z = \{\prod_j \mathfrak{U}_{1,j}^t\} \sigma_i^z, \tag{27}$$

where $\mathfrak{U}_{1,j}^t$ is constructed in the usual way from

$$U_{1,j}^t = \exp[-i\epsilon(1)\sigma_j^z \sigma_{j+1}^z t]. \tag{28}$$

For reasons quite analogous to those encountered above, (27) reduces to

$$\mathfrak{U}_1^t \sigma_i^z = \mathfrak{U}_{1,i-1}^t \mathfrak{U}_{1,i}^t \sigma_i^z, \tag{29}$$

which can be readily calculated (using the same technique as for the B dependence) as:

$$\begin{aligned} \mathfrak{U}_1^t \sigma_i^z = \sigma_i^z \cos^2 [2\epsilon(1)t] \\ - \tfrac{1}{2}\sigma_i^y(\sigma_{i-1}^z + \sigma_{i+1}^z) \sin [4\epsilon(1)t] \\ - \sigma_{i-1}^z \sigma_i^x \sigma_{i+1}^z \sin^2 [2\epsilon(1)t]. \end{aligned} \tag{30}$$

At this point it becomes more and more intricate to write the successive explicit expressions for

$$\mathfrak{U}_2^t \mathfrak{U}_1^t \sigma_i^z, \qquad \mathfrak{U}_3^t \mathfrak{U}_2^t \mathfrak{U}_1^t \sigma_i^z, \qquad \text{etc.}$$

However, we should remember that we are not interested in the evaluation of the operator (26) for itself, but in the expectation value (15), to which only the diagonal part of (26) contributes. A somewhat closer glimpse to the form of the successive $\mathfrak{U}_n^t$ shows that the only part of

$$U^{-t}\sigma_i^z U^t$$

which contributes to (15) is

$$\sigma_i^z \prod_n \cos^2 2\epsilon(n)t.$$

The reader can rapidly convince himself that

$$U^{-t}\sigma_i^y U^t$$

cannot contribute to (15). We immediately have the desired result:

$$\langle S^z \rangle(t) = \langle S^z \rangle(0) [\prod_{n>0} \cos^2 2\epsilon(n)t] \cos 2Bt. \tag{31}$$

One also obtains

$$\langle S^y \rangle(t) = -\langle S^z \rangle(0) [\prod_{n>0} \cos^2 2\epsilon(n)t] \sin 2Bt, \tag{32}$$

$$\langle S^x \rangle(t) = \langle S^x \rangle(0) = 0. \tag{33}$$

The generalization of the above result from a one-dimensional chain to an n-dimensional crystal is obvious and can be taken care of simply by replacing $\epsilon(n)$ by ϵ_{ik} and making the subsequent trivial changes. The resulting expression is known and provides a qualitative (if not quantitative) agreement with experiment[1] if the ϵ_{ik} are properly adjusted. We gave the calculation with some details for the case of a linear chain with the following reasons:

(1) To emphasize that no approximation and no supplementary "statistical" assumptions are involved in the derivation of (31)–(33) when one starts from

(a) the generalized Ising-model microscopic evolution as prescribed by (4) and (5);

(b) the initial condition (12).

(This remains true also for any n-dimensional Ising-model generalized in the sense described above.)

(2) To provide a basis (in a hopefully intuitive language) for the generalized considerations, which is the topic of the following section.

III RELEVANCE TO THE GENERAL THEORY OF NONEQUILIBRIUM

A. Particularization of the model

The main purpose of Sec. II was to exhibit the presence in Eqs. (31)–(32) of the function

$$f(t) = \prod_{n>0} \cos^2 2\epsilon(n)t. \tag{34}$$

In general, one could try to discuss the behavior of this function for any $\epsilon(n)$ satisfying the very weak conditions (6). However, our purpose is not to produce the oddest analytical time behavior one could imagine from a particular model, but rather to exploit the greater simplicity of the model as extensively as possible in order to gain some insight into the general theory and the pseudoparadoxes and difficulties usually met. The very problem in which we are interested is to see how and why a macroscopic approach to equilibrium is compatible with a purely quantum mechanical microscopic description, in order to bring to light the central role played by the lack of information involved in any classical measurement on a quantum system.

We therefore postulate $\epsilon(n)$ to have the form

$$\epsilon(n) = \epsilon_0 2^{-n}. \tag{35}$$

The advantage of this choice is that $f(t)$ has an exceedingly simple form which can be deduced from a formula due to Euler:

$$f(t) = [\sin(\epsilon_0 t)/(\epsilon_0 t)]^2. \tag{36}$$

This is a positive function which obviously leads (with an oscillatory approach) to stationary value of $\langle S \rangle$:

$$\lim_{t\to\infty} f(t) \text{ exists and is zero!} \tag{37}$$

This result corresponds to an infinite linear chain with the spin-spin interaction extending over *all* pairs of neighbors. In the case of a finite ring, or equivalently, of an infinite chain with a cut-off described by

$$\epsilon(n) = 0 \quad \text{for all} \quad n > N, \tag{38}$$

one can also calculate exactly the corresponding

$$f_N(t) = \prod_{n=1}^{N} \cos^2 2\epsilon(n)t = f(t)W_N^{-1}(t), \tag{39}$$

where

$$W_N(t) = [\sin \epsilon(N)t/\epsilon(N)t]^2. \tag{40}$$

The function $W_N^{-1}(t)$ therefore, takes care of the "finite-size effects" (for the finite rings) or for the finite-extension of the interaction (for both finite rings and infinite linear chain). It shows that, for times $t \ll T_N$ with

$$T_N \approx 2^N \pi/\epsilon_0, \tag{41}$$

these effects are negligible. This remark is relevant for the recurrence problems and is confirmed by the fact that for finite times

$$\lim_{N\to\infty} W_N(t) = 1. \tag{42}$$

This, moreover, shows explicitly that, in the infinite-"volume" limit, the recurrence paradox resolves itself naturally without any recourse to more or less ill-defined probabilistic statements such as "only small deviations with the tendency of approaching equilibrium occur often."

B. Generalization of the model

One of the reasons why we explicitly did the calculations of Sec. II was to make obvious the fact that the model, as presented there, is open to generalizations. The first point we wish to make is that the only property we used of the initial state $\rho(0)$, as defined by (12), was the fact that $\rho(0)$ is diagonal in $\{\Psi_{\{\alpha i\}}\}$-basis (16). Therefore, the result (31)–(33) will remain true without any modification for quite a large class of initial states, namely for the $\rho(0)$ which are diagonal in (16).

Secondly, these results would also remain true if, instead of the one observable S^z, we would have been interested in any observable A of the form

$$A = \sum_i a_i \sigma_i^z. \tag{43}$$

Besides S^z, which is obviously recognized as a macroscopic observable, there is, therefore, a wide class of observables which also exhibit an approach to equilibrium, and are indeed related to a much finer

description of the system than the original one provided by S^z alone. In fact, the set generated by all of the observables of the form (43) is maximal Abelian. In a less pedantic language, one could simply express this property by saying that the set of the observables (43) is a (over) complete set of commuting observables (csco). More simply, the simultaneous measurement of all the observables which generate this set would lead to an information which cannot be improved by any compatible quantum measurement performed at that same instant. Using now a language even more familiar to statistical mechanicists, we could say that this generalized version of our model leads to an approach to equilibrium which does not require any real coarse-graining, or expressed in a better form, that the macrocell defined by the observation[2] are all one-dimensional. This corresponds to the situation usually referred to as *fine-graining*. That a definite approach to equilibrium is compatible with fine-graining illustrates a remark already made by Pauli[4] a long time ago. He emphasized that one of the few fundamental differences between classical and quantum statistical mechanics is that an information which is complete at a given instant remains so in time in the former description, whereas it can be lost in the latter. This is essentially due to the fact that, when the Hamiltonian does not commute with the csco considered, a state ρ, initially diagonal in the proper basis of this csco, does not remain diagonal; however, the nondiagonal elements of ρ which appears in the course of the evolution are of no relevance for the determination of the instantaneous expectation values of the observables belonging to the csco of interest. The generalized version of the model discussed in this subsection precisely provides an illustration of this remark. Incidentally, this is related to the fact that the quantum master equations, either fine- or coarse-grained, are formally identical.[2] We should also recall at this point that no random phase assumption is needed in the fine-grained case. This is consistent with the fact that the coarse-graining projector $\mathfrak{D}$, introduced in Ref. 2, coincides in this case (and in this case only) with the projector introduced by von Neumann[3] in his discussion of the measuring process.

C. Illustration of the coarse-graining concepts

We now consider the particular case of a (large but) finite ring, comprising say M sites whose positions are denoted by the index $i = 0, \cdots, M-1$. Each site is occupied by a $\frac{1}{2}$-spin particle. The time evolution of this system is again assumed to be given by (4) and (5) with the supplementary condition (35), complemented by any cut-off compatible with the ring structure. We moreover assume for simplicity that $B = 0$ (at least for $t > 0$). Now, instead of being interested in S^z we consider the following particular observable of the family (43)

$$A = \sigma_0^z = \sigma^z \otimes I \otimes \cdots \otimes I, \tag{44}$$

where σ^z is the usual two-dimensional Pauli matrix. The spectral decomposition of A is thus

$$A = \sum_{\Delta} A(\Delta) E_{\Delta} \tag{45}$$

with Δ running over the two indices $+$ and $-$. One has obviously

$$\begin{aligned} A(\pm) &= \pm 1, \\ E_{\pm} &= \tfrac{1}{2}(I \pm A), \end{aligned} \tag{46}$$

where $E_{\pm}$ are two orthogonal projectors with the following properties:

$$E_{\pm} \mathfrak{H}^i \subseteq \mathfrak{H}^i \tag{47}$$

(the equality sign is valid for all i except $i = 0$) and

$$\dim E_{\pm} = 2^{M-1}.$$

One can now use the language systematically established in Ref. 2 and say that the $E_{\pm}$ are the two *macrocells* of our system: An observation based on A only can provide an information on the relative populations of these macrocells, but cannot lead to any information on the inside of them. One can furthermore introduce the maximal representative of the two macroscopic equivalence classes which occur in this model:

$$W_{\Delta} = E_{\Delta}/2^{M-1}. \tag{48}$$

The macroscopic state $\{p_{\Delta}^t\}$ corresponding to any given microscopic state W^t is defined by:

$$p_{\Delta}^t = \operatorname{Tr} W^t E_{\Delta}. \tag{49}$$

Following the considerations developed in Ref. 2 we are only interested (for the prediction of the evolution of the expectations values of A) in the time-dependent p_{Δ}^t with the initial condition

$$W^0 = \sum_{\Delta} p_{\Delta}^0 W_{\Delta} \tag{50}$$

and the microscopic evolution equation

$$W^t = U^{-t} W^0 U^t.$$

[4] W. Pauli, Nuovo Cimento, Suppl. 6, 166 (1949).

This problem can be solved directly (without the need to go through the master equation techniques) following the calculations of Sec. II. We obtain

$$p_{\pm}^{t} = \tfrac{1}{2}(1 \pm \alpha f_N(t)), \tag{51}$$

where α is determined by

$$p_{\pm}^{0} = \tfrac{1}{2}(1 \pm \alpha). \tag{52}$$

In this form the model will lead to further considerations.

D. Ergodicity

The first problem of interest in the discusssion of nonequilibrium problems is to check whether a given system approaches equilibrium in some very restricted sense, namely

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, \langle A\rangle(t) = \overline{\langle A\rangle}, \tag{53}$$

where $\overline{\langle A\rangle}$ is the *microcanonical* equilibrium value of the observable A. Although it is clear that the left-hand side of Eq. (53) has hardly any physical (operational) sense in connection with the problem of the true approach to equilibrium, its evaluation has some relevance to the problem, for the following reason: if $\langle A\rangle(t)$ ever approaches a limit as $t \to \infty$, then this limit should be equal to the ergodic average of $\langle A\rangle(t)$ which by definition is the left-hand side of (53) [There is no point to discuss here ergodic limits other than the ordinary $(C, 1)$—Cesaro average.]

One could calculate the left-hand side of (53) directly from (51). One would then see immediately that (53) is satisfied when the length of the cut-off (described by N) and subsequently the size of the ring (described by M) both go to infinity

$$\lim_{t\to\infty} \lim_{N\to\infty} f_N(t) = \lim_{t\to\infty} f(t) = 0, \tag{54}$$

and consequently

$$\lim_{t\to\infty} \lim_{N\to\infty} p_{\pm}(t) = \tfrac{1}{2}, \tag{55}$$

which corresponds to the microcanonical distribution.

We now discuss another approach which can be illustrated by the model in the form discussed in Sec. III.C. A general criterion for (53) has been proposed[2] in the case of discrete spectrum of the Hamiltonian. This criterion reads as follows [see also (68) below]:

$$\sum_{\mu} \mathrm{Tr}\,(P_\mu W_\Delta P_\mu E_{\Delta'}) = N_{\Delta'}/N_3 \quad \text{for all} \quad E_\Delta, E_{\Delta'} \subset S_3, \tag{56}$$

where W_Δ is defined as above [by (48) in the present case], $N_\Delta = \dim E_\Delta$, N_3 is the dimension of the energy-shell S_3, and P_μ are the eigenprojectors of the total Hamiltonian.

To make this condition clear, we first have to state precisely what we mean by energy-shell. An energy-shell is defined as any eigenprojector of the following operator

$$H_0 \equiv \sum_{\Delta} (\mathrm{Tr}\, H W_\Delta) E_\Delta, \tag{57}$$

to which we refer as the *macroscopic energy* for reasons which have been explained.[2]

In the present model

$$H_0 = 0, \tag{58}$$

and therefore there is only *one* energy-shell: the whole space $\mathfrak{H}$ defined by (3). We then have

$$N_3 = 2^M. \tag{59}$$

To see whether (56) is satisfied or not, we proceed as follows: First, we remark that the whole Hamiltonian of the system as it results from the definition (4), (5) is not relevant for the evolution of our model, but can be replaced by the effective Hamiltonian

$$H_e = \sigma_0^z \sum_{n>0} \epsilon(n)(\sigma_n^z + \sigma_{M-n}^z) \tag{60}$$

without changing anything to the evolution, as it should be clear from the detailed calculations of Sec. II. This Hamiltonian is diagonal in the basis of $\mathfrak{H}$ defined by

$$\Phi_{\{\beta_i\}} = \prod_i \otimes \varphi_{\beta_i}, \tag{61}$$

where

$$\sigma^z \varphi_{\beta_i} = \beta_i \varphi_{\beta_i} \tag{62}$$

with

$$\beta_i = \pm 1. \tag{63}$$

Let us now remark that

$$P_{\{\beta_i\}} W_\pm P_{\{\beta_i\}} = P_{\{\beta_i\}}/N_3 \quad \text{for all} \quad \{\beta_i\}, \tag{64}$$

where $P_{\{\beta_i\}}$ are the one-dimensional projectors on the respective pure states $\Phi_{\{\beta_i\}}$.

Therefore, if the spectrum of H_e were nondegenerate, (56) and consequently (53) would be satisfied. However, in the model considered here, H_e is degenerate and, consequently, there appear terms of the form

$$P_{\{\beta_i\}} W_\pm P_{\{\beta'_i\}} \tag{65}$$

in the summation (56), where $\{\beta_i\}$ and $\{\beta_i'\}$ would correspond to configurations having the same total energy.

Given any configuration $\{\beta_i\}$, let us denote by $\{\beta_i\}^*$ the configuration obtained from $\{\beta_i\}$ by changing only β_0 into $(-\beta_0)$. Let us, furthermore, denote by p_{β_i} (resp. w_Δ) the restriction of the operator $P_{\{\beta_i\}}$ (resp. W_Δ) to the subspace $\mathfrak{H}^t$ (resp. $\mathfrak{H}^0$). We can now remark that (65) vanishes unless

$$\{\beta_i'\} = \{\beta_i\}^* \tag{66}$$

and that (65) is then equal to

$$(p_{\beta_0} w_\Delta p_{\beta_0}) \otimes (\prod_{i \neq 0} \otimes p_{\beta_i}). \tag{67}$$

From the symmetry of the model, it is, furthermore, evident that two configurations $\{\beta_i\}$ and $\{\beta_i\}^*$ can correspond to the same eigenvalue of H_e only if this eigenvalue is zero. Therefore, for all eigenprojectors P_μ corresponding to a nonzero eigenvalue of the effective Hamiltonian, we have

$$P_\mu W_\Delta P_\mu = P_\mu / N_3 \tag{68}$$

valid for all P_μ except P_0 (the eigenprojector corresponding to the eigenvalue zero of H_e). At this point, we draw the attention of the reader to the importance of Eq. (68) (satisfied for *all* P_μ!) in the general theory developed in Ref. 2 for the existence and the uniqueness of an equilibrium state. The fact that P_0 does not satisfy (68) allows us, then, to evaluate the departure from ergodicity (in the sense of Ref. 2) in our model. It is, in fact, a simple matter of playing with the degenerate eigenstates corresponding to the eigenvalue zero to evaluate the contribution of this level to the ergodic average. One finally obtains

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, p_\pm^t = \tfrac{1}{2}[1 \pm (\tfrac{1}{2})^N (p_+^0 - p_-^0)]. \tag{69}$$

This result completes (55), which was obtained more directly. We remark that, besides its theoretical interest, this evaluation of the ergodic average in the case of a finite ring (or of an infinite linear chain with a cut-off in the interaction) can be generalized quite easily to the case where $\epsilon(n)$ has a more general form than that assumed in (35). This allows us, in particular, to determine, without calculating the exact time dependence of the expectation value of the magnetization, the value around which it oscillates and which is usually referred to as its equilibrium value.[2] In the case of an infinite chain with an interaction extending to infinity, this equilibrium value is zero, the magnetization then oscillates above its equilibrium value and, in fact, approaches it with a damped oscillating time behavior. That this equilibrium value corresponds to the microcanonical distribution is readily seen.

In the above remarks we were careful to describe the approach to equilibrium in terms of expectation values of macroscopic observables. It is not by accident that we avoided statements in terms of microscopic density operators. The reason for this is that it is only in the macroscopic sense that the states considered here approach equilibrium. More precisely, given any state W^0 of the form (50), we do not assert that W^t approaches, in general, the microcanonical distribution in the course of time. We only claim that the macroscopic equivalence class of W^t [defined through (49) as the set $\{p_\Delta^t\}$] approaches the macroscopic equivalence class (defined by $p_\Delta = N_\Delta/N_3$) of the microcanonical distribution. This statement is obviously sufficient to account for the *macroscopic approach to microcanonical equilibrium*, even if no such approach can be traced when one uses only arguments pertaining to the microscopic description. By these considerations we hope to strongly emphasize the central role of the concept of macroscopic equivalence classes of microscopic states inherited from the notion of coarse-graining (even in cases where not all the assumptions usually made[2] are satisfied!).

E. Zermelo's and Loschmidt's Paradoxes

We already mentioned in Sec. III.A how this model allows us, without any recourse to probabilistic arguments, to eliminate the recurrence paradox by a passage to the limit of infinite size and infinite extension of the interaction, and then obtain a bona fide approach to equilibrium. This takes care of Zermelo's paradox in the sense predicted by Boltzmann (and not only in a probabilistic sense!) in his well-known exclamation "Then, wait!" The expression (40) states precisely "how long!"

This model also shows how an approach to equilibrium can be compatible with the reversibility of the microscopic evolution.

This reversibility manifests itself through

$$f(t) = f(-t) \tag{70}$$

and, incidentally, also through

$$f_N(t) = f_N(-t) \tag{71}$$

(which, however, is less surprising). Equation (70) can be interpreted in two ways. One could say that the behavior for $t < 0$ is simply irrelevant since it cannot be produced in the laboratory. This is,

however, not quite satisfactory and we feel that a better interpretation seems to be the following: Given a situation at time $t = 0$, we have the same information on "which situation it comes from" as we have on "which situation it will evolve to." In short, the security of a prediction is the same as that of a postdiction. This illustrates also the challenge raised when Boltzmann reportedly answered to the critics of Loschmidt's type: "Go ahead, reverse them." Our model indicates a slightly more sophisticated answer which we believe to be of quite a general character: If we "reverse the time" at a given instant, say $t = t_0$, taking only into account the macroscopic knowledge we have of the system at *this particular instant* $t = t_0$, we shall also have an approach to equilibrium in the direction of the past, and the system will in general not pass again through its initial state (the verification of this last statement cannot be made since it was decided, at $t = t_0$, that any information besides the one at that time $t = t_0$ is forgotten). On the contrary, if we keep records of what happened on the system between $t = 0$ and $t = t_0$, and then reverse the time at $t = t_0$, taking into account all the information we now have, the system will return to its initial state (at time $t = 0$) and then decay to equilibrium in the direction of the past in agreement with (70).

F. The Non-Markovian Character of the Evolution

In statistical mechanics, one of the favorite ways to obtain (with relative ease) an equation describing (at least within some approximation) the actual approach to equilibrium is to make an assumption of the kind of Boltzmann's *Stosszahlansatz*. The quantum analog of this assumption is the so-called repeated random phase assumption. Pauli used it to derive his famous master equation, hereafter referred to as the PEM. The nonmechanical character of this assumption is evident. Its result is that the reduction of the microscopic unitary evolution to the macroscopic subspace[2] becomes a semigroup, the generator of which is dissipative in case of an actual approach to equilibrium. The evolution obtained in this way is usually referred to as a Markovian process. The validity of this assumption has been extensively questioned in the last ten years.

The main property of the present model, in the different variations presented here, is that no repeated random phase is required to solve it exactly and to obtain an actual approach to equilibrium. The fact that such an assumption is not needed in the derivation of the result still does not prove that the evolution is non-Markovian. To show this we proceed *ad absurdo*. Suppose then that the evolution were Markovian. Under the usual continuity assumptions (which are satisfied here!) the evolution of a Markov process is determined by the Chapman–Kolmogorov equations. These equations are nothing but the mathematical abstraction corresponding to the PEM. In a perhaps not quite orthodox[2] form (but equivalent to the usual one), the PEM can be written as

$$(d/dt)p^t = -\Lambda p^t \tag{72}$$

defined only for $t \geq 0$. Λ is an operator acting in the *macroscopic* Liouville subspace.[2] Λ is positive and Hermitian in the usual PEM. When Λ is bounded, (72) can be integrated (without any "if" and "but") to give

$$p^t = e^{-\Lambda t}p^0. \tag{73}$$

If, moreover, Λ has a discrete spectrum, the p_Δ^t are discrete superpositions of nonincreasing real exponentials.

In the variation of our model considered in Sec. III.C, the macroscopic space is two-dimensional. Therefore Λ is obviously bounded and has at most two distinct eigenvalues. As a consequence, the p_Δ^t, if described by a PEM, should be the superposition of at most two nonincreasing real exponentials. We, however, have the explicit form (51) of the p_Δ^t. Even in case where both the size (M) of the system and the extension (N) of the interaction are infinite, these p_Δ^t cannot be written as linear combination of two nondecreasing real exponentials. Consequently, by this reduction *ad absurdo*, we proved that this variation of the model presents a macroscopic approach to equilibrium which cannot be described as a Markovian process. This result apparently depends on the form of the function $\epsilon(n)$ as ascribed by (35). It is, in fact, the expression of a general theorem.[2] One could now still argue further. Very often one sees that a system, the evolution of which has to be described by a generalized master equation (GEM), can, however, present an evolution which, in the long-time limit, can be described by a PEM, at least if one does some kind of time-smoothing.[2] This, however, is not the case here, as might readily be seen from the time-smoothed, long-time limit of (51), even if one first carries out the limit of infinite extension of the interaction. This last result now depends *strongly* on the special assumption (35). In particular, it can be seen that there exist some other particular spatial dependence of the interaction (which are indeed closer to the actual spatial dependence of the dipole-dipole in-

teraction!) which would lead to an asymptotically exponential decay to equilibrium.

From the point of view of the axiomatics, the interest of this version of the model, as discussed above, is that it exhibits explicitly an approach to equilibrium which can be predicted exactly, and which can in *no sense* be described (even in some approximation) as a Markov process, the differential evolution equation of which is of the PEM type. Therefore, if one wants to describe the evolution of the model through a master equation, one has to consider a GEM which cannot be approximated by a PEM. One could obviously carry out the evaluation of the kernel of the GEM for this model. This would, however, turn out to be a somewhat tedious task, especially in the case of an interaction extending in space as (35). The simplicity of the model would anyway be lost in the process as one could already figure out by feeding the result (51) into the Laplace transform of the GEM. Similarly, one could also discuss on this model van Hove's conditions of diagonal singularity and interconnection of states. This would, however, not lead to an information deeper than that obtained from the general theory.[2] We therefore do not pursue further in this direction here.

IV CONCLUSIONS

The formal simplicity of a generalized Ising model allowed us to calculate exactly the time evolution of the transverse magnetization as well as some other related quantities, starting from a wide class of initial conditions. The exact solubility of the model was used to discuss several problems connected with the actual approach to equilibrium. We succeeded in stating precisely some of the persistently unclear statements related to the phenomena.

The two main ingredients of nonequilibrium statistical mechanics, namely the unitary, mechanistic, microscopic evolution and the partial macroscopic information connected already with any classical measurement on a quantum system, were proved to be sufficient (in principle) to ensure in some cases an actual approach to equilibrium if one evaluates correctly the infinite-size limit. In particular, the nonnecessity of any kind of repeated random phase assumption was exemplified by a truly non-Markovian model.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor W. E. Brittin for the hospitality extended to him as a Visiting Fellow of the Joint Institute for Laboratory Astrophysics of the University of Colorado and National Bureau of Standards at Boulder, where this work was undertaken. He also wishes to acknowledge the discussions he had during this period with Dr. E. G. D. Cohen, Dr. R. L. Peterson, and Dr. R. J. Swenson.[5]

Thanks are also due to Dr. E. W. Montroll and Dr. N. S. Goel for their comments as well as for their reading of the manuscript.

---

[5] When this work was completed, Dr. R. J. Swenson drew the author's attention to the following paper: S. P. Heims, Am. J. Phys. **33**, 722 (1965), the content of which overlaps with the first two sections of the present report. These two sections have been maintained in this publication because of the fact that the preliminary calculations are presented here in a way to provide a guide for the subsequent considerations.

On the Poles of the S Matrix for Long-Range Potentials*†

ERASMO M. FERREIRA AND ANTONIO F. F. TEIXEIRA
Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro
(Received 8 February 1965)

The analytic properties in the complex k plane of the S matrix for scattering by a screened Coulomb potential are studied. Particular attention is given to the limit as the screening radius tends to infinity, so as to show, in an explicit example, the effect of the tail of the potential on the properties of the analytically extended S matrix. It is shown that the pole configuration obtained in this way is different from that obtained in the usual description of the analytic properties of the Coulomb S matrix.

1. INTRODUCTION

IT is well known that the analytic behavior of the nonrelativistic S matrix for infinite range potentials depends fundamentally on the way the potentials behave at large distances. In particular, this analytic behavior is strongly affected by the existence of an infinite tail, however thin it may be.

It seems to us, however, that this dependence has never been shown explicitly and conveniently in a meaningful example. A rectangular barrier (or well), whose range tends to infinity[1] (thus establishing a constant potential in all space) does not provide us with an interesting situation. A soluble and convenient problem for this purpose is that of a screened Coulomb potential. By varying the radius of the screening layer, we tend continuously to the Coulomb potential case. However, as we may see, the properties of the S matrix in the complex momentum plane, as obtained by this limiting process, are not the same as those obtained by the usual analytic extension of the Coulomb S matrix to the complex momentum plane.[2] We believe that the procedure of analytic extension used here is a more natural one, and that, if nothing else, we can learn from this example something about the treatment of this sort of infinite range potentials.

We take a potential of the form

$$\begin{aligned} V(r) &= \mu C/r, \qquad r < b, \\ &= 0, \qquad\quad r > b, \end{aligned} \tag{1}$$

where C is a positive quantity and $\mu = +1$ for repulsive, and $\mu = -1$ for attractive, potentials. We choose as unity the quantity $mC/\hbar^2 = 1/a_0$, where a_0 is the first Bohr radius for an attractive potential of strength C, and m is the mass of the particle. Thus we measure lengths (the range b, for example) in units of a_0 and the momentum k in units of $1/a_0$. The dimensionless product kb is independent of the potential strength and is a convenient variable for many purposes. We also introduce the quantity $\lambda = i\mu/k$, where $i = (-1)^{\frac{1}{2}}$. We call Re $(k) = x$, Im $(k) = y$, so that $k = x + iy$.

The S_l function for the lth partial wave can be written down directly. It is

$$S_l(k, b) = -Y_l^{(2)}(k, b)/Y_l^{(1)}(k, b), \tag{2}$$

where

$$\begin{aligned} Y_l^{(j)}(k, b) \equiv{}& (l + \tfrac{1}{2})\{{}_1F_1(l + 1 + \lambda; 2l + 1; -2ikb) \\ &\times [H_{l+\frac{1}{2}}^{(j)}(kb) - iH_{l-\frac{1}{2}}^{(j)}(kb)] \\ &+ {}_1F_1(l + \lambda; 2l + 1; -2ikb) \\ &\times [H_{l+\frac{1}{2}}^{(j)}(kb) + iH_{l-\frac{1}{2}}^{(j)}(kb)]\} \\ \equiv{}& -ikb\,{}_1F_1(l + 1 + \lambda; 2l + 2; -2ikb) \\ &\times [H_{l+\frac{1}{2}}^{(j)}(kb) - iH_{l-\frac{1}{2}}^{(j)}(kb)] + (2l + 1) \\ &\times {}_1F_1(l + \lambda; 2l + 1; -2ikb) H_{l+\frac{1}{2}}^{(j)}(kb) \end{aligned} \tag{3}$$

with $j = 1, 2$. The last expression is obtained from the second one by well-known relations among the confluent hypergeometric functions.

From the facts that ${}_1F_1(a; c; x)$ is a real function of its arguments, that the complex conjugate to $H_p^{(2)}(z)$ is $H_p^{(1)}(z^*)$, that

$$H_{l+\frac{1}{2}}^{(2)}(-z) = \exp[i\pi(l + \tfrac{1}{2})] H_{l+\frac{1}{2}}^{(1)}(z),$$

and the use of the Kummer transformation of the confluent hypergeometric functions, we see immediately that (2) satisfies the well-known[3]

* Most of this work was done while the authors were at the Escuela de Física y Matemáticas, Universidad Central de Venezuela, Caracas. They are greatly indebted to the kind treatment received there. Special thanks are given to the staff of their Centro de Cálculo for extensive use of the IBM 1620 computer.

† Partially supported by Conselho Nacional de Pesquisas, Brazil.

[1] H. M. Nussenzveig, Nucl. Phys. 11, 499 (1959).

[2] C. Moller, Kgl. Danske Videnskab. Selskab, Mat.-Fys. Medd. 22, No. 19 (1946).

[3] R. G. Newton, J. Math. Phys. 1, 319 (1960); A. Martin, Nuovo Cimento 14, 403 (1959).

symmetry properties of the S_l function such as $S_l(k, b)S_l(-k, b) = 1$, $S_l^*(k, b) = S_l(-k^*, b)$, and $S_l(k, b)S_l^*(k^*, b) = 1$. These relations imply that the zeros and poles of $S_l(k, b)$ in the complex k plane are symmetric with respect to the imaginary axis, and that if k is a zero of $S_l(k, b)$, k^* is a pole of the same function. Thus we need only study the half-plane defined by Re $k \geq 0$.

We are now considering *finite* (nonzero, noninfinite) values of b. If Im $(k) \to +\infty$, $S_l(k, b)$ behaves like exp $[2b$ Im $(k)]$, thys presenting the well-known essential singularities for Im $(k) \to +\infty$. In the lower k plane, $S_l(k), b)$ is a meromorphic function, with infinitely many poles; the larger the value of |Re (k)|, the lower these poles are located in the k plane. For $k \to 0$, b finite, and also for $b \to 0$, k finite, we have $S_l(k, b) \to 1$, as shown later. There can be no poles of $S_l(k, b)$ in the upper half-plane, except on the imaginary axis. All these are well-known general properties of finite range potentials.[3]

In Sec. 2 of this paper we discuss the asymptotic behavior of the wavefunctions when $b \to \infty$. In Secs. 3–6 we discuss the distribution and displacements in the complex k plane of the poles of $S_l(k, b)$ as a function of the range b. Since $Y_l^{(i)}(k, b)$ is a regular function of k (except at the origin), the singularities of $S_l(k, b)$ will be poles due to zeros in the denominator $Y_l^{(1)}(k, b)$. Our general equation for the poles will then be $Y_l^{(1)}(k, b) = 0$.

2. ASYMPTOTIC COULOMB WAVEFUNCTIONS FOR COMPLEX k

For $|kb| \gg 1$ the asymptotic (large r) l wavefunction for the Schrödinger equation with potential (1) is

$$R_l(r) \sim \frac{1}{kr}\left\{e^{ik(r-b)}\left[\frac{e^{ikb}(2ikb)^{-l-\lambda}}{\Gamma(l+1-\lambda)} + \frac{e^{-ikb}\lambda(-2ikb)^{-l-1+\lambda}}{\Gamma(l+1+\lambda)}\right] + e^{-ik(r-b)}\left[\frac{e^{ikb}\lambda(2ikb)^{-l-1-\lambda}}{\Gamma(l+1-\lambda)} - \frac{e^{-ikb}(-2ikb)^{-l+\lambda}}{\Gamma(l+1+\lambda)}\right]\right\}. \quad (4)$$

If in (4) we just substitute $b = r$ and keep only the dominating terms we obtain the asymptotic Coulomb wavefunction

$$R_l^{\text{Coul}}(r) \sim \frac{1}{kr}\left[e^{ikr}\frac{(2ikr)^{-l-\lambda}}{\Gamma(l+1-\lambda)} - e^{-ikr}\frac{(-2ikr)^{-l+\lambda}}{\Gamma(l+1+\lambda)}\right] \quad (5)$$

and consequently the Coulomb S_l function

$$S_l^{\text{Coul}}(r) = \frac{\Gamma(l+1+\lambda)}{\Gamma(l+1-\lambda)}(-2ikr)^{-2\lambda}.$$

However, this substitution is quite arbitrary, since for $|kb| \gg 1$, (4) is valid for all $r > b$ and the definition of S_l is independent of r. Putting $r = b$ in (4) implies in that we have made r and b tend together to infinity, with $r - b = 0$. This is not the case in our problem, where we have the solution for a finite range potential whose range b is then allowed to increase. We might as well keep $r - b$, for instance, finite and nonzero. The two ways of taking the limit may lead to different properties of the resulting S_l function.

So, let us keep $r - b$ finite while $r, b \to \infty$.

If k is real the exponentials appearing in (4) are just oscillating functions and the limit $b \to \infty$ will give, selecting the dominating terms,

$$R_l(r) \sim \frac{1}{kr}\left[e^{ikr}\frac{(2ikb)^{-l-\lambda}}{\Gamma(l+1-\lambda)} - e^{-ikr}\frac{(-2ikb)^{-l+\lambda}}{\Gamma(l+1+\lambda)}\right], \quad \text{real } k, \quad (5')$$

which is again the asymptotic Coulomb wavefunction (put $b = r$ inside the brackets), and gives the usual Coulomb S_l function. So, nothing is different for the case k is real.

Let us now assume that Im $(k) \neq 0$ in (4). Now, there will be real exponentials, and, at least for values of k not in the neighborhood of poles of $\Gamma(l + 1 + \lambda)$ or $\Gamma(l + 1 - \lambda)$, we find, when taking the limits $r, b \to \infty$, the dominating terms to be those containing positive exponents. Thus we will have

$$R_l(r) \sim \frac{1}{kr}\left[e^{ikr}\frac{\lambda(-2ikb)^{-l-1+\lambda}e^{-2ikb}}{\Gamma(l+1+\lambda)} - e^{-ikr}\frac{(-2ikb)^{-l+\lambda}}{\Gamma(l+1+\lambda)}\right] \quad (6)$$

for Im $(k) > 0$ and

$$R_l(r) \sim \frac{1}{kr}\left[e^{ikr}\frac{(2ikb)^{-l-\lambda}}{\Gamma(l+1-\lambda)} + e^{-ikr}\frac{\lambda(2ikb)^{-l-1-\lambda}e^{2ikb}}{\Gamma(l+1-\lambda)}\right] \quad (7)$$

for Im $(k) < 0$ with the possible exception of isolated points for which $\Gamma(l + 1 + \lambda)$ and $\Gamma(l + 1 - \lambda)$ increase without limit.

None of these two formulas (6) and (7) for Im $(k) \neq 0$ reproduces the Coulomb wavefunction. One of them, Eq. (6), contains the Coulomb incoming

wave, combined with a different outgoing part, the other (7) contains the Coulomb outgoing wave combined with a different incoming part. These wavefunctions (6) and (7) or the corresponding S_l functions show the behavior we obtain if we solve the problem for finite range b, then taking larger and larger values of b. Perhaps this would correspond more appropriately to a physical situation, where a truly infinite Coulomb potential cannot really exist.

We can easily see that the expressions (6) and (7) for the asymptotic wavefunctions do not allow for poles in the corresponding S_l function. Thus, as $b \to \infty$ poles can only be located in points where these expressions are not valid.

Let us return to (4). The position of the poles of S_l will be determined by equating to zero the coefficient of $\exp(-ikr)$, that is, by

$$e^{2ikb}\lambda(2ikb)^{-2\lambda-1}\Gamma(l+1+\lambda) = (-1)^{-l+\lambda}\Gamma(l+1-\lambda). \quad (8)$$

Using this relation to evaluate the coefficient of $\exp(+ikr)$ in (4), we then obtain the fact that, near a pole of S_l, the asymptotic wavefunction in the limit $b \to \infty$ behaves like

$$R_l^{\text{pole}}(r) \sim \frac{1}{kr}\left\{e^{ikr}\frac{(2ikb)^{-l-\lambda}}{\Gamma(l+1-\lambda)} + e^{-ikr}\left[\frac{e^{2ikb}\lambda(2ikb)^{-l-1-\lambda}}{\Gamma(l+1-\lambda)} - \frac{(-2ikb)^{-l+\lambda}}{\Gamma(l+1+\lambda)}\right]\right\}. \quad (9)$$

If $Im\ (k) > 0$, Eq. (8) has solutions only for points in the k plane which tend to the poles of $\Gamma(l+1+\lambda)$ as $b \to \infty$. For Im $(k) < 0$ the poles of S_l tend to the poles of $\Gamma(l+1-\lambda)$. Thus the structure of the pole distribution for the S_l function obtained this way is different from that obtained with formula (5). We now have the fact that, for an attractive potential, there will be poles in both the positive and negative imaginary axis, for

$$\text{Im}\ (k) = \pm(l+1+n)^{-1},$$

where $n = 0, 1, 2 \cdots$. For a repulsive potential, the poles can only move to the origin in the k plane, where the above formulas are not valid in general, as we see later. We remind here that, in the usual discussion of the Coulomb potential problem,[4] there appear poles in the positive imaginary axis for the attractive potential, and in the negative imaginary axis for the repulsive potential. We must remark that there is no different behavior concerning the bound-state poles (those on the positive imaginary axis).

3. THE LIMIT b → 0

From (2) it is easy to see that when the range b tends to zero with k noninfinite, $S_l(k, b)$ tends to one. This follows immediately from the fact that ${}_1F_1(a^{-1}; c; ax) \to 1$ as $x \to 0$ and that $H_p^{(1)}(z)/H_p^{(2)}(z)$ goes to 1 when $z \to 0$. Thus, when the range of the potential tends to zero, there can be poles of the S_l function only for $|k| \to \infty$.

We can now ask what happens in the variable kb. To keep kb finite when $b \to 0$ we must have $|k| \to \infty$, but taking this limit $|k| \to \infty$ in (3) is equivalent to make $\lambda \to 0$, or in other words, to reduce to zero the strength of the potential. We expect that S_l goes to zero in this limit. In the Appendix we show that this is in fact true.

Thus, there can be no poles in the finite kb plane when $b \to 0$, the poles are being pushed to infinity when this limit is taken. To find how the poles displace themselves, we can use well-known asymptotic expressions, valid for large kb, for the Hankel and hypergeometric functions occurring in $Y_l^{(1)}(k, b)$. The asymptotic formula for the confluent hypergeometric function is

$${}_1F_1(a; c; z) \underset{|z|\to\infty}{\to} \frac{\Gamma(c)}{\Gamma(c-a)}(-z)^{-a}[1 + O(z^{-1})] + \frac{\Gamma(c)}{\Gamma(a)}e^z z^{a-c}[1 + O(z^{-1})]. \quad (10)$$

We obtain after a lengthy but straightforward calculation

$$Y_l^{(1)}(k, b) \sim (2/\pi kb)^{\frac{1}{2}}\exp(-ikb)[(2l+1)!/l!]i^{-l}k^{-1} \times (2ikb)^{-l-1}[(-1)^l(2k^2b - \mu\exp(2ikb)]. \quad (11)$$

The equation for the poles is then

$$\mu\exp(2ikb) = (-1)^l 2k^2 b, \quad (12)$$

which has solutions

$$xb = [2n + l + \tfrac{1}{2}(\mu+1)](\tfrac{1}{2}\pi), \quad (13)$$

$$yb \to -\infty, \quad \text{with} \quad \exp(-2yb) = 2y^2 b, \quad (13')$$

where n is a positive or negative integer, or zero.

This result is similar to the one obtained by Nussenzveig[1] in the case of a rectangular well or barrier and by Humblet[5] in a more general case. It says that, in the limit of very short range or very weak screened Coulomb potential, the poles tend to $yb \to -\infty$, approaching the asymptote lines

[4] H. M. Nussenzveig, *Analytic Properties of Non-Relativistic Scattering Amplitudes*, Lecture Notes, Escuela Latino americana de Física, Universidad de México, 1962.

[5] J. Humblet, Mém. Soc. Roy. Sci. Liège 12, No. 4, 70 (1952).

$xb = \pm N(\frac{1}{2}\pi)$, where N is odd for l odd (even) and N is even for l even (odd) in the case of attractive (repulsive) potentials. Thus the odd l-wave poles in the attractive and the even l-wave poles in the repulsive potentials both tend to the lines $xb = \pm\frac{1}{2}\pi$, $\pm\frac{3}{2}\pi$, $\cdots$, while in the other cases the poles tend to $xb = 0, \pm\pi, \cdots$. This is in fact a particular example of a quite general property of weak potentials.[4] This behavior is shown in the curves of Secs. 5 and 6, where trajectories described by the poles when b varies are drawn.

The asymptotic behavior of the numerator $Y_l^{(2)}(k, b)$ when $b \to 0$ is given by

$$Y_l^{(2)}(k, b) \sim (2/\pi kb)^{\frac{1}{2}} \exp(-ikb)[(2l+1)!/l!]i^l k^{-1} \times (-2ikb)^{-l-1}[(-1)^l 2k^2 b - \mu \exp(-2ikb)] \quad (14)$$

and then

$$S_l(k, b) \sim -i/[2b(k - k_0)] \quad (15)$$

in the neighborhood of a pole k_0. Then in the plane kb the residues of the poles tend to $-\frac{1}{2}i$, and in the k plane they increase as $-i/2b$ when $b \to 0$. It is interesting to note that, for the residues, these results do not depend on l. Also Eq. (13′), which determines how fast the poles are pushed to infinity, does not depend on the particular pole (that is, on N) neither on l. This means that the centrifugal barrier at the origin has no effect in the strength of the poles for a limiting short-range potential nor in the speed with which they are sent away to infinity. Since the centrifugal barriers are just equivalent to a potential behaving like C/r^2 at the origin, we can predict that, for this sort of potentials whose range is made to approach zero, the S_l function will have the same pole structure as given by (13), (13′), and (15), except that the asymptote lines will be given by (13) with l modified so as to include the effect of the potential, that is $l(l+1)$ is substituted by $l(l+1) + C$. We must notice that, in this case of potentials behaving like C/r^2 at the origin, we obtain different behavior of the poles in the two cases where we reduce the strength and where we reduce the range to zero.

4. THE LIMIT $b \to \infty$

When taking the limit $b \to \infty$ we have several cases to consider. We may have k finite (nonzero) and thus $|kb| \to \infty$, or we may have $k \to 0$. In this last case we may have kb finite or kb increasing to infinity. Let us consider each of these possibilities separately. This is essential, as the asymptotic behavior of the functions involved are different under these different conditions.

(a) We keep k finite, nonzero, and let b increase without limit. Since $|kb| \to \infty$ and all other parameters are limited, we may use well-known asymptotic expressions for the Hankel and hypergeometric functions. We now show that the S_l function tends to limits which, in general, are different from the usual Coulomb S_l function. The results here indicated are in agreement with those of Sec. 2.

Writing $Y_l^{(i)}(k, b)$ in the form

$$Y_l^{(i)}(k, b) = \{-ikb[H_{l+\frac{1}{2}}^{(i)}(kb) - iH_{l-\frac{1}{2}}^{(i)}(kb)] + (l+\lambda)H_{l+\frac{1}{2}}^{(i)}(kb)\}{}_1F_1(l+1+\lambda; 2l+2; -2ikb) + (l+1-\lambda)H_{l+\frac{1}{2}}^{(i)}(kb){}_1F_1(l+\lambda; 2l+2; -2ikb), \quad (16)$$

we obtain, for $|kb| \gg 1$,

$$Y_l^{(2)}(k, b) \sim (2/\pi kb)^{\frac{1}{2}} \exp(-ikb)i^{l+1} \times \{\lambda_1 F_1(l+1+\lambda; 2l+2; -2ikb) + (l+1-\lambda){}_1F_1(l+\lambda; 2l+2; -2ikb)\} \quad (17)$$

and

$$Y_l^{(1)}(k, b) \sim (2/\pi kb)^{\frac{1}{2}} \exp(-ikb)i^{-l-1} \times \{(l+1+\lambda){}_1F_1(l-\lambda; 2l+2; 2ikb) - \lambda_1 F_1(l+1-\lambda; 2l+2; 2ikb)\}. \quad (18)$$

We can prove that the asymptotic forms of the ${}_1F_1$'s which occur in both expressions (17) and (18) are such that the first one predominates over the second when Im $(k) > 0$ and the second ${}_1F_1$ dominates when Im $(k) < 0$. This dominance may not be true for the isolated points where the dominating ${}_1F_1$ happens to be zero (this will happen in the poles and in the zeros of S_l). In these cases we have to keep the second dominating terms in the expressions for $Y_l^{(1)}$ and $Y_l^{(2)}$. The expression for the S_l function which is valid in all cases is

$$S_l(k) = -\frac{[(2ikb)^{-l-\lambda}/\Gamma(l+1-\lambda)] + [\lambda(-2ikb)^{-l-1+\lambda}\exp(-2ikb)/\Gamma(l+1+\lambda)]}{[(-2ikb)^{-l+\lambda}/\Gamma(l+1+\lambda)] - [\lambda(2ikb)^{-l-1-\lambda}\exp(2ikb)/\Gamma(l+1-\lambda)]}. \quad (19)$$

For Im $(k) > 0$ and for $|kb| \gg 1$, $S_l(k, b)$ behaves like

$$S_l(k, b) \underset{b\to\infty}{\sim} -\lambda(-2ikb)^{-1}\exp(-2ikb), \text{ Im } (k) > 0 \quad (19')$$

in all points not in the neighborhood of the zeros of the denominator in (19). Thus, $S_l(k, b)$ explodes exponentially as $b \to \infty$ in the whole upper k plane. On the other hand, for Im $(k) < 0$, except in the neighborhood of zeros of the denominator in (19), $S_l(k, b)$ behaves like

$$S_l(k, b) \underset{b\to\infty}{\sim} \lambda^{-1}(2ikb) \exp(-2ikb), \qquad \text{Im}(k) < 0. \tag{19''}$$

Thus, when $b \to \infty$, S_l tends to zero for every finite k in the lower k plane, with the possible exception of isolated points.

Let us discuss the behavior of the poles of $S_l(k, b)$ as b increases to infinity. For Im $(k) > 0$ the exponential in the denominator of (19) will tend to zero as b gets large, and the poles will then move to the points where $\Gamma(l + 1 + \lambda)$ also gets large, that is, to the points such that $l + 1 + \lambda = -n$, with $n = 0, 1, \cdots$. For Im $(k) < 0$ the exponential in the denominator in (19) increases with b, and the poles must tend to the points for which $\Gamma(l + 1 - \lambda)$ is also large. Thus the poles in the lower k plane will move to the points given by $l + 1 - \lambda = -n$. Since $\lambda = i\mu/k$, μ being the sign of the potential, we see that in the attractive case ($\mu = -1$) there will be poles in both positive and negative imaginary axis, while for a repulsive ($\mu = 1$) potential in the limit $b \to \infty$ there will be no poles with finite nonzero k. This is in agreement with what has been said in Sec. 2. We should remark that (19) still obeys the symmetry properties $S_l(k, b) = S_l^{-1}(-k, b)$ and $S_l^*(k, b) = S_l(-k^*, b)$.

The residues of the bound-state poles are given by

$$\frac{iy^2(2yb)^{2y-1}}{(y^{-1} - l - 1)!\,(y^{-1} + l)!}, \qquad y = \text{Im}(k) > 0,$$

and they increase as b increases. The residues of the poles in the negative imaginary axis for large b values are given by

$$\frac{-iy^4(2yb)^{2-2y-1}\exp(4yb)}{(-y^{-1} - l - 1)!\,(-y^{-1} + l)!}, \qquad y = \text{Im}(k) < 0,$$

and thus they tend to zero as b increases.

(b) Now, let us consider the case in which k goes to zero while $b \to \infty$, with kb being finite. In these conditions the first parameter in ${}_1F_1(a; c; z)$ is large, while the second parameter and the variable are limited. We must then use the asymptotic expressions valid under these conditions.[6] We obtain for $|k| \ll 1$, $|kb|$ finite,

$$\begin{aligned} Y_l^{(j)}(k, b) &\approx (2l + 1)!\,\pi^{-\frac{1}{2}} \exp(-ikb)(2b)^{-l-\frac{1}{2}} \\ &\times \exp[-i(l + \tfrac{1}{4})(3 + \mu)(\tfrac{1}{2}\pi)](kb[H_{l+\frac{1}{2}}^{(j)}(kb) \\ &+ iH_{l-\frac{1}{2}}^{(j)}(kb)]\{\exp[2i(-2\mu b)^{\frac{1}{2}}] \\ &+ \exp(i3\pi/2 - 2i[(-2\mu b)^{\frac{1}{2}}]\} + H_{l+\frac{1}{2}}^{(j)}(kb) \\ &\times \exp[i(1 + u)\pi/4](2b)^{\frac{1}{2}}\{\exp[2i(-2\mu b)^{\frac{1}{2}}] \\ &+ \exp[i\pi/2 - 2i(-2\mu b)^{\frac{1}{2}}]\}). \end{aligned} \tag{20}$$

It is now easy to see that in both cases of $\mu = \pm 1$, the second part in the above expression dominates over the first. This means that, in the conditions studied here, we may simplify $Y_l^{(j)}(k, b)$ to

$$\begin{aligned} Y_l^{(j)}(k, b) &\approx (2l + 1)\exp(-2ikb) \\ &\times {}_1F_1(l - \lambda; 2l + 1; 2ikb)H_{l+\frac{1}{2}}^{(j)}(kb), \quad |k| \ll 1. \end{aligned} \tag{21}$$

The S_l function will then be

$$S_l(k, b) \approx -H_{l+\frac{1}{2}}^{(2)}(kb)/H_{l+\frac{1}{2}}^{(1)}(kb). \tag{22}$$

The result is that when $b \to \infty$ there will be poles in the plane kb at the points given by the roots $H_{l+\frac{1}{2}}^{(1)}(kb) = 0$. The residues of these poles are not zero, since the roots of $H_{l+\frac{1}{2}}^{(1)}(kb)$ and of $H_{l+\frac{1}{2}}^{(2)}(kb)$ never coincide (they are symmetric with respect to the origin). $H_{l+\frac{1}{2}}^{(1)}(kb)$ has roots only for Im $(kb) < 0$ and $H_{l+\frac{1}{2}}^{(2)}(kb)$ only for Im $(kb) > 0$. For $l = 0$ there will be no such poles. For $l = 1$ we have a pole at $kb = -i$, for $l = 2$ there are poles at $kb = \mp\frac{3}{2} - i\sqrt{3}/2$, and so on. There will be l poles of this kind for a given l wave. For odd l values one pole will be on the imaginary axis and the $(l - 1)$ remaining ones will be distributed symmetrically with respect to this axis. For even l values there will be no poles on the imaginary axis, since $H_{l+\frac{1}{2}}^{(1)}(kb)$ does not admit double roots. In the k plane all these poles tend to the origin. This pole structure does not depend on whether the potential is attractive or repulsive. These conclusions are confirmed by the numerical calculations and are shown clearly in figures of the next sections. These poles are responsible for the essential singularity of S_l in the Coulomb case.

(c) Only one possibility remains to be discussed in this limit $b \to \infty$, that in which $k \to 0$ and $|kb| \to \infty$. We now have both the first parameter a and the variable z in the functions ${}_1F_1(a; c; z)$ increase without limit. The asymptotic expressions for the attractive and repulsive cases are different, and also different expressions have to be used for different regions of the complex plane.[7] We then

[6] Bateman Manuscript Project, *Higher Transcendental Functions*, (McGraw-Hill Book Company, Inc., New York, 1953), Vol. I., Chap. VI.

[7] L. J. Slater, *Confluent Hypergeometric Functions* (Cambridge University Press, New York, 1960), p. 85.

go directly to the points we wish to demonstrate, avoiding more general calculations which are not absolutely necessary.

As we see in the next section, the behavior of the poles in the attractive potential is already completely described in terms of the cases previously analyzed. We then specialize to the repulsive case. We show that there exists an infinite number of poles such that $|k^2b|$ is kept finite while $k \to 0$.

We want to study the behavior of the S_l function in the right-hand side of the lower half-plane of the variable k, that is, we have

$$-\tfrac{1}{2}\pi < \arg(k) < 0, \qquad 0 < \arg(2ikb) < \tfrac{1}{2}\pi.$$

In the denominator of the S_l function enters two hypergeometric functions,

$$F_1 = {}_1F_1(l + 1 - \lambda; 2l + 2; 2ikb)$$

and

$$F_2 = {}_1F_1(l - \lambda; 2l + 1; 2ikb).$$

By using the appropriate asymptotic expressions (7), we can prove, after a rather long but straightforward calculation, that kbF_1 and F_2 are of the same order of magnitude in the limits considered, as long as k^2b is a finite quantity, that is we have $O(kbF_1/F_2) = (k^2b)^{\frac{1}{2}}$. Since F_1 and F_2 appear in $Y_l^{(1)}(k, b)$ in the combination

$$Y_l^{(1)}(k, b) = \exp(-2ikb)\{ikb[H_{l+\frac{1}{2}}^{(1)}(kb) + iH_{l-\frac{1}{2}}^{(1)}(kb)]F_1 + (2l + 1)H_{l+\frac{1}{2}}^{(1)}(kb)F_2\}, \tag{23}$$

and since

$$H_{l+\frac{1}{2}}^{(1)}(kb)/[H_{l+\frac{1}{2}}^{(1)}(kb) + iH_{l-\frac{1}{2}}^{(1)}(kb)] \approx ikb/l,$$

when $(kb) \to \infty$, the term containing F_2 will dominate over that containing F_1.

A similar evaluation can be performed with the terms contributing to $Y_l^{(2)}(k, b)$. We now take the form

$$Y_l^{(2)}(k, b) = \exp(-2ikb)\{-ikb[H_{l+\frac{1}{2}}^{(2)}(kb) - iH_{l-\frac{1}{2}}^{(2)}(kb)]F_1 + (2l + 1)H_{l+\frac{1}{2}}^{(2)}(kb)F_3\}, \tag{24}$$

where $F_3 = {}_1F_1(l + 1 - \lambda; 2l + 1; 2ikb)$ and prove that $O(kbF_1/F_3) = (k^2b)^{\frac{1}{2}}$.

Now, since $H_{l+\frac{1}{2}}^{(2)}(kb)$ is larger than $H_{l+\frac{1}{2}}^{(2)}(kb) - iH_{l-\frac{1}{2}}^{(2)}(kb)$ when $|kb| \to \infty$, the term containing F_3 dominates over the other, and the S_l function becomes

$$S_l(k, b) \underset{\substack{b\to\infty \\ k^2b \text{ finite}}}{\to} H_{l+\frac{1}{2}}^{(2)}(kb)\,{}_1F_1(l + \lambda; 2l + 1; -2ikb)/[\exp(-2ikb)H_{l+\frac{1}{2}}^{(1)}(kb)\,{}_1F_1(l - \lambda; 2l + 1; 2ikb)]. \tag{25}$$

This satisfies the symmetry properties known for the S matrix, mentioned in Sec. 1.

With this result our problem reduces to the search for the existence of zeros of ${}_1F_1(l - \lambda; 2l + 1; 2ikb)$ in the conditions considered, that is, with $k \to 0$, $b \to \infty$, $t = 2ikb/[4(\lambda + \frac{1}{2})]$ finite, $0 \leq \arg(i/k + \frac{1}{2}) \leq \frac{1}{2}\pi$. We can then use the appropriate asymptotic formula for ${}_1F_1$, and equate it to zero.

We have[7]

$$F_2 \underset{\substack{k\to 0 \\ k^2b \text{ finite}}}{\approx} \exp(2ikb)(2ikb)^{-l-\frac{1}{2}}\Gamma(2l + 1) \times [1 - (2\lambda + 1)/ikb]^{-\frac{1}{2}}A_l(k, b), \tag{26}$$

where

$$A_l(k, b) = (\lambda + \tfrac{1}{2})^{-\lambda-\frac{1}{2}} \exp(E + \lambda + \tfrac{1}{2})/\Gamma(l - \lambda) + (\lambda + \tfrac{1}{2})^{\lambda+\frac{1}{2}} \exp(-E - \lambda - \tfrac{1}{2}) \times \exp[i\pi(l - \lambda)]/\Gamma(l + 1 + \lambda) \tag{27}$$

and

$$E = (2\lambda + 1)\{t^{\frac{1}{2}}(t - 1)^{\frac{1}{2}} + \log[t^{\frac{1}{2}} - (t - 1)^{\frac{1}{2}}]\}. \tag{28}$$

Using the asymptotic limit

$$\Gamma(z) \approx (2\pi)^{\frac{1}{2}}z^{z-\frac{1}{2}}e^{-z} \quad (\text{large } |z|)$$

for the Γ functions and noticing that λ is large [so that $(\lambda + \frac{1}{2})^{\lambda+\frac{1}{2}} \approx \lambda^{\lambda+\frac{1}{2}}e^{\frac{1}{2}}$] we obtain

$$A_l(k, b) \approx (2\pi)^{-\frac{1}{2}}\lambda^{-l}(-1)^l \times \exp(-i\pi\lambda - E)[1 + \exp(2E - \tfrac{1}{2}i\pi)]. \tag{29}$$

Thus the necessary condition for the existence of poles in the S_l function is then

$$\text{Re}(E) = 0,$$
$$\text{Im}(E) = 3\pi/4 + m\pi, \qquad m \text{ integer}.$$

We first notice that these equations determining the poles are independent of l. We now look for solutions of these equations such that $y/x \to 0$ while $x, y \to 0$. In these conditions the system of equations is

$$(\tfrac{1}{2} + y/x^2)[(\tfrac{1}{2}bx^2)^{\frac{1}{2}}(\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}} + \log|(\tfrac{1}{2}bx^2)^{\frac{1}{2}} - (\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}}|] - (1 + 4y/x^2)(\tfrac{1}{2}bx^2)^{\frac{1}{2}}(\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}} = 0, \tag{30}$$

$$(2/x)[(\tfrac{1}{2}bx^2)^{\frac{1}{2}}(\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}} + \log|(\tfrac{1}{2}bx^2)^{\frac{1}{2}} - (\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}}|] + (\tfrac{1}{2}x)(\tfrac{1}{2} + y/x^2)(1 + 4y/x^2)(\tfrac{1}{2}bx^2)^{\frac{1}{2}} \times (\tfrac{1}{2}bx^2 - 1)^{\frac{1}{2}} = 3\pi/4 + m\pi. \tag{31}$$

As $x \to 0$ the first term in (31) increases without limit, unless $\frac{1}{2}bx^2$ tends fast to one. In this case (31) becomes simply

$$(\tfrac{1}{2}bx^2 - 1)/x = 3\pi/4 + m\pi \qquad (32)$$

which gives an explicit solution

$$xb = (2b)^{\frac{1}{2}} + 3\pi/4 + m\pi \qquad (33)$$

for one of the coodinates of the poles. Taking (33) into (30) we obtain

$$yb \to -\tfrac{1}{2}. \qquad (34)$$

Thus the result is that, for repulsive potentials, we have an infinite number of poles which, as b increases, tend to infinity in the kb plane, approaching asymptotically the line $yb = -\frac{1}{2}$. The existence and position of these poles are independent of the value of the angular momentum. All this can be seen in the curves of Sec. 6.

5. THE POLES OF THE S MATRIX FOR ATTRACTIVE POTENTIALS

We have seen that (13) and (13′) show that for increasing b the poles approach lines parallel to the imaginary axis in the kb plane. For curves with even l values the poles approach the lines $xb = N(\frac{1}{2}\pi)$, with $N = 0, \pm 2, \pm 4$, and so on. For the odd waves the asymptotes are given by the same formula, but with $N = \pm 1, \pm 3, \pm 5$, and so on. We label a pole by the number N that defines the asymptote of its trajectory when $b \to 0$.

From (13′) it can be shown that as $b \to 0$ we have $d(yb)/d(\log b) = \frac{1}{2}$. This means that dividing b by a given factor implies that all the poles go down the same vertical distance in the plane kb. This can easily be observed in Figs. 1, 3, and 5, where the trajectories of the poles for the attractive s, p, and d waves are shown respectively. Since Eq. (13′) is also valid for repulsive potentials, this behavior can also be observed in the corresponding curves for the repulsive case (Fig. 7).

As b increases, the poles move upwards in the kb plane, and in the attractive case they turn so as to run towards the imaginary axis.

(a) The s-wave poles

We first describe in detail the behavior of the s-wave poles. There is a pole ($N = 0$) which moves along the imaginary axis, from $kb = -i\infty$ to $kb = +i\infty$. The pair of poles ($N = \pm 2$) coming from the asymptotes $xb = \pm\pi$ reach the imaginary axis in the point $yb = -1.5774$ for $b = 3.4115$.

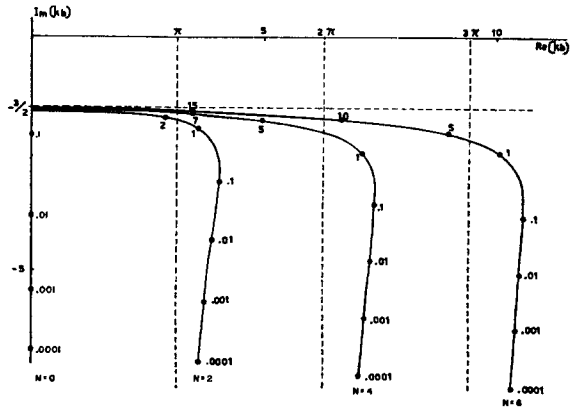

FIG. 1. Poles for the S matrix for s-wave scattering by a screened attractive Coulomb potential of range b. The values of b are shown on the curves. The trajectory $N = 0$ is always on the imaginary axis. Each trajectory has a symmetric one for negative values of Re(kb). After two symmetric poles reach the imaginary axis as b increases, one moves upwards and the other moves downwards along the imaginary axis.

All pairs of symmetric poles meet at the imaginary axis; as b is further increased one pole goes up along the axis towards $yb = \infty$, while the other moves downwards to $yb = -\infty$. The poles $N = \pm 4$ reach the imaginary axis in the point $yb = 1.5227$ for $b = 8.9781$. For $N = \pm 6$ this happens for $yb = -1.5101$ and $b = 17.001$.

The most distant (large $|N|$) poles reach the imaginary axis for larger and larger values of b. We now prove that as $|N| \to \infty$, the poles reach the axis in points closer and closer to $yb = -1.5$. The denominator of the S_l function for $l = 0$ is $Y_0^{(1)}(k, b) = {}_1F_1(i/k; 1; 2ikb)$. Poles on the imaginary axis are determined by ${}_1F_1(1/y; 1; -2yb) = 0$. This gives y as a function of b. For the points in which the poles leave the imaginary axis, we must have $db/dy \equiv -(\partial F/\partial y)/(\partial F/\partial b) = 0$, and thus these points are determined by the simultaneous solution of these two equations. We want to study the distant poles, i.e., these which enter the axis for large b. We can try a solution with $y \to 0$, yb finite. In these conditions the hypergeometric functions can be expanded in terms of a series of Bessel functions,

$${}_1F_1(1/y; 1; -2yb) \approx \exp(-yb)(J_0\{2[(2 - y)b]^{\frac{1}{2}}\} + (b^{\frac{1}{2}}2^{\frac{1}{2}}/6)y^2 J_1\{2[(2 - y)b]^{\frac{1}{2}}\}). \qquad (35)$$

Thus for small y the "distant" poles must satisfy

$$J_0\{2[(2 - y)b]^{\frac{1}{2}}\} = 0. \qquad (36)$$

From (35) we obtain

$$\partial {}_1F_1/\partial y = \exp(-yb)(\tfrac{1}{2}b)^{\frac{1}{2}} \times (1 + \tfrac{2}{3}by)J_1\{2[(2 - y)b]^{\frac{1}{2}}\}. \qquad (37)$$

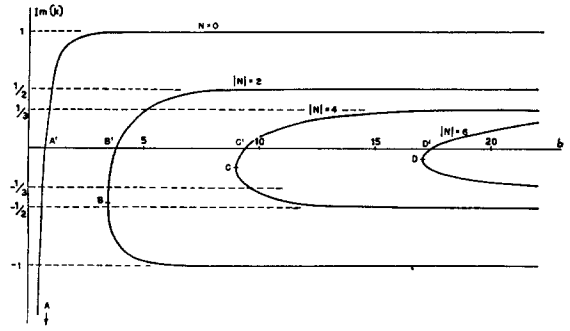

FIG. 2. Purely imaginary poles for s-wave scattering by an attractive screened Coulomb potential. The points A, B, C, D show the values for which the poles reach the imaginary axis. A', B', C', D' correspond to bound states of zero binding energy. The asymptotes Im $(k) = 2/(|N| + 2)$ give the binding energies of the Rydberg formula. The poles that tend to Im $(k) = -2/|N|$ are not present in the usual analytic extension of the Coulomb S matrix

Since J_0 and J_1 cannot be simultaneously zero, we must then have $yb = -\frac{3}{2}$ as the only possibility to satisfy simultaneously the two equations. This is what we wanted to prove. All the poles (except the pole $N = 0$) enter the imaginary axis for values of yb between -1.5774 and -1.5. From (36) we then obtain that the values of b for which the distant poles enter the imaginary axis are approximately given by the larger roots of $J_0[2(2b + \frac{3}{2})^{\frac{1}{2}}] = 0$. The pole $N = 0$ can be said to enter the imaginary axis for $yb = -\infty$.

When a pole crosses the origin and enters the positive imaginary axis a new bound state is formed. The values of b for which this happens can be determined in the following way. For $y \to 0$, b finite we have

$$ {}_1F_1(1/y; 1; -2yb) \approx J_0[2(2b)^{\frac{1}{2}}] + y(\tfrac{1}{2}b)^{\frac{1}{2}}J_1[2(2b)^{\frac{1}{2}}]. \quad (38)$$

Thus, new bound states arise whenever $2(2b)^{\frac{1}{2}}$ reaches a root of the Bessel function of order zero. b is measured in units of the Bohr radius $a_0 = \hbar^2/mZe^2$.

The displacement of the poles along the imaginary axis can be better observed in Fig. 2, where Im (k) is plotted against b. Only the purely imaginary poles are represented.

For values of b corresponding to the vertices A, B, C, D new poles reach the imaginary axis. Every time one of these values is reached, two new purely imaginary poles arise. This always happens for a value of Im (k) such that $2/|N| <$ Im $(k) < 0$. One of these poles moves downwards along the imaginary axis and tends asymptotically to the point Im $(k) = -2/|N|$ as $b \to \infty$. The "twin" pole moves upwards, and as $b \to \infty$ it approaches asymptotically the point Im $(k) = +2/(|N| + 2)$. Thus in the attractive Coulomb potential limit the pole configuration in the k plane in the s-wave case is the following: there are poles in the points of the positive imaginary axis, corresponding to the usual bound states [Im $(k) = 2/(|N| + 2)$, $|N| = 0, 1, 2 \cdots$] and in points of the negative imaginary axis [given by Im $(k) = -2/|N|$, $|N| = 1, 2, \cdots$]. We again remark that this is not the same on the usual pole description of the S_0 function for Coulomb potential, where the poles in the negative imaginary axis are not present.

It is interesting to note how the bound-state poles tend to the points determined by the Rydberg formula as the range b increases. According to (38) new bound states (with zero binding energy) appear for $b = 0.74a_0$, $3.74a_0$, $9.33a_0$, $17.35a_0$, that is, for values of the range close to the values of the Bohr radius n^2a_0. As b increases from this value, the binding energy tends to the maximum values, given by Rydberg formula. In the first level (the ground state, $N = 0$) the maximum is reached rapidly: for b equal to $2a_0$ the binding energy is less than 0.5% different from the limit value. Thus this binding energy is not much affected by the existence of a tail in the potential. The same is true of the other bound states: if the range of the potential is twice as large as the range necessary to create the bound state, the binding energy and the position of the two "twin" poles are almost the same as if the tail were complete.

From (38), which is valid for finite b and small y, we can see that $(\partial F/\partial y)_{y=0} \neq 0$. This means that the s-wave poles for attractive potential do not enter the imaginary axis in the origin $k = 0$. In other words, the vertices A', B', C', D' of the curves in Fig. 2 do not coincide with the points A, B, C, D where bound states are formed. Comparing the equation $J_0[2(2b + \frac{3}{2})^{\frac{1}{2}}] = 0$ which determines the vertices for distant poles and the equation $J_0[2(2b)^{\frac{1}{2}}] = 0$ which determines the values of b for which the poles cross the origin, we see that the two values of b tend to differ by 0.75 for the very distant poles.

(b) The p-wave poles

According to (13), for small values of b the poles are close to the lines

$$\text{Re } (kb) = \tfrac{1}{2}N\pi \quad (N = \pm 1, \pm 3, \pm 5, \cdots).$$

As b increases the poles move towards the origin. This is shown in Fig. 3, where the trajectories in

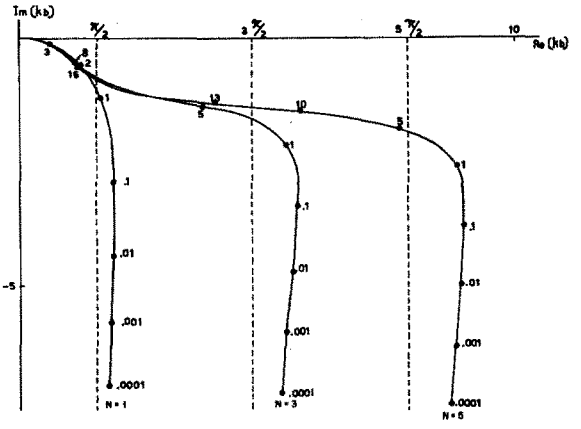

FIG. 3. Trajectories of the poles for p-wave scattering by attractive potentials. The values of the range b are indicated on the curves. All trajectories enter the imaginary axis at the origin, and then run upwards and downwards along the imaginary axis.

the kb plane are drawn. For certain values of b two symmetric poles reach the imaginary axis at the origin.

With increasing b one pole moves upwards along the positive imaginary axis to $yb = +\infty$, which other moves downwards to $yb = -\infty$. That the entry point of all the poles is at the origin can be proved in the following way. For small y the pole defining equation $Y_l^{(1)}(k, b) = 0$ becomes (if $l \neq 0$),

$$J_{2l}[2(2b)^{\frac{1}{2}}] + y^2 p(l, b) J_{2l+1}[2(2b)^{\frac{1}{2}}] = 0, \qquad (39)$$

where

$$p(l, b) = (l + 1)(4l^2 - 1 + 4b)(2b)^{\frac{1}{2}}/[12(2l - 1)]. \qquad (40)$$

Thus the values of b for which the poles pass the origin are given by the roots of

$$J_{2l}[2(2b)^{\frac{1}{2}}] = 0. \qquad (41)$$

Due to the square dependence on y, we have that in that points also $(\partial Y_l^{(1)}(k, b)/\partial y)_{y=0} = 0$. This means that the origin is also the point in which the trajectory enter the imaginary axis. Clearly these results are valid for all $l \neq 0$.

Call b_n a solution of $J_{2l}[2(2b)^{\frac{1}{2}}] = 0$. Since $p(l, b)$ is a positive definite quantity and since

$$J_{2l}[2(2b)^{\frac{1}{2}}] \underset{b \to b_n}{\longrightarrow} -[2l/(2b_n)^{\frac{1}{2}}](b - b_n) J_{2l+1}[2(2b_n)^{\frac{1}{2}}],$$

we can conclude that the curves $y = y(b)$ defined by (39) have, for small y, concavities directed towards the large values of b. Clearly these results are valid for all $l \neq 0$. All this can be better seen in Fig. 4, where Im (k) is plotted against b, and the purely imaginary poles are indicated. Near the points where the curves $y = y(b)$ cross the b axis, they are symmetric with respect to y (due to the square dependence mentioned above). Thus the vertices coincide with the axis.

The lower parts of the curves in Fig. 4 present plateaux which all occur in the neighborhood of $kb = -i$. This point has the character of a "sink," attracting the poles: large variation in the value of b is necessary to remove a pole from the neighborhood of this point of the kb plane. All the poles are "attracted" by this point. For b large enough, we can say that almost always there will be a pole around $kb = -i$. This is the pole corresponding to the solution of $H_{3/2}^{(1)}(kb) = 0$, as found in Sec. 4b.

The bound-state poles again tend rapidly to the Rydberg values Im $(k) = 2/(|N| + 3)$ as b increases. Unlike the s-wave case, the "twin" poles tend to symmetric points Im $(k) = -2/(|N| + 3)$.

(c) The d-wave poles

We have here a few complications as compared to the previous cases.

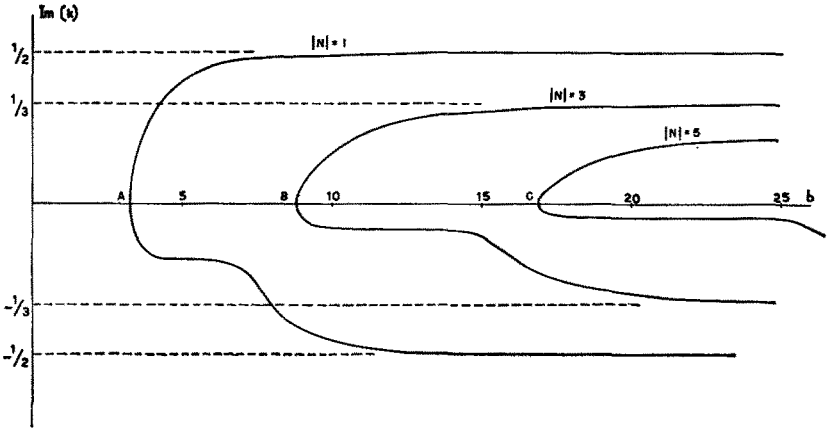

FIG. 4. Purely imaginary poles for p-wave scattering by attractive potentials. The points A, B, C give the values of b for which the poles reach the origin and new bound states are formed. As $b \to \infty$ the bound-state poles tend to the values of Im (k) corresponding to the binding energies given by Rydberg formula. In the Coulomb limit there are symmetric poles in the negative imaginary axis.

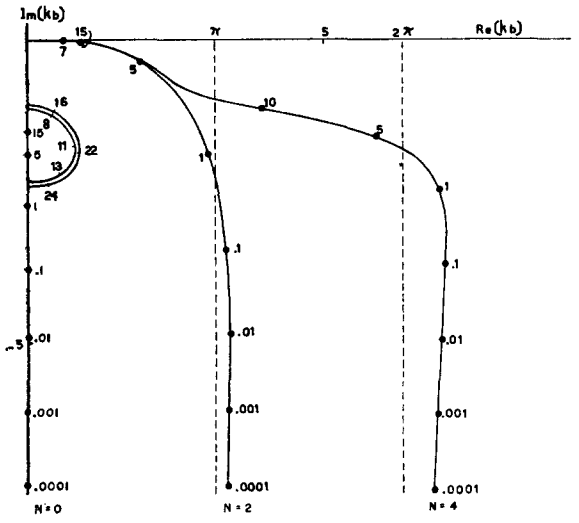

FIG. 5. d-wave poles for attractive potentials. As b increases the complex poles move towards the origin, and then follow the imaginary axis. Those which run along the negative imaginary axis pass again to the complex plane, describing "semicirles," and turning back to the imaginary axis. When $b \to \infty$ there will be poles on the imaginary axis and at the points $\mp \frac{1}{2}\sqrt{3} - \frac{3}{2}i$. See text for detailed description.

According to (13), for small b the poles are close to lines $xb = \frac{1}{2}N\pi$, with $N = 0, \pm 2, \pm 4$, etc. For increasing b the complex poles ($N \neq 0$) move in the complex plane (see Fig. 5) and reach the origin in pairs, for values of b given by (41).

While one of the two poles $|N| = 2$ moves along the positive imaginary axis to $yb = \infty$, the other moves downwards along the negative imaginary axis. In the meantime the pole $N = 0$ is climbing up the negative imaginary axis. For $b = 7.54$ the two poles meet each other in the point $yb = -1.08$. As b is further increased, they leave the imaginary axis and pass to the complex plane, one for each side, symmetrically. The symmetry of the S_l function is thus not destroyed. The two poles describe two 'semicircles," and join again in the imaginary axis in the point $yb = -2.35$ for $b = 13.6$. Now, one of them chooses to go down the imaginary axis, running to $yb = -\infty$, $y = -\frac{1}{3}$, as b increases. The other one (we cannot tell which one) goes up the imaginary axis, traveling towards the origin. In the meantime, the pair of poles with $|N| = 4$ has arrived at the origin, and decided that one of them would go to $yb = +\infty$ ($y = +\frac{1}{4}$) and the other would travel down the imaginary axis. When $b = 15.6$ this pole reaches the point $yb = -1.04$, and there it meets the pole which was just going up the imaginary axis after having described the semicircles already described. Since these two poles meet there for the same value of b, they can pass symmetrically to the complex plane without destroying the right–left symmetry of the S_l function. They do it, describing semicircles in the complex plane. The process is thus repeated continuously.

It can be remarked that the poles, when describing the semicircles, are very much slowed down when near the points $kb = \mp\frac{1}{2}\sqrt{3} - \frac{3}{2}i$. This effect is more and more important as $|N|$ becomes larger. This is analogous to what happened with the point $kb = -i$ in the p-wave case, and we can say that as $b \to \infty$, there will always be found a pole in those points. This is in agreement with the results obtained in Sec. 4b.

In Fig. 6 this rather complicated situation can be observed in a different way. There we plot Im (k) against b. The poles on the imaginary axis are represented by full lines, the complex poles by dotted lines. We have an infinite number of valid labeling of certain lines, since after two poles join each other we cannot tell which is which. Two of the simplest descriptions are attempted in Fig. 6. In one of them the pole $N = 0$ never goes to infinity: it is always describing semicircles, taking charge of meeting other poles to keep symmetry of the S_l function. As $b \to \infty$ it tends to be retained by one of the two points which are the roots of $H^{(1)}_{5/2}(kb) = 0$. In the other interpretation each pole $N \neq 0$ describes semicircles twice: one time symmetrically to a pole of smaller $|N|$, another time to the next larger one. In this case the $N = 0$ describes only one semicircle.

6. POLES FOR REPULSIVE POTENTIALS

For repulsive potentials the behavior of the poles is rather simple. In Fig. 7 are drawn the trajectories of the s-, p-, and d-wave poles in the kb plane.

For small values of b the trajectories are close to the vertical lines $xb = \frac{1}{2}N\pi$, in agreement with (13′). As b increases an infinite number of poles

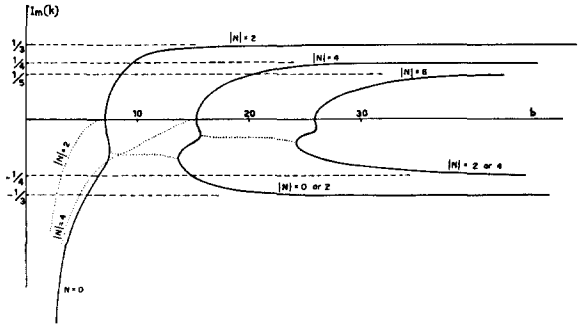

FIG. 6. Poles of the S-matrix for d-wave scattering by an attractive potential. In full lines are represented the poles on the imaginary axis, in dotted lines the imaginary part of the complex poles.

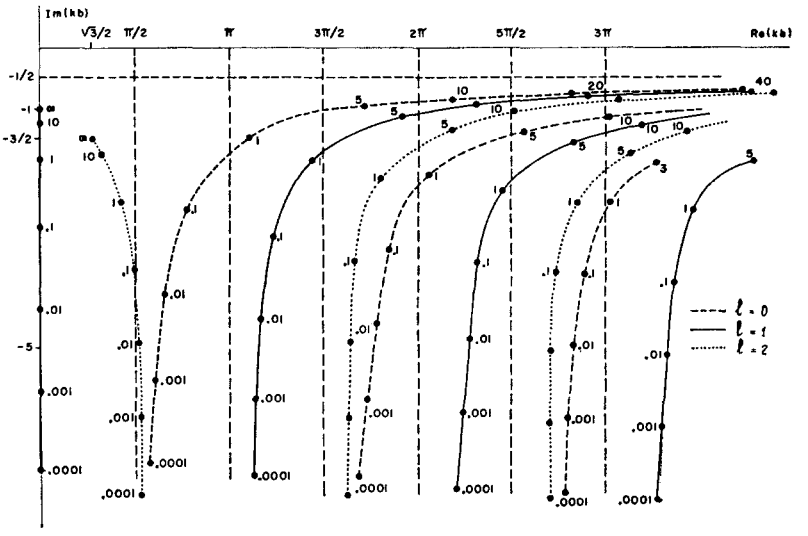

FIG. 7. s-, p-, and d-wave poles for repulsive screened Coulomb potentials. All poles which move to infinity tend asymptotically to the line Im $(kb) = -0.5$, with $k^2b \to 2$ and Re $(kb) = (2b)^{1/2} + \frac{3}{4}\pi + m\pi$ $(m = 1, 2,...)$. There are poles which tend to the points defined by $H^{(1)}_{l+1/2}(kb) = 0$ as $b \to \infty$.

turn towards large values of Re (kb), approaching asymptotically the line $yb = -0.5$. In all of them $k^2b \to 2$ as $b \to \infty$. They form groups, consisting of one trajectory for each l value which go together to infinity. Each group tending to keep a distance π (in the kb plane) from the next one. All this is in agreement with the results of item c in Sec. 4.

For each value of l there are l trajectories (counting those in both left- and right-hand side planes) with an special behavior: they remain in the finite kb plane as $b \to \infty$, tending asymptotically to the points which are the roots of $H^{(1)}_{l+\frac{1}{2}}(kb) = 0$. These solutions have been mentioned in Sec. 4b, and are clearly shown in Fig. 7. Only in these trajectories do the several waves differ from each other when b is large.

In the k plane all the poles go asymptotically to the origin as $b \to \infty$. Those which correspond to the trajectories which go to infinity in the kb plane, approach the origin taking the x axis as a tangent, since for them $y/x \to 0$. The l trajectories which end in points with finite $|kb|$ are the only ones which do not take the x axis as a tangent direction.

APPENDIX I. PROOF THAT $S_l(k,b) \to 1$ WHEN $b \to 0$ WITH kb FINITE

Taking the functional relation among confluent hypergeometric functions

$$ {}_1F_1(a+1;2a+1;z) + {}_1F_1(a;2a+1;z) = 2\,{}_1F_1(a;2a;z) $$

multiplying by

$$ (\tfrac{1}{2}z)\,{}_1F_1(a+1;2a+2;z) = (a+\tfrac{1}{2})[{}_1F_1(a+1;2a+1;z) - {}_1F_1(a;2a+1;z)] $$

and rearranging terms we obtain

$$ -\frac{{}_1F_1(a+1;2a+1;z)}{{}_1F_1(a;2a+1;z)} = \frac{(\frac{1}{2}z)\,{}_1F_1(a+1;2a+2;z) + (2a+1)\,{}_1F_1(a;2a;z)}{(\frac{1}{2}z)\,{}_1F_1(a+1;2a+2;z) - (2a+1)\,{}_1F_1(a;2a;z)}. $$

The convenience of this is that we obtained a relation between the two F's that appear in $S_l(k, b; \lambda = 0)$ (put $a = l$, $z = -2ikb$) and functions of the type ${}_1F_1(p; 2p; z)$ which can be expressed in terms of Bessel functions. We obtain

$$ -\frac{{}_1F_1(a+1;2a+1;z)}{{}_1F_1(a;2a+1;z)} = \frac{J_{a+\frac{1}{2}}(\frac{1}{2}iz) + iJ_{a-\frac{1}{2}}(\frac{1}{2}iz)}{J_{a+\frac{1}{2}}(\frac{1}{2}iz) - iJ_{a-\frac{1}{2}}(\frac{1}{2}iz)}, \qquad \text{(A1)} $$

which can also be written

$$ -\frac{{}_1F_1(a+1;2a+1;z)}{{}_1F_1(a;2a+1;z)} = \frac{[H^{(1)}_{a+\frac{1}{2}}(\frac{1}{2}iz) + iH^{(1)}_{a-\frac{1}{2}}(\frac{1}{2}iz)] + [H^{(2)}_{a+\frac{1}{2}}(\frac{1}{2}iz) + iH^{(2)}_{a-\frac{1}{2}}(\frac{1}{2}iz)]}{[H^{(1)}_{a+\frac{1}{2}}(\frac{1}{2}iz) - iH^{(1)}_{a-\frac{1}{2}}(\frac{1}{2}iz)] + [H^{(2)}_{a+\frac{1}{2}}(\frac{1}{2}iz) - iH^{(2)}_{a-\frac{1}{2}}(\frac{1}{2}iz)]}, $$

which is enough to prove that $S_l(k, b) \to 1$ when $\lambda \to 0$. Since this corresponds to $|k| \to \infty$, it also implies in that $b \to 0$, since $|kb|$ is supposed to be kept finite.

Kinks*

DAVID FINKELSTEIN†
Belfer Graduate School of Science, Yeshiva University, New York

(Received 22 December 1965)

In sufficiently nonlinear field theories there are extended objects whose number is strictly conserved because of continuity of the underlying field as a function of space. We call these *kinks*. Kinks provide a covariant description of extended but indestructible particles. We give the properties the field theory must possess in order for kinks to exist, and the circumstances under which kinks can have spin ½.

THE strict conservation laws of elementary particle physics correspond to exact symmetries of the underlying quantum field. In some cases the symmetries have an obvious universal significance, and the correspondence is an especially fruitful one, such as for angular momentum and charge, understood in terms of Lorentz and gauge invariance. In other cases the correspondence serves as little more than a convenient transcription of our experimental knowledge. The three conserved fermion numbers A, N_e, N_μ (atomic or baryon number, electron number, and muon number) are of this more mysterious kind. The gauge transformations they generate may have no corresponding gauge field.

In classical particle mechanics, the conservation of the number of particles is different from other conservation laws. There, particle conservation is a consequence of the continuity of the particle trajectories, and is thus valid in a much wider variety of circumstances than other conservation laws.

In sufficiently nonlinear field theories, there are also objects whose number is strictly conserved because of continuity, (a continuity, however, of the basic fields rather than of trajectories). We call these conserved objects *kinks*, and seek properties that the underlying field must possess for kinks to exist and to possess half-integer spin and Fermi–Dirac statistics.

Kinks provide a description of particles with an internal structure, distributed over a finite volume rather than concentrated at one point. Many attempts at such a description within nonlinear field theory rely on a tendency of the field to clot or clump as a result of self-interactions, a nonlinear dynamical process. Such theories of particles, however, do not seem to account for the conservation of certain particle numbers, which are remarkably exact empirical conservation laws. Given enough energy in such theories, one clump can eventually become two, and there is no discontinuity dividing the one from the two. Not so with kinks.

The simplest example of a kink was already noted by Skyrme[1] in a one-dimensional model arising in his nonlinear theory of strong interactions. A three-dimensional kink occurs in general relativity,[2] and Skyrme's one-dimensional theory was rediscovered in the course of studying this gravitational kink. Skyrme's one-dimensional theory arose yet again in the work of Enz,[3] who regarded it as a field theory of spin waves, rather than isospin waves. Skyrme has also investigated a three-dimensional kink.[1]

To avoid misunderstanding, I explicitly abjure two heresies:

The topological heresy. In this theory, particles are not topological deformities of space–time. To be sure, kinks are topological in some sense. But the kinks treated here have nothing to do with any topological deformity of space–time. On the other hand, kinks still occur and indeed in greater variety, if there are admitted topological space–time deformities.

The quantum heresy. This is not a classical theory of quantum effects. To be sure, kinks are "quantized" in some sense: they are discrete in number. But their behavior exhibits no trace of quantum mechanics when it is described in a completely classical field theory. On the other hand, kinks still occur if complementarity is explicitly taken into account and the field is treated quantum mechanically; then, of course, they exhibit the usual quantum effects.

* This work was supported by the National Science Foundation. I am also grateful to Tougaloo College for the hospitality afforded me during part of this work.

† Young Men's Philanthropic League Professor of Physics.

[1] T. H. R. Skyrme, Proc. Roy. Soc. (London) **A 252**, 236 (1959). For subsequent work of Skyrme see Nucl. Phys. **31**, 556 (1962). I am indebted to Julio Rubinstein for these references, and for frequent discussions.

[2] D. Finkelstein and C. W. Misner, Ann. Phys. (N. Y.) **6**, 230 (1959). General criteria for the existence and spin of kinks are stated here.

[3] V. Enz, Phys. Rev. **131**, 19392 (1963).

I. CLASSICAL FIELDS

1. First, we required some concepts concerning a classical field $\varphi(x)$. At each point x of space–time the field assumes values φ forming an n-dimensional "field manifold" Φ. The value of the field φ may be represented by n real numbers φ_α forming a coordinate system on Φ. We do *not* assume that one coordinate system φ_α is able to cover the entire field manifold Φ; this assumption would restrict us to theories that are trivial from the point of view of the topological considerations presented here. On the other hand, we *do* assume that the space–time variable x ranges over the four-dimensional number space $R^4 \equiv X$.

What is important here is the topology of the dependent (field) variable, not the independent (space–time) variable.

In most field theories treated in the literature, Φ is assumed to be not merely a manifold but also a vector space. Whenever we use perturbation theory and expand about one point (the "origin") of Φ, we thereby effectively replace Φ by a vector space, the tangent space to Φ at the origin. Thus the global structure of Φ is actually relatively unexplored at present and cannot be explored by perturbation methods. The properties of fields we are concerned with here depend entirely on the global structure of Φ.

2. We suppose that the classical field equations define a continuous evolution $\varphi(x) = \varphi(\mathbf{x}, t)$ of the field starting from any sufficiently differentiable initial values $\varphi(\mathbf{x}, 0), \dot{\varphi}(\mathbf{x}, 0), \cdots$ of the field and enough of its derivatives at time $t = 0$. The derivatives $\dot{\varphi}(\mathbf{x}, 0)$ are objects of the differential geometry on the manifold Φ and define tangent vectors to the Φ manifold at $\varphi(x, 0)$.

3. Boundary conditions are essential for the conservation laws we are dealing with, since they are for the more familiar ones, but these boundary conditions can be relatively weak. We suppose all physical fields $\varphi(x)$ approach some fixed value φ_0 as $|\mathbf{x}| \to \infty$. Other boundary conditions would also serve our purpose, which is to prevent the escape of the kink to infinity, resulting in an apparent nonconservation.

4. In calling φ a classical field we also mean to imply that φ has a single-valued law of transformation under the Poincaré group. Indeed, a main goal of the present work is to ascertain which classical fields can give rise to spin, in the sense of a double-valued law of rotation, when the canonical quantization is performed.

II. PARTICLE NUMBERS

1. We wish now to define *homotopic conservation laws*. By way of prelude we go back to the particle number of classical mechanics.

If Q^n is the configuration space of n particles, then the configuration space of an unspecified number of particles is the union

$$Q = \bigcup_n Q_n.$$

In this space the particle number n is a nontrivial dynamical variable, i.e., nonconstant with values 0, 1, $\cdots$ in the corresponding regions of Q. However, the configuration space Q is disconnected, and there is no continuous curve that joins a point of Q to one with a different number of particles. In particular, a trajectory, assumed continuous, cannot pass from one value of n to another.

This re-expression of the trivial fact that n does not change in the usual classical mechanics is intended to alert us to the possibility and the significance of a disconnected configuration space. Whenever the configuration space Q admits a resolution $Q = \bigcup Q_n$ into components Q_n as in the above example, there is present a strong conservation law, independent of symmetries of the classical theory, and dependent only on continuity.

2. In casting classical field theory as Lagrangian theory, we employ as configuration space Q a function space whose points are functions $\varphi(\mathbf{x})$ representing possible fields. We are therefore forced to study the connectivity of a function space. The tool for this is *homotopy theory*. We designate the space of continuous fields by Φ^X, giving the domain X of the field (three-dimensional space) and the range Φ of the field (the φ manifold of Sec. I.1). Within Φ^X we define a subset $\Phi^X(\varphi_0)$ of fields obeying the boundary conditions $\varphi(\mathbf{x}) \to \varphi_0$ as $|\mathbf{x}| \to \infty$. It is this subset we take as configuration space Q. The question now is whether $Q = \Phi^X(\varphi_0)$ is connected, that is, whether it is possible to find a path, i.e., a continuous function $\varphi(\mathbf{x}, \alpha)$ $(0 \leq \alpha \leq 1)$, "joining" any two points $\varphi_1(\mathbf{x})$, $\varphi_2(\mathbf{x})$ of Q. Inasmuch as our purpose is to deduce conservation laws, we require $\varphi(\mathbf{x}, \alpha)$ to obey the continuity and boundary conditions of a possible evolution $\varphi(\mathbf{x}, t)$, namely:

(a) $\varphi(\mathbf{x}, \alpha)$ is continuous in the joint variables $(\mathbf{x}, \alpha)$.

(b) For all α, $0 \leq \alpha \leq 1$, $\varphi(\mathbf{x}, \alpha) \to \varphi_0$ as $|x| \to \infty$.

(c) $\varphi(\mathbf{x}, 0) = \varphi_1(\mathbf{x})$, $\varphi(\mathbf{x}, 1) = \varphi_2(\mathbf{x})$.

If such a function $\varphi(\mathbf{x}, \alpha)$ exists, then $\varphi_1(\mathbf{x})$ and $\varphi_2(\mathbf{x})$ are *homotopic*, written $\varphi_1 \sim \varphi_2$. The function

$\varphi(\mathbf{x}, \alpha)$ is called a *homotopy*. Evidently $\varphi_1 \sim \varphi_2$ means that in a sense φ_1 and φ_2 can be joined by "curves" in the configuration space Q. Homotopy is an equivalence relation; its equivalence classes are called homotopy classes. We write $\eta = \hom \varphi(\mathbf{x})$ for the class containing $\varphi(\mathbf{x})$. The homotopy classes are analogous to the connected components Q_n in the decomposition $Q = \bigcup Q_n$ given before for the classical configuration space of an unspecified number of particles.

3. The decompositions into connected components of the phase space and of the space of paths of the physical system are similar to that of the configuration space, there being one-to-one correspondences between the homotopy classes of each of these spaces. To see this in the case of the phase space, it is helpful to represent points of the phase space as couples $[\varphi_\alpha(\mathbf{x}), \dot{\varphi}_\alpha(\mathbf{x})]$ of fields and their velocities (time derivative), recalling that the $\dot{\varphi}_\alpha(x)$ belong to a linear vector space. In the case of the paths $\varphi(\mathbf{x}, t)$, it is important that no boundary conditions are imposed in t, although there are in $\mathbf{x}$. Then, it is easy to make a homotopy of every path $\varphi(\mathbf{x}, t)$ to an associated "constant path" $\varphi'(\mathbf{x}, t) = \varphi(\mathbf{x}, 0)$ again reducing the problem of classification to that of the configuration space.

4. A dynamical variable f that depends only on the homotopy class of $\varphi(\mathbf{x})$ will be called a *homotopy variable.*

5. Let $\varphi_0(\mathbf{x})$ designate the constant field $\varphi_0(\mathbf{x}) \equiv \varphi_0$. If another field $\varphi(\mathbf{x})$ is *not* homotopic to $\varphi_0(\mathbf{x})$, we say that there is a *kink* in $\varphi(\mathbf{x})$.

6. Now, let us see how to *count* kinks. Heuristically speaking, the number of kinks in $\varphi(\mathbf{x})$ is the number of times that φ winds about Φ as $\mathbf{x}$ ranges over X.

The homotopy classes of $\Phi^X(\varphi_0)$ form a group in a natural way. Let us define the composition $\oplus$,

$$\eta_1 \oplus \eta_2 = \eta_3,$$

as follows. By a homotopy, if necessary, we may construct a representative field $\varphi_1(\mathbf{x})$ of η_1 which takes on the value φ_0 not only for $|\mathbf{x}| \to \infty$ but also for the half-space $x_3 \geq 0$, where x_3 is the third component of $\mathbf{x}$:

$$\varphi_1(\mathbf{x}) = \varphi_0, \qquad x_3 \geq 0.$$

Similarly, we may construct a representative $\varphi_2(\mathbf{x})$ of η_2 such that

$$\varphi_2(\mathbf{x}) = \varphi_0, \qquad x_3 \leq 0.$$

We take the "sum" of the two classes to be the class η_3 containing

$$\varphi_3(\mathbf{x}) \equiv \varphi_1(\mathbf{x}), \qquad x_3 \leq 0.$$
$$\equiv \varphi_2(\mathbf{x}), \qquad x_3 \geq 0.$$

Briefly, we push the nontrivial parts of η_1 and η_2 to opposite sides of the plane $x_3 = 0$ and then sew them together at this plane. We will adopt the notation $\pi_3(\Phi, \varphi_0)$ or occasionally $\pi_3(\Phi)$ for the group whose elements are the homotopy classes of $\Phi^X(\varphi_0)$ and whose composition is defined above. $\pi_3(\Phi, \varphi_0)$ is isomorphic to what is usually called the third homotopy group of the space Φ. The subscript 3 on $\pi_3(\Phi)$ reminds us that the variable $\mathbf{x}$ ranges over a 3-cell. If instead $\mathbf{x}$ varies over an n-cell with boundary value φ_0 imposed on the surface of the n-cell, we obtain the group $\pi_n(\Phi)$. This group would occur if we dealt with kinks in a field of n(rather than 3)-dimensional space. The most familiar case is $n = 1$, corresponding to a one-dimensional field theory: $\pi_1(\Phi)$ is called the Poincaré or fundamental group of the space Φ. Its elements are homotopy classes of mappings $\varphi = \varphi(x)$ with $\varphi(-\infty) = \varphi(+\infty)$. This mapping is evidently a closed path or (for brevity) a *loop* in Φ. The homotopic conservation laws depend on the structure of the group $\pi_3(\Phi, \varphi_0)$. If $\pi_3(\Phi, \varphi_0)$ is the group ∞ of the integers, then kinks add like particles when they are juxtaposed, the group generators corresponding to one kink (or antikink, as preferred). It is readily seen that if $\varphi(\mathbf{x})$ describes one kink then $\varphi(-\mathbf{x})$ describes one antikink, the inverse to $\varphi(\mathbf{x})$ with respect to the group operation $\oplus$. We shall later take up several examples in which this group may be computed, but first we establish its relevance to quantum physics.

III. QUANTUM FIELDS

1. We now consider the quantum field theory. The operators representing dynamical variables are supposed to act on functionals $\Psi[\varphi(\mathbf{x})]$, which associate a complex number or "amplitude" with each classical field function $\varphi(\mathbf{x})$. The dynamical variables F of the classical field theory that do not involve the time derivatives of the field are represented in the quantum field theory by multiplication, as is usual:

$$F[\varphi] : \Psi[\varphi] \to F[\varphi]\Psi[\varphi].$$

2. Let f be any homotopy variable of the field φ (Sec. II.4). Then

$$df/dt = 0$$

holds both in the classical field theory and in the canonical quantum field theory.

The classical part of the theorem has already been made clear. For the quantum part, we suppose that the evolution of the system is given by an integral over histories of the type introduced by Feynman:

$$\Psi_t[\varphi(\mathbf{x}, t)] = \int_0^t \exp\{(iS[\varphi])\Psi_0[\varphi(\mathbf{x}, 0)]\, d\mu\, [\varphi]\}.$$

Here, the integration is over all continuous histories beginning with the initial configuration $\varphi(\mathbf{x}, 0)$ and ending with the final configuration $\varphi(\mathbf{x}, t)$. $S[\varphi]$ is the classical action for the history φ, and $\mu[\varphi]$ is a measure. Consider a wave functional Ψ at time 0 "belonging" to one of the homotopy classes of $\Phi^X(\varphi_0)$, say that of $\varphi_1(\mathbf{x})$. That is, suppose only fields homotopic to φ_1 have nonzero amplitude:

$$\Psi_0[\varphi(\mathbf{x}, 0)] = 0 \quad \text{or} \quad \varphi(\mathbf{x}, 0) \sim \varphi_1.$$

The contribution to $\Psi_t[\varphi(\mathbf{x}, t)]$ arises from an integral over continuous histories leading from $\varphi(\mathbf{x}, 0)$ to $\varphi(\mathbf{x}, t)$. However, there are no such histories unless $\varphi(\mathbf{x}, 0) \sim \varphi_1$. Therefore

$$\Psi[\varphi(\mathbf{x}, t)] = 0 \quad \text{or} \quad \varphi(\mathbf{x}, t) \sim \varphi_1.$$

That is, the final wave functional belongs to the same homotopy class as the original. Define the dynamical variable $C[\varphi(\mathbf{x}) \sim \varphi_1(\mathbf{x})]$ to be 1 if the indicated homotopy is valid, 0 if not. The dynamical variable C is the characteristic functional of the homotopy class $\varphi_1(x)$. From what we have said, it follows that if $C\Psi = \lambda\Psi$ holds at $t = O(\lambda = 0, 1)$ then it holds at any later time with the same λ. Therefore

$$dC/dt = 0.$$

The homotopy variable f is a linear combination of such characteristic functionals and therefore we have the conservation law

$$df/dt = 0.$$

Evidently this result depends essentially on the continuity of the histories over which this integral is taken.

The conservation laws for these homotopy variables form the *homotopic conservation laws* of the field theory. In the simplest cases there is either no significant homotopic law, or else the conserved quantity is an integer $N = 0, \pm 1, \cdots$, or possibly an integer modulo 2, depending on the way in which kinks combine.

IV. EXAMPLES OF HOMOTOPY CONSERVATION LAWS

In the following examples, no proofs of topological theorems are given. Only the needed results, taken from Steenrod or Hu,[4] are mentioned.

1. Vector fields: $\Phi = R^n$

There is no significant homotopic conservation law for linear field theories. Indeed, if Φ is a linear vector space, then $\pi_3(\Phi, \varphi_0) = 1$, the group of one element, and the homotopic dynamical variables reduce to $f = \text{const}$.

2. Unit 3-vector field: $\Phi = S^2$

If φ is a unit vector field with three components, $\varphi = (\varphi_1, \varphi_2, \varphi_3)$, $\varphi_1^2 + \varphi_2^2 + \varphi_3^2 = 1$, then Φ is the 2-dimensional sphere (surface) S^2. Then the topological result needed (cf. Table I) is[4]

$$\pi_3(S^2) = \infty,$$

the infinite cyclic group. Therefore, there is a single homotopic variable $N = 0, \pm 1, \pm 2, \cdots$ in terms of which all others can be expressed and which combines additively under the group operation $\oplus$. We call N the homotopic number.

In this example it is possible to visualize the conserved structure such that the homotopic number counts. We first use three Pauli matrices ρ_1, ρ_2, ρ_3 to

TABLE I. $\pi_n(S^m)$. The entry p designates the additive group of integers modulo p; ∞, the additive group of integers; blanks, the group with one element ($p = 1$). The groups $\pi_n(S^m)$, $\pi_{n+1}(S^m)$ are required in the kink and spin analysis of a field theory in n-dimensional space with a field variable belonging to an m-dimensional sphere S^m, as for example to $SU_2 = S^3$. For sufficiently large n, $\pi_{n+r}(S^n)$ becomes a function of r alone (independent of n), and the asterisk indicates that this limiting value has been achieved. The other limiting values for $r \leq 8$ are $\pi_{12}(S^7) = 1^*$, $\pi_{12}(S^5) = 2^*$, $\pi_{16}(S^9) = 240^*$, $\pi_{18}(S^{10}) = 2 \oplus 2^*$. This information is taken from Hu (see Ref. 4).

				m				
n	1	2	3	4	5	6	7	8
1	∞							
2	1	∞						
3	1	∞	∞					
4	1	2*	2*	∞				
5	1	2	2	2	∞			
6	1	12	12	2	2	∞		
7	1	2	2	$\infty \oplus 12$	2	2	∞	
8	1	2	2	$2 \oplus 2$	24*	2	2	∞
9	1	3	3	$2 \oplus 2$	2	24	2	3

[4] All the results of topology needed for the present work can be found in N. E. Steenrod, *Theory of Fibre Bundles* (Princeton University Press, Princeton, N. J., 1951), or S. Hu, *Homotopy Theory* (Academic Press Inc., New York, 1959). It is a pleasure to acknowledge the benefit of frequent discussions of these problems with C. W. Misner.

supply space with a point at infinity through the transformation

$$\mathbf{x} \to (\rho\cdot\mathbf{x} + i)(\rho\cdot\mathbf{x} - i)^{-1} \tag{1}$$

associating with each finite $\mathbf{x}$ a unitary 2×2 matrix and with ∞ the matrix 1. This transformation is akin to the stereographic projection of $\mathbf{x}$-space onto a 3-sphere S^3, the complex unitary 2×2 matric group SU_2 having the same topology as S^3. The operator u is a spin-$\frac{1}{2}$ representation of a rotation about a radial axis passing through $\mathbf{x}$. The angle of rotation varies from 0 at $|\mathbf{x}| = \infty$ to 2π at $\mathbf{x} = 0$. The second column of the matrix u is redundant. We obtain a simpler stereographic coordinate by taking the first column of u setting

$$v = u\uparrow,$$

where $\uparrow = \binom{1}{0}$. The spinor v evidently ranges over S^3, since $v = \binom{a+ib}{c+id}$, $a^2 + b^2 + c^2 + d^2 = 1$, with $v = \binom{1}{0}$ corresponding to $\mathbf{x} = \infty$ and $v = -\binom{1}{0}$ to $\mathbf{x} = 0$. We substitute v for $\mathbf{x}$ as independent variable. A φ field with one kink is now a unit 3-vector depending on v, and may be defined by the important correspondence

$$\varphi = v^+\varrho v \equiv \varphi_1(v) \tag{2}$$

equivalent to the *Hopf mapping*[4] of S^3 into S^2. A sort of spatial reflection of $\varphi_1(v)$,

$$\varphi_{-1}(v) \equiv \varphi_1(v^\lambda)$$

[v^λ = Hermitian adjoint of v, resulting from $\mathbf{x} \to -\mathbf{x}$ in (1)] defines a field with one antikink.

3. A rotator field: $\Phi = P^3$

The three-dimensional projective space P^3 is defined as the space of rays in R^4, i.e., a point of P^3 is the set of all multiples of some nonzero vector in R^4. P^3 occurs as the range Φ of the field variable in some field theories. A "rotator" field, whose field variable is an element of $SO(3, R)$ (the rotation group in three dimensions) is an example, since $SO(3, R) \sim P^3$. Because[6] $\pi_3(P^3) = \infty$, there is a single nontrivial homotopic number. The kink in this field is best described after the next example.

4. A unit 4-vector field: $\Phi = S^3$

The three-sphere S^3 is the covering space of the projective sphere P^3 of the previous example, and $\pi_3(S^3) = \infty$ (Table 1). S^3 is the range of the field variable in a theory of strong interactions of Skyrme[1] which suggests that the fundamental variable is not an isovector $\pi = (\pi^1, \pi^2, \pi^3)$ but an element Π of SU_2, to which an isovector π can be related according to

$$\Pi = \exp(i\pi\cdot\tfrac{1}{2}\tau) \tag{3}$$

for Π sufficiently close to 1. Here τ is the usual isovector operator (τ^1, τ^2, τ^3) generating the algebra of infinitesimal spin-$\frac{1}{2}$ rotation operators. Thus, Skyrme's theory, in addition to implying multipion interactions because of the dynamical nonlinearity, implies the existence of a strictly conserved homotopic quantity because of the topological nonlinearity. The class of unit homotopic number is represented in the stereographic coordinates (1) by the identity mapping

$$\Pi = \Pi(u) = u.$$

With this result we return to Example 3. A rotator field $\varphi_1(v)$ of unit homotopic number can be constructed from $\Pi_1(v)$. Let $D(\Pi)$ be the spin 1 representation of the group SU_2, defined by

$$D(\Pi)\tau = \Pi\tau\Pi^{-1}.$$

Then

$$\varphi_1(v) = D[\Pi_1(v)].$$

5. General Relativity[3]

Let us postpone the consideration of topologically nontrivial space–times and continue to describe space–time by four coordinates x^μ ranging over R^4, with the boundary conditions

$$g_{\mu\nu} \to \gamma_{\mu\nu}, \quad (x^1)^2 + (x^2)^2 + (x^3)^2 \to \infty$$

imposed on a distant cylinder. The coordinate x^0 is supposed to be timelike on this distant cylinder; indeed we take

$$\gamma_{\mu\nu} = \text{diag}\,(1, -1, -1, -1).$$

The previous discussion of conservation of homotopic quantities in classical field theories can now be applied, provided we replace the concept of time by the coordinate x^0 throughout. The conserved quantity is associated with a cross section $x^0 = \text{const}$ of space–time; this cross section is not in general spacelike throughout, and in general there is no cross section which is spacelike throughout. The field variable φ is now a symmetric tensor $g_{\mu\nu}$ of signature $+1{-}3$, and the space Φ of these tensors has known homotopy groups. In particular[4]

$$\pi_3(\Phi) = \infty.$$

Therefore, there is again just one conserved homotopic number, called metricity in an earlier work.[2] The heuristic argument concerning the definition and the conservation of these homotopic quantities

in quantum field theory is somewhat less than convincing for this case, because the quantization process is not at all understood when the direction of time is as tangled as it is in most of the metrical universes introduced in this computation. For example, the field of time-directions in a gravitational kink is shown in Ref. 2.

6. Displacement field

Suppose a continuous medium undergoes a continuous differentiable deformation $\mathbf{x} \to \mathbf{x}' = \mathbf{f}(\mathbf{x})$, which can be regarded as an idealized crystal lattice deformation. Then, at each point, $\mathbf{x}$ is defined a Jacobian matrix $\varphi(x) \equiv \partial \mathbf{x}'/\partial \mathbf{x}$ belonging to $GL(3, R)$. We suppose $\varphi(\mathbf{x}) \to 1$ as $\mathbf{x} \to \infty$. Are there kinks in the field φ? We are led to seek $\pi_3[GL(3, R)]$. The polar factorization

$$\varphi = \rho\omega$$

expresses φ in terms of a real positive definite symmetric ρ and a real orthogonal ω, defining a homeomomorphism

$$GL(3R) \sim \{\rho\} \otimes 0_3.$$

The exponential form

$$\rho = e^{\sigma}$$

defines a homeomorphism of $\{\rho\}$ onto the real symmetric matrices $\{\sigma\}$, which form a vector space and may be discarded. Thus

$$\pi_3[GL(3, R)] = \pi_3(0_3) = \infty.$$

There exists one kind of kink in the field φ. This analysis does not take into account the integrability conditions obeyed by φ, which in principle makes a still finer analysis possible. If $\mathbf{f}(\mathbf{x})$ is required to possess a continuous inverse (noninterpenetrability) then the number of kinks must be 0 or 1.

V. SPIN AND DOUBLE VALUEDNESS

1. The operation W of a continuous rotation through 2π will be called spin parity. W appears to return every physical variable to its original value and yet W changes the sign of the state (wavefunction) of some systems. This spin double valuedness must cause anguish in any physicist, who is choosing a candidate for a truly fundamental field. Geometrical type arguments generally lead to tensorial or at any rate single-valued fields with $W \doteq +1$, such as the metric tensor, the affinity, or the electromagnetic vector potential. On the other hand, there are such algebraic arguments as the facts that a vector can be formed from a spinor field but not conversely, suggesting that the fundamental field have a built-in double-valuedness with $W \doteq -1$. Some physicists have followed the "geometric" path, others the "algebraic".

2. Here, we show that certain appropriately nonlinear but single-valued fields $\varphi(\mathbf{x})$ admit double-valued but continuous wave-functionals $\Psi[\varphi(\mathbf{x})]$ that have transformation properties belonging to $W \doteq -1$. We shall inspect each of the kinks we have mentioned to see whether it admits odd-spin parity in this way. We shall also see examples of fields which do not admit kinks but still admit structures of odd-spin parity. The connection between spin partity and statistics are discussed in a later paper.

3. The tool for the study of multivaluedness is the homotopy theory, since a multivalued function on a configuration space Q is one whose value depends not merely on a point q in Q but also on how that point is arrived at, and therefore on a path in Q with homotopic paths being assigned the same value.

4. By a multivalued wavefunction we really mean a function on the universal covering space of the configuration space, whose definition we will now recall. If Q is the configuration space, its (universal) covering space is represented by choosing in each connected component Q^n of Q a base point x_0^n and considering paths $x(s)$, $0 \leq s \leq 1$, starting at a base point

$$x(0) = x_0^n.$$

There is an equvalence relation $\sim$ defined between such paths obtained when two paths are homotopic with end points fixed during the homotopy. The resulting equivalence classes each consist of paths all having a common end point $x(1)$, but there may be more than one class for a given end point. Each class is a point of the universal covering space CQ, and we will write Cx for the set of points of CQ associated with the end point x in Q. Cx is said to cover x. The class containing the path $x(s)$ is written $\{x(s)\}$. Inverse to the covering correspondence C is the natural projection p that maps CQ onto Q by associating with each curve $x(s)$ in Q representing a point of CQ its end point $x(1)$; $p(y) = x$ for y in Cx. The natural projection turns dynamical variables f that are functions of q into functions Cf on the covering space CQ through

$$Cf(p) = f(q) \text{ for } p \text{ in } Cq.$$

In the following we do not distinguish between f and Cf related in this manner.

5. The operation of a rotation on a point of the covering space CQ is not uniquely defined by its effect on the points of Q; there is an intrinsic multivaluedness. For the uniqueness of definition we must also go to the covering space of the rotation group. If G is any continuous connected group, we turn the covering space of G into a group by considering paths that begin at the identity 1 in G and defining the product of two such paths $g(s)$ and $g'(s)$ through point-by-point multiplication:

$$g(s)\cdot g'(s) \equiv g''(s).$$

It is important, of course, that the equivalence class of $g''(s)$ is determined by those of $g(s)$ and $g'(s)$. Some well-known groups and their covering groups are

$$CR = R, \quad CS^1 = R, \quad CSO_3 = SU_2,$$

where R is the additive group of the real numbers and SO_3 is the rotation group in three (real) dimensions.

Suppose G operates on Q according to some rule $g \circ q$. We can now define the operation of CG on CQ through the point-by-point operation of G on Q. If $h = \{g(s)\}$ and $p = \{q(s)\}$ then we define

$$h \circ p \equiv \{g(s) \circ q(s)\} = \{q'(s)\}.$$

This makes $h \circ p$ an element of $C(g \circ q)$. Since $q'(0) = q(0)$, the path $q'(s)$ defines a point of CQ which is obviously determined by the equivalence classes of $g(s)$ and $q(s)$. [An equally natural definition of $Cg \circ Cq$ uses the path consisting of $q(2s)$, $0 \leq s \leq \frac{1}{2}$, from q_0 to $q(1)$, followed by the path $g(2s-1) \circ q(1)$, $\frac{1}{2} \leq s \leq 1$, from $q(1) \circ q(1)$. This is equivalent.]

6. When we speak of multivalued representations of a group G, we really mean single-valued representations of CG. The multiplicity of the representation of G (the number of values at a point g of G) is at most the number of points of CG covering each point of G, which is a topological property of G. For the rotation group $G = SO_3(R)$, CG covers G twice so that 2 is the highest multiplicity possible. This multiplicity is attained for the multivalued wavefunction $\Psi[p]$ if points p and $1' \circ p$ are different points of CQ, where $1'$ represents the element of $C1$ which is *not* the identity of CG; because then it is possible to construct a continuous function (al) Ψ $[p]$ which takes on different values on the two different points p and $1' \circ p$ covering the same point q of the underlying configuration space Q. The paths in $C1$ are loops, where their end and initial points coincide.

7. This implies the criterion for the existence of negative spin parity stated previously[2,5]:

The loop $g(s) \circ q$ must be nontrivial when the loops $g(s)$ is nontrivial. In other words, $g(s) \circ q$ must determine an element of order 2 of the fundamental group $\pi_1(Q)$. We shall call negative spin parity ($W = -1$) simply *spin* when this does not lead to confusion.

8. To compute $\pi_1(Q)$ where $Q = \Phi^X(\varphi_0)$ we proceed as follows. A loop in Q is a mapping such that

(a) $\varphi(\mathbf{x}, s)$ is continuous in $\mathbf{x}$ and s for $\mathbf{x}$ in X, $0 \leq s \leq 1$;
(b) $\varphi(\infty, s) = \varphi_0$;
(c) $\varphi(\mathbf{x}, 0) = \varphi(\mathbf{x}), \varphi(\mathbf{x}, 1) = \varphi(\mathbf{x}).$ (2)

It is therefore equivalent to a mapping of a 4-cell $X \times I$ or I^4 with given boundary values on the boundary ∂I^4. We require the homotopy classification of such mappings. When a 2π rotation acts on a field $\varphi(\mathbf{x})$ it results in a loop in Q

$$\varphi_W(\mathbf{x}, s) = g(s) \circ \varphi[g^{-1}(s)\mathbf{x}],$$

where $g(s)$ traverses a loop in the rotation group G of order 2 in $\pi_1(G)$. The transformation laws of both space points $\mathbf{x}$ and field values φ figure in this loop. We require the order of $\varphi_W(\mathbf{x}, s)$ in $\pi_1(Q)$: is it 1 or 2?

9. In general, it seems that the order of $\varphi_W(\mathbf{x}, s)$ could depend on the choice of $\varphi(\mathbf{x})$, but an example of this has not been established. If the space Q is connected, there is no such dependence, because homotopic fields yield homotopic 2π rotations. This is the situation when kinks are not possible, though spin (Sec. V.7) may be.

10. Even if kinks are admitted and Q is disconnected, the order may not depend on the initial field, i.e., on the presence and nature of its kinks. For example, if Φ admits a group structure (φ_0 as identity, without loss of generality), then this group structure enables a homeomorphism to be established between any component Q_n of Q and the component Q_0 containing the constant field $\varphi(\mathbf{x}) = \varphi_0$, so that $\pi_1(Q_n) \sim \pi_1(Q_0)$. If moreover $\pi_1(Q_0) = 1$ it follows that $\pi_1(Q_n) = 1$.

11. There is a heuristic argument which suggests that either spin-$\frac{1}{2}$ wavefunctionals can be constructed with any connected component of Q as support, or else there are none at all. If any kind of kink admits odd-spin parity W, a combination of that kink in an odd-W state with its antikink

[5] This spin criterion is applied to numerous mechanical (nonfield) systems in D. Finkelstein, Phys. Rev. 100, 924 (1955).

in an even-W state should give an odd-W state with no kinks. A combination of this state with an even-W state of any number of any kind of kinks then should give an odd-W state with the same kinks.

12. We now take initial fields $\varphi(\mathbf{x}) = \varphi_0$ in (2). Then $\varphi(\mathbf{x}, s) = \varphi_0$ for

$$|\mathbf{x}| \to \infty, \quad 0 \leq s \leq 1$$
$$\text{any } \mathbf{x}, \ s = 0,$$
$$\text{any } \mathbf{x}, \ s = 1.$$

Now, the loop $\varphi_W(\mathbf{x}, s)$ in Q defines an element of $\pi_4(\Phi, \varphi_0)$, which is either the identity or of order 2 by Sec. V. 8.

So the first test for spin is to find $\pi_4(\Phi, \varphi_0)$ and see if it contains an element of order 2. If so, this shows that the theory admits double-valued wavefunctionals describing zero kinks. It is then necessary to test whether the loop $\varphi_W(\mathbf{x}, s)$ is of order 2. If not, it is then necessary to consider wavefunctionals describing nonzero kinks, for which the considerations of Secs. V. 9–11 are helpful. This is the criterion formulated previously for application to general relativity,[2] and we now apply it to a wider range of theories.

13. In each case we have to ask how the rotation group G acts on Φ. It is frequently natural to suppose that the boundary value φ_0, which φ approaches as $|\mathbf{x}| \to \infty$, is invariant under G, in order to guarantee isotropy of the theory. Then the field $\varphi(\mathbf{x}) \equiv \varphi_0$, typical of the zero-kinks case, is spherically symmetric, the loop $\varphi_W(\mathbf{x}, s)$ is trivial, and there is no spin. That leaves only the case where kinks are present to be investigated.

14. Furthermore, if all physical quantities depend on φ not directly but only through invariants formed from φ and the derivatives of φ, isotropy is still possible with $g \circ \varphi_0 \neq \varphi_0$, and the field with no kinks may have spin.

VI. EXAMPLES OF SPIN CALCULATIONS

1. Vector fields

$\Phi = R^n$, $\pi_4(\Phi, \varphi_0) = 1$. No spin here.

2. Unit 3-vector field

$\Phi = S^2$. $\pi_4 = 2$ according to Table I, so that double-valued wavefunctionals are admitted by this theory. For the second test we need the action of the rotation group G on Φ. First (cf. Sec. V. 13) the boundary value φ_0 is assumed to be invariant under G. It can readily be shown that this implies that Φ is invariant under $G : g \circ \varphi = \varphi$, as in the case of an isovector field. Since the constant field $\varphi(\mathbf{x}) = \varphi_0$ is spherically symmetric, the loop $\varphi_W(\mathbf{x}, s)$ is just one point, therefore trivial, and the states of zero kinks possess even-spin parity. But the one-kink field (2) is not spherically symmetric, and we do not know whether the corresponding loop is trivial (cf. Sec. V.11). Conclusion: For a unit isovector field, odd-spin parity is excluded in the absence of kinks but has not been excluded in the presence of a kink.

On the other hand, for a unit spatial vector, $\varphi(\mathbf{x}) = \varphi_0$ is not spherically symmetric and $\varphi_W(\mathbf{x}, s)$ is not independent of s, but may readily be shown to be trivial. (Deform the axis of rotation to lie along φ_0.) The conclusion is thus the same for a vector as an isovector.

3. Unit 4-vector field

$\Phi = S^3$, $\pi_4(\Phi, \varphi_0) = 2$, (Table I), showing the possibility of double-valued wavefunctionals. We now ask the action of G on Φ. The first choice that comes to mind is to treat the 4-vector φ as a Lorentz 4-vector, transforming only three components $\varphi_1\varphi_2\varphi_3$ and leaving φ_0 alone. This law of transformation leaves the constant field $\varphi(\mathbf{x}) = (1, 0, 0, 0)$ fixed, so the zero-kink states have even-spin parity. By Sec. V. 10, since S^3 is a group (SU_2 or the unit quaternions), the n-kink states also have even-spin parity. If φ is treated as an isovector, $g \circ \varphi = \varphi$, the result is the same, according to Secs. V. 10, V. 13.

4. Rotator field

$\Phi = SO_3 \sim P^3$. $\pi_4(P^3) = \pi_4(S^3) = 2$ (Table I) so we must consider the action of G on Φ. If $g \circ \varphi \equiv \varphi$, ("isorotator") there is no odd W in Q_0. If φ is regarded as an element of G and $g \circ \varphi$ is the group product, then, for $\varphi(x) = \varphi_0 \equiv 1$ (without loss of generality), $\varphi_W(\mathbf{x}, s) = g(s)$, a 2π rotation. This is obviously nontrivial. Conclusion: A rotator field exhibits odd-spin parity.

5. General Relativity

The spherical symmetry of the Minkowski metric shows that there is no odd-spin parity in the absence of kinks. But the kink in this field itself admits a spherically symmetric representation, and so cannot possess odd-spin parity.[2]

6. Displacement field

The reduction of this problem to the rotator problem was already performed in the analysis of the kinks in the displacement field, and enables us to limit our inspection to a field $\varphi(\mathbf{x})$ with φ in SO_3. The action of a rotation g on φ is readily seen to be $g \circ \varphi = g\varphi$. This is the second of the cases treated in Sec. VI. 4 and was found there to admit spin, and so the displacement field likewise admits spin.

Solution of very Singular N/D Equations—An Approach to Nonrenormalizable Field Theory*

MARTIN B. HALPERN†
Department of Physics, Harvard University, Cambridge, Massachusetts
(Received 18 November 1965)

In this paper, we find closed form solutions to unsubtracted and once-subtracted N/D equations, both nonrelativistic and relativistic, with pathological left-hand absorptive parts that increase at large ν like $\alpha(-\nu) \sim -\lambda\nu^m$, m not necessarily integral. The solutions allow dispersive, unitary and regulator-free calculation in nonrenormalizable field theory. For example, N/D equations with exchange of a higher spin particle (spin $\frac{3}{2}$ etc.) can be solved. In general, the technique allows the use of a large class of divergent graphs as input, that is, those whose left-hand absorptive parts are finite, although asymptotically ill behaved. For example, in the W theory of leptonic weak interactions, one of the possible inputs is any number of ladder graphs cut so that all the bosons are on the mass shell. Many nonplanar graphs can be included as well. In much the same way, we can also calculate in the Fermi theory, (multiple exchanges in) theories of higher spin in general, and theories with derivative coupling.

There are an infinite number of solutions to these singular integral equations, each of which is characterized by a branch point at $g^2 = 0$, and a branch point of oscillatory nature at infinite $|\nu|$. Out of these, we pick the solution which sums the iterative expansion of the equations as most meaningful. It is seen explicitly that the unitary requirement generates its own regulation in the form of rapid oscillations at large unphysical energies, thus eliminating the need for any regulator limiting process. The oscillations are associated with an infinite number of ghosts, but these stay very far from the physical region. In addition, the solutions violate unitarity in the cross channels, so we expect the program to be useful at most in and near the physical region. As shorter range forces are systematically included, there is some indication that the program may converge rapidly for small physical energies.

INTRODUCTION

As has been previously discussed,[1,2] many individual terms in the perturbation expansion of a nonrenormalizable field theory, although divergent, have finite absorptive parts. For example, in the W-meson theory of leptonic weak interactions, each of the ladder graphs cut parallel to the lepton lines (Fig. 1) is finite, because there remain no closed loops from which a divergent integration could occur. In the partial waves of the s channel, this means that the discontinuity across each of the left-hand cuts associated with multiple W-meson exchange is finite. The nonrenormalizability of the theory is reflected in the fact that these discontinuities $\alpha(\nu)$ increase far to the left like powers of ν. (ν is the center of mass 3-momentum squared.) For the exchange of N W mesons, it was shown that asymptotically

$$\alpha(-\nu) \sim \lambda\nu^{N-1}.$$

The possibility of summing the left-hand discontinuities has been discussed in the previous references. Here we want to actually find solutions to N/D equations whose inputs are pathological. The idea would be to solve for some finite number of W exchanges, and study the convergence of the solution as shorter and shorter range forces are included. There is some indication that the result in the low-energy physical region may essentially be independent of N for fairly large N.

In general, the solutions obtained will allow us to use as input into the N/D equations any set of graphs, regardless of divergence, whose left-hand cuts (in general asymptotically ill behaved) are finite. This includes the exchange of one or more particles of higher spins (e.g., or spin $\frac{3}{2}$ exchange). Many nonplanar graphs and subsets of graphs in the Fermi theory and derivative coupling theories can be included as well. (It is clear that any ladder-type

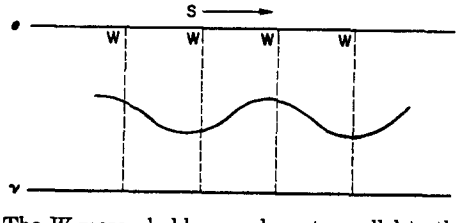

FIG. 1. The W-meson ladder graphs cut parallel to the lepton lines.

* Based in part on the Ph. D. thesis "S-Matrix Theory and Higher Order Connections to Weak Interactions," Harvard University, October 1964, during which time the author was a National Science Foundation Pre-doctoral Fellow.

† Present address: Physics Department, University of California, Berkeley 94720.

[1] M. B. Halpern, Doctoral thesis. Harvard University, 1964.

[2] M. B. Halpern, Phys. Rev. 140, B1570 (1965). See this reference for related references in nonrenormalizable field theory.

graph can be included because, as above, cutting all the exchange lines leaves no possibility of divergence.)

A few preliminary comments are in order. The first is that, as we shall see, all four types of N/D equations considered here (homogeneous and inhomogeneous, relativistic and nonrelativistic) involve essentially the same principles and difficulties in their solutions. The second comment concerns the order of presentation of the material. We could simply have begun with the relativistic once-subtracted equations, which, according to discussion in the previous references are the ones that we want (at least for the W theory). But, because the nonrelativistic homogeneous (unsubtracted) equations involve most of the same principles and avoid the "frosting on the cake" complications of the inhomogeneity and the relativistic phase-space factor, we felt it would be better to begin with it. The unsubtracted, relativistic and nonrelativistic cases are the subject of Part A. The inhomogeneous (once-subtracted equations) are studied in Part B. The third comment is that the Fourier transform methods we employ for the solution in all these cases are on the borderline of being ill defined. At least when everything is discussed in terms of integral representations in some ranges of the parameters involved, it is necessary to introduce a convergence factor, analogous to the ordinary exp $(-\epsilon\,|x|)$ convergence factor of usual Fourier transform theory, and appropriate to the rapid oscillations found in these solutions. We emphasize that we shall find this convergence factor to be self-justifying in that its use leads to solutions that explicitly satisfy the original integral equations. We remark also that, in all cases, there are an infinite number of solutions to the equations. Out of these we pick the solution which sums the iterative solution, which, as discussed in the previous references, we feel is the most physical. This is the subject of Part C.

Finally we emphasize that all the solutions to these singular N/D equations involve an infinite number of ghosts very far to the left. Because of these ghosts, our program is evidently having some of the same sort of trouble as we have found in the Feinberg–Pais (FP) program.[1,2] We advance arguments as to why our ghosts may be less serious than the failures of the Feinberg–Pais program, and even try to interpret the ghosts as indicating an oscillatory left-hand cut in the full theory. However, when faced with repeated breakdown of field theoretic postulates like this, one cannot help but also wonder if it is even possible to require all the postulates of local field theory in the presence of nonrenormalizability.

A. HOMOGENEOUS N/D EQUATIONS

We first consider the nonrelativistic (NR) unsubtracted N/D equations and later generalize to the relativistic (and inhomogeneous) cases.

Nonrelativistic, Unsubtracted, Homogeneous Case

We are interested then in the solution of the coupled integral equations[3]

$$N(s) = \int_{-\infty}^{-1} \frac{d\bar{s}}{\bar{s} - s}\, D(\bar{s})\alpha(\bar{s}),$$
$$D(s) = -\frac{1}{\pi}\int_0^{\infty} d\bar{s}\, \frac{\bar{s}^{\frac{1}{2}}N(\bar{s})}{s - s} \tag{1}$$

with inputs of asymptotic form $\alpha(-s) \sim f(g^2)s^{\frac{1}{2}(m-1)}$, $s \to +\infty$, $m \geq 1$. We have already scaled the (dimensionless) center-of-mass energy s in units of the lowest possible mass exchange—for the lepton ladder this is essentially $(M)^2$. We would like to point out that there is no known reason for the N/D equations with these singular inputs not to have solutions. That is, $\alpha(s)$ and $D(s)$ always appear multiplied together, and, hopefully, the D function will correct for the ill behavior of the input.

If solutions to (1) exist, we can say ahead of time that the amplitude must exhibit a branch point of an oscillatory nature at $|s| = \infty$, and an infinite number of ghosts. That ∞ must be a branch point follows simply from having the right- and left-hand cuts go all the way to ∞.[4] That it must be of an oscillatory nature comes about in the following way: the asymptotic modulus of the amplitude is required to be bounded from below on the left by $|s|^{\frac{1}{2}(m-1)}$ (because that is the behavior of the absorptive part), and from above on the right by $s^{-\frac{1}{2}}$ (unitarity). This requires the asymptotic modulus of the solution to be, in general, different along different rays toward ∞. This implies that the branch point must be associated either with exponential increase {for example, $(\ln s)[\exp(s)]$, $\exp(s^{\frac{1}{2}})$, etc.} or oscillations [for example, $\exp i(\ln s)$, $\exp i(\ln^2 s)$, etc.] because ordinary branch points at ∞ (e.g., $\ln s$, s^m, etc.) are characterized by having the same asymptotic behavior of their modulus as the point at ∞ is approached from different directions. The assumption of dispersion

[3] R. Blankenbecler, M. L. Goldberger, N. N. Khuri, and S. B. Treiman, Ann. Phys. 10, 62 (1960).

[4] Whenever either a right cut or a left cut goes all the way to the point at ∞, the singularity at ∞ cannot be an essential singularity, as this term is properly reserved only for single-valued functions.

relations forces the branch point to be oscillatory, as Hilbert transforms do not admit exponential increase. If $A = N/D$ has the oscillatory branch point, then at least one of N and D must also. In fact, since N and D are, respectively, each other's spectral functions, if one oscillates, so does the other. The oscillations (zeros on the left) of D will correspond to ghosts in the solution.

We will solve (1) by expanding about the "most singular" part of the input in a unitary manner. That is, we write

$$\alpha(s) = \alpha_0(s) + \Delta\alpha(s), \quad \alpha_0(s) = (-s)^{\frac{1}{2}(m-1)} f(g^2), \qquad \Delta\alpha(s) \equiv \alpha(s) - \alpha_0(s), \tag{2}$$

where g^2 is the weak coupling constant, and take as the "most singular" approximation

$$N_0(s) = \int_{-\infty}^{0} \frac{d\bar{s}}{\bar{s} - s} D_0(\bar{s})\alpha_0(\bar{s}), \qquad D_0(s) = -\frac{1}{\pi}\int_0^{\infty} \frac{d\bar{s}}{\bar{s} - s}\frac{\bar{s}^{\frac{1}{2}} N_0(\bar{s})}{\bar{s} - s}, \tag{3}$$

which maintains the unitarity. Although we have kept the high-energy pathology of the model, we have let the left-hand cut come down to $s = 0$. Because of this and the neglect of all the low-energy information in $\alpha(s)$, we expect the solution N_0/D_0 to be poor near threshold. Assuming solutions to (3) can be found, we can iterate around $D_0(s)$ in a unitary way, just as one ordinarily iterates around 1 in the once-subtracted N/D equations.[5] That is, the first iterated solution is

$$N_1(s) = \int_{-\infty}^{-1} \frac{ds'}{s' - s} D_0(s')\alpha(s'), \qquad D_1(s) = -\frac{1}{\pi}\int_0^{\infty} \frac{ds'(s')^{\frac{1}{2}} N_1(s')}{s' - s} \tag{4}$$

and so on in the ordinary way. Notice that we have the correct gap to the left of threshold: The primary function of the iteration scheme is to correct near threshold the two defects in that region of the "most singular" N_0 and D_0. Further, we have every reason to believe that, if N_0 and D_0 exist, so also do the solutions to the full equations (1) as well, since the former include all the pathology of the latter. (Certainly any finite number of iterations exist.) We comment finally that, just as in any N/D iteration scheme, at any finite level of the iteration we do not have the exact left-hand cut. In (4) we have, instead of $\alpha(s)$, the discontinuity

$$[D_0(s)/D_1(s)]\alpha(s).$$

[5] For example, see F. Zachariason and C. Zemach, Phys. Rev. 128, 849 (1962).

The equation for the D function, in this most singular approximation is, submerging the index zero

$$D(-s) = \int_0^{\infty} \frac{D(-\bar{s})\alpha(-\bar{s})\, d\bar{s}}{(\bar{s})^{\frac{1}{2}} + s^{\frac{1}{2}}}. \tag{5a}$$

In the region $s < 0$, we introduce the change of variable $-s = x^2$, $D(-s) \equiv \mathfrak{D}(x)$, $\alpha(-s) \equiv A(x)$, in terms of which we find

$$\mathfrak{D}(x) = 2\int_0^{\infty} \frac{y\, dy\, A(y)\mathfrak{D}(y)}{x + y}. \tag{5b}$$

The further change of variable

$$x = e^{\xi}; \qquad e^{\xi/2}\mathfrak{D}(e^{\xi}) \equiv \Phi(\xi), \qquad B(\eta)e^{\eta} = xA(x) = f(g^2)e^{m\eta} \tag{6}$$

results in

$$\Phi(\xi) = f(g^2)\int_{-\infty}^{+\infty} d\eta\, e^{m\eta}\, \operatorname{sech}\tfrac{1}{2}(\eta - \xi)\Phi(\eta). \tag{7}$$

We present here a general method for finding solutions to the class of homogeneous integral equations

$$\Phi(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} K(\xi - \eta)\Phi(\eta)e^{\eta m}\, d\eta \tag{8}$$

when such exist. In Fourier transform space, we obtain the finite difference equation

$$\Phi(\omega) = K(\omega)\Phi(\omega + mi), \qquad F(\omega) \equiv \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\xi\omega}F(\xi)\, d\xi. \tag{9}$$

The inverse Fourier transform of a solution to (9) does not automatically satisfy (8).[6] It is easy to

[6] For example, Eq. (7) goes over in Fourier transform space to $\Phi(\omega) = 2\pi f(g^2)\operatorname{sech}\pi\omega\,\Phi(\omega + im)$. We can easily guess a particular solution to the difference equation whose inverse Fourier transform does not satisfy (7). Since the solution to $Q(\omega) = \lambda Q(\omega + im)$, λ constant, is $Q(\omega) = \lambda^{i\omega/m}$, and because $\operatorname{sech}\pi\omega$ is almost a constant under translation through the period strip, we are tempted to guess a solution of the form $\Phi(\omega) = (2\pi g^2 \operatorname{sech}\pi\omega)^{i\omega/m}F(\omega)$. This results in the simpler equation $F(\omega) = (-1)^{i(\omega+im)}F(\omega + im)$. The further guess $F(\omega) = (-1)^{\omega^2/2m}g(\omega)$ leaves only $g(\omega) = (-1)^{3m/2}g(\omega + im)$, which is solved immediately as in the first example. Thus, we have found the particular solution to the difference equation

$$\Phi_P(\omega) = [2\pi f(g^2)\operatorname{sech}\pi\omega]^{i\omega/m}(-1)^{\frac{1}{2}i\omega}(-1)^{\omega^2/2m}.$$

But this does not satisfy (11a) because it has poles in the period strip, and its inverse transform is not a solution of (7). It is easy to show that $\Phi_P(\omega)\,\mathfrak{a}(\omega)$ is also a solution of the difference equation if $\mathfrak{a}(\omega)$ is a function of period im. We could attempt to find some $\mathfrak{a}(\omega)$ so that $\mathfrak{a}(\omega)\Phi_P(\omega)$ satisfied (11), but our Fourier transform technique does this for us. To anticipate slightly, we notice that even this $\Phi_P(\omega)$ manifests the very important $e^{i\omega^2}$ behavior; it was generated because $\operatorname{sech}\pi\omega$ changes sign under translation through the period strip.

show by substitution back into (8) that a sufficient condition for

$$\int_{-\infty}^{+\infty} \Phi(\omega) e^{i\xi\omega}\, d\omega$$

to be a solution is

$$\int_{-\infty}^{+\infty} \Phi(\omega) e^{i\xi\omega}\, d\omega = \int_{-i\infty+im}^{+i\infty+im} \Phi(\omega) e^{i\xi\omega}\, d\omega. \tag{10}$$

That is, we must be able to translate the contour through the "period strip." To guarantee this, it is necessary that $\Phi(\omega)$ is

(a) analytic in the period strip $0 < \operatorname{Im}\omega < m$,
(b) such that the line integrals closing the contour at $\operatorname{Re}\omega = \pm\infty$ (Fig. 2) can be neglected. (11)

We can guarantee condition (11a) by solving (9) in the following manner. Define a new function $F(\omega)$ by

$$\Phi(\omega) = \exp F(\omega), \qquad F(\omega + im) - F(\omega) = -\ln K(\omega). \tag{12}$$

In Fourier transform space, this becomes

$$F(x) = \frac{1}{1 - e^{mx}} \int_{-\infty}^{+\infty} d\omega\, e^{i\omega x} \ln K(\omega) + \lambda\, \delta(x), \tag{13}$$

where λ is an arbitrary constant that results from the division by the singular structure $1 - e^{mx}$. It corresponds to the multiplicative constant that always remains undetermined in a homogeneous equation. We may leave the singularity at $x = 0$ unspecified, remembering that a particular choice ($\pm i\epsilon$ etc.,) can only change $\Phi(\omega)$ by a constant factor. Then, a particular solution to (9) is

$$\Phi(\omega) = \exp\left\{\int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{1 - e^{mx}} \times \left[\int_{-\infty}^{+\infty} d\bar{\omega}\, e^{i\bar{\omega}x} \ln K(\bar{\omega})\right]\right\}. \tag{14}$$

The factor $(1 - e^{mx})$ guarantees (11a) as promised; whether or not (11b) is satisfied has to be checked for each particular $K(\omega)$. If so, then the particular solution of (8) is

$$\Phi(\xi) = \int_{-\infty}^{+\infty} e^{i\omega\xi}\, d\omega \exp\left\{\int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{1 - e^{mx}} \times \left[\int_{-\infty}^{+\infty} d\omega\, e^{i\omega x} \ln K(\omega)\right]\right\}. \tag{15}$$

Depending on the particular kernel, there may be other solutions to (8). If $\Phi(\omega)$ is a solution of (9), certainly any function of the form $\mathfrak{A}(\omega)\Phi(\omega)$, where $\mathfrak{A}(\omega) = \mathfrak{A}(\omega + im)$ is also. Generally, however, such

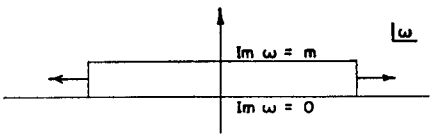

FIG. 2. Closing the contour at $\operatorname{Re}\omega = \pm\infty$.

functions, for example, $\cosh 2\pi\omega/m$, $\operatorname{sech} 2\pi\omega/m$, will either put poles in the period strip, violating (11a), or cause exponential increase at large ω, violating (11b). We postpone the important discussion of the uniqueness of the solutions until later in the paper.

In the particular case of Eq. (7), we find

$$K(\omega) = 2\pi f(g^2) \operatorname{sech} \pi\omega. \tag{16}$$

The solution depends then on the evaluation of the Fourier transform[7]

$$\int_{-\infty}^{+\infty} e^{i\omega x} \ln\left(\frac{\operatorname{sech}\pi\omega}{2}\right) d\omega = \frac{P\pi}{ix} \int_{-\infty}^{+\infty} \tanh \pi\omega\, e^{i\omega x}\, d\omega = P\frac{\pi}{x \sinh \frac{1}{2}x}. \tag{17}$$

In ω space then we find

$$\Phi(\omega) = \exp\left\{\frac{i\omega}{m} \ln [4\pi f(g^2)] + \frac{P}{2} \int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{x(1 - e^{mx}) \sinh \frac{1}{2}x}\right\}, \tag{18}$$

where we have taken $(1 - e^{mx})^{-1} \to P(1 - e^{mx})^{-1}$ for simplicity. Observe that the original symmetry of (9) with this particular $K(\omega)$, namely, $\Phi(\omega) = \Phi^*(-\omega)$ for real ω, is here guaranteed by the principal part. This preserves the reality of the D function on the left. We have received a small bonus in analyticity: (18) is analytic in the strip

$$-\tfrac{1}{2} < \operatorname{Im}\omega < m + \tfrac{1}{2}.$$

The integral that appears in (18) can be evaluated in terms of well-known functions, in particular, logarithms, dilogarithms, and hypergeometric functions. This is done in Appendix III. The explicit form is of use in applications of the theory, but the integral representation is more compact and convenient for the theoretical development of the program.

To decide whether

$$D(-s) = (s)^{-1} \int_{-\infty}^{+\infty} e^{i\omega \ln s^2}\, d\omega \left[\exp\left\{\frac{i\omega}{m} \ln [4\pi(g^2)] + \frac{P}{2} \int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{x(1 - e^{mx}) \sinh \frac{1}{2}x}\right\}\right] \tag{19}$$

[7] An example of this kind of integration is given in Appendix III.

actually solves (5), we must determine whether or not (11b) is satisfied. The behavior of $\Phi(\omega)$ at large $|\omega|$ is dominated by the third-order pole at $x = 0$. A simple calculation yields the asymptotic behavior for large $|\omega|$

$$\Phi(\omega) \sim e^{-(\pi i/2m)\epsilon(\mathrm{Re}\,\omega)(\omega^3 - im\omega)} e^{(i\omega/m)\ln 4\pi/(\omega^2)}. \quad (20)$$

We see that $\Phi(\omega)$ is exponentially damped for $\mathrm{Im}\,\omega < \frac{1}{2}m$, but in the region $\frac{1}{2}m < \mathrm{Im}\,\omega < m$, $|\mathrm{Re}\,\omega| \to \infty$, there is both exponential increase and extremely rapid oscillation. In fact, the vertical line integrals go to the curiously ill-defined form

$$e^{-i\epsilon(\omega)\infty^3} e^{\infty}/\infty^2 \quad (21)$$

as $|\mathrm{Re}\,\omega| \to \infty$. Moreover, the integral along the top of the period strip has the peculiar form

$$\int_{-\infty}^{+\infty} e^{\frac{1}{2}\pi|\omega|}\, d\omega\, e^{i\omega\xi} e^{-(i\pi\omega^3/2m)\epsilon(\omega)} \quad (22)$$

whose integrand has an exponentially increasing modulus! Both structures (21) and (22) lack definition as they stand, and it is not clear whether or not $D(-s)$, as given in (19), formally solves (5).

We need to discuss this seeming lack of definition in some detail. We show that the original equations (3) have built into them a natural definition for these ill-defined structures. In particular, we show that the prescription

$$\Phi(\omega) \to \Phi(\omega) e^{-\epsilon\omega^3} \quad (23)$$

actually yields an $N(s)$ and $D(s)$ that explicitly have the correct discontinuity relations, taking the $\epsilon = 0$ limit after all integrations are performed. This is done in the section below entitled *The N function and the full amplitude N/D: The validity of the convergence factor*. In this way, the results of the convergence factor will justify its use (and raise itself above the level of the FP regulation procedure which, as was discussed in the previous references, creates a solution where there really is none). We postpone the explicit demonstration that the N and D obtained with the convergence factor in fact have the correct discontinuities until we have introduced the relativistic equations which are closer to our direct interest.

Notice that the use of the convergence factor (23) yields exactly our solution (19) because, for $\epsilon \neq 0$, the sufficient conditions (11) are satisfied:

(a) The vertical line integrals at $|\mathrm{Re}\,\omega| = \infty$ go to zero; thus we learn that the convergence factor sets such ill-defined structures to zero.

(b) The $\epsilon = 0$ limit of

$$D_\epsilon(-s) = s^{-\frac{1}{2}} \int_{-\infty}^{+\infty} \exp(i\omega \ln s^{\frac{1}{2}}) \Phi(\omega) e^{-\epsilon\omega^3}\, d\omega \quad (24)$$

is smooth and goes to (19) because (19) was well-defined in the first place.

(c) The inverse transform with contour along the top of the period strip is defined.

In statement (c), we mean that, after the integral is done, the ϵ limit is smooth and finite. For example, consider the family of functions whose integrands (without the convergence factors) have exponentially increasing moduli:

$$G_\epsilon^{(m)}(x) = \int_{-\infty}^{+\infty} e^{+\omega m} e^{+i\omega^3} e^{i\omega x} e^{-\epsilon\omega^3}\, d\omega. \quad (25)$$

Even when m is a positive integer, $G_\epsilon^{(m)}(x)$ is certainly well defined for $\epsilon \neq 0$. The question is whether or not, after evaluating such an integral, the $\epsilon = 0$ limit is smooth and finite. It is simple to show that the answer is in the affirmative. We can rotate the contour of integration in (25) by $+\frac{1}{4}\pi$ to obtain

$$G_\epsilon^{(m)}(s) = \exp(+\tfrac{1}{4}i\pi) \int_{-\infty}^{+\infty} \exp(+mre^{\frac{1}{4}i\pi}) \times \exp(-R^2) \exp(-\epsilon R^2 i) \exp(ixRe^{+\frac{1}{4}i\pi})\, dR. \quad (26)$$

In the rotation of Fig. 3 the factors $e^{+i\omega^3} e^{-\epsilon\omega^3}$ allow us to drop the contribution at large $|\omega|$. The well-defined (26) has an obviously smooth and finite limit in ϵ, *thanks to the tremendous (super-exponential) damping power of* $e^{i\omega^3}$ *along this new contour*:

$$G_0^{(m)}(x) = \exp(+\tfrac{1}{4}i\pi) \int_{-\infty}^{+\infty} \exp(mRe^{\frac{1}{4}i\pi}) \times \exp(-R^2) \exp(ixRe^{+\frac{1}{4}i\pi})\, dR. \quad (27)$$

It is easy to show that, if ϵ is set to zero before the rotation, the contribution at large $|\omega|$ goes to $e^{i\omega^3} e^{\infty}/\infty^2$. Again we see that the convergence factor defines this sort of structure to zero. Actually, $G_\epsilon^m(x)$ can be explicitly evaluated by completing the square and using the convergence factor to translate the contour. We find, for example,

$$G_\epsilon^{(1)}(x) = \exp\left\{-(i-\epsilon)\left[\frac{1+ix}{2(i-\epsilon)}\right]^2\right\} \times \int_{-\infty}^{+\infty} \exp[(i-\epsilon)\omega^2]\, d\omega. \quad (28)$$

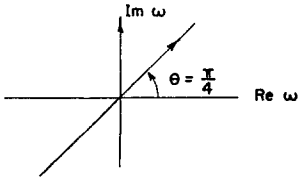

FIG. 3. Rotation that proves the convergence of $G_\epsilon^{(m)}$.

The ϵ limit is smooth and finite because the ϵ always appears in the combination $i - \epsilon$ (the i is from the $e^{i\omega^2}$!). We find

$$G_0^{(1)}(x) = (1+i)(\tfrac{1}{2}\pi)^{\frac{1}{2}} \exp\left[\tfrac{1}{4}(-ix^2 - 2x + i)\right]. \quad (29)$$

We have learned then in both methods of evaluation of $G^{(m)}$ that, although the convergence factor $e^{-\epsilon\omega^2}$ will define, for $\epsilon \neq 0$, any integral the modulus of whose integrand increases exponentially, only the powerful oscillations of the $e^{i\omega^2}$ factor, and its associated damping power, allows a smooth finite $\epsilon = 0$ limit after the integration! Without the $e^{-i\omega^2}$, the $e^{-\epsilon\omega^2}$ is powerless to bring about any lasting convergence; for example, although

$$\int_{-\infty}^{+\infty} e^{+\omega} e^{ix\omega} e^{-\epsilon\omega^2}\, d\omega = \exp\left[+\epsilon\left(\frac{1+ix}{2\epsilon}\right)^2\right] \int_{-\infty}^{+\infty} e^{-\epsilon\omega^2}\, d\omega \quad (30)$$

is well defined for $\epsilon \neq 0$, its $\epsilon = 0$ limit blows up.

Two final comments are in order here about integrals like $G^{(m)}(x)$: first, we note that we can, in general, expect exponential increase in x from integrals of this kind. For example, as we see in (29), the modulus of $G_0^{(1)}(x)$ increases exponentially for large negative x. Second, it is worth noting that considerations similar to those just given can be extended to include a very large class of curious integrals. Consider

$$G_\epsilon^{m,n}(x) = \int_{-\infty}^{+\infty} F^{m,n}(\omega) e^{i\omega x}\, d\omega, \qquad F^{m,n} \sim e^{(m\omega)^n} e^{i\omega^{n+1}}. \quad (31)$$

Despite terrible superexponential increase of the moduli of these integrands, the $G^{m,n}(x)$ have smooth, finite limits with the convergence factors

$$e^{-\epsilon|\omega|^{n+1}}.$$

The proof of this for all n is analogous to that given for $n = 1$. For general n, the necessary rotation is $\Theta_n = \pi/2(n+1)$. The case $n = 0$ is, in fact, simply the ordinary δ-function definition! Our convergence factor is in this sense a generalization of the ordinary $e^{-\epsilon|\omega|}$ that we would use naturally in Fourier transforming an ordinary oscillatory function like $\cos\omega$. The simple convergence factor $e^{-\epsilon|\omega|}$ can, of course, be put on a rigorous footing,[8] but we cannot do anything so simple for $n > 0$. Presumably the use of $e^{-\epsilon\omega^2}$ etc., could be made rigorous in distribution theory, but as stated above, our justification for its use is the fact that it will lead us to N and D functions which explicitly satisfy the discontinuity relations.

[8] E. C. Titchmarsh, *Theory of Fourier Integrals* (Oxford Clarendon Press, New York, 1950).

Relativistic N/D

The method can easily be extended to the relativistic homogeneous N/D equations. The situation in this case is essentially the same as for the NR equations. Consider

$$N(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1} \frac{D(\nu')\alpha(\nu')\, d\nu'}{\nu' - \nu}, \qquad D(\nu) = -\frac{1}{\pi}\int_0^{\infty} \frac{\rho(\nu')N(\nu')\, d\nu'}{\nu' - \nu}, \quad (32)$$

where ν is the dimensionless center-of-mass (momentum)^2. We take as the "most singular" approximation to this,

$$N_0(\nu) = \int_{-\infty}^{0} \frac{D_0(\nu')\alpha_0(\nu')\, d\nu'}{\nu' - \nu}, \qquad D_0(\nu) = -\frac{1}{\pi}\int_0^{\infty} \frac{N_0(\nu')\, d\nu'}{\nu' - \nu}, \quad (33)$$

where, consistent with keeping only

$$\alpha_0(\nu) = f(g^2)(-\nu)^m,$$

we have set $\rho(\nu) = 1$. This approximation preserves only high-energy unitarity, but we enforce exact unitarity with the iteration scheme: for example,

$$N_1(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1} \frac{D_0(\nu')\alpha(\nu')\, d\nu'}{\nu' - \nu}, \qquad D_1(\nu) = -\frac{1}{\pi}\int_0^{\infty} \frac{\rho(\nu')N_1(\nu')\, d\nu'}{\nu' - \nu}. \quad (34)$$

Defining, in Eq. (33)

$$D_0(-\nu) = \bar{D}(\nu), \quad \alpha(-\nu) = \bar{\alpha}(\nu), \quad \nu = e^{\xi}, \quad \nu > 0, \qquad e^{\frac{1}{2}\xi}N_0(e^{\xi}) = \chi(\xi), \quad e^{\frac{1}{2}\xi}\bar{D}(e^{\xi}) = \Phi(\xi), \quad (35)$$

we obtain the two coupled integral equations for N_0, D_0,

$$\chi(\xi) = -\frac{f(g^2)}{2\pi}\int_{-\infty}^{+\infty} d\eta\, \operatorname{sech}\tfrac{1}{2}(\eta - \xi)\Phi(\eta)e^{m\eta}, \qquad \Phi(\xi) = -\frac{1}{2\pi}\int_{-\infty}^{+\infty} d\eta\, \operatorname{sech}\tfrac{1}{2}(\eta - \xi)\chi(\eta). \quad (36)$$

These go over, in Fourier transform space, into the two coupled finite difference equations,

$$\chi(\omega) = -f(g^2)\operatorname{sech}\pi\omega\Phi(\omega + im), \qquad \Phi(\omega) = -\operatorname{sech}\pi\omega\chi(\omega). \quad (37)$$

The equation for the D function is

$$\Phi(\omega) = f(g^2)\,\mathrm{sech}^2\,\pi\omega\Phi(\omega + im), \qquad K(\omega) = f(g^2)\,\mathrm{sech}^2\,\pi\omega. \tag{38}$$

Using the general technique proposed above, we obtain

$$D(-\nu) = \frac{1}{\nu^{\frac{1}{2}}}\int_{-\infty}^{+\infty} d\omega\, e^{i\omega \ln\nu} e^{(i\omega/m)\ln f(g^2)4} \times \exp\left[P\int_{-\infty}^{+\infty}\frac{dx\, e^{-i\omega x}}{x(1-e^{mx})\sinh\frac{1}{2}x}\right]. \tag{39}$$

This is almost identical to the nonrelativistic solution, and all the previous discussion about the convergence factor, etc., may be taken over here in toto. It should be mentioned that here too we find $\Phi(\omega)$ analytic in the upper half-plane. This time

$$\Phi(\omega) \sim \exp\{2\ln[\omega - i(m+\tfrac{1}{2})]\} \sim [\omega - i(m+\tfrac{1}{2})]^2 \tag{40}$$

near $\omega = i(m + \frac{1}{2})$. In the relativistic case, the zero is second order, in contrast with the first-order zero found in the nonrelativistic case. Subsequent application of the difference equation (38) proves the analyticity in the upper half-plane, just as in the NR case. Again, the lower half-plane is analytic except for poles of increasing order on the negative imaginary axis. The principal part integral of (39) is evaluated in Appendix III.

The N function and the full amplitude $A = N/D$: the validity of the convergence factor

In this section, we explicitly establish that the ϵ convergence factor prescription gives the correct right- and left-hand discontinuities. In this, we concern ourselves primarily with the relativistic equations, but our remarks apply in principle to the nonrelativistic case as well.

Knowledge of the function $\Phi(\omega)$ solves the scattering problem completely because, for example, using Eq. (37) and Eq. (38), we can find the N function

$$N(\nu) = -\frac{1}{\nu^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln\nu}\cosh\pi\omega\Phi(\omega), \quad \nu > 0, \tag{41}$$

which saves us doing the Hilbert transformation that is usually necessary to find N from D. The Fourier transform structure naturally incorporates the real analyticity of N and D, as is clear on taking ν complex:

$$D(r,\vartheta_D) = \frac{e^{-\frac{1}{2}i\vartheta_D}}{r^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln r}e^{-\omega\vartheta_D}\Phi(\omega)\,d\omega, \tag{42a}$$

$$N(r,\vartheta_N) = -\frac{e^{-\frac{1}{2}i\vartheta_N}}{r^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln r}e^{-\omega\vartheta_N} \times \cosh\pi\omega\Phi(\omega)\,d\omega, \tag{42b}$$

where $\nu = re^{i\vartheta}$ and the cuts and phases are taken as in Fig. 4.

Notice from Eq. (42) that D on the right and left, and N on the right, are convergent integrals. On the other hand, N on the left, when written in the form of this integral representation, needs the convergence factor for convergence.

Above and below the cuts, we find

$$D(\nu \pm i\delta) = \pm\frac{i}{\nu^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln\nu}e^{\pm\omega\pi}\Phi(\omega)\,d\omega, \quad \nu > 0, \tag{43a}$$

$$N(\nu \pm i\delta) = \pm\frac{i}{(|\nu|)^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln|\nu|}\,d\omega\, e^{\mp\omega\pi} \times \cosh\pi\omega\Phi(\omega), \quad \nu < 0. \tag{43b}$$

Also, it is simple to verify from Eqs. (42) and (43) that for $\nu > 0$

$$N(\nu) = (1/2i)[D(\nu - i\delta) - D(\nu + i\delta)]. \tag{44a}$$

This is the ordinary discontinuity relation on the right, reflecting unitarity. Actually, because N and D on the right are convergent integrals, Eq. (44a) follows without use of the convergence factor. For $\nu < 0$, using Eq. (38)

$$\begin{aligned}\mathrm{Im}\,N(\nu + i\delta) &= \frac{1}{2i}[N(re^{i\pi}) - N(re^{-i\pi})]\\ &= \frac{1}{r^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln r}\cosh^2\pi\omega\Phi(\omega)\,d\omega\\ &= \frac{f(g^2)}{r^{\frac{1}{2}}}r^m\int_{-\infty+im}^{+\infty+im} e^{i\omega\ln r}\Phi(\omega)\,d\omega\\ &= f(g^2)r^m D(\nu).\end{aligned} \tag{44b}$$

In this, we have used our ability to translate through the strip in the last step. *To do this, the convergence factor is necessary.* Equation (44b) is the ordinary discontinuity relation on the left, the statement about the left-hand input. Thus, we find that, using the convergence factor when we need it,

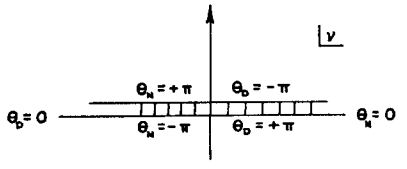

FIG. 4. Cuts and phases for N and D.

automatically yields the correct discontinuity relations. This proves then the validity of the convergence factor. Below we find the explicit forms of N and D, at least asymptotically, and verify explicitly the preservation of the discontinuity relations.

We have occasion below to use the further relations derivable from Eq. (42):

$$N(\nu) = \tfrac{1}{2}i[D(\nu) + D(\nu e^{2\pi i})], \tag{45}$$

$$A(\nu) = \begin{cases} \dfrac{i\int_{-\infty}^{+\infty} e^{i\omega \ln r} \cosh \pi\omega \Phi(\omega)\, d\omega}{\int_{-\infty}^{+\infty} e^{i\omega \ln r} e^{\omega\pi} \Phi(\omega)\, d\omega}, & \nu > 0, \\ \dfrac{i\int_{-\infty}^{+\infty} e^{i\omega \ln r} e^{-\omega\pi} \cosh \pi\omega \Phi(\omega)\, d\omega}{\int_{-\infty}^{+\infty} e^{i\omega \ln r} \Phi(\omega)\, d\omega}, & \nu < 0. \end{cases} \tag{46}$$

The meaning of the convergence factor

Before going any further it proves very instructive to study the nature of the $e^{-\epsilon\omega^2}$ factor more closely. We find that it corresponds to solving a modified problem in which the original input is multiplied by an oscillatory factor which goes to unity as $\epsilon \to 0$.

To show this, we consider, instead of Eq. (3), the slightly modified set of N/D equations

$$N_\epsilon(\nu) = \frac{1}{\pi}\int_{-\infty}^{0} \frac{D_\epsilon(\nu')\bar{\alpha}(\nu')\, d\nu'}{\nu' - \nu}, \qquad D_\epsilon(\nu) = -\frac{1}{\pi}\int_0^{\infty} \frac{N_\epsilon(\nu')\, d\nu'}{\nu' - \nu}, \tag{47}$$

where the absorptive part on the left, $\bar{\alpha}(\nu)$, is defined to be

$$\bar{\alpha}(\nu) = \alpha(\nu)\, D_\epsilon(\nu^\epsilon)/D_\epsilon(\nu) \tag{48}$$

and $\alpha(\nu)$ is the original input. $D_\epsilon(\nu)$ is the solution itself and ϵ is a small positive parameter. We are studying then a self-consistent problem of sorts in which the discontinuity on the left is determined by the solution, which determines the discontinuity on the left, and so forth. In the limit $\epsilon = 0$, Eqs. (47) reduce to Eq. (3). In these slightly modified equations, e.g., in the equation for D,

$$D_\epsilon(-\nu) = \frac{1}{\pi^2}\int_0^{\infty} \frac{D_\epsilon(-\nu' e^\epsilon)\alpha(-\nu')}{\nu' - \nu} \ln\frac{\nu'}{\nu}\, d\nu', \tag{49}$$

we are simply refusing to evaluate the two D_ϵ's at exactly the same value of s until the end of the calculation. In Fourier transform space, we have

$$\Phi_\epsilon(\omega) = e^{i\epsilon\omega} f(g^2)\Phi_\epsilon(\omega + im) \operatorname{sech}^2 \pi\omega, \tag{50}$$

whose solution, in terms of Eq. (18) is simply

$$\Phi_\epsilon(\omega) = \Phi(\omega) e^{-\epsilon\omega^2/2m} e^{\frac{1}{2}i\omega\epsilon}. \tag{51}$$

Thus, the modified input (48) is completely equivalent to the convergence factor prescription (23). It is quite clear that (51) satisfies (11) and we obtain, for $\epsilon > 0$,

$$D_\epsilon(-\nu) = \nu^{-\frac{1}{2}} \int_{-\infty}^{+\infty} e^{i\omega \ln \nu}\, d\omega \times \exp\left[\frac{i\omega}{m}\ln f(g^2)4 - \frac{\omega^2\epsilon}{2m} + \frac{i\omega\epsilon}{2} + P\int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{x(1 - e^{mx})\sinh \frac{1}{2}x}\right]. \tag{52}$$

The smooth $\epsilon \to 0$ limit again yields exactly our formal solution (19).

What does this small translation $\nu \to \nu e^\epsilon$ in (47) mean physically? We know that it corresponds to a modified left-hand discontinuity, the high-energy behavior of which we want to study here. It turns out that $D_\epsilon(\nu e^\epsilon)/D_\epsilon(\nu)$ is oscillatory for $|\nu| \to \infty$, $\epsilon \neq 0$, thus helping to tone down the terrible inputs of the original problem. We are interested then in finding the high-energy limit of $\Phi_\epsilon(\xi)$.

Writing

$$\Phi_\epsilon(\xi) = \int_{-\infty}^{+\infty} e^{i\xi\omega}\Phi(\omega)e^{-\epsilon\omega^2}\, d\omega \tag{53}$$

with $\Phi(\omega)$ as given in (18), and using (20), we find two saddle points in the integrand at

$$\omega_\pm = \frac{1}{\pm 1 - (im\epsilon/\pi)} \times \left[\pm\frac{im}{2} + \frac{m\xi}{2\pi} + \frac{1}{2\pi}\ln 4f(g^2)\right],$$

$$\operatorname{Im}\omega_\pm = \frac{1}{\pm 1 + (m^2\epsilon^2/\pi^2)} \times \left\{\frac{m}{2} + \frac{m\epsilon}{2\pi^2}[m\xi + \ln 4f(g^2)]\right\}. \tag{54}$$

The motion of these for fixed ξ, as $\epsilon \to 0$, are shown in Fig. 5. Presumably there is also some sort of singularity at $\operatorname{Im}\omega = m + \frac{1}{2}$, along which the Fourier transform in (18) ceases to converge. It

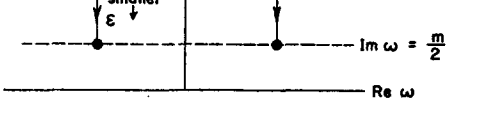

FIG. 5. Motion of the saddle points in the integrand of $\Phi\epsilon(\xi)$.

appears at first glance that the saddle points will dominate the behavior as long as $\text{Im}\ \omega_\pm < m + \frac{1}{2}$. However, for higher ξ, i.e., $\xi \geq (m+1)\pi^2/m^2\epsilon$, one expects the singularity at $\text{Im}\ \omega = m + \frac{1}{2}$ to dominate, as the saddle points at very high energy are above that line. Actually, this is not the case; as we show, in fact, doing the integral in $\Phi(\omega)$ yields a function analytic in the upper half-plane—so that the contribution from the saddle points dominates at high energies.

We can translate the contour of (53) up to $\text{Im}\ \omega = m + \frac{1}{2}$ to find

$$\Phi_\epsilon(\xi) = e^{-(m+\frac{1}{2})\xi} \int_{-\infty}^{+\infty} e^{i\omega\xi}\Phi_\epsilon[\omega + i(m+\tfrac{1}{2})]\, d\omega, \quad (55)$$

$$\Phi_\epsilon[\omega + i(m+\tfrac{1}{2})] = \exp\left\{\frac{-\epsilon}{2m^4}[\omega + i(m+\tfrac{1}{2})]^2\right\} \times \exp\left\{\frac{i[\omega + i(m+\frac{1}{2})]}{m}\ln 4f(g^2) - 2P\int_{-\infty}^{+\infty}\frac{e^{-ix\omega}\,dx}{x(1-e^{-mx})(1-e^{-x})}\right\}.$$

Certainly the integral in the exponent has a singularity at $\omega = 0$. An elementary calculation yields

$$P\int_{-\infty}^{+\infty}\frac{e^{-ix\omega}\,dx}{x(1-e^{-mx})(1-e^{-x})} \sim -\ln\omega, \quad \omega \to 0. \quad (56)$$

Thus, near $\omega = 0$, we find

$$\Phi[\omega + i(m+\tfrac{1}{2})] \sim \omega^2. \quad (57)$$

That is, the exponentiation of the logarithmic singularity has completely removed it, leaving the boundary of the period strip free of singularities! Moreover, since

$$\Phi(\omega + im) = [f(g^2)]^{-1}\cosh^2\pi\omega\Phi(\omega) \quad (58)$$

then analyticity in the strip $0 < \text{Im}\ \omega \leq m + \frac{1}{2}$ is sufficient to guarantee analyticity in the entire upper half-plane, except for the point at infinity. This is not to say that the lower half-plane is analytic, because to relate down a period strip, one must multiply by factors of $\text{sech}^2\ \pi\omega$, which will generate an infinite set of poles of higher and higher order along the negative imaginary axis. Except for these poles, the lower half-plane is also analytic.

With $\epsilon = 0$, it is a straightforward matter to translate the integral (53) to the line $\text{Im}\ \omega = \text{Im}\ \omega_\pm$ and approximate by the method of stationary phase along that path. Dropping terms in ϵ^2 and keeping only terms linear or quadratic in ξ, we find the dominant contribution at large ξ to be

$$\Phi(\xi) \sim e^{-\xi v_+}e^{-\epsilon\xi^2(m^2/4\pi^2)}e^{(m\xi\epsilon/\pi)[1-1/2\pi\ \ln 4f(g^2)]} \times \cos\left[-\frac{\pi}{m}u_+^2 + \xi u_+ - mu_+\epsilon + \frac{u_+}{m}\ln 4f(g^2)\right], \quad (59)$$

where $\omega_+ = u_+ + iv_+$. We emphasize that the analyticity in the upper half-plane is independent of ϵ, and is found in the $\epsilon = 0$ limit as well. Only when $\epsilon \neq 0$, however, can we translate the transform contour freely through the upper half-plane, neglecting the contributions from the vertical integrations at the ends. This ability to translate guarantees formally that $\Phi(\xi)$ decreases faster than any exponential in ξ. In particular, we see in (59) that both of the first two factors exhibit such behavior, that is decrease like $e^{-\epsilon\xi^2}$. Thus, we learn that with $\epsilon \neq 0$, both $\Phi_\epsilon(\xi)$ and $\Phi_\epsilon(\omega)$ decrease faster than any exponential, due to the presence of the convergence factors $e^{-\epsilon\xi^2} = e^{-\epsilon\ln^2\nu}$ and $e^{-\epsilon\omega^2}$.

Our immediate interest is the ratio of the two D functions. Using (59), we find

$$\frac{D(-\nu e^{\epsilon})}{D(-\nu)} \sim e^{-\epsilon v_+}(\nu)^{-\epsilon^2(m^2/2\pi^2)}e^{(m\epsilon^2/\pi)[1-(1/2\pi)\ \ln 4f(g^2)]} \times \frac{\cos[\Psi(\ln\nu + \epsilon\ln\nu)]}{\cos[\Psi(\ln\nu)]}, \quad (60)$$

$$\Psi(\ln\nu) = -\frac{\pi}{m}u_+^2 + \epsilon(\ln\nu)u_+ + \frac{u_+}{m}\ln 4f(g^2),$$

which for high ν implies,

$$\bar{\alpha}(-\nu) \sim \alpha(-\nu)\frac{\cos[\Psi(\ln\nu) + \epsilon u_+(\ln\nu)]}{\cos[\Psi(\ln\nu)]}. \quad (61)$$

The oscillatory $\bar{\alpha}$ gives better defined equations than α itself. This is the nature of the $e^{-\epsilon\omega^2}$ convergence factors. Notice that $\bar{\alpha}(\nu)$ contains ghosts, that is, poles near $\nu = -\infty$, from the zeros of $\cos[\Psi(\ln\nu)]$ in its denominator. Thus, the integral $\int \bar{\alpha}\, D_\epsilon$ is well defined only if D_ϵ develops zeros to cancel these, which, of course, it does. We discuss these ghosts later.

We might have anticipated that multiplying $\Phi(\omega)$ by $e^{-\epsilon\omega^2}$ corresponded to a high-energy regulation process: notice that the substitution

$$\Phi(\omega) \to \Phi(\omega e^{-i\,\delta\epsilon(\omega)}), \quad 1 \gg \delta > 0;$$
$$\epsilon(\omega) = \begin{cases} +1, & \omega > 0, \\ -1, & \omega < 0, \end{cases} \quad (62)$$

results in

$$e^{-i\omega^2\epsilon(\omega)} \to e^{-i\omega^2\epsilon(\omega)}e^{-\delta\omega^2}, \quad (63a)$$

$$\Phi[\omega e^{-i\,\delta\epsilon(\omega)}] = \int_{-\infty}^{+\infty} e^{-ix\omega}e^{-\delta|\omega|\xi}\Phi(\xi)\, d\xi, \quad (63b)$$

which links the $e^{-\epsilon\omega^2}$ factor with a curious high-energy ($\xi \to +\infty$) regulation.

We emphasize that since we have shown the convergence factor prescription to give the correct right and left discontinuities, all the structure discussed in this section is naturally implied in the problem. After we get some more feeling for the solutions themselves, and solve the inhomogeneous equations, we shall make some final comments on the convergence factor.

High-energy behavior of $D(\nu)$ on the left

The convergence factor prescription yields the ϵ-free well-defined (19) for $D(\nu)$. What is its high-energy behavior on the left? The third-order pole at $x = 0$ dominates again, and we find

$$\Phi(\omega) \sim \exp\left[(-\pi i/m)\epsilon(\mathrm{Re}\,\omega)(\omega^2 - im\omega) + (i\omega/m)\ln 4f(g^2)\right]. \quad (64)$$

Thus, the integrand has two saddle points at

$$\omega_+ = -\omega_-^* = \frac{im}{2} + \frac{\xi m}{2\pi} + \frac{1}{2\pi}\ln 4f(g^2). \quad (65)$$

Since $\Phi(\omega)$ decays exponentially for $0 < \mathrm{Im}\,\omega < \frac{1}{2}m$, $|\mathrm{Re}\,\omega| \to \infty$, we can translate the contour up to $\mathrm{Im}\,\omega = \mathrm{Im}\,\omega_\pm$ with inpunity and use the method of stationary phase to obtain the asymptotic behavior of D:

$$D(-\nu) \sim \nu^{-\frac{1}{2}}\nu^{-\frac{1}{2}m}2[f(g^2)]^{-\frac{1}{2}} \times \cos\left[\frac{\pi}{m}u_+^2 - u_+\ln\nu - \frac{u_+}{m}\ln 4f(g^2)\right]$$
$$u_+ = \frac{m}{2\pi}\ln\nu + \frac{1}{2\pi}\ln 4f(g^2) = \mathrm{Re}\,\omega_+. \quad (66)$$

Notice then that the integral $\int^\infty \alpha D$ goes like $\int^\infty \nu^{+\frac{1}{2}m}\cos(\ln^2\nu + \cdots)$. We found above that there are convergence factors in both ω and ξ, so such integrals in ν space should be taken as $\int^\infty \nu^{+\frac{1}{2}m}\cos(\ln^2\nu + \cdots)e^{-\epsilon\ln^2\nu}$ which, again, has a finite smooth limit in ϵ.

Asymptotic behavior in the physical region

Far to the right, writing the integrand of (43a) as $e^{iJ_+(\omega,\xi)}$, we find

$$J_+(\omega,\xi) = \omega\xi - (\pi\omega^2/m) + (\omega/m)\ln 4f(g^2) \quad (67)$$

from which we find a saddle point at $\bar{\omega}_+$ [see (66)]. There is also a negative saddle point, this time off the real axis, which is dominated by the contribution from $\bar{\omega}_+$. The method of stationary phase along the real axis yields, dropping all constants common to N and D,

$$D(\nu + i\epsilon) \sim \frac{i}{\nu^{\frac{1}{2}}}\exp iZ, \quad (68)$$
$$Z = \bar{\omega}_+\ln\nu - \frac{\pi\bar{\omega}_+^2}{m} + \frac{\bar{\omega}_+}{m}\ln 4f(g^2)$$
$$= \frac{m\ln^2\nu}{4\pi} + \frac{\ln\nu}{2\pi}\ln f(g^2)4 + \frac{\ln^2 4f(g^2)}{4\pi m}.$$

A similar analysis on (42b) yields the *explicitly unitary* amplitude (N/D)

$$A(\nu) \sim +i\cos Ze^{-iZ}. \quad (69)$$

Opposite sign of the input

Before going on, we want to note that the general technique can be used to find a solution for $f(g^2) < 0$ as well. We find on evaluation of (14) in this case, simply (39) with

$$e^{(i\omega/m)\ln f(g^2)} \to e^{(i\omega/m)\ln|f(g^2)|}e^{\pm\pi\omega/m} \quad (70a)$$

depending on the branch we choose for $\ln(-1)$. Clearly, to maintain the reality of the D function, we must choose the linear combination of these two solutions

$$e^{(i\omega/m)\ln|f(g^2)|}\cosh(\pi\omega/m). \quad (70b)$$

[Having $\Phi(\omega) = \Phi^*(-\omega^*)$ guarantees a real D.]

The $e^{i\omega^2}$ oscillation (with the convergence factor) defines this solution as well! It is a general feature of these equations that solutions exist for either sign of $f(g^2)$.

B. INHOMOGENEOUS N/D EQUATIONS

As discussed in Refs. 1 and 2 in order to guarantee (in the W theory) that the Born term appear explicitly in the iterative solution of the N/D equations, we want to write them with one subtraction at threshold in N and D, with the subtraction constant in N set to zero in all partial waves. The method proposed above for the solution of the homogeneous equations can easily be extended to this case. Consider the relativistic inhomogeneous equation for the D function in this case in the "most singular" approximation

$$D(-\nu) = 1 + \frac{\nu}{\pi^2}\int_0^\infty \frac{\alpha(-\nu')\,D(-\nu')\,d\nu'}{\nu'(\nu' - \nu)}\ln\left(\frac{\nu'}{\nu}\right) \quad (71)$$

[i.e., we have taken, just as before, $\rho(\nu) = 1$, and let the left cut come up to threshold]. The substitutions

$$\alpha(-\nu) = f(g^2)\nu^m;\ D(-\nu) = \Phi(\xi),\ \nu = e^\xi,\ \nu > 0 \quad (72)$$

yield

$$\Phi(\xi) = 1 + \frac{f(g^2)}{\pi^2}\int_{-\infty}^{+\infty}\frac{e^{\eta m}\Phi(\eta)(\eta - \xi)}{e^{\eta-\xi} - 1}\,d\eta. \quad (73)$$

It is clear that an attempt to solve (73) by iteration yields a termwise divergent series. Thus, we expect the solution to be singular at $f(g^2) = 0$.

We present here a general method for the solution of the class of inhomogeneous integral equations

$$\Phi(\xi) = 1 + \frac{1}{2\pi}\int_{-\infty}^{+\infty} K(\xi - \eta)\Phi(\eta)e^{m\eta}\,d\eta, \qquad (74)$$

In Fourier transform space this corresponds to an inhomogeneous finite difference equation

$$\Phi(\omega) = \delta(\omega) + K(\omega)\Phi(\omega + im). \qquad (75)$$

To solve this equation, we guess a "variation of parameter"-type solution of the form

$$\Phi(\omega) = \Phi_H(\omega)G(\omega). \qquad (76)$$

where $\Phi_H(\omega)$ is the solution of the auxiliary homogeneous equation,

$$\Phi_H(\omega) = K(\omega)\Phi_H(\omega + im). \qquad (77)$$

This is soluble by our previous method. On substitution of (76) into (75), we find that $G(\omega)$ satisfies the "Green's function" difference equation

$$G(\omega + im) - G(\omega) = -\Phi_H^{-1}(0)\,\delta(\omega) \qquad (78)$$

as long as $\Phi_H(0)$ is finite and nonzero. (We find that this is generally the case.) This is algebraic in Fourier transform space with the immediate solution

$$G(\omega) = \frac{\Phi_H^{-1}(0)}{2\pi}\int_{-\infty}^{+\infty}\frac{e^{-i\omega x}\,dx}{1 - e^{mx}}. \qquad (79)$$

As in all Green's functions, we have a choice of three ways to go around the pole at $x = 0$. Thus, we have three Green's functions

$$\Phi_H(0)G_P(\omega) \equiv P\int_{-\infty}^{+\infty}\frac{dx}{2\pi}\frac{e^{-ix\omega}}{1 - e^{mx}}$$

$$= \frac{i}{2m}\coth\frac{\pi}{m}(\omega + i\epsilon), \qquad (80a)$$

$$\Phi_H(0)G_\pm(\omega) = \int_{-\infty\pm i\epsilon}^{+\infty\pm i\epsilon}\frac{dx}{2\pi}\frac{e^{-ix\omega}}{1 - e^{mx}}$$

$$= \frac{i}{2m}\operatorname{csch}\frac{\pi}{m}(\omega + i\epsilon)e^{\pm\pi\omega/m}. \qquad (80b)$$

Which of these solutions to use in any particular case is clear on demanding that the final $\Phi(\omega)$ have the symmetry property $\Phi(\omega) = \Phi^*(-\omega)$ for real ω (to guarantee the reality of the D function). For example, if $\Phi_H(\omega)$ comes out symmetric, i.e., $\Phi_H(\omega) = \Phi_H^*(-\omega)$, then we clearly want to multiply by the symmetrical Green's function

$$G_P(\omega)[G_P(\omega) = G_P^*(-\omega)],$$

because this will maintain the symmetry of $\Phi(\omega)$. If $\Phi_H(\omega)$ comes out asymmetrical, and sometimes it will, we want to use one of the "asymmetrical" Green's functions $G_\pm(\omega)$.

We notice that (80) solves (78) only if we agree to approach singularities on the boundary of the strip from within, i.e., if we interpret (78) as $G(\omega + im - i\epsilon) - G(\omega + i\epsilon) = -\Phi_H^{-1}(0)\delta(\omega)$. This prescription is inherent in the entire Fourier transform technique. In particular, it tells us that periodic functions with no singularities on the boundary of the period strip satisfy simple periodicity relations like $P(\omega) = P(\omega + im)$, but if the periodic function has a singularity on the boundary of the strip, it satisfies an inhomogeneous Green's function relation like (78). This is an important distinction in our later discussion of the uniqueness of the solutions.

Both $\Phi_H(\omega)$ and $G(\omega)$, and hence $\Phi(\omega)$, satisfy (11a) by explicit construction. If condition (11b) is satisfied, we find by inverse Fourier transform

$$\Phi(\xi) = \frac{i}{2m}\Phi_H^{-1}(0)\int_{-\infty}^{+\infty} e^{i\xi\omega}\,d\omega\coth\frac{\pi}{m}(\omega + i\epsilon)$$

$$\times\exp\left\{\int_{-\infty}^{+\infty}\frac{dx\,e^{-i\omega x}}{1 - e^{mx}}\left[\int_{-\infty}^{+\infty} d\tilde{\omega}\,e^{i\tilde{\omega}x}\ln K(\tilde{\omega})\right]\right\}, \qquad (81)$$

where

$$\Phi_H(0) = \exp\left\{\int_{-\infty}^{+\infty}\frac{dx}{1 - e^{mx}}\right.$$

$$\left.\times\left[\int_{-\infty}^{+\infty} d\omega\,e^{i\omega x}\ln K(\omega)\right]\right\} \qquad (82)$$

and we have assumed $G(\omega) = G_P(\omega)$ for simplicity.

In the case of interest, Eq. (73), we find the difference equation

$$\Phi(\omega) = \delta(\omega) - f(g^2)\operatorname{csch}^2\pi(\omega - i\epsilon)\Phi(\omega + im). \qquad (83)$$

Necessary in the derivation is the identity

$$\int_{-\infty}^{+\infty}\frac{e^{-i\xi\omega}\xi\,d\xi}{e^{-\xi} - 1} = \pi^2\operatorname{csch}^2\pi(\omega - i\epsilon). \qquad (84)$$

The method of solution outlined above leads to

$$\Phi_H(\omega) = \exp\left\{\frac{i\omega}{m}\ln 4f(g^2)\right.$$

$$\left. + 2P\int_{-\infty}^{+\infty}\frac{e^{-ix\omega}}{(1 - e^{-x})(1 - e^{mx})}\right\}. \qquad (85)$$

To obtain this, one needs the identity[9]

$$2\int_{-\infty}^{+\infty} d\omega\,e^{ix\omega}\ln[\operatorname{csch}\pi(\omega - i\epsilon)]$$

$$= P\frac{4\pi}{x(1 - e^{-x})} + (2\pi\ln 4 + 2\pi^2 i)\,\delta(x)$$

[9] See Appendix I.

or

$$\int_{-\infty}^{+\infty} d\omega\, e^{iz\omega} \ln\left[-\operatorname{csch} \pi(\omega - i\epsilon)\right] = P\,\frac{2\pi}{x(1-e^{-x})}. \tag{86}$$

This implies a choice of branch in the logarithm. Other choices correspond to other solutions. We discuss the multiplicity of solutions to these equations in a later section. The fact that $\Phi_H(\omega) = \Phi_H^*(-\omega)$ requires the use of the symmetrical Green's function $G_P(\omega)$, and we obtain

$$D(-\nu) = \frac{i}{2m}\,\Phi_H^{-1}(0)\int_{-\infty}^{+\infty} d\omega\, e^{i\omega \ln \nu} \times \coth\frac{\pi}{m}(\omega + i\epsilon)\exp\left[\frac{i\omega}{m}\ln 4f(g^2) + 2P\int_{-\infty}^{+\infty}\frac{dx}{x}\,\frac{e^{-ix\omega}}{(1-e^{-x})(1-e^{mx})}\right] \tag{87}$$

(using a convergence factor to prove that (11) is satisfied), where

$$\Phi_H(0) = \exp\left[2P\int_{-\infty}^{+\infty}\frac{dx}{x(1-e^{mx})(1-e^{-x})}\right] \tag{88}$$

is indeed finite and nonzero. This is true because of the small bonus in analyticity evident in (87): $\Phi_H(\omega)$ is analytic in the strip $-1 < \operatorname{Im}\omega < m$, thus preventing $\Phi_H(\omega)$ from being singular. In Appendix II, the rather surprising identity,

$$\exp\left[-2P\int_{-\infty}^{+\infty}\frac{dx}{x(1-e^{mx})(1-e^{-x})}\right] = m = \Phi_H^{-1}(0) \tag{89}$$

is established, which simplifies (87). Actually, because $\Phi_H(0)$ is a constant, it will appear in the N function as well, and thus will divide out of the amplitude. Again, the fact that $\Phi(\omega)$ satisfies (11) depends on the appearance of the exp $(i\omega^2)$ oscillations (and the associated convergence factor).

Finally, we mention that, just as discussed above in the homogeneous case, the amplitude can easily be shown to have the correct discontinuity relations (thus justifying the convergence factor).

The inhomogeneous once-subtracted NR equations

$$N_0(s) = \frac{s}{\pi}\int_{-\infty}^{0}\frac{D_0(s')\alpha_0(s')\,ds'}{s'(s'-s)}\,D_0(s) = 1 - \frac{s}{\pi}\int_0^{\infty}\frac{N_0(s')(s')^{\frac{1}{2}}\,ds'}{s'(s'-s)} \tag{90}$$

can be solved in much the same way as the relativistic. With the substitutions

$$\alpha_0(-s) = -f(g^2)s^{\frac{1}{2}(m-1)}; \quad D_0(-s) = \Phi(\xi), \quad N_0(s) = \chi(\xi), \quad s = e^{2\xi}, \quad s > 0, \tag{91}$$

we find the coupled difference equations for N and D

$$\Phi(\omega) = \delta(\omega) - i\operatorname{csch}\tfrac{1}{2}\pi(\omega - i\epsilon)\chi(\omega + i), \tag{92a}$$

$$\chi(\omega) = +if(g^2)\operatorname{csch}\tfrac{1}{2}\pi(\omega - i\epsilon) \times \Phi[\omega + i(m-1)]. \tag{92b}$$

With (92a), we understand the sufficient condition on $\chi(\omega)$ that the inverse Fourier transform contour be translatable within the strip $0 < \operatorname{Im}\omega < 1$ [just like (11), but only over this strip]; with (92b), we understand possible translation of the contour of the inverse Fourier transform of $\Phi(\omega)$ through the strip $0 < \operatorname{Im}\omega < m - 1$. The resultant equation for $\Phi(\omega)$

$$\Phi(\omega) = \delta(\omega) - 2if(g^2)\operatorname{csch}\pi(\omega - i\epsilon)\Phi(\omega + im) \tag{93}$$

yields, using our general procedure for solution of difference equations of this type

$$\Phi(\omega) = \frac{i}{2m}\,\Phi_H^{-1}(0)\coth\frac{\pi}{m}(\omega + i\epsilon)\exp\left[\frac{i\omega}{m}\ln 4f(g^2) + P\int_{-\infty}^{+\infty}\frac{dx\,e^{-i\omega x}}{x(1-e^{mx})(1-e^{-x})}\right]; \quad \Phi_H^{-1}(0) = m^{\frac{1}{2}}, \tag{94}$$

which is analytic in the strip $-1 < \operatorname{Im}\omega < m$. Again, $G_P(\omega)$ is used because $\Phi_H(\omega)$ turns out to be symmetric.

The difference equation (93) for D is true only in the sense of an analytic continuation in ω. That is, the actual D function integral equation

$$\Phi(\xi) = 1 + \frac{2f(g^2)}{\pi^2}\int_{-\infty}^{+\infty}\frac{e^{m\eta}\Phi(\eta)\,d\eta}{e^{2(\eta-\xi)}(e^{\xi-\eta}+1)} \tag{95}$$

is not in this case directly Fourier transformable, in that the Fourier transform of the kernel is divergent. The translation

$$\operatorname{csch}\tfrac{1}{2}\pi(\omega - i\epsilon + i) \to -i\operatorname{sech}\tfrac{1}{2}\pi\omega,$$

involved in the derivation of (93) from (92a, *b*), corresponds to evaluating this divergent Fourier transform by analytic continuation; that is, one writes

$$L(\omega) = \int_{-\infty}^{+\infty}\frac{e^{-i\xi\omega}}{e^{\xi}+1}\,d_\xi \tag{96}$$

and calculates $L(\omega + 2i)$ by first evaluating the integral and then continuing. The proof of the validity of the analytic continuation is that the

solutions obtained under its assumption explicitly satisfy the discontinuity relations, as discussed above in the homogeneous case.

We might have expected that if our method yielded a solution of the difference equation (93), the results would satisfy the original integral equations, because (93) automatically guarantees translatability of the inverse Fourier transform contour of $\Phi(\omega)$ at least through the full strip $0 < \mathrm{Im}\,\omega < m$, which is more than the sufficient analyticity demanded in (92b).

Alternately, one can bypass (93) by merely looking for solutions to (92a, b) directly in the form

$$\Phi_H(\omega) = \exp[F_\Phi(\omega)], \qquad \chi_H(\omega) = \exp[F_\chi(\omega)]. \tag{97}$$

This leads directly to (94).

Moreover, in the cases of both the NR and relativistic inhomogeneous equations, the functions $\Phi(\omega)$ are analytic in the upper half-plane. For example, we find in the former that

$$\Phi_H(\omega) \sim (\omega - im), \qquad \text{near}\,\omega = im, \tag{98}$$

which just cancels the pole of $\coth \pi\omega/m$. The difference equation for Φ_H

$$\Phi_H(\omega + im) = \frac{\sinh \pi(\omega - i\epsilon)}{-2\pi i f(g^2)}\,\Phi_H(\omega) \tag{99}$$

guarantees the continued cancellation of the poles of $\coth \pi\omega/m$ higher in the upper half-plane. In the relativistic case, we find

$$\Phi_H(\omega) \sim (\omega - im)^2, \qquad \Phi_H(\omega + im) = \frac{\sinh^2 \pi(\omega - i\epsilon)}{f(g^2)}\,\Phi_H(\omega), \tag{100}$$

which damps the poles of $\coth \pi\omega/m$ even more strongly.

In summary, we have found $\Phi(\omega)$ analytic in the upper half-plane for all four cases, that is NR and relativistic, homogeneous and inhomogeneous. Also in each case, the lower half-plane is analytic except for an infinite set of poles of increasingly higher order along the negative imaginary axis.

High-energy behavior

We can now study the high-energy behavior of the solutions to the inhomogeneous equations. The high-energy ($\nu \to -\infty$) behavior of the relativistic D is dominated by the contribution from saddle points at large $|\omega|$, just as in the homogeneous cases. To find these, we need the large $|\omega|$ behavior of the $\Phi(\omega)$ corresponding to Eq. (87). This is dominated by the third-order pole at $x = 0$:

$$2P\int_{-\infty}^{+\infty} \frac{dx}{x}\,\frac{e^{-ix\omega}}{(1 - e^{-x})(1 - e^{mx})} \sim \mp\frac{\pi l\omega^2}{m} \mp \frac{\pi\omega}{m}(m - 1) \pm \frac{\pi i}{m}\left(\frac{m^2}{6} - \frac{m}{2} + \frac{1}{6}\right) \tag{101}$$

as $\omega \to \pm\infty$. Using this and noting that for large values of ω, $\coth(\pi/m)(\omega + i\epsilon) \sim 1$, we find saddle points in the integrand of (87) at

$$\omega_\pm = \pm\frac{m\hat{\xi}}{2\pi} + i\left(\frac{m-1}{2}\right), \quad \hat{\xi} \equiv \ln\{\nu[f(g^2)]^{1/m}\}. \tag{102}$$

Because the pole of the coth is below the real axis, we can translate the original contour up to $\mathrm{Im}\,\omega = m - \frac{1}{2}$. (The integrand is exponentially damped at the ends of this substrip, so that the convergence factor is not needed.) Using the method of stationary phase along this new contour to pick up the contribution from the two saddle points, we find

$$D(-\nu) \sim m^{\frac{1}{2}}\hat{\xi}^{-\frac{1}{2}(m-1)} \sin[J(\bar{\omega}_+, \hat{\xi}) - \tfrac{1}{4}\pi], \qquad J(\bar{\omega}_+, \hat{\xi}) = \bar{\omega}_+\hat{\xi} - (\pi/m)\bar{\omega}_+^2 - (\pi/12)[m + (1/m)], \quad \bar{\omega}_+ = \mathrm{Re}\,\omega_+, \tag{103}$$

which is very similar to the sort of high-energy behavior found in the homogeneous cases. We see that the inhomogeneity is oscillated to death at high energies.[10] This is what makes it possible now to obtain N by doing a Hilbert transform over αD!

At no time in this paper have we considered the equation for the N function itself—rather we have insisted on calculating it from D. This is not just whimsy; it is a necessity. Any attempt to obtain the equation for N itself, in either the homogeneous or inhomogeneous cases, and especially in the latter, necessitates an (invalid) interchange of integration order, resulting in a divergent structure like $\int^\infty \alpha(s)\,ds$. The remark applies in ω space as well. That this interchange of integration order is not allowed (due to the nonuniformity of convergence of these oscillatory integrands) is implicitly demonstrated in our explicit construction of N.

This damping of the inhomogeneity at high energies also regulates the iteration scheme to correct N_0 and D_0. For example, the "first-iterated" solution of

$$N(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1} \frac{d\nu'}{\nu' - \nu}\,D(\nu')\alpha(\nu'), \qquad D(\nu) = 1 - \frac{1}{\pi}\int_0^\infty \frac{d\nu'}{\nu' - \nu}\,N(\nu')\rho(\nu') \tag{104}$$

[10] Presumably that, in much the same way, replacing $g^2 \to g^2 - g^6\nu^2/2 + \ldots = g^2\cos g^2\nu$ suppresses the high-energy behavior of the first term.

is, in our sense,

$$N_1(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1}\frac{d\nu'}{\nu'-\nu}D_0(\nu')\alpha(\nu'),$$
$$D_1(\nu) = 1 - \frac{1}{\pi}\int_0^{\infty}\frac{d\nu'}{\nu'-\nu}N_1(\nu')\rho(\nu'). \tag{105}$$

Notice we are iterating around D_0, not 1! It appears at first glance that the next iteration for N will not be finite, because of the inhomogeneity in D_1

$$N_2(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1}\frac{d\nu'}{\nu'-\nu}\alpha(\nu') - \frac{1}{\pi^2}\int_{-\infty}^{-1}\frac{d\nu'}{\nu'-\nu}\alpha(\nu') \times\int_0^{\infty}\frac{d\nu''}{\nu''-\nu'}N_1(\nu'')\rho(\nu''). \tag{106}$$

We emphasize that actually the inhomogeneity is no more dangerous here than in the N_0D_0 equations, and that the sum of the two integrals in (106) is quite finite. To show this, we need only rewrite (105) in a form that absorbs the inhomogeneity into D_0:

$$N_1(\nu) = N_0(\nu) + \Delta N_0(\nu), \qquad D_1(\nu) = D_0(\nu) - \frac{1}{\pi}\int_0^{\infty}\frac{\Delta N_0(\nu')\,d\nu'\,\rho(\nu')}{\nu'-\nu},$$
$$\Delta N_0(\nu) = \frac{1}{\pi}\int_{-\infty}^{-1}d\nu'\,\frac{D_0(\nu')\,\Delta\alpha(\nu')}{\nu'-\nu} - \frac{1}{\pi}\int_{-1}^{0}\frac{D_0(\nu')\alpha_0(\nu')}{\nu'-\nu}\,d\nu'. \tag{107}$$

We see that the higher iterations, e.g.,

$$N_2(\nu) = N_0(\nu) - \frac{1}{\pi^2}\int_{-\infty}^{-1}\frac{d\nu'}{\nu'-\nu}\int_0^{\infty}\frac{d\nu''\,\Delta N_0(\nu'')}{\nu''-\nu'}\rho(\nu'') \tag{108}$$

are all finite by virtue of the oscillations in D_0.

It cannot be overemphasized that the unitarity is here generating its own regulation, and one is tempted to conjecture that the true theory may have a self-regulating mechanism of this sort. It is strongly suggested here that the requirement of unitarity can, and will, always obviate the need for regulator procedures.

Opposite signs of the coupling

In the previous section we studied particular signs of the coupling, i.e. (R = relativistic),

$$\text{R:}\quad \alpha_0(-\nu) = +f(g^2)\nu^m; \qquad \text{NR:}\quad \alpha_0(-s) = -f(g^2)s^{\frac{1}{2}(m-1)}, \tag{109a}$$

where $f(g^2) > 0$. In this section we want to solve the equations with the signs reversed, i.e.,

$$\text{R:}\quad \alpha_0(-\nu) = -f(g^2)\nu^m; \qquad \text{NR:}\quad \alpha_0(-s) = +f(g^2)s^{\frac{1}{2}(m-1)} \tag{109b}$$

and $f(g^2) > 0$. We see that the essential difference is only to break the symmetry $[\Phi_H(\omega) = \Phi_H^*(-\omega)]$, found for the inputs (109a), thus requiring multiplication by a corresponding antisymmetric $G(\omega)$.

In the relativistic case, the input (109b) yields the difference equation for the D function

$$\Phi(\omega) = \delta(\omega) + f(g^2)\,\operatorname{csch}^2\pi(\omega + i\epsilon)\Phi(\omega + im), \tag{110}$$

which differs from (83), the case with the opposite sign of the input, only in the sign of the second term on the right. The general procedure goes through as above, and one finds

$$\Phi_H^{(-)}(\omega) = \Phi_H^{(+)}(\omega)e^{\pm\pi\omega/m}, \tag{111}$$

where the $(-)$ superscript corresponds to the negatively weighted input (109b) and the $(+)$ to the positively weighted input (109a). $\Phi^{(+)}(\omega)$ appears explicitly in (87). A quick way of getting (111) is to guess a $\Phi^{(-)}(\omega)$ of the form $\Phi_H^{(-)}(\omega) = \Phi_H^{(+)}L(\omega)$. Substituting into (110) and using (83) gives

$$L(\omega) = -L(\omega + im) \Rightarrow L(\omega) = e^{\pm\pi\omega/m}. \tag{112}$$

To obtain a symmetric $\Phi(\omega)$, we need multiply $\Phi_H^-(\omega)$ by $G_\mp(\omega)$. Finally, in this case then, one finds

$$\Phi(\omega) = \tfrac{1}{2}i\,\operatorname{csch}\frac{\pi}{m}(\omega + i\epsilon)e^{(i\omega/m)\ln 4f(g^2)} \times\exp\left[2P\int_{-\infty}^{+\infty}\frac{dx}{x}\frac{e^{-ix\omega}}{(1-e^{-x})(1-e^{mx})}\right]. \tag{113}$$

The situation is the same in the NR case, corresponding to the input (109b). We find

$$\Phi_H(\omega) = \exp\left[\frac{i\omega}{m}\ln 4f(g^2) + P\int_{-\infty}^{+\infty}\frac{dx}{x}\frac{e^{-i\omega x}}{(1-e^{mx})(1-e^{-x})} \mp \frac{\pi\omega}{m}\right]. \tag{114}$$

So that to obtain a symmetric $\Phi(\omega)$, we need multiply by $G_\pm(\omega)$.

We mention that this method of patching together a symmetric $\Phi(\omega)$ is different from that used in the homogeneous cases above, but that method works here as well. For example, the two forms of $\Phi_H(\omega)$ given in (111) could be added to form first a symmetric $\Phi_H(\omega)$, then $G_P(\omega)$ could be used; this would result in a solution identical with (113).

except that

$$\operatorname{csch}\frac{\pi}{m}(\omega+i\epsilon)\to\cosh^2\frac{\pi\omega}{m}\operatorname{csch}\frac{\pi}{m}(\omega+i\epsilon). \quad (115)$$

Clearly, the solutions we are finding are not unique (for either sign of the input). We again postpone the discussion of the multiplicity of the solutions. It should be mentioned, however, that we are taking the trouble to exhibit these particular solutions explicitly [i.e., Eqs. (87), (114)] because, as shown below, these are, in fact, the iterative solutions of the equations (which we feel are the most physical.)

The amplitude far to the left: the ghosts

We have noted in all our solutions that $D(\nu)$ oscillates for $\nu\to-\infty$. How do these zeros affect the amplitude? Since we expect the solutions to maintain $\alpha(\nu)$ on the left, we are first tempted to expect that at these zeros the N function vanishes as well. We see that this is not so and that even while maintaining Im $N=\alpha D$, the solutions involve an infinite number of ghost poles near $\nu=-\infty$.

We discuss explicitly the relativistic homogeneous case, but essentially identical arguments can be given in each of the other cases. From (45), we see that the N function is dominated asymptotically on the left by

$$N(re^{i\pi})\sim\frac{i}{2r^{\frac{1}{2}}}\int_{-\infty}^{+\infty}e^{i\omega\ln r}e^{-2\pi\omega}\Phi(\omega)\,d\omega. \quad (116)$$

Using (101), we find saddle points in the integrand at ω_+ and ω_-, as shown in Fig. 6. Even though $e^{i\omega\ln r}$ is increasing in the lower half-plane, the behavior of $e^{+i\omega^2}e^{-\epsilon\omega^2}$ allows us to translate to Im $\omega=-\frac{1}{2}im$ and pick up the contribution at ω_-, which dominates anything else from Im $\omega>-\frac{1}{2}m$ (such as the contribution from the poles along the negative imaginary axis). We find

$$N(re^{i\pi})\sim\frac{i[f(g^2)]^{\frac{1}{2}}}{2r^{\frac{1}{2}}}e^{\frac{1}{2}m\xi}e^{-iZ} \quad (117)$$

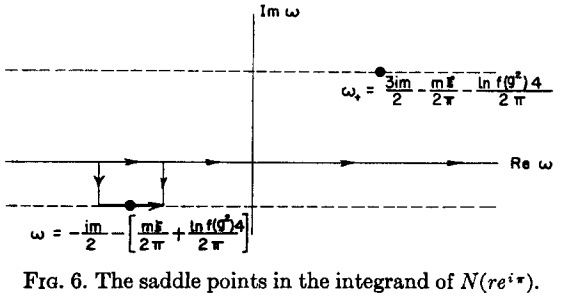

FIG. 6. The saddle points in the integrand of $N(re^{i\pi})$.

[which is exponentially increasing for $\xi>0$(!)], and thus, using (66),

$$A(-\nu)\sim i\nu^m f(g^2)e^{-iZ}\sec Z,\qquad \nu>0. \quad (118)$$

The N function fails to cancel the zeros of D, which persist as ghosts in the amplitude. (They correspond to particles with large imaginary masses.) The asymptotic form (118) explicitly maintains Im $N=\alpha D$, but because of the ghosts, $\mathrm{Im}_L\,A(\nu)=\alpha(\nu)$ is not maintained. In particular, the imaginary part of (118) has, in addition to the input, an infinite series of δ functions. Thus, just as in the $g^2>1$ and $g^2<0$ cases of the one W exchange N/D equations discussed in Ref. 1, no solutions exist to the partial wave dispersion relation itself.

In these more singular cases, we find the ghosts, and hence the inconsistency of the partial wave dispersion relation, regardless of the sign of the input: the solutions for $f(g^2)\to-f(g^2)$ are obtained by the prescription (70b) in (87); we obtain the same behavior as $\nu\to+\infty$

$$\frac{N(-\nu)}{D(-\nu)}\sim\frac{i\alpha(-\nu)\nu^2e^{iZ}}{\nu^{-(m+2)}\cos Z}. \quad (119)$$

Similarly, the ghosts are present in every one of our homogeneous, inhomogeneous, relativistic, and nonrelativistic solutions. As pointed out earlier in the chapter, these ghosts are expected on general grounds. The nature of the argument given there (depending only on the asymptotic behavior of $\alpha(\nu)$ and unitarity) indicates, in fact, that the ghosts are expected to occur independently of the number of subtractions.

Another way of saying all this is the following: there exists no Hilbert-transformable solution to the problem of analytically continuing to the right, in a unitary fashion, an imaginary part which goes asymptotically on the left like $\alpha(-\nu)\sim\lambda\nu^m$, $m>0$. (Our solutions, because of the ghosts, simply do not satisfy the partial wave dispersion relation for $A=N/D$ with the original input.) If a solution to the problem of continuing these inputs to the right exists at all, it must then be non-Hilbert-transformable; that is, it must exhibit exponential increase at ∞. We remind the reader that very similar conclusions were reached in the $m=0$ (vector-meson exchange) case discussed in Ref. 1. We emphasize that the ghosts, and hence the exponential increase of the "true" solution, are expected on general grounds. We do not attempt to find this nondispersive "true" solution.

A simple calculation reveals that the residues at these new poles are *opposite* in sign to the sign

of the input. Thus, we see that unitarity generates the ghosts in an attempt to subtract away the worst part of the asymptotically ill-behaved input. (The infinite series of δ functions gives a kind of oscillatory nature to the left-hand cut.) This is generally the reason for the appearance of either bound states or ghosts in solutions to the N/D equations.[11] As discussed in the section on the physical meaning of the convergence factor, these ghosts were actually introduced in the "modified" input $\bar{\alpha}(\nu)$, and the D function had to develop these zeros just to cancel those of the input.

If the calculations of Bjorken and Goldberg[12] on the exponential potential have any relevance here, the appearance of the δ function modifications of the original input might be taken as an indication of an oscillatory left-hand cut in the true theory: theirs and similar calculations, e.g., the calculations in which "dynamical" bound states arise in the solutions (which also "modify" the original discontinuity relations), have hinted toward a sixth sense for the N/D equations, in that when an N/D calculation yields a modified input in one way or another, one should tend to take the modified input (original plus δ functions) more seriously than the original itself. We feel that this is an attractive viewpoint.

We emphasize that the ghosts are very far out to the left indeed: we found the oscillations by making the assumption that

$$2\pi\bar{\omega}_+ = \omega \ln \nu + \ln f(g^2) \gg 0.$$

This is so, provided $|\nu| \gg |f(g^2)|^{-1/m}$. In the weak interactions, we know from Ref. 1 that $\alpha(-\nu) \sim (g^2)^{m+1}\nu^m$, so, remembering ν is scaled in W masses, we are finding the zeros in the region

$$|\nu| \gg (1/g^2)(W \text{ masses})^2. \tag{120}$$

The ghosts are extremely far from the physical region and stay there for higher m(!), thus giving our scheme some chance for accuracy in the physical region.

The ϵ-convergence factors and definition by analytic continuation

Looking back, we find that the convergence factors have been used essentially in two places only,

(a) to prove formally that our D functions solved their integral equations, and

(b) to define N on the left—that is, N on the right and D, as given in (42), are quite well defined without convergence factors.

We find it extremely interesting to note that, at least in the case of (b) above, the convergence factors can, and in fact should, be interpreted in terms of an exactly equivalent analytic continuation.

Consider for simplicity the (well-behaved) D function in the relativistic homogeneous case ($\nu > 0$),

$$D(-\nu) = \frac{1}{\nu^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{i\omega \ln \nu} \Phi(\omega)\, d\omega,$$

$$|\Phi(\omega)| \sim e^{-|\omega|\pi}. \tag{121}$$

According to (45), we can obtain the N function on the left from the knowledge of $D(\nu)$ and $D(\nu e^{2\pi i})$. Of course, when $D(\nu e^{2\pi i})$ is written in the Fourier transform representation,

$$D(-\nu e^{2\pi i}) = \frac{1}{\nu^{\frac{1}{2}}} \int_{-\infty}^{+\infty} e^{i\omega \ln \nu} e^{-2\pi\omega} \Phi(\omega)\, d\omega \tag{122}$$

its integrand has an exponentially increasing modulus ($\omega \to -\infty$). We know that we can define (122), even with this pathology, with a convergence factor. Actually, however, what we should do to find $D(-\nu e^{2\pi i})$ is to evaluate the integral representation (121) and then continue the evaluation to $\nu e^{2\pi i}$. We can do essentially this by rotating each half of the integral in (121) down by $\pi/4$ to the new contour C', shown in Fig. 7. We obtain

$$D(-\nu) = \frac{1}{\nu^{\frac{1}{2}}} \Big\{ \exp(-\tfrac{1}{4}i\pi) \int_0^{\infty} \exp(iRe^{-\frac{1}{4}i\pi} \ln \nu) \times \Phi[R \exp(-\tfrac{1}{4}i\pi)]\, dR + \exp(+\tfrac{1}{4}i\pi) \times \int_0^{\infty} \exp(-iRe^{+\frac{1}{4}i\pi} \ln \nu) \Phi[-R \exp(+\tfrac{1}{4}i\pi)]\, dR \Big\}. \tag{123}$$

To do this, we need notice that $\Phi(\omega)$, although not completely analytic in the lower half-plane, has singularities only on the imaginary axis. The fact that $\Phi(\omega) \sim e^{-i\omega^2\epsilon(\omega)}$ allows us to drop the contributions at infinity. Along C', we can be sure that $\Phi(\omega)$ is damped very powerfully by the usual e^{-R^2} coming from the $e^{-i\epsilon(\omega)\omega^2}$. We emphasize that this rotation is valid without any convergence factor because D is convergent as it stands. Now

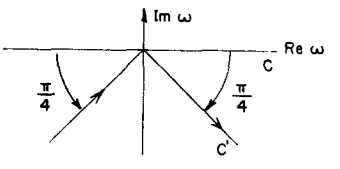

FIG. 7. Alternate inverse fourier transform contour.

[11] S. C. Frautschi, *Regge Poles and S-Matrix Theory* (W. A. Benjamin Company, Inc., New York, 1963).
[12] J. D. Bjorken and A. Goldberg, Nuovo Cimento 16, 539 (1960).

we can continue (123) to $\nu e^{2\pi i}$, obtaining quite rigorously,

$$D[-\nu \exp(2\pi i)] = \frac{1}{\nu^{\frac{1}{2}}}\left\{\exp(-\tfrac{1}{4}i\pi)\int_0^\infty \exp(iRe^{-\frac{1}{4}i\pi}\ln\nu) \times \exp(-Re^{-\frac{1}{4}i\pi}2\pi)\Phi[R\exp(-\tfrac{1}{4}i\pi)]\,dR + \exp(+\tfrac{1}{4}i\pi)\int_0^\infty \exp(-iRe^{+\frac{1}{4}i\pi}\ln\nu) \times \exp(+Re^{\frac{1}{4}i\pi}2\pi)\Phi[-R\exp(\tfrac{1}{4}i\pi)]\,dR\right\}, \quad (124)$$

which is quite convergent, the exp $(-R^2)$ of Φ killing the exponential increase of the exp $(+Re^{\frac{1}{4}i\pi}2\pi)$ factor.

We wish to make two important points here. First, notice that, in the form (123), $D(-\nu)$ has no singularity at $\nu = e^{i\pi}$, although it seems that the original form (121) of D might cease to converge at just that point, since that is where the modulus of the integrand just begins to grow exponentially. This is a very strange analytic continuation then, because, although doing an integral usually analytically continues it beyond its original region of convergence, there is always a pole or some other singular reminder that the integral representation ceased to converge at that point. Here we find no such singularity! This suggests that the integral representation (121) was actually well defined above $\nu = e^{i\pi}$. We have not tried to prove this rigorously, although probably it could be done in the sense of distribution theory. Such remains to be investigated. Second, we note that the form (124) for $D(-\nu e^{2\pi i})$ obtained by "doing" the integral (121) and then analytically continuing it, is exactly what one would get with an ϵ-convergence factor used directly on (122). Consider

$$D(-\nu e^{2\pi i}) = \frac{1}{\nu^{\frac{1}{2}}}\int_{-\infty}^{+\infty} e^{i\omega\ln\nu}e^{-2\pi\omega}\Phi(\omega)e^{-\epsilon\omega^2}\,d\omega. \quad (125)$$

We can "fold" the contour to C' in this integral as well; this time the convergence factor damps out the contributions at infinity. The result is exactly (124).

Thus, we have interpreted the convergence factor in terms of an analytic continuation, at least in the case of (b) above. It is clear that the two methods of defining such pathological structures as (122) are completely equivalent. On the other hand, we have not yet been able to use an analytic continuation to prove (a). At any rate, since both methods give the same (correct) results, we can only conclude that both are implied naturally in the equations.

In summary, we comment that the unitarity requirement can be thought of then as having generated its own regulation—by spawning the very rapid $e^{i\omega^2}$ oscillations. Two particular forms of this regulation are the convergence factors and the analytic continuation method just discussed. Thus, no F–P regulator limit is really needed. It cannot be overemphasized that the solutions we obtain indeed solve the original equations, whereas the solutions of F–P do not, in general, solve theirs (after the regulator limit). This is a strong indication that the requirement of unitarity can, in general, always generate its own regulation in so natural a manner as found here; we feel that unitarity may be the key to future regulator-free calculations in the weak interactions.

Uniqueness of solutions

As anticipated above, our solutions to the N/D equations are not unique. If $\Phi(\omega)$ is a solution of the homogeneous difference equation whose inverse Fourier transform solves the integral equation in ξ space, then the inverse transform of $\Phi(\omega)\mathfrak{a}(\omega)$ does also, provided $\mathfrak{a}(\omega)$ has the following properties,

$$\begin{aligned}&\text{(a)}\ \ \mathfrak{a}(\omega) = \mathfrak{a}(\omega + im),\\ &\text{(b)}\ \ \mathfrak{a}(\omega)\Phi(\omega)\ \text{must satisfy (11)},\\ &\text{(c)}\ \ \mathfrak{a}(\omega) = \mathfrak{a}^*(-\omega^*).\end{aligned} \quad (126)$$

The first proviso implies that $\mathfrak{a}(\omega)$ has no singularities on the strip boundary; as we saw above, functions singular on the boundary satisfy a $\delta(\omega)$ equation. Proviso (c) guarantees the reality of D on the left.

How much ambiguity does this leave? Functions like $e^{2\pi n\omega/m}$ are excluded by (c). Superdamping factors like $e^{-\cosh 2\pi\omega/m}$ are excluded by (b); this function has a terrible singularity like $e^{\mathrm{Re}\,\omega\,\mathrm{Re}^{\,\omega}}$ at Im $\omega = \frac{1}{2}m$. Requirements (a) and (b) together, eliminate elliptic functions. Requirement (b) eliminates functions like sech $2\pi\omega/m$, which have singularities in the period strip. Proviso (a) eliminates functions singular on the strip boundary, e.g., i coth $\pi\omega/m$, because these functions really satisfy δ-function inhomogeneous difference equations like (78). We find then that the ambiguities can be taken in the form

$$\mathfrak{a}_m(\omega) = \left\{\cosh\frac{2n\pi\omega}{m},\ \cosh^2\frac{n\pi\omega}{m},\ i\sinh\frac{2\pi n\omega}{m},\ \cdots\right\} \quad (127)$$

and sums and products of these. Despite the exponential growth of the $\Phi(\omega)$'s (at large $|\omega|$) associated with these factors, the exp $(i\omega^2)$ oscillations will define the solutions, as discussed above.

In the case of the inhomogeneous equations, there are ostensibly two sources of ambiguity. (a) The first is that, to our direct exponentiation-and-Fourier-transform solution $\Phi^{EFT}(\omega)$, one can add any solution of the homogeneous equation. The discussion of the ambiguities then follows the previous case. (b) A second source seems to be that one can multiply the solution to the homogeneous equation obtained directly by our exponentiation-and-Fourier-transform technique $[\Phi_H^{EFT}(\omega)]$ by the functions $\mathfrak{a}_m(\omega)$ before applying the Green's function. It is shown in Appendix IV that the functions in the set $\mathfrak{a}_m(\omega)$ which vanish at $\omega = 0$, e.g., $i \sinh 2\pi n\omega/m$ can effectively be omitted from the set in this case, in that no new solutions, above and beyond those obtainable with the cosh-type factors, can be obtained with their use. In our proof below that $\Phi^{EFT}(\omega)$ is the solution in coordinate space which corresponds to the iterative solution in energy space, we are thus able to consider the set $\mathfrak{a}_m(\omega)$ as being only those functions with the properties (126), which satisfy the constraint $\mathfrak{a}_m(0) = 1$. (This particular normalization is without loss of generality, as the application of the Green's function immediately divides out whatever normalization we choose.) Using the $\mathfrak{a}(\omega)$ in this sense then, we find finally that, if the inverse Fourier transform of $G(\omega)\Phi_H^{EFT}(\omega)$ satisfies the original integral equation, then so also does that of

$$\Phi(\omega) = \mathfrak{a}_m(\omega)G(\omega)\Phi_H^{EFT}(\omega). \tag{128}$$

This is because the δ-function inhomogeneity of the Green's function equation only "senses" the value of $\Phi_H(\omega)$ at $\omega = 0$.

Actually these two ostensibly different forms of ambiguity are entirely equivalent. We can show this most clearly by simply constructing corresponding $\Phi_H(\omega)$'s and $\mathfrak{a}_m(\omega)$'s in terms of one another. For example, given a solution $\mathfrak{a}_m(\omega)\Phi^{EFT}(\omega)$, we can construct it as the sum of $\Phi^{EFT}(\omega)$ and a solution of the homogeneous equation by taking this latter to be

$$\Phi_H(\omega) = [\mathfrak{a}_m(\omega) - 1]\Phi^{EFT}(\omega). \tag{129}$$

Using the inhomogeneous difference equation [e.g. (83)] and the facts that $\mathfrak{a}(\omega + im) = \mathfrak{a}(\omega)$, $\mathfrak{a}(0) = 1$, it is easy to show that the right-hand side of (129) indeed satisfies the corresponding homogeneous difference equation.

On the other hand, given $\Phi_H(\omega)$, one can calculate the corresponding $\mathfrak{a}(\omega)$ through the formula

$$\mathfrak{a}(\omega) = 1 + [\Phi_H(\omega)/\Phi^{EFT}(\omega)]. \tag{130}$$

We can easily show that the right-hand side of (130) has all the expected properties of $\mathfrak{a}_m(\omega)$: using (in the relativistic case)

$$\Phi_H(\omega + im) = -[f(g^2)]^{-1} \sinh^2 \pi\omega\Phi_H(\omega), \tag{131a}$$

$$\Phi^{EFT}(\omega + im) = -[f(g^2)]^{-1} \sinh^2 \pi\omega\Phi^{EFT}(\omega), \tag{131b}$$

it follows that $\mathfrak{a}(\omega + im) = \mathfrak{a}(\omega)$. [We have set $\sinh^2\pi\omega$ $\delta(\omega)$ to zero in (131b).] It should be noted that (131) follows only because the kernel $\operatorname{csch}^2 \pi(\omega + i\epsilon)$ is free of zeros in the period strip. To prove that the value of the right-hand side of (130) is unity at $\omega = 0$, one need only notice that $\Phi^{EFT}(\omega) \sim \omega^{-1}$ near $\omega = 0$, while all $\Phi_H(\omega)$ go to a constant or zero at $\omega = 0$. This latter follows from noting that $\Phi_H^{EFT}(\omega)$ goes to a constant at $\omega = 0$, and there are no members of $\mathfrak{a}(\omega)$, as given in (126), which are singular at the origin. Finally, we see that $\mathfrak{a}(\omega)$ is analytic in the period strip. This follows from the analyticity of $\Phi_H(\omega)$ and $\Phi^{EFT}(\omega)$ in the period strip, and the fact that $\Phi^{EFT}(\omega)$ is free of zeros in the period strip [which is guaranteed by the explicit exponential form of $\Phi^{EFT}(\omega)$].

Thus, we have explicitly established the equivalence of the two types of ambiguities.

The asymptotic behavior of these solutions can be investigated as above with the method of stationary phase. They all have the same wild oscillation for large $|\nu|$, and one finds that essentially the only change is that the power of ν multiplying the oscillations in N and D increases as the corresponding $\mathfrak{a}(\omega)$ become more strongly exponential at $\omega = \infty$. These powers cancel in the ratio, however, and the asymptotic amplitude is, interesting enough, the same for all $\mathfrak{a}(\omega)$. In particular, all the solutions exhibit the ghosts discussed above.

C. THE SOLUTION THAT SUMS THE ITERATIVE EXPANSION

In the relativistic case we have presented two ways (homogeneous and inhomogeneous N/D) of constructing a unitary amplitude from these asymptotically ill-behaved inputs. We found an infinite number of solutions in each of these cases. Out of this multitude of solutions, which one should we pick as being most physical?

We make our choice on the basis of our desire to use the dispersion methods to sum as much of the ordinary perturbation series as possible. According

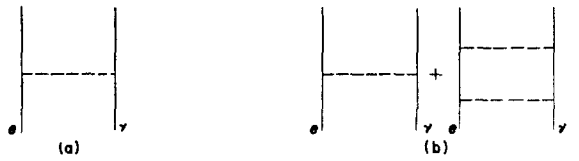

FIG. 8. Simple inputs for the N/D equations.

to this criterion, it was pointed out in Ref. 1 that the N/D equations with input like Fig. 8(a) or Fig. 8(b) should be written (in all partial waves) with one subtraction at threshold in N and D, with D normalized to unity at threshold, and the subtraction constant in N set to zero. In this way, because the Born term ($1W$ exchange) can also be written as a Hilbert transform once subtracted at threshold, with zero subtraction constant, we formally include the Born term itself in the iterative solution of the equations. [The first iterated N function will always contain the Born term. More particularly, with Fig. 8(a) as input, the first iterated N function is the Born term; for inputs like Fig. 8(b), the first iterated N function contains the Born term formally.] The remaining question is, from among the infinity of solutions to this particular form of the N/D equations, how do we find the one that sums the iterative expansion?

We do this by regulating the N/D equations in a simple way, so that we can discuss a perturbation expansion with finite terms. Although we still find an infinite number of solutions even in the regulated case, there will be only one of these, the direct exponentiation- and Fourier-transform solution, whose perturbative expansion coincides with the (regulated) iterative solution. The (smooth) infinite regulator limit of this solution is our above-exhibited solution $\Phi^{EFT}(\omega)$. The other solutions of the regulated problem go over smoothly, in the regulator limit, to our "other" solutions. The crucial property of $\Phi^{EFT}(\omega)$, which singles it out in this way from all the other solutions, is that $\Phi_H^{EFT}(\omega)$ is the only solution to the homogeneous equation which is free of zeros in the period strip. Out of all the four families of solutions exhibited in this paper then, we feel that the solution $\Phi^{EFT}(\omega)$, because it is the smooth regulator limit of the regulated iterative solution (in energy space), is the most meaningful solution.

Our method of regulation is to consider Eq. (73) with the inhomogeneity cut off above $\lambda > 0$:

$$\Phi(\xi, \lambda) = \vartheta(\lambda - \xi) + \frac{f(g^2)}{\pi^2} \int_{-\infty}^{+\infty} \frac{d\eta}{e^{\eta-\xi} - 1} (\eta - \xi) e^{m\eta} \Phi(\eta, \lambda). \quad (132a)$$

In coordinate space this goes over to

$$\Phi(\omega, \lambda) = -\frac{e^{-i\lambda\omega}}{2\pi i(\omega + i\epsilon)} - f(g^2) \times \operatorname{csch}^2 \pi(\omega - i\epsilon) \Phi(\omega + im, \lambda). \quad (132b)$$

Before going any further, observe that the term-by-term Fourier transform of the iterative solution of (132a) is not, in general, the iterative solution of (132b):

$$\Phi(\xi, \lambda) = \vartheta(\lambda - \xi) + \frac{f(g^2)}{\pi^2} \int_{-\infty}^{\lambda} \frac{d\eta}{e^{\eta-\xi} - 1} (\eta - \xi) e^{m\eta} + \left[\frac{f(g^2)}{\pi^2}\right]^2 \int_{-\infty}^{+\infty} \frac{d\eta\, e^{m\eta}}{e^{\eta-\xi} - 1} (\eta - \xi) \times \int_{-\infty}^{\lambda} \frac{d\bar{\eta}}{e^{\bar{\eta}-\eta} - 1} (\bar{\eta} - \eta) e^{m\bar{\eta}} + \cdots, \quad (133a)$$

$$\Phi(\omega, \lambda) = \frac{e^{-i\lambda\omega}}{2\pi i(\omega + i\epsilon)} + f(g^2) \times \operatorname{csch}^2 \pi(\omega - i\epsilon) \frac{e^{i\lambda\omega} e^{\lambda m}}{2\pi i(\omega + im)} + [-f(g^2)]^2 \times \operatorname{csch}^4 \pi(\omega - i\epsilon) \frac{e^{i\lambda\omega} e^{2\lambda m}}{2\pi i(\omega + 2im)} + \cdots. \quad (133b)$$

We see that the Fourier transform of the first two terms of (133a) yield the first two terms of (133b). However, for $\lambda < \infty$

$$\int_{-\infty}^{\lambda} \frac{d\eta}{e^{\eta-\xi} - 1} (\eta - \xi) e^{m\eta} \to \xi/m \quad \text{as} \quad \xi \to +\infty, \quad (134)$$

so that the third term of (133a) is divergent and has a divergent Fourier transform, while the $[f(g^2)]^2$ contribution to (133b) is finite and has a finite inverse Fourier transform. The fact that the iterative solutions in the two spaces are not connected does not bother us at all; we are interested really only in the iterative solution in energy space, and have no interest in the iterative solution in coordinate space: each of the terms of the latter has a pole in the period strip (from the kernel) and one can show that so also does their sum. Thus, the inverse Fourier transform of the exact iterative solution in coordinate space does not solve the original integral equation. It is clear then that none of the solutions in ω space considered here correspond to the iterative solution of the difference equation. We also see from (133a) that our regulation is not a very good one because all the terms in $[f(g^2)]^n$ ($n \geq 2$) are still divergent. Still, this much regulation proves sufficient to identify $\Phi^{EFT}(\omega, \lambda)$ as the sum of the (regulated) iterative expansion: it will turn out that,

of all the solutions to (132b) (whose inverse Fourier transforms satisfy the original integral equation), there is a unique solution whose inverse Fourier transform agrees to order $f(g^2)$ with (133a). [This will be $\Phi^{EFT}(\omega, \lambda)$.] Thus, the higher-order terms will not be needed to pick out the desired solution.

To see this, we find, in essentially the usual way, all the solutions of (132b). This time we make a slightly different factorization, leaving the $f(g^2)$ dependence in the Green's function:

$$\Phi(\omega, \lambda) = G(\omega, \lambda)\Phi_H(\omega) \tag{135a}$$
$$\Phi_H(\omega) = -\operatorname{csch}^2 \pi(\omega - i\epsilon)\Phi_H(\omega + im),$$

$$f(g^2)G(\omega + im) - G(\omega) = \frac{e^{-i\lambda\omega}}{2\pi i(\omega + i\epsilon)} \frac{1}{\Phi_H(\omega)}. \tag{135b}$$

Both of these are soluble in the usual way. We exhibit the set of solutions in the following form:

$$\Phi(\omega, \lambda) = -\mathfrak{A}(\omega)\Phi_H^{EFT}(\omega) \int_{-\infty}^{+\infty} \frac{dx}{2\pi} \frac{e^{-i\omega x}}{1 - f(g^2)e^{mx}} \times \int_{-\infty}^{+\infty} \frac{d\bar{\omega}}{2\pi i} \frac{e^{-i\lambda\bar{\omega}}}{\bar{\omega} + i\epsilon} \frac{e^{ix\bar{\omega}}}{\mathfrak{A}(\bar{\omega})\Phi_H^{EFT}(\bar{\omega})}, \tag{136}$$

where $\mathfrak{A}(\omega)$ has the properties (126) and, as discussed above, we take $\mathfrak{A}(0) = 1$; $\Phi_H^{EFT}(\omega)$ is the direct exponentiation- and Fourier-transform solution to (135a)—i.e., Eq. (85) without the $f(g^2)$ term.

We can formally expand the set of solutions (136) in a power series in $f(g^2)$

$$\Phi(\omega, \lambda) = \sum_{n=0}^{\infty} [f(g^2)]^n \Phi_n(\omega, \lambda), \tag{137}$$
$$\Phi_n(\omega, \lambda) = -\mathfrak{A}(\omega)\Phi_H^{EFT}(\omega) \int_{-\infty}^{+\infty} \frac{dx}{2\pi} e^{-i\omega x} e^{nmx} \times \int_{-\infty}^{+\infty} \frac{d\bar{\omega}}{2\pi i} \frac{e^{-i\lambda\bar{\omega}}}{\bar{\omega} + i\epsilon} \frac{e^{ix\bar{\omega}}}{\mathfrak{A}(\bar{\omega})\Phi^{EFT}(\bar{\omega})}.$$

We see immediately that

$$\Phi_0(\omega, \lambda) = -e^{i\lambda\omega}/2\pi i(\omega + i\epsilon) \tag{138}$$

independent of $\mathfrak{A}(\omega)$. All the solutions have the correct term independent of $f(g^2)$, so that one cannot distinguish between them on this basis. Now ask about $\Phi_1(\omega, \lambda)$. For which $\mathfrak{A}(\omega)$ do the integrals even converge? We can find this out by looking at the analyticity in $\bar{\omega}$. First take $\mathfrak{A}(\omega) = 1$ (the *EFT* solution). We know from our previous discussion that $\Phi_H^{EFT}(\omega)$ is analytic in the period strip with its first zeros at the strip boundaries (on the imaginary axis). This is guaranteed by its explicit exponential structure. In fact, we know that $\Phi_H^{EFT}(\omega)$ is analytic in the upper half-plane (with a family of zeros along the positive imaginary axis), so that $[\Phi_H^{EFT}(\omega)]^{-1}$ is analytic in the upper half-plane except for a family of poles along the positive imaginary axis at $\omega_n = inm$, $n \geq 1$. Further, we know that for large $|\omega|$

$$[\Phi_H^{EFT}(\omega)]^{-1} \sim e^{\epsilon(\omega)(\pi/m)[i\omega^2 + \omega(m-1)]}. \tag{139}$$

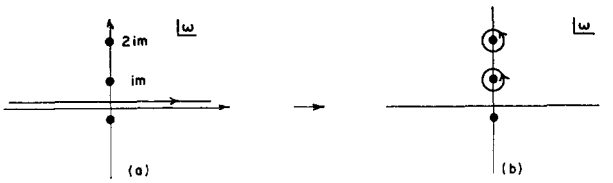

FIG. 9. Method of evaluating Eq. (140).

Thus, we can close the contour above in the $\bar{\omega}$ plane, picking up only the poles at $\bar{\omega}_n$ as shown in Fig. 9.

The lowest-lying pole at $\omega_1 = im$ gives the asymptotic behavior of the $\bar{\omega}$ integral for large x. That is,

$$\int_{-\infty}^{+\infty} \frac{d\bar{\omega}}{2\pi i} \frac{e^{-i\lambda\bar{\omega}}}{\bar{\omega} + i\epsilon} \frac{e^{ix\bar{\omega}}}{\Phi_H^{EFT}(\omega)} \sim e^{-xm}, \qquad x \to \infty. \tag{140}$$

Thus, in the case $\mathfrak{A}(\omega) = 1$ (the *EFT* case), $\Phi_1(\omega)$ converges but $\Phi_n (n > 1)$ diverge. In the other cases, i.e., $\mathfrak{A}(\omega) \neq 1$, we find a pole in the period strip of the $\bar{\omega}$ plane. For example, in the case $\mathfrak{A}(\omega) = \cosh 2n\pi\omega/m$, there is a pole at $\bar{\omega} = im/4n$. This allows the $\bar{\omega}$ integral to fall off no faster than $\exp(-xm/4n)$ at large positive x. In all these cases then Φ_1 diverges. Only the *EFT* solution converges to order $f(g^2)$. It is straightforward to show in fact that[13]

$$-\Phi_H^{EFT}(\omega) \int_{-\infty}^{+\infty} \frac{dx}{2\pi} e^{-i\omega x} e^{xm} \times \int_{-\infty}^{+\infty} \frac{d\bar{\omega}}{2\pi i} \frac{e^{-i\lambda\bar{\omega}}}{\bar{\omega} + i\epsilon} \frac{e^{ix\bar{\omega}}}{\Phi_H^{EFT}(\bar{\omega})} = -\operatorname{csch}^2 \pi(\omega - i\epsilon) \frac{e^{\lambda m} e^{i\lambda\omega}}{2\pi i(\omega + im)}, \tag{141}$$

whose inverse Fourier transform is just the second term of (133a). This establishes that the solution $\Phi^{EFT}(\omega, \lambda)$ is the sum of the regulated iterative expansion. This *EFT* solution clearly goes over

[13] To do this, study

$$F(\omega) = \int \frac{dx}{2\pi} e^{-i\omega x} e^{mx} \int \frac{d\bar{\omega}}{2\pi i} \frac{e^{-i\lambda\bar{\omega}}}{\bar{\omega} + i\epsilon} \frac{e^{ix\omega}}{\Phi_H^{EFT}(\omega)}$$

by first evaluating

$$F(\omega - im) = \frac{e^{-i\omega\lambda}}{2\pi i(\omega + i\epsilon)} \frac{1}{\Phi_H^{EFT}(\omega)}$$

and then continuing to find $F(\omega)$. Since the integrals actually converge, we are guaranteed the validity of the continuation technique.

smoothly as $\lambda \to \infty$, into the *EFT* solution in the unregulated case. The other solutions displayed in (136) go over smoothly into the other solutions in the unregulated case. Thus, we have been able to show that $\Phi^{EFT}(\omega)$ is the smooth regulator limit of the regulated iterative solution. For this reason, as discussed above, we feel it is the most physical, and should be used in physical applications of the theory.

Nonrenormalizability of solutions

We have not discussed the behavior of the amplitude near zero coupling. Because $f(g^2)$ and ν generally appear in the solutions as the product

$$\ln \{\nu[f(g^2)]^{1/m}\} \tag{142}$$

small coupling is linked with low energy. It turns out that there is a logarithmic singularity at zero coupling. These topics will be the subject of a future paper. For now, it suffices to mention that the solutions are clearly singular at zero coupling [no perturbation expansion in $f(g^2)$]. For example, in the case of the inhomogeneous solution (87), we see that already in ω space

$$\frac{i}{2m} \coth \frac{\pi}{m}(\omega + i\epsilon) \exp\left[\frac{i\omega}{m} \ln f(g^2)\right] = \frac{P}{2\pi} \int_{-\infty}^{+\infty} \frac{d\xi\, e^{-i\xi\omega}}{1 - f(g^2)e^{\xi m}} \equiv \sum_{n=0}^{\infty} [f(g^2)]^n\, \delta(\omega + imn); \tag{143}$$

$$\delta(\omega + imn) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} e^{mnx}\, dx.$$

D. SUMMARY AND COMMENTS

We have solved the nonrelativistic, relativistic, homogeneous, and inhomogeneous N/D equations with the singular inputs that occur in nonrenormalizable field theory, and picked the single solution we felt was physically most meaningful. It is the direct exponentiation- and Fourier-transform solution to the relativistic inhomogeneous case—because this one corresponds to the sum of the (regulated) perturbation expansion. Since the iteration process systematically includes long-range information, our program reduces some large part of the weak interaction problem to quadratures. The reader is reminded that in Appendix III we do the integrals *inside* the final inverse transform, leaving only one integration, the final inverse transform itself, to be done. Thus, we can calculate nonrenormalizable, unitary, and dispersive generalizations whenever an absorptive part, while not itself divergent, has bad asymptotic behavior.

We hasten to repeat our awareness that the terrible behavior of the inputs far to the left lead to solutions that violate unitarity in the cross channel; we know that such left-hand cuts are forbidden for this reason, in a full theory, yet we feel that the approach may still be valuable in the physical region.

We want to discuss this in some detail in the next few paragraphs.

As explained in Ref. 1, using the first N ladder graphs as input into our program simply guarantees the exact ladder graphs up to that order in the solution. Above that the N/D has picked up just those bits and pieces of the higher-order graphs that are needed to make the first N graphs finite and unitary. This is a remarkable feat and, certainly, as N increases, we introduce a very great deal of information into the amplitude.

The point is that, *if* a theory of the W meson exists that adds up the finite absorptive parts of the perturbation theory (which the F–P theory does not!), then as N increases we are feeding in the *exact*[14] left-hand cut (no mass approximations are necessary in the program) further and further to the left, and each new graph actually specifies the absorptive part out a very large increment to the left. For example, with N ladder graphs as input, we would have the exact left-hand cut of the ladder out to $s = -(NM)^2$, where M is the W mass.

This much is close in spirit to the usual S matrix "nearest-singularity" philosophy. A seeming objection to the applicability of such a philosophy in this context is that, still, beyond $s = -(NM)^2$, the input cut increases like a power of s (whereas, presumably, the real cut eventually begins to oscillate)—and distant parts of the cut appears to be heavily weighted. In answer to this, we emphasize that *the program contains its own (self-consistent) damping mechanism* (the oscillations generated by the unitarity requirement), *which tends to pay less and less attention to whatever is far out on the left* anyway! Furthermore, we have seen that the damping becomes progressively *more* severe as the inputs misbehave violently. For example, note the term cos $(m \ln^2 \nu)$ in Eq. (69). This damping of the input far to the left lends credibility to the nearest singu-

[14] The left-hand cuts of the (mass shell) ladder graphs are generated strictly from many-meson intermediate states (the mixed meson-lepton intermediate states can only contribute off the mass shell) and these are *all* of finite discontinuity.

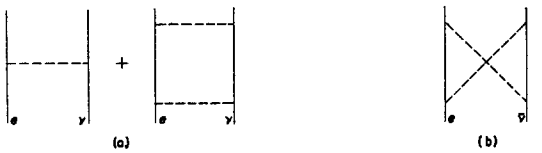

FIG. 10. $e - \nu$ channel input; (b) $e - \bar{\nu}$ channel input.

larity aspects of the program. Investigations presently in progress seem to indicate that in fact the program may actually converge rapidly in the low-energy physical region (as the input is improved by the addition of progressively shorter-range forces).

Possible applications of the program are legion:

For experimental purposes, it will be interesting to determine the low-energy predictions of the program using, for example, some finite number of ladder graphs as input. For example, the simplest inputs for the left-hand cuts in the $e - \nu$ and $e - \bar{\nu}$ channels would be those of Fig. 10(a,b). A W-meson bootstrap is now possible in the $e - \bar{\nu}$ channel, e.g. with the input Fig. 10(b). (This would be the first step toward a self-consistent S-matrix theory of weak interactions.) The problem is, in principle, reduced to quadratures here, but the self-consistent solution, if such exists, is still a difficult numerical task.

Besides the ladder graphs, some nonplanar (crossed) graphs can also be included as input (e.g., that of Fig. 11). In general, unfortunately,

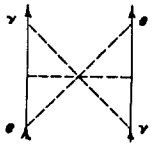

FIG. 11. Nonplanar graph input.

vertex and self-energy corrections like those of Fig. 12 cannot be included, because they have some divergent left-cut contributions.

The partial waves of a combined strong-weak S-matrix can be studied. Of particular interest would be the effect of both higher-order weak and strong corrections on the G_β/G_μ ratio. It would also be interesting to study weak mass corrections in general: in a nonrenormalizable theory, there is really no simple perturbational argument to predict magnitudes of weak mass shifts ahead of time.

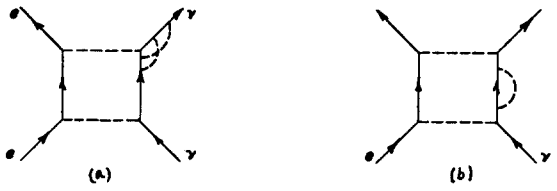

FIG. 12. Vertex and self-energy correction inputs.

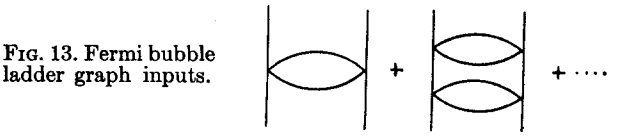

FIG. 13. Fermi bubble ladder graph inputs.

Finally, we emphasize that the program proposed here is of value beyond lepton–lepton scattering with W mesons. For example, the Fermi-bubble ladder graphs of Fig. 13 also have finite left-hand cuts, and can be used as input information. Further, ignoring the associated infrared divergence, the method can be extended to linearized gravitational theory, in which multiple graviton exchange graphs, shown in Fig. 14, have ill-behaved left-hand cuts of just the type treated here. Other higher spin exchange N/D calculations, such as spin $\frac{3}{2}$ exchange, are also possible.

All these applications are currently under investigation. The program unfortunately cannot be used for calculation in the singular potential theories, in the case of which the left-hand discontinuities are themselves infinite in perturbation theory.[15]

It is difficult to see how one might improve the program, for example, how one might include unitarity in the cross channels. Of course, in principle, the knowledge of the physical region in the partial waves of the $e - \nu$ channel determines the (now oscillatory) left-hand cuts in the cross channels. The N/D equations could be solved in these channels and the results in these physical regions used to redetermine the left-hand cuts in the $e - \nu$ channel, and so on. In practice, this is out of the question at the present time.

We would like to emphasize strongly, however, our feeling based on the results here that the requirement of unitarity is a very powerful and a very wise constraint in these problems, and that, in particular, it may be capable, in general, of obviating the need for any regulator procedure. We

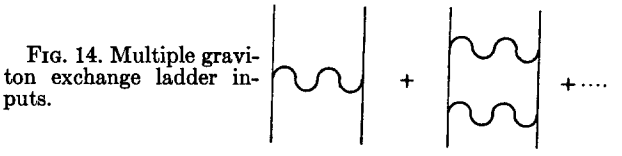

FIG. 14. Multiple graviton exchange ladder inputs.

[15] To see this, note that the singular potentials $[V(r) \sim r^{-n}]$ can be written as superpositions of Yukawa potentials (see Ref. 3), heavily weighted towards the short-range force $[\sigma(\mu) \sim \mu^{n-2}]$. Then the ordinary expression for the double spectral function of the box graph in terms of $\sigma(\mu)$.

$$\rho^{(2)}(s, t) = \frac{\pi}{4s^{\frac{1}{2}}} \int_0^\infty dt'\, dt''\, \sigma(t')\sigma(t'')\, \frac{1}{K(t, t', t'')}$$

is divergent; this is certainly not the case in the W theory.

feel strongly that unitarity may be the key to getting at the correct theory.

ACKNOWLEDGMENTS

The author would like to thank Professor Walter Gilbert for many hours of helpful discussion. Several clarifying discussions with J. Rix are also gratefully acknowledged. Some of the work and the final writing were done as a NATO Postdoctoral Fellow (1964–1965). The author thanks Professor L. Van Hove for the hospitality extended to him at CERN during this period.

APPENDIX I

An important identity in the text is Eq. (86). We prove it here as an example of how this type of integral can be evaluated. We want to evaluate

$$L(x) = \int_{-\infty}^{+\infty} d\omega\, e^{ix\omega} \ln [\operatorname{csch} \pi(\omega - i\epsilon)]. \tag{I1}$$

To do this, multiply both sides by x, change the x on the right to a derivative on the exponential, integrate by parts and use the identity (obtainable from residues)

$$\int_{-\infty}^{+\infty} e^{ix\omega}\, d\omega \coth \pi(\omega - i\epsilon) = P\frac{2i}{1 - e^{-x}}$$

to obtain

$$L(x) = [P2\pi/x(1 - e^{-x})] + \lambda\, \delta(x), \tag{I2}$$

where the λ is a parameter associated with dividing out the x. To determine λ, we insist that

$$\int_{-\infty}^{+\infty} \frac{dx}{2\pi} e^{-ix\omega}\left[P\frac{2\pi}{x(1 - e^{-x})} + \lambda\, \delta(x)\right] = \ln [\operatorname{csch} \pi(\omega - i\epsilon)]. \tag{I3}$$

Asymptotically, the right side of (I3) goes to

$$\begin{cases} -\pi\omega + \ln 2, & \text{as } \omega \to +\infty, \\ +\pi\omega + i\pi + \ln 2, & \text{as } \omega \to -\infty. \end{cases} \tag{I4}$$

The asymptotic behavior of the left side is dominated by the δ function and the pole at $x = 0$; it is

$$\mp \pi\omega \mp \tfrac{1}{2}\pi i + \lambda/2\pi, \quad \text{as } \omega \to \pm\infty. \tag{I5}$$

This allows us to identify $\lambda = 2\pi \ln 2 + \pi^2 i$, which establishes (86) of the text.

APPENDIX II

In this section we want to prove the identity

$$\exp\left[2P\int_{-\infty}^{+\infty} \frac{dx}{x(1 - e^{mx})(1 - e^{-x})}\right] = \frac{1}{m}. \tag{II1}$$

We start by establishing an interesting lemma

$$\exp\left(2P\int_{-\infty}^{+\infty} \frac{dx}{x}\frac{e^{-ix\omega}}{1 - e^{mx}}\right) = -4\sinh^2\frac{\pi\omega}{m}. \tag{II2}$$

The simplest way to prove this is to show that, on at least one branch of the logarithm,

$$2P\int_{-\infty}^{+\infty} \frac{dx}{x}\frac{e^{-ix\omega}}{1 - e^{mx}} = \ln\left(-4\sinh^2\frac{\pi\omega}{m}\right). \tag{II3}$$

To do this, we use the already known

$$2\int_{-\infty}^{+\infty} d\omega\, e^{i\omega x} \ln\sinh\frac{\pi}{m}(\omega + i\epsilon) = \frac{P4\pi}{x(1 - e^{mx})} - (2\pi \ln 4 - 2\pi^2 i)\, \delta(x) \tag{II4}$$

[which, within a change of variable and a complex conjugation, is just the identity proven in Appendix I]. Inverse Fourier transformation of (II4) gives exactly (II2). In the same way, one can easily show that

$$\exp\left(P\int_{-\infty}^{+\infty} \frac{dx}{x}\frac{e^{-ix\omega}}{1 - e^{mx}}\right) = -2i\sinh\frac{\pi\omega}{m} \tag{II5}$$

[which is essentially the square root of (II2)]. Now we can rewrite the identity (II5) in the form

$$\exp\left[P\int_{-\infty}^{+\infty} \frac{dx}{x}\frac{e^{-i\omega x}}{(1 - e^{mx})(1 - e^{-x})}\right] = -2i\sinh\frac{\pi\omega}{m} \times \exp\left[P\int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{x(1 - e^{mx})(e^{x} - 1)}\right]. \tag{II6}$$

The limit of the left-hand side at $\omega = 0$ is simply

$$\exp\left[P\int_{-\infty}^{+\infty} \frac{dx}{x(1 - e^{mx})(1 - e^{-x})}\right]. \tag{II7}$$

To obtain the value of the right-hand side at $\omega = 0$, we note that

$$M(\omega) = \exp\left[P\int_{-\infty}^{+\infty} \frac{dx\, e^{-i\omega x}}{x(1 - e^{mx})(e^{x} - 1)}\right] \times e^{(i\omega/m)\ln 2} \tag{II8}$$

satisfies the homogeneous difference equation

$$M(\omega) = i\operatorname{csch}\pi(\omega + i\epsilon)M(\omega + im). \tag{II9}$$

Thus, near the origin,

$$M(\omega) \sim [i/\pi(\omega + i\epsilon)]M(im). \tag{II10}$$

This gives us finally that, at $\omega = 0$, the right side of (II6) goes to

$$\frac{1}{m}\exp\left[P\int_{-\infty}^{+\infty} \frac{dx}{x(e^{-mx} - 1)(e^{x} - 1)}\right]. \tag{II11}$$

Rearranging the $\omega = 0$ limit of Eq. (II6), i.e., (III1) and (II7), gives immediately the identity (II1).

APPENDIX III

We want to show here that the integrals appearing inside the final inverse transforms in the paper can be expressed in terms of well-known functions, in particular, logarithms, dilogarithms, and hypergeometric functions. The two integrals of interest are

$$L_1^{(m)}(\omega) = P\int_{-\infty}^{+\infty} \frac{e^{-i\omega x}\,dx}{x(1-e^{mx})(1-e^{-x})},$$
$$L_2^{(m)}(\omega) = P\int_{-\infty}^{+\infty} \frac{e^{-i\omega x}\,dx}{x(1-e^{mx})\sinh \tfrac{1}{2}x}. \tag{III1}$$

The former appears in the solutions to the inhomogeneous equations, the latter in the homogeneous equations. We only need to calculate one of them because they are related by

$$L_2^{(m)}(\omega) = 2L_1^{(m)}(\omega - \tfrac{1}{2}i). \tag{III2}$$

We calculate $L_1^{(m)}(\omega)$ only.

The integrand of $L_1^{(m)}(\omega)$ displays an infinite of poles along the imaginary x axis. In particular, there is one third-order pole at $x = 0$, second-order poles at $x_n = 2\pi i n$ (n is any integer) and first-order poles at $x_{n'} = 2\pi i n'/m$ (n' is any integer not a multiple of m). We do the calculation for $\omega < 0$. $\omega > 0$ follows analogously. The residues at each of the poles in the upper half-plane are

$$R_0 = +\frac{1}{2m}\left[\omega^2 - i\omega(m-1) + \frac{m}{2} - \frac{m^2}{6} - \frac{1}{6}\right],$$
$$R_n = \frac{1}{2\pi i n m} e^{2\pi\omega n}\left(i\omega + \frac{1}{2\pi i n} + \frac{m-1}{2}\right), \tag{III3}$$
$$R_{n'} = -\frac{1}{2\pi i n'}\left[\frac{e^{2\pi\omega n'/m}}{(1-e^{-2\pi i n'/m})}\right].$$

Closing the contour above, ($\omega < 0$), we obtain the contributions to $L_1^{(m)}(\omega)$:

Third-order pole

$$\frac{\pi i}{2m}\left[\omega^2 - i\omega(m-1) + \frac{m}{2} - \frac{m^2}{6} - \frac{1}{6}\right]. \tag{III4}$$

Second-order poles

$$\frac{1}{m}\sum_{n=1}^{\infty}\frac{1}{n} e^{2\pi n\omega}\left(i\omega + \frac{1}{2\pi i n} + \frac{m+1}{2}\right)$$
$$= -\frac{1}{m}\left(i\omega + \frac{m-1}{2}\right)\ln(1-e^{2\pi\omega}) + \frac{1}{2\pi i m} Li_2(e^{2\pi\omega}), \tag{III5}$$

where we have recognized the dilogarithm, $Li_2(x) = \sum_{n=1}^{\infty}(x^n/n^2)$.

First-order poles (putting $n' = lm + r$; $l = 0, 1, \cdots$; $r = 1, \cdots m-1$)

$$-\sum_{r=1}^{m-1}\sum_{l=0}^{\infty}\frac{1}{lm+r}\left[\frac{e^{(2\pi\omega/m)(lm+r)}}{1-e^{-2\pi i r/m}}\right]$$
$$= \frac{1}{m}\sum_{r=1}^{m-1}\frac{e^{2\pi\omega r/m}}{1-e^{-2\pi i r/m}}\Phi\left(e^{2\pi\omega}, 1, \frac{r}{m}\right)$$
$$= \sum_{r=1}^{m-1}\frac{e^{2\pi\omega r/m}}{r(1-e^{-2\pi i r/m})}\,{}_2F_1\left(1, \frac{r}{m}; 1+\frac{r}{m}; e^{2\pi\omega}\right). \tag{III6}$$

In (III6), we have used the Bateman Manuscript Project (*Higher Transcendental Functions*, pp. 27 and 30).

We exhibit, for example, the particular cases $m = 1$ and $m = 2$:

$$L_1^{(1)}(\omega) = \frac{\pi i}{2}(\omega^2 + \tfrac{1}{6}) - i\omega\ln(1-e^{2\pi\omega}) + \frac{1}{2\pi i} Li_2(e^{2\pi\omega}), \tag{III7}$$

$$L_1^{(2)}(\omega) = \frac{\pi i}{4}(\omega^2 - i\omega + \tfrac{1}{6}) - \tfrac{1}{2}(i\omega + \tfrac{1}{2})\ln(1-e^{2\pi\omega}) + \frac{1}{4\pi i} Li_2(e^{2\pi\omega}) + \tfrac{1}{4}\ln\left(\frac{1+e^{\pi\omega}}{1-e^{\pi\omega}}\right). \tag{III8}$$

To obtain (III8), we used Bateman, Vol. I, p. 102.

APPENDIX IV. THE FUNCTIONS $\mathfrak{a}_m(\omega)$ WHICH VANISH AT $\omega \to 0$

In the text, it was pointed out that there are two types of ambiguities in the solution of the inhomogeneous cases. We restate here the one with which we are concerned in this section. After obtaining $\Phi_H^{EFT}(\omega)$, one may multiply by any of the functions $\mathfrak{a}(\omega)$ [see (126)], before applying the Green's function. We want to show here that the use of functions $\mathfrak{a}(\omega)$, in this way, which vanish at $\omega = 0$, such as, for example, $i\sinh 2\pi\omega/m$, do not yield any new solutions. We mean new in the sense that the solution generated might be different from some solutions generated with $\mathfrak{a}(\omega)$'s which are finite at $\omega = 0$.

Suppose we have found $\Phi_H^{EFT}(\omega)$, and for definiteness, that it is symmetric as well. And suppose that, before applying the Green's function, we first multiply by $i\sinh 2\pi\omega/m$. The behavior of the resulting solution to the homogeneous equation is, near $\omega = 0$,

$$\Phi_H(\omega) \sim i(2\pi\omega/m)\Phi_H^{EFT}(0). \tag{IV1}$$

The equation for the associated Green's function is then

$$\hat{G}(\omega + im) - \hat{G}(\omega) = (m/2\pi i)[\Phi_H^{EFT}(0)]^{-1}\ \delta'(\omega) - \lambda\ \delta(\omega), \qquad \text{(IV2)}$$

where λ is an arbitrary parameter that arises from the multiplication by the singular structure $-i$ csch $2\pi\omega/m$. We see that the zero at $\omega = 0$ introduced into the homogeneous solution by our choice of $\mathfrak{a}(\omega)$ has led to a more singular Green's function equation. This can be solved immediately by Fourier transform

$$\Phi_H^{EFT}(0)\hat{G}(\omega) = \frac{1}{4m}\,\mathrm{csch}^2\,\frac{\pi}{m}\,(\omega + i\epsilon) + \frac{\lambda' i}{2m}\coth\frac{\pi}{m}\,(\omega + i\epsilon). \qquad \text{(IV3)}$$

Thus, for the solution to the inhomogeneous equation, we have

$$\Phi(\omega) = \frac{i}{2m}\coth\frac{\pi}{m}\,(\omega + i\epsilon)[\Phi_H^{EFT}(0)]^{-1}\Phi_H^{EFT}(\omega) + \lambda''\cosh^2\frac{\pi\omega}{m}\,\Phi_H^{EFT}(\omega). \qquad \text{(IV4)}$$

This is exactly what would have been obtained by application of the ordinary Green's functions, e.g., (80), directly to $\Phi_H^{EFT}(\omega)$. The term proportional to λ'' corresponds in this latter case simply to adding a multiple of the homogeneous equation.

Similar results are obtained whenever an $\mathfrak{a}(\omega)$ vanishes at $\omega = 0$. This includes families of functions like $i^l \sinh^l 2\pi\omega n/m$, $i^l \sinh^l 2\pi\omega n/m \cosh^{l'} 2\pi\omega n/m$, etc. We conclude that particular $\mathfrak{a}(\omega)$'s which vanish at $\omega = 0$ do not, in general, generate solutions over and above those obtainable with the $\mathfrak{a}(\omega)$'s finite at $\omega = 0$. This justifies our taking the set $\{\mathfrak{a}(\omega)\}$ in the text, to be composed only of those periodic functions finite at the origin.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 7 JULY 1966

Wick Polynomials at a Fixed Time

ARTHUR JAFFE*
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 13 November 1965)

In general, a Wick polynomial must be smeared with a test function depending on both time and space in order to yield an operator in Hilbert space. However, in space time of two dimensions, it is sufficient to smear in the space direction alone. This statement is proved by an application of Weinberg's asymptotic theorem. Operators formed in this manner are candidates for approximate interaction Hamiltonians.

I. INTRODUCTION

An investigation of the old-fashioned Hamiltonian formalism in field theory leads to a study of energy densities which are local functions of free fields. We know from Haag's theorem that such densities do not yield operators when integrated over all space at a given time. Hence a natural approximate Hamiltonian to study would be gotten by smearing a density at a fixed time with a suitably smooth test function which depends on the space variables. Such a procedure may or may not lead to a well-defined operator.

In general, a quantum field $\phi(\mathbf{x}, t)$ is a distribution in space–time; when smeared with a test function $f(\mathbf{x}, t)$ depending on both space and time it yields an operator in Hilbert space. Borchers[1] has shown that it is sufficient to use a test function depending on the time alone. The resulting smeared field operators (when applied to a suitable invariant, dense set of vectors) are infinitely differentiable in spacelike directions. On the other hand, it is only under quite special circumstances that smearing in spacelike directions alone will do, and except in such special cases, the Hamiltonian needs further butchering.

For free fields, it is known to be the case that they may be taken at a sharp time. In fact, for $f(\mathbf{x})$ in the Schwartz space $\mathcal{S}$, the free field $\phi(f, t)$ is an infinitely differentiable function of the time.

* N.A.S.–N.R.C. Postdoctoral Fellow, supported by the Air Force Office of Scientific Research.

[1] H. J. Borchers, Nuovo Cimento **33**, 1600 (1964).

The equation for the associated Green's function is then

$$\hat{G}(\omega + im) - \hat{G}(\omega) = (m/2\pi i)[\Phi_H^{EFT}(0)]^{-1}\ \delta'(\omega) - \lambda\ \delta(\omega), \qquad \text{(IV2)}$$

where λ is an arbitrary parameter that arises from the multiplication by the singular structure $-i$ csch $2\pi\omega/m$. We see that the zero at $\omega = 0$ introduced into the homogeneous solution by our choice of $\mathfrak{a}(\omega)$ has led to a more singular Green's function equation. This can be solved immediately by Fourier transform

$$\Phi_H^{EFT}(0)\hat{G}(\omega) = \frac{1}{4m}\,\mathrm{csch}^2\,\frac{\pi}{m}\,(\omega + i\epsilon) + \frac{\lambda' i}{2m}\coth\frac{\pi}{m}\,(\omega + i\epsilon). \qquad \text{(IV3)}$$

Thus, for the solution to the inhomogeneous equation, we have

$$\Phi(\omega) = \frac{i}{2m}\coth\frac{\pi}{m}\,(\omega + i\epsilon)[\Phi_H^{EFT}(0)]^{-1}\Phi_H^{EFT}(\omega) + \lambda''\cosh^2\frac{\pi\omega}{m}\,\Phi_H^{EFT}(\omega). \qquad \text{(IV4)}$$

This is exactly what would have been obtained by application of the ordinary Green's functions, e.g., (80), directly to $\Phi_H^{EFT}(\omega)$. The term proportional to λ'' corresponds in this latter case simply to adding a multiple of the homogeneous equation.

Similar results are obtained whenever an $\mathfrak{a}(\omega)$ vanishes at $\omega = 0$. This includes families of functions like $i^l \sinh^l 2\pi\omega n/m$, $i^l \sinh^l 2\pi\omega n/m \cosh^{l'} 2\pi\omega n/m$, etc. We conclude that particular $\mathfrak{a}(\omega)$'s which vanish at $\omega = 0$ do not, in general, generate solutions over and above those obtainable with the $\mathfrak{a}(\omega)$'s finite at $\omega = 0$. This justifies our taking the set $\{\mathfrak{a}(\omega)\}$ in the text, to be composed only of those periodic functions finite at the origin.

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 7 JULY 1966

Wick Polynomials at a Fixed Time

ARTHUR JAFFE*
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received 13 November 1965)

In general, a Wick polynomial must be smeared with a test function depending on both time and space in order to yield an operator in Hilbert space. However, in space time of two dimensions, it is sufficient to smear in the space direction alone. This statement is proved by an application of Weinberg's asymptotic theorem. Operators formed in this manner are candidates for approximate interaction Hamiltonians.

I. INTRODUCTION

An investigation of the old-fashioned Hamiltonian formalism in field theory leads to a study of energy densities which are local functions of free fields. We know from Haag's theorem that such densities do not yield operators when integrated over all space at a given time. Hence a natural approximate Hamiltonian to study would be gotten by smearing a density at a fixed time with a suitably smooth test function which depends on the space variables. Such a procedure may or may not lead to a well-defined operator.

In general, a quantum field $\phi(\mathbf{x}, t)$ is a distribution in space–time; when smeared with a test function $f(\mathbf{x}, t)$ depending on both space and time it yields an operator in Hilbert space. Borchers[1] has shown that it is sufficient to use a test function depending on the time alone. The resulting smeared field operators (when applied to a suitable invariant, dense set of vectors) are infinitely differentiable in spacelike directions. On the other hand, it is only under quite special circumstances that smearing in spacelike directions alone will do, and except in such special cases, the Hamiltonian needs further butchering.

For free fields, it is known to be the case that they may be taken at a sharp time. In fact, for $f(\mathbf{x})$ in the Schwartz space $\mathcal{S}$, the free field $\phi(f, t)$ is an infinitely differentiable function of the time.

* N.A.S.–N.R.C. Postdoctoral Fellow, supported by the Air Force Office of Scientific Research.

[1] H. J. Borchers, Nuovo Cimento **33**, 1600 (1964).

However, the nonlinear functions of a free field, such as Wick polynomials, are another story. In particular it took an involved argument by Gårding and Wightman[2] in order to show that Wick polynomials of a free field actually define operators when smeared over *both* time and space. If one tries to restrict the test function to the space variables only, it is easy to see that in three or more dimensional space–time *no* Wick polynomial of degree two or more defines an operator applicable to vectors in the cyclic subspace generated from the Fock vacuum. In fact, expressions such as $:\phi^n:(f, t)$ cannot even be applied to the vacuum Ω_0, for they yield states whose norm squared is proportional to $\int |\tilde{f}(\mathbf{p}_1 + \cdots + \mathbf{p}_n)|^2 \, d\Omega(\mathbf{p}_1) \cdots d\Omega(\mathbf{p}_n) = \infty$.

On the other hand, if we work in two-dimensional space–time, $\int :\phi(x, t)^n: f(x)\, dx$ does define an operator, since the energy denominator in the invariant phase-space volume provides a natural cutoff. More generally, one can take at a fixed time any Wick polynomial of the free field which does not involve derivatives. Any such Wick polynomial yields an operator for test functions in $\mathcal{S}(x)$, the Schwartz space in the space variable. The resulting operator is continuous in the time and has a domain D_1, left invariant by the application of such operators. The end result is

Theorem 1: If $\phi(x, t)$ is a massive, free scalar field in two-dimensional space–time and

$$A(x, t) = \sum_{n=0}^{Q} a_n :\phi(x, t)^n: \tag{1}$$

is a Wick polynomial of $\phi(x, t)$ without derivatives, then

$$A(f, t) = \int A(x, t) f(x)\, dx, \qquad f \in \mathcal{S}, \tag{2}$$

is an operator defined on a dense domain $D_1 \subset H$. The domain D_1 can be chosen to be independent of t and invariant under the field,

$$A(f, t) D_1 \subset D_1. \tag{3}$$

The vacuum expectation values arising from such Wick polynomials are jointly continuous in the time and in the test function variables. Using the $\mathcal{S}$ norms defined in Eq. (17), the vacuum expectation values satisfy the following continuity conditions: For any ϵ and μ such that

$$\epsilon > 0, 1 > \mu > 0, \quad [\Omega_0, A_1(f_1, t_1) A_2(f_2, t_2) \cdots \cdots A_s(f_s, t_s)\Omega_0] \leq M(\epsilon, s)\, ||f_1||_\epsilon \cdots ||\tilde{f}_s||_\epsilon, \tag{4a}$$

and

$$\begin{aligned} &|[\Omega_0, A_1(f_1, t_1) \cdots A_s(f_s, t_s)\Omega_0] \\ &\quad - [\Omega_0, A_1(f_1, t_1') \cdots A_s(f_s, t_s')\Omega_0]| \\ &< M(\epsilon, s)\, ||\tilde{f}_1||_\epsilon \cdots ||\tilde{f}_s||_\epsilon \\ &\quad \times \left\{ \sum_{i=1}^{s} |t_i - t_i'|^{1-\mu} \, ||\tilde{f}_i||_{(1-\mu)/\mu} \right\}. \end{aligned} \tag{4b}$$

Here $M(\epsilon, s)$ is a constant and a tilde denotes the Fourier transform.

A possible domain D_1 is just given by all polynomials in smeared Wick polynomials $A(f, t)$ applied to the vacuum Ω_0. Then in terms of operators, the continuity estimates mean that for any Ω in D_1,

$$||A(f, t_1)\Omega - A(f, t_2)\Omega|| \leq M_1(\epsilon, s)\, ||\tilde{f}||_{\epsilon + (1-\mu)/\mu} \, |t_1 - t_2|^{\frac{1}{2}(1-\mu)}. \tag{5}$$

II. DIGRESSION ON $||\int: \phi(x,0)^n: f(x)\, dx\, \Omega_0||$

In this section we set

$$A = \int :\phi(x, 0)^n: f(x)\, dx, \tag{6}$$

and we see that the expression formally given by

$$\begin{aligned} ||A\Omega_0||^2 &= n! \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[\frac{1}{i} \Delta^{(+)}(x_1 - x_2) \right]^n \\ &\qquad \times \overline{f(x_1)} f(x_2)\, dx_1\, dx_2 \\ &= n! (2\pi)^{-n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\tilde{f}(p_1 + \cdots + p_n)|^2 \\ &\qquad \times \prod_{i=1}^{n} dp_i \, [p_i^2 + m^2]^{-\frac{1}{2}} \end{aligned} \tag{7}$$

actually defines a distribution in $\mathcal{S}'(x_1, x_2)$. This example is not directly relevant to the proof of Theorem 1, but only serves to give an idea of the more general method which is used to prove the theorem.

We wish to study the transformation $g \to Tg$ defined by

$$(Tg)(x) = \int_{-\infty}^{\infty} g(x + q)[q^2 + m^2]^{-\frac{1}{2}}\, dq. \tag{8}$$

Let $\mathfrak{F}_\mu^1$ be the space of all continuous functions $g(x)$ for which there exist real positive constants M, β, and x_0 such that

$$|g(x)| < Mx^{-\mu} \, |\log x|^\beta \quad \text{for} \quad |x| > x_0. \tag{9}$$

Now we apply the asymptotic theorem of Weinberg[3] to deduce that for all $g \in \mathfrak{F}_\mu^1$, where $0 < \mu \leq 1$,

[2] A. S. Wightman and L. Gårding, Arkiv Fysik **28**, 129 (1964).

[3] S. Weinberg, Phys. Rev. **118**, 838 (1960).

the integral which defines Tg is absolutely convergent and, furthermore,

$$T: \mathfrak{F}_\mu^1 \to \mathfrak{F}_\mu^1. \tag{10}$$

The theorem is applied by noting that the asymptotic coefficient α for any line in the space $\mathbf{R}^2(x, q)$ is $-(1 + \mu)$, except for the line in the direction $(1, -1)$ which has the coefficient -1. Furthermore, $\alpha(\mathbf{R}^2) = -(1 + \mu)$ and $D_I = -\mu$. Hence, the integral in question exists and has the asymptotic coefficient $-\mu$ as given by Weinberg's formula (12). This shows that Tg has the correct asymptotic behavior. We must still prove that it is continuous in order to show that it belongs to $\mathfrak{F}_\mu^1$. For this purpose we write

$$(Tg)(x) - (Tg)(x + h) = \left(\int_{-\infty}^{-M} + \int_M^{\infty}\right) \times [g(x + q) - g(x + h + q)](q^2 + m^2)^{-\frac{1}{2}}\, dq + \int_{-M}^{M} [g(x + q) - g(x + h + q)](q^2 + m^2)^{\frac{1}{2}}\, dq.$$

The first two terms can be made arbitrarily small by choosing M sufficiently large, since the integral defining $T(g)$ is absolutely convergent. In order to deal with the last term, we note that since $g(x)$ is continuous, it is uniformly continuous on the compact $[x - |\delta| - M, x + |\delta| + M]$. Hence this term can be made small by choosing $|h|$ small, which demonstrates that $Tg \in \mathfrak{F}_\mu^1$. This completes the proof of (10).

We see that

$$||A\Omega_0||^2 = n!\,(2\pi)^{-n}[T^n(|\tilde{f}|^2)](0). \tag{11}$$

Since $f \in \mathfrak{S}$, so are $\tilde{f}$ and $|\tilde{f}|^2$ also elements of $\mathfrak{S} \subset \mathfrak{F}_\mu^1$. Hence, $T^n(|\tilde{f}|^2) \in \mathfrak{F}_\mu^1$, $||A\Omega_0||^2$ is well defined for all test functions in $\mathfrak{S}$.

Let us now use the norms

$$||g||_r = \sup_x (1 + x^2)^{r/2}\, |g(x)|. \tag{12}$$

We see that for every $\epsilon > 0$, $r > 0$ such that $r + \epsilon < \mu \leq 1$ and $g \in \mathfrak{F}_\mu^1$, then

$$||Tg||_r \leq M(\epsilon)\, ||g||_{r+\epsilon}, \tag{13}$$

where $M(\epsilon)$ is a constant depending on ϵ but independent of g. This follows from the fact that

$$|(1 + x^2)^{r/2}(Tg)(x)| = \left|\int_{-\infty}^{\infty} dq\, (1 + x^2)^{\frac{1}{2}r}[1 + (x + q)^2]^{\frac{1}{2}r + \frac{1}{2}\epsilon} \times [1 + (x + q)^2]^{-\frac{1}{2}(r+\epsilon)}(q^2 + m^2)^{-\frac{1}{2}} g(x + q)\right|$$

$$\leq ||g||_{r+\epsilon} \int_{-\infty}^{\infty} dq\, (1 + x^2)^{r/2}(q^2 + m^2)^{-\frac{1}{2}} \times [1 + (x + q)^2]^{-\frac{1}{2}(r+\epsilon)}. \tag{14}$$

Since $h_{r+\epsilon}(x) = (1 + x^2)^{-\frac{1}{2}(r+\epsilon)} \in \mathfrak{F}_{r+\epsilon}^1$, it follows that

$$\int_{-\infty}^{\infty} dq\, (q^2 + m^2)^{-\frac{1}{2}}[1 + (x + q)^2]^{-\frac{1}{2}(r+\epsilon)} = [T(h_{r+\epsilon})](x)$$

is also in $\mathfrak{F}_{r+\epsilon}^1$. Hence,

$$\sup_x (1 + x^2)^{\frac{1}{2}r}(Th_{r+\epsilon})(x) = M(\epsilon) < \infty, \tag{15}$$

and so inequality (13) is proved. From this we deduce that for all $\epsilon > 0$,

$$|[T^n(g)](0)| \leq ||T^n(g)||_0 \leq M(\epsilon)^n\, ||g||_{n\epsilon}. \tag{16}$$

But if $g = |\tilde{f}|^2$, then $g \in \mathfrak{F}_\mu^1$ for all $0 < \mu \leq 1$ and

$$||g||_r \leq (||\tilde{f}||_{\frac{1}{2}r})^2,$$

so that for all $\epsilon > 0$ such that $n\epsilon < 1$,

$$||A\Omega_0||^2 \leq (2\pi)^{-n} n!\, [M(\epsilon)]^n (||\tilde{f}||_{\frac{1}{2}n\epsilon})^2,$$

which proves that expression (7) defines a distribution of the required form.

III. THE SPACES $\mathfrak{S}_\mu^s$, $\mathfrak{F}_\mu^s$ AND THE OPERATORS T_{ij}'

In this section we introduce spaces of and operators on functions of s variables analogous to those for one variable studied in Sec. II. In what follows we often regard functions $g(x_1, \cdots, x_s)$, which depend on s scalar variables, as functions of one s-vector variable $\mathbf{x}$. We now make the following definitions:

The Spaces $\mathfrak{S}_\mu^s$. Let r stand for the Euclidean distance in $\mathbf{R}^s(x_1, \cdots, x_s)$. $r^2 = \sum_{i=1}^s x_i^2$. Let $||\cdot||_\mu$ denote the family of norms on functions of s variables given by

$$||g||_\mu = \operatorname*{ess.\,sup}_{\mathbf{x} \in \mathbf{R}^s} (1 + r^2)^{\frac{1}{2}\mu}\, |g(\mathbf{x})|. \tag{17}$$

Let $\mathfrak{S}_\mu^s$ be the space of all continuous functions of s variables with finite μ norm. Note that

$$\mathfrak{S}_{\mu_1}^s \subset \mathfrak{S}_{\mu_2}^s, \quad \text{for} \quad \mu_1 > \mu_2. \tag{18}$$

The Spaces $\mathfrak{F}_\mu^s$. A function of s variables $g(\mathbf{X})$ belongs to the space $\mathfrak{F}_\mu^s$ if it is continuous and for any subspace $S \subset \mathbf{R}^s$ of dimension r, and for any basis $\mathbf{L}_1, \cdots, \mathbf{L}_r$ which spans S, $g(\mathbf{X})$ satisfies

$$|g(\mathbf{L}_1\eta_1\eta_2 \cdots \eta_r + \mathbf{L}_2\eta_2 \cdots \eta_r + \cdots + \mathbf{L}_r\eta_r)| = O[\eta_1^{-\mu}\eta_2^{-2\mu} \cdots \eta_r^{-r\mu}(\log \eta_1)^{\beta_1} \cdots (\log \eta_r)^{\beta_r}], \tag{19}$$

for some $\beta_1, \cdots, \beta_r$ when $\eta_1, \cdots, \eta_r$ tend independently to infinity.

Note that $\mathfrak{S}_\mu^s \subset \mathfrak{F}_\mu^s$ and

$$\mathfrak{F}_{\mu_1}^s \subset \mathfrak{F}_{\mu_2}^s \quad \text{for} \quad \mu_1 > \mu_2. \tag{20}$$

The Integral Operators T_{ij}^s. We define a set of integral operators T_{ij}^s for $i \neq j$ which act on functions of s variables by

$$(T_{ij}^s g)(x_1, \cdots, x_s) = \int_{-\infty}^{\infty} dq\, (q^2 + m^2)^{-\frac{1}{2}} \times g(x_1, \cdots, x_i + q, \cdots, x_j - q, \cdots, x_s). \tag{21}$$

Note that for a given s, the various T_{ij}^s formally commute.

Theorem 2: I. If $g \in \mathfrak{F}_\mu^s$, $0 < \mu$, then the integral which defines $T_{ij}^s(g)$ converges absolutely.

II. $T_{ij}^s : \mathfrak{F}_\mu^s \to \mathfrak{F}_\mu^s$ for $0 < \mu \leq 1$.

III. For fixed s, the various T_{ij}^s commute on $\mathfrak{F}_\mu^s$ if $\mu > 0$.

IV. For any $\epsilon > 0$, $\mu > 0$ such that $\mu + \epsilon \leq 1$, T_{ij}^s is a bounded operator from $\mathfrak{S}_{\mu+\epsilon}^s$ to $\mathfrak{S}_\mu^s$. Hence there is a constant $M(\epsilon, \mu)$ depending only on μ and ϵ such that

$$||T_{ij}^s(g)||_\mu \leq M(\epsilon, \mu)\, ||g||_{\mu+\epsilon}, \qquad \mu + \epsilon \leq 1. \tag{22}$$

Remarks: 1. Statement II implies that if $g \in \mathfrak{F}_\mu^s$, $T_{ij}^s(g)$ is continuous.

2. From (22), we see that T_{ij}^s has a unique extension to the closure of $\mathfrak{S}_\mu^s$ in the μ norm. So extended it is a continuous linear operator into the closure of $\mathfrak{S}_{\mu-\epsilon}^s$ in the $\mu - \epsilon$ norm.

Proof: The convergence of the integral defining $T_{ij}^s(g)$ I and the rate of decrease of the function $T_{ij}^s(g)$ (part of II) both follow by a direct application of Weinberg's theorem (3). The verification of the hypotheses is just as straightforward, but far more tedious than, the discussion of T on $\mathfrak{F}_\mu^1$ given above, and merely uses the definitions of T_{ij}^s and $\mathfrak{F}_\mu^s$. In order to finish the proof of II, we must show that $T_{ij}^s(g)$ is continuous,

$$\begin{aligned}(T_{ij}^s g)(\mathbf{x}) - (T_{ij}^s g)(\mathbf{x} + \mathbf{h}) &= \int_{-\infty}^{\infty} dq\, [q^2 + m^2]^{-\frac{1}{2}} \{g(x_1, \cdots, x_i + q, \cdots, x_s) - g((x_1 + h_1, \cdots, x_s + h_s)\} \\ &= \left(\int_{-\infty}^{-M} + \int_{M}^{\infty}\right) dq\, [q^2 + m^2]^{-\frac{1}{2}} \{g(x_1, \cdots, x_i + q, \cdots, \cdots, x_j - q, \cdots, x_s) \\ &\quad - g(x_1 + h_1, \cdots, x_i + h_i + q, \cdots, x_j + h_j - q, \cdots, x_s)\} \\ &\quad + \int_{-M}^{M} dq\, [q^2 + m^2]^{\frac{1}{2}} \{g(x_1, \cdots, x_i + q, \cdots, x_j - q, \cdots, x_s) \\ &\quad - g(x_1 + h_1, \cdots, x_i + h_i + q, \cdots, x_j + h_j - q, \cdots, x_s)\}. \end{aligned} \tag{23}$$

The first two terms of (23) can be made arbitrarily small by choosing $M(\mathbf{x})$ large enough, since the integrals defining $T_{ij}^s(g)$ are absolutely continuous. On the compact

$$\begin{aligned}&([x_1 - |\delta_1|, x_1 + |\delta_1|], \cdots, \\ &\quad \times [x_i - |\delta_i| - M, x_i + |\delta_i| + M], \cdots, \\ &\quad \times [x_j - |\delta_j| - M, x_j + |\delta_j| + M], \cdots, \\ &\quad \times [x_s - |\delta_s|, x_s + |\delta_s|]),\end{aligned}$$

$g(\mathbf{x})$ is uniformly continuous. Henceby choosing $|h_i| < |\delta_i|$ sufficiently small, the third term of (23) can be made arbitrarily small. This completes the proof of II.

In order to demonstrate the commutativity of the various T_{ij}^s, we need only apply the Weinberg theorem to the integrand corresponding to the product of the two integral operations in question. For $g \in \mathfrak{F}_\mu^s$, $\mu > 0$, the theorem tells us that the double integral is absolutely convergent. Hence by Fubini's theorem, the two iterated integrals agree and equal the double integral.

We now proceed to prove IV.

$$\begin{aligned}|(1 + r^2)^{\frac{1}{2}\mu}(T_{ij}^s g)(\mathbf{x})| = \Big| &\int_{-\infty}^{\infty} dq\, [q^2 + m^2]^{-\frac{1}{2}} (1 + r^2)^{\frac{1}{2}\mu} [1 + (x_i + q)^2 + (x_j - q)^2 + \sum_{\substack{k \neq i \\ k \neq j \\ 1 \leq k \leq s}} x_k^2]^{\frac{1}{2}(\mu+\epsilon)} \\ &\times [1 + (x_i + q)^2 + (x_j - q)^2 + \sum_{\substack{k \neq i \\ k \neq j \\ 1 \leq k \leq s}} x_k^2]^{-\frac{1}{2}(\mu+\epsilon)} g(x_1, \cdots, x_i + q, \cdots, x_j - q, \cdots, x_s)\Big|\end{aligned}$$

$$\leq \sup_{q'} \{[1 + (x_i + q')^2 + (x_j - q')^2 \sum_{\substack{k\neq i\\ k\neq j\\ 1\leq k\leq s}} x_k^2]^{\frac{1}{2}(\mu+\epsilon)} |g(x_1, \cdots, x_i + q', \cdots, x_j - q', \cdots, x_s)|\}$$

$$\times \int_{-\infty}^{\infty} dq\, [q^2 + m^2]^{-\frac{1}{2}}[1 + r^2]^{\frac{1}{2}\mu}[1 + (x_i + q)^2 + (x_j - q)^2 + \sum_{\substack{k\neq i\\ k\neq j\\ 1\leq k\leq s}} x_k^2]^{-\frac{1}{2}(\mu+\epsilon)}. \tag{24}$$

The sup term on the right-hand side of (24) is majorized by $||g||_{\mu+\epsilon}$. The remaining integral term could also be written as $(1 + r^2)^{\frac{1}{2}\mu}[T^s_{ij}(h^s_{\mu+\epsilon})](\mathbf{x})$, where

$$h^s_l(\mathbf{x}) = (1 + r^2)^{-\frac{1}{2}l} \in \mathfrak{F}^s_l. \tag{25}$$

Hence, by (ii) we know that $T^s_{ij}(h^s_{\mu+\epsilon}) \in \mathfrak{F}^s_\lambda$, where $\lambda = \min(\mu + \epsilon, 1)$, so that if $\mu + \epsilon \leq 1$,

$$\sup_{\mathbf{x}\in\mathbf{R}^s} (1 + r^2)^{\frac{1}{2}\mu}[T^s_{ij}(h^s_{\mu+\epsilon})](\mathbf{x}) = M(\epsilon, \mu) < \infty.$$

Hence,

$$||T^s_{ij}(g)||_\mu = \sup_{\mathbf{x}\in\mathbf{R}^s} |(1 + r^2)^{\frac{1}{2}\mu}[T^s_{ij}(g)](\mathbf{x})|$$

$$\leq M(\epsilon, \mu)\, ||g||_{\mu+\epsilon} \quad \text{for} \quad \mu + \epsilon \leq 1.$$

IV. PROOF OF THEOREM 1

We note that, in order to prove Theorem 1, it is sufficient to prove that the vacuum expectation values satisfy (4a) and (4b). From these estimates, it follows that one can use the reconstruction theorem of Wightman[4] to ensure the existence of field operators $A(f, t)$ which are defined on the domain D_1 described in Theorem 1 and which leave D_1 invariant. Furthermore, the reconstruction theorem tells us that the vacuum expectation values of these operators $A(f, t)$ will just be those with which we begin.

Let us use the notation for Wightman functions,

$$W(\mathbf{f}, \mathbf{t}) = [\Omega_0, A_1(f_1, t_1) \cdots A_s(f_s, t_s)\Omega_0]. \tag{26}$$

Now $W(\mathbf{f}, \mathbf{t})$ can be expanded in a finite sum of terms, each one in the form

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big[\prod_{1\leq i<j\leq s} \Big\{\frac{1}{i}\Delta^{(+)}[(x_i, t_i) - (x_j, t_j)]\Big\}^{r_{ij}} \times \prod_{k=1}^{s} f(x_k)\, dx_k\Big\]. \tag{27}$$

(For the exact expansion, see the remark following Theorem A.1 of Ref. 5.) We now use a graph to represent each of these typical terms. The points of the graph are labeled 1, 2, $\cdots$, s corresponding to $x_1, x_2, \cdots, x_s$. There is a directed line corresponding to each factor $(1/i)\Delta^{(+)}(x_k - x_l)$ which runs from k to t, where $k < t$. The elements of the incidence matrix for the graph ϵ_{il} is labeled by points $i = 1, 2, \cdots, s$ and lines $l = 1, 2, \cdots, N$, where $N = \sum_{i<j} r_{ij}$. Then

$$\epsilon_{il} = \begin{cases} +1 & \text{if line } l \text{ runs from vertex } i,\\ -1 & \text{if line } l \text{ runs to vertex } i,\\ 0 & \text{if line } l \text{ does not enter vertex } i. \end{cases}$$

To each line l we assign a momentum variable P_l which ranges over $(-\infty, \infty)$. Then the typical term (27) contributing to $W(\mathbf{f}, \mathbf{t})$ can be written

$$(2\pi)^{-N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\Big\{-i \sum_{i=1}^{s}\sum_{l=1}^{N} \epsilon_{il} t_i \omega_l\Big\} \times \Big[\prod_{i=1}^{s} f_i\Big(\sum_{l=1}^{N} \epsilon_{il} P_l\Big)\Big] \prod_{l=1}^{N} \omega_l^{-1}\, dP_l, \tag{28}$$

where $\omega_l = (P_l^2 + m^2)^{\frac{1}{2}}$. In order to prove (4a), we note that (28) can be majorized by

$$(2\pi)^{-N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{s} \Big|\tilde{f}_i\Big(\sum_{l=1}^{N} \epsilon_{il} P_l\Big)\Big| \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}}\, dP_l \leq (2\pi)^{-N} \sup_{P_l'\in\mathbf{R}^N} \prod_{i=1}^{s}\Big[1 + \Big(\sum_{l=1}^{N} \epsilon_{il} P_l'\Big)^2\Big]^{\frac{1}{2}\epsilon}$$

$$\times \Big|\tilde{f}_i\Big(\sum_{l=1}^{N} \epsilon_{il} P_l'\Big)\Big| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{s}\Big[1 + \Big(\sum_{l=1}^{N} \epsilon_{il} P_l\Big)^2\Big]^{-\frac{1}{2}\epsilon} \times \prod_{l=1}^{N} (P_l^2 + m^1)^{-\frac{1}{2}}\, dP_l.$$

Now

$$\sup_{P\in\mathbf{R}^N} \prod_{i=1}^{s}\Big[1 + \Big(\sum_{l=1}^{N} \epsilon_{il} P_l\Big)^2\Big]^{\frac{1}{2}\epsilon} \Big|\tilde{f}_i\Big(\sum_{l=1}^{N} \epsilon_{il} P_l\Big)\Big| = \prod_{i=1}^{s} ||\tilde{f}_i||_\epsilon \tag{29}$$

and

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{s}\Big[1 + \Big(\sum_{i=1}^{N} \epsilon_{il} P_l\Big)^2\Big]^{-\frac{1}{2}\epsilon} \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}}\, dP_l = [\{\prod_{1\leq i<j\leq s} (T^s_{ij})^{r_{ij}}\}(k^s_\epsilon)](\mathbf{0}) = M.(\epsilon, s) < \infty, \tag{30}$$

[4] A. S. Wightman, Phys. Rev. **101**, 860 (1956).
[5] A. M. Jaffe, Ann. Phys. (N. Y.) **32**, 127 (1965).

where

$$k_\epsilon^s(x_1 \cdots, x_s) = \prod_{i=1}^{s} (1 + x_i^2)^{-\frac{1}{2}\epsilon},$$

and we have used

$$k_\epsilon^s \in \mathfrak{S}_\epsilon^s,$$

and Theorem (2.III) and (2.IV). Hence expression (28) is majorized by

$$(2\pi)^{-N} M(\epsilon, s) \prod_{i=1}^{s} ||\tilde{f}_i||_\epsilon,$$

and $W(\mathbf{f}, \mathbf{t})$ is majorized by a finite sum of such terms.

In order to prove (4b), we return to expression (28) for a typical contribution to $W(\mathbf{f}, \mathbf{t})$. A typical contribution to $W(\mathbf{f}, \mathbf{t}) - W(\mathbf{f}, \mathbf{t}')$ is

$$(2\pi)^{-N} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\exp\left\{ -i \sum_{j=1}^{s} \sum_{l=1}^{N} \epsilon_{jl} t_j \omega_l \right\} - \exp\left\{ -i \sum_{j=1}^{s} \sum_{l=1}^{N} \epsilon_{jl} t_j' \omega_l \right\} \right] \times \prod_{i=1}^{s} \tilde{f}_i\left(\sum_{l=1}^{N} \epsilon_{il} P_l \right) \times \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}} \, dP_l. \tag{31}$$

Since

$$A_1 A_2 \cdots A_s - B_1 B_2 \cdots B_s = (A_1 - B_1) A_2 \cdots A_s + B_1 (A_2 - B_2) A_3 \cdots A_s + \cdots + B_1 B_2 \cdots B_{s-1} (A_s - B_s),$$

we can set $C_j = \sum_{l=1}^{N} \epsilon_{jl} (P_l^2 + m^2)^{\frac{1}{2}}$,

$$A_j = \exp\{-iC_j t_j\},$$

and

$$B_j = \exp\{-iC_j t_j'\}.$$

Thus (31) splits into a sum of s integrals where in each of these integrals one term $(A_j - B_j)$ appears. Let us now split each of these integrals in two, according to whether $|C_j| < |t_j - t'|^{-\mu}$ or $|C_j| > |t_j - t_j'|^{-\mu}$. Call these regions, respectively, I_j and II_j. In region I_j, majorize $|A_j - B_j|$ by

$$|A_j - B_j| = \left| iC_j \int_{t_j'}^{t_j} e^{-iC_j s_j} \, ds_j \right| \leq |C_j| \, |t_j - t_j'| \leq |t_j - t_j'|^{1-\mu},$$

and majorize $|A_k|$ and $|B_k|$ by 1, where $k \neq j$. This yields a term majorized by

$$|t_j - t_j'|^{1-\mu} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{i=1}^{s} \left| \tilde{f}_i\left(\sum_{l=1}^{N} \epsilon_{il} P_l \right) \right| \times \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}} \, dP_l < M(\epsilon, s) \, |t_j - t_j'|^{1-\mu} \prod_{i=1}^{s} ||\tilde{f}_i||_\epsilon. \tag{32}$$

In region II_j we write $|A_j - B_j| \leq 2$, and hence majorize that integral by

$$\begin{aligned}
&\int_{II_j} \prod_{i=1}^{s} \left| \tilde{f}_i\left(\sum_{l=1}^{N} \epsilon_{il} P_l \right) \right| \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}} \, dP_l \\
&\leq \sup_{p' \in \mathbf{R}^N} \left(\left[1 + \left(\sum_{l=1}^{N} \epsilon_{jl} P_l' \right)^2 \right]^{(1-\mu)/2\mu} \prod_{i=1}^{s} \left\{ \left[1 + \left(\sum_{l=1}^{N} \epsilon_{il} P_l' \right)^2 \right]^{\frac{1}{2}\epsilon} \left| \tilde{f}_i\left(\sum_{l=1}^{N} \epsilon_{il} P_l' \right) \right| \right\} \right) \\
&\quad \times \int_{II_j} \left[1 + \left(\sum_{l=1}^{N} \epsilon_{jl} P_l \right)^2 \right]^{-(1-\mu)/2\mu} \prod_{i=1}^{s} \left[1 + \left(\sum_{l=1}^{N} \epsilon_{il} P_l \right)^2 \right]^{-\frac{1}{2}\epsilon} \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}} \, dP_l \\
&\leq ||\tilde{f}_1||_\epsilon \cdots ||\tilde{f}_j||_{\epsilon + (1-\mu)/\mu} \cdots ||\tilde{f}_s||_\epsilon \sup_{P' \in II_j} \left\{ 1 \Big/ \left[1 + \left(\sum_{l=1}^{N} \epsilon_{jl} P_l' \right)^2 \right] \right\}^{(1-\mu)/2\mu} \\
&\quad \times \int_{II_j} \prod_{i=1}^{s} \left[1 + \left(\sum_{l=1}^{N} \epsilon_{il} P_l \right)^2 \right]^{-\frac{1}{2}\epsilon} \prod_{l=1}^{N} (P_l^2 + m^2)^{-\frac{1}{2}} \, dP_l \\
&\leq |t_j - t_j'|^{1-\mu} \, ||\tilde{f}_1||_\epsilon \cdots ||\tilde{f}_j||_{\epsilon + (1-\mu)/\mu} \cdots ||\tilde{f}_s||_\epsilon \, M(\epsilon, s) C.
\end{aligned} \tag{33}$$

In estimate (32) we can replace $||\tilde{f}_j||_\epsilon$ by the larger quantity $||\tilde{f}_j||_{\epsilon + (1-\mu)/\mu}$. Combining estimates (32) and (33) and then summing over $j = 1, 2, \cdots, s$ yields an upper bound for the modulus of (31) which is of the form required on the right-hand side of (46). Since $W(\mathbf{f}, \mathbf{t}) - W(\mathbf{f}, \mathbf{t}')$ is just a finite sum of terms, each with the form (31), we can get estimate (46) by replacing $M(\epsilon, s)$ in (32) and (33) by the largest constant coming from the typical terms, times the number of terms. This completes the proof of Theorem 1.

The author thanks O. E. Lanford for remarks.

Exact Invariants for a Class of Nonlinear Wave Equations

MARTIN D. KRUSKAL
Princeton University Plasma Physics Laboratory, Princeton, New Jersey
AND
NORMAN J. ZABUSKY
Bell Telephone Laboratories, Inc., Whippany, New Jersey
(Received 14 July 1965; final manuscript received 2 September 1965)

A method is given for obtaining explicitly an *infinite* number of exact invariants for a physical system described by the coupled set of first-order hyperbolic partial differential equations

$$\partial u_i/\partial t = A_{ij}(u_0, u_1, \cdots u_n)\, \partial u_j/\partial x \qquad (i, j = 0, 1, \cdots, n). \tag{A1}$$

Temporal and spatial invariants are constructed as integrals of temporally and spatially invariant densities T and X, over appropriate spatial and temporal intervals, respectively. For physical systems the energy and momentum densities are temporal invariant densities. These invariant densities are solutions of the hodograph transformed equations corresponding to (A1). For the case $n = 1$ every invariant density T satisfies an equation in *conservation form*: $(T_{u_0})_t - (T_{u_1})_x = 0$. The methods are applied to the equation

$$y_{tt} - (1 + \epsilon y_x)^\alpha y_{xx} = 0, \tag{A2}$$

and a denumerable infinity of invariant densities, each expressible as a polynomial, are calculated in two equivalent cases: the first when (A2) (with $\alpha = 1$) is expressed in *zero diagonal* form $u_t = v_x$, $v_t = (1 + \epsilon u)u_x$, where $u = y_x$ and $v = y_t$; and the second when (A2) is expressed in *diagonal* form $r_t - \Phi r_x = 0$, $s_t + \Phi s_x = 0$, where $\Phi = (\frac{1}{2}\epsilon(2 + \alpha)(r + s))^{\alpha/(2+\alpha)}$ and r and s, the Riemann invariants of (A2), are related to u and v. A theorem of Noether is used to construct from the invariant densities *continuous transformation* groups that leave the action functional invariant. Using the methods of Kruskal we derive the *adiabatic invariant* for the continuous system (A2) ($\alpha = 1$) which has nearly periodic solutions. To order ϵ the adiabatic invariant is identical with one of the exact invariants and gives no new information about the system.

1. INTRODUCTION

THE use of invariant or conserved quantities of physical systems may provide one with a deeper understanding of their dynamics. The gross features of the behavior of a physical system can sometimes be visualized in terms of invariant quantities as, for example, energy, momentum, etc., and adiabatically invariant quantities as, for example, the magnetic moment of a charged particle in a magnetic field. The study of the properties of one-dimensional nonlinear continua has been augmented, in recent years, by formulating the problem in terms of coupled partial differential equations expressed in *conservation form.*[1-5] For example, one uses the equations of conservation of mass, momentum, and energy to describe inviscid hydrodynamics problems.

Loewner[3] and Rozhdestvenskii[4] have considered the problem of transforming a set of partial differential equations of first order to conservation form. Loewner was mainly interested in elliptic systems and applied his results to a stationary, two-dimensional, compressible fluid-flow boundary-value problem. Rozhdestvenskii treated hyperbolic systems of partial differential equations and applied his results to the problem of the uniqueness of generalized solutions of the Cauchy problem. They both showed that certain *pairs* of coupled first-order partial differential equations could be replaced by an infinite set of equations in conservation form; and Rozhdestvenskii showed that systems of three or more partial differential equations generally cannot be represented by equivalent sets of equations in conservation form.

In the present work, we approach the problem differently, and in Sec. 2 we show that one can derive an infinite family of *exact invariants* for a system of two first-order nonlinear partial differential equations. These invariants are spatial integrals of linearly independent "invariant densities." Each is a solution of the linear, second-order, partial differential equation obtained by applying the *hodograph transformation*[6,7] to the given pair

[1] R. Courant, *Methods of Mathematical Physics, II (Partial Differential Equations)* (Interscience Publishers, Inc., New York, 1962). Appendix 2 of Chap. II and Chap. V, Paragraph 9.

[2] P. D. Lax, Commun. Pure Appl. Math. 10, 537 (1957), and "Nonlinear Hyperbolic Systems of Conservation Laws", published in *Nonlinear Problems*, edited by R. E. Langer (University of Wisconsin Press, Madison, Wisconsin, 1963), p. 3.

[3] C. Loewner, J. Ratl. Mech. Anal. 2, 537, (1953). The authors are indebted to P. Lax for this reference.

[4] B. L. Rozhdestvenskii, Soviet Math. Uspekhi 2, 53, (1959). See Secs. 6, 7, and 12.

[5] S. K. Godunov, Soviet Math. Doklady 2, 947 (1962).

[6] These invariants were found while working toward an understanding of the Fermi, Pasta, Ulam (FPU) phenomena. N. J. Zabusky, "Phenomena Associated with the Oscillations of a Nonlinear Model String" in *Proceedings of the Conference on Mathematical Models in the Physical Sciences*, S. Drobot, Ed. (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963), and M. D. Kruskal and N. J. Zabusky, J. Math. Phys. 5, 231 (1964).

[7] N. J. Zabusky, J. Math. Phys. 3, 1028 (1962) Paragraph 3. The hodograph transformation interchanges the roles of the independent and dependent variables and in our problems linearizes the equations of motion.

of equations. We then show that the invariant densities satisfy equations in conservation form, thereby giving an alternate derivation of a result of Rozhdestvenskii. In Sec. 3 we apply these general considerations to a one-dimensional nonlinear hyperbolic equation which describes, for example, longitudinal wave propagation along an elastic bar.[8] We show that each invariant density contains a *finite* number of terms of the form $y_x^l y_t^m$, where $y_x = u$ is the elastic strain and $y_t = v$ is the time rate of change of the displacement. In Table I we separate the invariant densities into two classes and display the first six of each. We note that the second entries in class "a" and class "b" are proportional to the energy density and momentum density, respectively.

In Sec. 4 we use E. Noether's theorem (1918) to illustrate the relationship between equations in conservation form and continuous transformation groups which leave the action functional invariant. Finally, in Sec. 5, we calculate the *adiabatic invariant* of this *nearly periodic continuous* system to an appropriate low order and show that it is identical to the exact invariant to that order. In fact, we discovered the exact invariants by using the methods developed by Kruskal[9] to calculate the adiabatic invariant.

2. THE PARTIAL DIFFERENTIAL EQUATIONS FOR THE EXACT INVARIANTS

A. Preliminaries

To start, we will consider $n + 1$ coupled first-order partial differential equations

$$u_{i;t} = A_{ij}u_{j;x} \qquad (i, j = 0, 1, \cdots, n), \tag{2.1}$$

where

$$u_{i;t} \equiv \partial u_i(x, t)/\partial t \quad \text{and} \quad u_{i;x} \equiv \partial u_i(x, t)/\partial x, \tag{2.2}$$

and where the summation convention is understood for repeated subscripts. For this system to yield hyperbolic equations, the matrix

$$A_{ij} = A_{ij}(u_0, u_1, \cdots, u_n)$$

must have distinct real eigenvalues. We consider an initial-value problem for a function periodic in space, whose fundamental period is the interval $-1 \leq x < 1$.

B. The Temporal Invariant

We now consider the integral

$$I = \int_{-1}^{1} dx\, T(u_0, u_1, \cdots, u_n). \tag{2.3}$$

If I is a temporal invariant of the physical system, then

$$0 = \frac{dI}{dt} = \int_{-1}^{1} T_{u_i}u_{i;t}\, dx = \int_{-1}^{1} T_{u_i}A_{ij}u_{j;x}\, dx, \tag{2.4}$$

where we have used the equations of motion (2.1) for $u_{i;t}$. Substituting $du_j = u_{j;x}dx$ we obtain

$$\int_{-1}^{1} T_{u_i}A_{ij}\, du_j = 0. \tag{2.5}$$

The integral vanishes if the integrand is a perfect differential,

$$T_{u_i}A_{ij}\, du_j = dV = V_{u_j}\, du_j, \tag{2.6}$$

where V is a function of the periodic functions u_j, so that

$$V\,|_{x=-1} = V\,|_{x=1}. \tag{2.7}$$

We thus obtain the set of $n + 1$ coupled equations for the dependent variables T and V

$$V_{u_j} = T_{u_i}A_{ij} \qquad (j = 0, 1, \cdots, n). \tag{2.8}$$

We could also have chosen the range of integration $-\infty \leq x \leq +\infty$ with $V|_{x=-\infty} = V|_{x=\infty}$.

We can eliminate V from (2.8) by cross-differentiation and obtain $\frac{1}{2}n(n + 1)$ coupled second-order partial differential equations of the form

$$(T_{u_i}A_{ij})_{u_k} = (T_{u_i}A_{ik})_{u_j} \qquad (j < k), \tag{2.9}$$

for the unknown temporal invariant densities T.

One could readily discuss the properties of the *linear* equations given in (2.9) as a function of properties of the matrix A_{ij}; however, we will go directly to the problem at hand and consider a 2×2 matrix ($n = 1$) in *diagonal* ($A_{01} = A_{10} = 0$) or *zerodiagonal* ($A_{00} = A_{11} = 0$) form. For distinctness we write $\bar{T}$ instead of T for the diagonal case.

If the system of Eqs. (2.1) is hyperbolic and the matrix A has distinct eigenvalues, then it can always be reduced to a diagonal or "normal" form by introducing the Riemann invariants.[10] With the diagonal form and the given form one can always obtain a zerodiagonal form. In the present work we find that the zerodiagonal form is easier for computation in special cases and that the diagonal form is easier for computation in more general cases and permits a concise presentation of results. The two forms are of course equivalent.

Thus, for $n = 1$, Eq. (2.9) yields

Diagonal:

$$\bar{T}_{u_0u_1}(A_{00} - A_{11}) + \bar{T}_{u_0}A_{00;u_1} - \bar{T}_{u_1}A_{11;u_0} = 0; \tag{2.10}$$

[8] R. Courant, Ref. 1, Chap. V, Paragraph 3.
[9] M. Kruskal, J. Math. Phys. 3, 806 (1962).
[10] R. Courant, Ref. 1, Chap. V, Paragraph 2. Also, P. D. Lax, J. Math. Phys. 5, 611 (1964).

Zerodiagonal:

$$T_{u_0u_0}A_{01} - T_{u_1u_1}A_{10} + T_{u_0}A_{01;u_0} - T_{u_1}A_{u_110;} = 0. \tag{2.11}$$

Equations (2.10) *and* (2.11) *are identical with the "t" hodograph equations*[7] *obtained by applying the hodograph transformation to the original partial differential equations.* Hence, the exact temporal invariants are obtained by integrating solutions of the "t" hodograph equation over the fundamental period $-1 \leq x < 1$.

C. Partial Differential Equations in Conservation Form for the Temporally Invariant Densities

For the linear system described by the wave equation

$$y_{tt} = y_{xx},$$

with $u_1 = y_t$ and $u_0 = y_x$, the energy density is

$$\tfrac{1}{2}(y_t^2 + y_x^2) = \tfrac{1}{2}(u_1^2 + u_0^2).$$

It is obvious that the right side of this equation satisfies the zerodiagonal form, $T_{u_0u_0} - T_{u_1u_1} = 0$.

The Lagrangian density for the linear system

$$\mathcal{L} = \tfrac{1}{2}(y_t^2 - y_x^2) = \tfrac{1}{2}(u_1^2 - u_0^2),$$

satisfies the relations

$$\partial\mathcal{L}_{u_0}/\partial t + \partial\mathcal{L}_{u_1}/\partial x = 0,$$

and

$$\partial\mathcal{L}_{u_1}/\partial t + \partial\mathcal{L}_{u_0}/\partial x = 0,$$

where the first equation is an identity when expressed in terms of y. This suggests that we inquire whether also the solutions $\bar{T}$ or T will satisfy an equation of the conservation form

$$-(\partial/\partial t)(\alpha_0 T_{u_0} + \alpha_1 T_{u_1}) + (\partial/\partial x)(\beta_0 T_{u_0} + \beta_1 T_{u_1}) = 0, \tag{2.12}$$

where α_0, α_1, β_0, and β_1 are functions of u_0 and u_1. Equation (2.12) has the form of the Euler–Lagrange equation for a certain class of physical systems. We expand (2.12) and for the diagonal case we replace $u_{0;t}$ by $A_{00}u_{0;x}$ and $u_{1;t}$ by $A_{11}u_{1;x}$. Gathering terms we obtain

$$0 = [-(\alpha_0\bar{T}_{u_0} + \alpha_1\bar{T}_{u_1})_{u_0}A_{00} + (\beta_0\bar{T}_{u_0} + \beta_1\bar{T}_{u_1})_{u_0}]u_{0;x} + [-(\alpha_0\bar{T}_{u_0} + \alpha_1\bar{T}_{u_1})_{u_1}A_{11} + (\beta_0\bar{T}_{u_0} + \beta_1\bar{T}_{u_1})_{u_1}]u_{1;x}. \tag{2.13}$$

Since $u_{0;x}$ and $u_{1;x}$ are locally independent of u_0 and u_1 and of each other, each bracket in Eq. (2.13) must vanish identically. This vanishing should be a consequence of (2.10), so each bracket should be a multiple of the left side of (2.10), say by the factors $R_0(u_0, u_1)$ and $R_1(u_0, u_1)$, respectively. By carrying out the differentiations and comparing, we obtain the conditions which the α and β must satisfy:

Conditions Imposed on the Coefficient of $u_{0;x}$		Conditions Imposed on the Coefficient of $u_{1;x}$	
$\beta_0 - \alpha_0 A_{00} = 0$	(a)	$\beta_1 - \alpha_1 A_{11} = 0$	(e)
$\beta_1 - \alpha_1 A_{00} = (A_{00} - A_{11}) R_0$	(b)	$\beta_0 - \alpha_0 A_{11} = (A_{00} - A_{11}) R_1$	(f)
$\beta_{0;u_0} - \alpha_{0;u_0} A_{00} = A_{00;u_1} R_0$	(c)	$\beta_{1;u_1} - \alpha_{1;u_1} A_{11} = - A_{11;u_0} R_1$	(g)
$\beta_{1;u_0} - \alpha_{1;u_0} A_{00} = - A_{11;u_0} R_0$	(d)	$\beta_{0;u_1} - \alpha_{0;u_1} A_{11} = A_{00;u_1} R_1$	(h)

(2.14)

Note that $A_{00} \not\equiv A_{11}$, for otherwise the original system of equations would be degenerate.

Substituting (2.14a) into (2.14f) and (2.14e) into (2.14b) we find

$$R_0 = -\alpha_1 \quad \text{and} \quad R_1 = \alpha_0. \tag{2.15}$$

Substituting (2.14e) into (2.14d) and (2.14a) into (2.14h) we find

$$\alpha_{1;u_0}(A_{11} - A_{00}) = 0 \quad \text{and} \quad -\alpha_{0;u_1}(A_{11} - A_{00}) = 0. \tag{2.16}$$

Thus, we conclude

$$\alpha_0 = \alpha_0(u_0) \quad \text{and} \quad \alpha_1 = \alpha_1(u_1). \tag{2.17}$$

Substituting (2.14a) into (2.14c) and (2.14e) into (2.14g) we obtain

$$\alpha_0 A_{00;u_0} = -\alpha_1 A_{00;u_1} \quad \text{and} \quad \alpha_1 A_{11;u_1} = -\alpha_0 A_{11;u_0}. \tag{2.18}$$

Thus, we see that

$$-\frac{\alpha_0}{\alpha_1} = \frac{A_{00;u_1}}{A_{00;u_0}} = \frac{A_{11;u_1}}{A_{11;u_0}},$$

or $A_{00;u_0}A_{11;u_1} - A_{00;u_1}A_{11;u_0} = 0$, so that A_{00} and

A_{11} *must* be functionally related if we are to be able to satisfy the imposed conditions. For example, if

$$\alpha_0 = -\alpha_1 = 1, \text{ then } \beta_0 = A_{00}, \text{ and } \beta_1 = A_{11}, \tag{2.19}$$

and hence

$$-(\bar{T}_{u_0} - \bar{T}_{u_1})_t + (A_{00}\bar{T}_{u_0} - A_{11}\bar{T}_{u_1})_x = 0, \tag{2.20}$$

provided that Eq. (2.18) is satisfied.

For the zerodiagonal case, we proceed in a similar fashion and find the following imposed conditions:

Condition Imposed on the Coefficient of $u_{0;x}$		Conditions Imposed on the Coefficient of $u_{1;x}$	
$\beta_0 = -A_{01} R_0$	(a)	$\beta_1 = -A_{10} R_1$	(f)
$\beta_1 = \alpha_0 A_{10}$	(b)	$\beta_0 = \alpha_1 A_{01}$	(g)
$\alpha_1 A_{10} = -A_{10} R_0$	(c)	$\alpha_0 A_{01} = -A_{01} R_1$	(h)
$\beta_{0;u_0} - \alpha_{0;u_1} A_{10} = -A_{10;u_0} R_0$	(d)	$\beta_{1;u_1} - \alpha_{1;u_0} A_{01} = -A_{10;u_1} R_1$	(i)
$\beta_{1;u_0} - \alpha_{1;u_1} A_{10} = A_{10;u_1} R_0$	(e)	$\beta_{0;u_1} - \alpha_{0;u_0} A_{01} = A_{01;u_0} R_1.$	(j)

(2.21)

From (2.21c) and (2.21h) we see that

$$R_0 = -\alpha_1 \quad \text{and} \quad R_1 = -\alpha_0 \tag{2.22}$$

and thus (2.21a) is consistent with (2.21g) and (2.21f) is consistent with (2.21b). Furthermore, if (2.21a) is substituted into (2.21d) and (2.21f) into (2.21i), they yield the same relation

$$\alpha_{1;u_0} A_{01} = \alpha_{0;u_1} A_{10}. \tag{2.23}$$

Substituting (2.21f) into (2.21e) and (2.21a) into (2.21j) yields two relations,

$$\begin{aligned} (\alpha_0 A_{10})_{u_0} &= \alpha_{1;u_1} A_{10} - \alpha_1 A_{10;u_1}, \\ (\alpha_1 A_{01})_{u_1} &= \alpha_{0;u_0} A_{01} - \alpha_0 A_{01;u_0}. \end{aligned} \tag{2.24}$$

For example, if $A_{01} = 1$ and $A_{10;u_1} = 0$ and if we take $\alpha_0 = \beta_1 = 0$ and $\alpha_1 = \beta_0 = 1$, then all the above relations are satisfied and the conservation equation becomes

$$-(T_{u_1})_t + (T_{u_0})_x = 0. \tag{2.25}$$

Thus, we have found under what conditions linear combinations of the functional partial derivatives of the invariant density T or $\bar{T}$ satisfy partial differential equations in conservation form.

D. The Spatial Invariant

In Sec. 2B we emphasized the temporally invariant densities because we had the initial-value problem in mind. For boundary-value problems, in which u_i and $u_{i;x}$ are given at the same x and periodic solutions in t are required, the spatial invariant densities X are natural analogs. If J is a spatial invariant,

$$J = \int_{-1}^{1} dt\, X(u_0, u_1, \cdots, u_n), \tag{2.26}$$

then $dJ/dx = 0$ or

$$0 = \int_{-1}^{1} X_{u_i} u_{i;x}\, dt = \int_{-1}^{1} X_{u_i} (A^{-1})_{ij} u_{j;t}\, dt, \tag{2.27}$$

where the integration is over a periodic time interval. The matrix $(A^{-1})_{ij}$ is the inverse of A_{ij}, so that

$$(A^{-1})_{ij} A_{jk} = \delta_{ik}. \tag{2.28}$$

Equation (2.27) is satisfied identically if the spatial invariant densities satisfy

$$[X_{u_i}(A^{-1})_{ij}]_{u_k} = [X_{u_i}(A^{-1})_{ik}]_{u_j}, \qquad j < k. \tag{2.29}$$

This is the analog of (2.9). For $n = 1$, we obtain the diagonal case from (2.10) by replacing A_{00} by $1/A_{00}$, A_{11} by $1/A_{11}$, and $\bar{T}$ by $\bar{X}$, and the zerodiagonal case from (2.11) by replacing A_{01} by $1/A_{10}$, A_{10} by $1/A_{01}$, and T by X.

Diagonal:

$$\bar{X}_{u_0 u_1}\left(\frac{1}{A_{00}} - \frac{1}{A_{11}}\right) + \bar{X}_{u_0}\left(\frac{1}{A_{00}}\right)_{u_1} - \bar{X}_{u_1}\left(\frac{1}{A_{11}}\right)_{u_0} = 0; \tag{2.30}$$

Zerodiagonal:

$$X_{u_0 u_0}\left(\frac{1}{A_{10}}\right) - X_{u_1 u_1}\left(\frac{1}{A_{01}}\right) + X_{u_0}\left(\frac{1}{A_{10}}\right)_{u_0} - X_{u_1}\left(\frac{1}{A_{01}}\right)_{u_1} = 0. \tag{2.31}$$

Equations (2.30) and (2.31) are identical with the "x" hodograph equation[7] obtained by applying the hodograph transformation to the original partial differential equations, (2.1) with $n = 1$. X will also satisfy an equation in conservation form. This equation is obtained in an entirely analogous manner to that given in the previous section.

The polynomial solutions of the temporal invariant density equations (2.10) and (2.11) are given in Sec. 3. In an analogous manner one can determine the spatial polynomial invariant densities. For a special case, we will exhibit the second spatial invariant density, that is the one which is analogous

to the energy density of the system. If we multiply the hyperbolic equation

$$y_{tt} - y_{xx}(1 + \epsilon y_x) = 0 \tag{2.32}$$

by y_x and integrate over the *temporal* interval where the solution is periodic, we obtain

$$0 = \int_{-1}^{1} dt(y_x y_{tt} - y_x y_{xx} - \epsilon y_x^2 y_{xx}) = -\frac{d}{dx}\int_{-1}^{1} dt(\tfrac{1}{2}y_t^2 + \tfrac{1}{2}y_x^2 + \tfrac{1}{3}\epsilon y_x^3).$$

Thus,

$$X_2(y_x, y_t) = \tfrac{1}{2}y_t^2 + \tfrac{1}{2}y_x^2 + \tfrac{1}{3}\epsilon y_x^3 \tag{2.33}$$

is the second spatial invariant density and it differs from the energy or second temporal invariant density $\frac{1}{2}T_2(y_t, y_x) = \frac{1}{2}y_t^2 + \frac{1}{2}y_x^2 + \frac{1}{6}\epsilon y_x^3$ in the coefficient of the last term. This result can be obtained directly from (2.31), if we rewrite (2.32) as two coupled first-order equations

$$u_{0;t} = u_{1;x}, \qquad u_{1;t} = (1 + \epsilon u_0)u_{0;x}, \tag{2.34}$$

where $u_0 = y_x$ and $u_1 = y_t$. Thus, $A_{01} = 1$ and $A_{10} = 1 + \epsilon u_0$ and Eq. (2.31) becomes

$$[X_{u_0}/(1 + \epsilon u_0)]_{u_0} - X_{u_1 u_1} = 0. \tag{2.35}$$

It is evident that $X_2(u_0, u_1) = \frac{1}{2}u_1^2 + \frac{1}{2}u_0^2 + \frac{1}{3}\epsilon u_0^3$ is a solution of (2.35).

The precise meaning of these spatial invariants is unclear at present, but they also should enable one to determine more properties of the solution of the equation than one could do with only the temporal invariants. It should be emphasized that these derivations are applicable to any set of hyperbolic equations which can be put in the form (2.1), and apply not only to elastic longitudinal wave propagation but also to comparable hydrodynamic and electrodynamic equations.

3. CALCULATION OF THE EXACT TEMPORAL INVARIANTS

A. The Hodograph-Transformed Equations, An Example

We will now find *polynomial* solutions for the exact temporal invariant of the diagonal and zero-diagonal partial differential equations. The two forms are equivalent, but in special cases one solution will be easier to derive than will the other.

Elsewhere,[6,7] we treated in detail properties of the wave equation,

$$y_{tt} = \Phi^2(y_x)y_{xx}, \qquad \Phi^2 = (1 + \epsilon y_x)^\alpha. \tag{3.1}$$

We can write (3.1) in diagonal or zerodiagonal form. The zerodiagonal form is obtained directly from (3.1) by introducing

$$u = y_x, \qquad v = y_t. \tag{3.2}$$

Hence,

$$u_t = v_x, \qquad v_t = \Phi^2(u)u_x, \tag{3.3}$$

and by identification with (2.1): $u_0 = u$, $u_1 = v$, $A_{00} = A_{11} = 0$, $A_{01} = 1$, $A_{10} = \Phi^2(u)$. These satisfy the conditions (2.21). The hodograph equation (2.11) that corresponds to (3.3) is

$$\Phi^2 T_{vv} - T_{uu} = 0, \tag{3.4}$$

which in our example is

$$(1 + \epsilon u)^\alpha T_{vv} - T_{uu} = 0. \tag{3.5}$$

The diagonal form is obtained by constructing the Riemann-invariant equations equivalent to (3.3), namely

$$r_t = \Phi r_x, \qquad s_t = -\Phi s_x, \tag{3.6}$$

where

$$\Phi = (1 + \epsilon u)^{\frac{1}{2}\alpha}$$

and

$$\begin{bmatrix} r \\ s \end{bmatrix} = \pm\tfrac{1}{2}v + \frac{1}{2}\int^u \Phi(u')\,du',$$

or

$$\begin{bmatrix} r \\ s \end{bmatrix} = \pm\tfrac{1}{2}v + \epsilon^{-1}(2 + \alpha)^{-1}(1 + \epsilon u)^{\frac{1}{2}(2+\alpha)}, \tag{3.7}$$

and u and v are given by (3.2). Note that (3.7) defines u as a function of $(r + s)$,

$$u = (1/\epsilon)\{(\tfrac{1}{2}\epsilon(2 + \alpha)(r + s))^{2/(2+\alpha)} - 1\}.$$

In this case we identify

$$\mu_0 = r,\ \mu_1 = s,\ A_{01} = A_{10} = 0,$$
$$A_{00} = -A_{11} = (\tfrac{1}{2}\epsilon(2 + \alpha)(r + s))^{\alpha/(2+\alpha)}.$$

These satisfy the conditions (2.14). The hodograph equation (2.10) that corresponds to (3.6) is

$$2\Phi\bar{T}_{rs} + \Phi_s\bar{T}_r + \Phi_r\bar{T}_s = 0 \qquad (\Phi_r = \Phi_s), \tag{3.8}$$

which in our case is the Euler–Poisson–Darboux equation,

$$\bar{T}_{rs} + [n/(r + s)](\bar{T}_r + \bar{T}_s) = 0, \tag{3.9}$$

where $\bar{T} = \bar{T}(r, s)$ and $n = \alpha/2(2 + \alpha)$. It will be easier to deal with (3.9) than with (3.5) when α is nonintegral or large.

B. Polynomial Solutions for the Zerodiagonal Representation with $\alpha = 1$

The solution of (3.9) or (3.5) for boundary conditions given in either the r, s, or the u, v plane

TABLE I. The polynomial invariant densities and their u partial derivatives for $(1 + \epsilon u)\, T_{vv} - T_{uu} = 0$.

"a" Invariant Density

$$T_1 = v$$
$$T_2 = v^2 + u^2 + \epsilon[u^3/3]$$
$$T_3 = v^3 + 3vu^2 + \epsilon[vu^3]$$
$$T_4 = v^4 + 6v^2u^2 + u^4 + \epsilon[2v^2u^3 + 4u^5/5] + \epsilon^2[2u^6/15]$$
$$T_5 = v^5 + 10v^3u^2 + 5vu^4 + \epsilon[10v^3u^3/3 + 4vu^5] + \epsilon^2[2vu^6/3]$$
$$T_6 = v^6 + 15v^4u^2 + 15v^2u^4 + u^6 + \epsilon[5v^4u^3 + 12v^2u^5 + 9u^7/7] + \epsilon^2[2v^2u^6 + \tfrac{1}{2}u^8] + \epsilon^3[u^9/18]$$

$$T_{1;u} = 0$$
$$T_{2;u} = 2\{u + \epsilon u^2/2\}$$
$$T_{3;u} = 3\{2vu + \epsilon[vu^2]\}$$
$$T_{4;u} = 4\{3v^2u + u^3 + \epsilon[3v^2u^2/2 + u^4] + \epsilon^2[u^5/5]\}$$
$$T_{5;u} = 5\{4v^3u + 4vu^3 + \epsilon[2v^3u^2 + 4vu^4] + \epsilon^2[4vu^5/5]\}$$
$$T_{6;u} = 6\{5v^4u + 10v^2u^3 + u^5 + \epsilon[5v^4u^2/2 + 10v^2u^4 + 3u^6/2] + \epsilon^2[2v^2u^5 + 2u^7/3] + \epsilon^3[u^8/12]$$

"b" Invariant Density

$$T_1 = u$$
$$T_2 = 2vu$$
$$T_3 = 3v^2u + u^3 + \epsilon[u^4/2]$$
$$T_4 = 4v^3u + 4vu^3 + \epsilon[2vu^4]$$
$$T_5 = 5v^4u + 10v^2u^3 + u^5 + \epsilon[5v^2u^4 + u^6] + \epsilon^2[5u^7/21]$$
$$T_6 = 6v^5u + 20v^3u^3 + 6vu^5 + \epsilon[10v^2u^4 + 6vu^6] + \epsilon^2[10vu^7/7]$$

$$T_{1;u} = 1$$
$$T_{2;u} = 2v$$
$$T_{3;u} = 3\{v^2 + u^2 + \epsilon[2u^3/3]\}$$
$$T_{4;u} = 4\{v^3 + 3vu^2 + \epsilon[2vu^3]\}$$
$$T_{5;u} = 5\{v^4 + 6v^2u^2 + u^4 + \epsilon[4v^2u^3 + 6u^5/5] + \epsilon^2[u^6/3]\}$$
$$T_{6;u} = 6\{v^5 + 10v^3u^2 + 5vu^4 + \epsilon[20v^2u^3/3 + 6vu^5] + \epsilon^2[5vu^6/3]\}$$

has attracted some interest.[7,11] Here, we shall be concerned with polynomial solutions in the two representations. We substitute

$$T = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} a_{l,m} u^l v^m \tag{3.10}$$

into (3.5) with $\alpha = 1$ and obtain

$$(1 + \epsilon u) \sum \sum m(m-1) a_{l,m} u^l v^{m-2} - \sum \sum l(l-1) a_{l,m} u^{l-2} v^m = 0,$$

where we require that

$$a_{l,m} = 0 \quad \text{for} \quad l \text{ or } m < 0. \tag{3.11}$$

We make each term bear the common factor $u^l v^m$ by appropriately shifting indices of summation, and obtain

$$\sum \sum u^l v^m [(m+2)(m+1)(a_{l,m+2} + \epsilon a_{l-1,m+2}) - (l+2)(l+1) a_{l+2,m}] = 0. \tag{3.12}$$

The recursion formula defined by (3.12) can be written as

$$a_{l+2,m} = \frac{(m+2)(m+1)}{(l+2)(l+1)} (a_{l,m+2} + \epsilon a_{l-1,m+2}),$$

or

$$a_{l+2,p-l} = \frac{(p-l+2)(p-l+1)}{(l+2)(l+1)} [a_{l,p-l+2} + \epsilon a_{l-1,p-l+2}], \tag{3.13}$$

where $p = l + m$. For each value of $p = p_0$ we have two *finite* polynomial series: the "a" series beginning with v^{p_0} and ending with

(constant) $\epsilon^{\frac{1}{2}p_0} u^{\frac{1}{2}p_0}$ for p_0 even,

(constant) $\epsilon^{\frac{1}{2}(p_0-1)} v u^{\frac{1}{2}(p_0-1)}$ for p_0 odd;

and the "b" series beginning with $p_0 u v^{p_0-1}$ and ending with

(constant) $\epsilon^{\frac{1}{2}(p_0-2)} v u^{\frac{1}{2}(3p_0-4)}$ for p_0 even,

(constant) $\epsilon^{\frac{1}{2}(p_0-1)} u^{\frac{1}{2}(3p_0-1)}$ for p_0 odd.

The even a series has $\frac{1}{8}(p_0+2)(p_0+4)$ terms and the even b series has $\frac{1}{8}p_0(p_0+2)$ terms. The odd a and odd b series have $\frac{1}{8}(p_0+1)(p_0+3)$ terms each. These results are readily obtained by observing the property of the recursion relation (3.13) in the (l, p) plane. For a given $p = p_0$ the sum of the integral exponents of u and v for each term $u^l v^m = u^l v^{p-l}$ fall on the boundary of and within the triangular region bounded by $p = p_0$ and:

$p = l$	and	$p = l/3 + p_0$,	a even;
$p = l + 1$	and	$p = l/3 + p_0$,	a odd;
$p = l + 1$	and	$p = (l-1)/3 + p_0$,	b even;
$p = l$	and	$p = (l-1)/3 + p_0$,	b odd.

Table I gives the polynomials and their u derivatives for values of p_0 up to 6. $T_{1,a}$ and $T_{1,b}$ are trivial. $T_{2,a}$ is twice the *energy density* of the non-

[11] E. T. Copson, Proc. Roy. Soc. (London) **A216**, 539 (1953); A. G. Mackie, J. Ratl. Math. Mech. 4, 733 (1955).

linear system and $T_{2,b}$ is twice the *momentum* density. At present a physical interpretation of the higher invariant densities has not been effected. One can establish that the polynomials satisfy the relations:

$$(T_n)_v = nT_{n-1}, \qquad \text{for } a \text{ and } b; \tag{3.14a}$$

$$(T_n)_u = n(n-1)\int_0^u (1+\epsilon u)T_{n-2}\,du, \text{ for } a; \tag{3.14b}$$

$$(T_n)_u = nv^{n-1} + n(n-1)\int_0^u (1+\epsilon u)T_{n-2}\,du, \quad \text{for } b. \tag{3.14c}$$

Furthermore, one observes that terms not involving ϵ, that is, polynomials derived from a linear wave equation, are obtained from the relation

$$\begin{bmatrix} T_{n,a} \\ T_{n,b} \end{bmatrix} = \tfrac{1}{2}(v+u)^n \pm \tfrac{1}{2}(v-u)^n. \tag{3.15}$$

Hence, the nonzero numerical coefficients of the terms not involving ϵ are those of the binomial expansion.

D. Polynomial Solutions for the Diagonal Representation

The polynomial invariant densities obtained from the diagonal equation apply to a more general case and can be expressed in a more concise form than those obtained from the zerodiagonal equation.

Equation (3.9) has a singularity at $r+s=0$ and, by analogy with the method of Frobenius for ordinary differential equations, we seek a power-series solution of the form

$$\bar{T} = \sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} a_{l,m}(r+s)^{\gamma+l}(r-s)^m, \tag{3.16}$$

where $a_{l,m}=0$ for l or $m<0$ and γ is to be determined as a solution of an indicial equation. We substitute (3.16) into (3.9) and for the convenience of the reader we exhibit separately the terms containing a_{00}, a_{01}, and a_{10}.

$$\begin{aligned} 0 &= (r+s)\bar{T}_{rs} + n(\bar{T}_r + \bar{T}_s) \\ &= a_{00}\gamma(\gamma+2n-1)(r+s)^{\gamma-1} + (r+s)^{\gamma-1} \\ &\times [a_{01}\gamma(\gamma+2n-1)(r-s) + a_{10}(\gamma+1)(\gamma+2n)(r+s)] \\ &+ \sum_l \sum_m{}' (r+s)^{\gamma+l+1}(r-s)^{m-2} \\ &\times [(\gamma+l+2)(\gamma+l+2n+1)a_{l+2,m-2} - m(m-1)a_{l,m}], \end{aligned} \tag{3.17}$$

where the prime indicates summation over all but the three terms that are separately exhibited. The vanishing of the coefficient of $a_{00}(r+s)^{\gamma-1}$ yields the *indicial equation*, and leads to two values of γ,

$$\gamma = 0 \quad \text{and} \quad \gamma = 1-2n = 2/(2+\alpha), \tag{3.18}$$

while the third term of (3.17) yields (at least if $n \geq 2$)

$$a_{10} = 0. \tag{3.19}$$

The recurrence relation for the remaining terms is obviously

$$a_{l+2,m-2} = \frac{m(m-1)}{(\gamma+l+2)(\gamma+l+2n+1)}\,a_{l,m} \qquad (l+m \geq 2), \tag{3.20}$$

or

$$a_{p-m+2,m-2} = \frac{m(m-1)}{(\gamma+p-m+2)(\gamma+p-m+2n+1)}\,a_{p-m,m} \qquad (p \geq 2). \tag{3.21}$$

Since (3.21) goes in steps of 2 and γ has two possible values, it might seem that, for a given order $p = l+m$, we would have twice as many invariants as we did for the zerodiagonal representation. The invariants sought have the form

$$\bar{T}_a = a_{0,p}(r+s)^{\gamma}(r-s)^p + \cdots + a_{p,0}(r+s)^{\gamma+p}, \tag{3.22a}$$

$$\bar{T}_b = a_{1,p-1}(r+s)^{\gamma+1}(r-s)^{p-1} + \cdots + a_{p-1,1}(r+s)^{\gamma+p-1}(r-s). \tag{3.22b}$$

[This is for even p; for odd p the final terms of (3.22a) and (3.22b) should be interchanged.] However, if we substitute $m = p+1$ into (3.21), we find that $a_{1,p-1}=0$, and so

$$0 = a_{1,p-1} = a_{3,p-3} = \cdots. \tag{3.23}$$

[$m = p+2$ does not give a similar result because the denominator in (3.21) then vanishes by (3.18).] That is, $\bar{T}_b = 0$, and we have only the same number of invariants as before. Consideration of the recurrence formula (3.21) shows that the polynomials in the diagonal representation are given by the formula

$$\bar{T}_p = \sum_{q=0}^{[\frac{1}{2}p]} g_q^{-1}\,\frac{(r+s)^{\gamma+2q}(r-s)^{p-2q}}{(p-2q)!}, \tag{3.24}$$

where $g_0 = 1$,

$$g_q = (\gamma+2)(\gamma+4)\cdots(\gamma+2q) \times (\gamma+1+2n)(\gamma+3+2n)\cdots(\gamma+2q-1+2n),$$

for $q > 1$, and [c] designates the integer part of c.

Substituting $r + s = (2/3\epsilon)(1 + \epsilon u)^{\frac{1}{2}}$ and $r - s = v$ into the expression for $\bar{T}_2$ as given in (3.24), we find

$$\bar{T}_2 \,|_{\gamma=0} = T_{2,a} + (u/\epsilon) + (1/3\epsilon^2),$$

and

$$\bar{T}_2 \,|_{\gamma=\frac{1}{2}} = (2/3\epsilon)^{\frac{1}{2}} \times [T_{2,a} + (\epsilon/3)T_{3,b} + (2u/3\epsilon) + (1/6\epsilon^2)].$$

In general, if one excludes constant and linear terms in u, one can express the invariant densities of the diagonal representation as linear combinations of the zerodiagonal invariant densities.

4. NOETHER'S THEOREM AND THE RELATION BETWEEN EQUATIONS IN CONSERVATION FORM AND CONTINUOUS TRANSFORMATION GROUPS

In 1918, Emmy Noether[12] published a theorem which affords a general method for deriving conservation laws and constructing temporally invariant functionals for physical systems or fields. That is, if one has a continuous transformation group which leaves the action functional invariant, one can derive conservation laws. In what follows, we use the reverse approach, and for the zerodiagonal case we will construct transformation groups from the conservation laws and invariants which we have derived in the previous sections. With these transformation groups, one should be able to obtain more information concerning the solutions of specific problems.

In the nomenclature of this paper, Noether's theorem[13] can be stated thus:

If the functional

$$A[y] = \int_R \mathfrak{L}(t, x, y, y_t, y_x)\, dt\, dx \tag{4.1}$$

is invariant under the family of transformations

$$\left.\begin{aligned} t \to t^* &= t + \eta\varphi_0(t, x, y, y_t, y_x) + o(\eta), \\ x \to x^* &= x + \eta\varphi_1(t, x, y, y_t, y_x) + o(\eta), \end{aligned}\right\} \tag{4.2}$$

$$y \to y^* = y + \eta\,\psi(t, x, y, y_t, y_x) + o(\eta) \tag{4.3}$$

as $\eta \to 0$ for an arbitrary region R in (t, x) space, then, on each extremal surface of $A[y]$,

$$(\bar{\psi}\,\partial\mathfrak{L}/\partial y_t + \mathfrak{L}\varphi_0)_t + (\bar{\psi}\,\partial\mathfrak{L}/\partial y_x + \mathfrak{L}\varphi_1)_x = 0, \tag{4.4}$$

where

$$\bar{\psi} = \psi - \varphi_0 y_t - \varphi_1 y_x. \tag{4.5}$$

We identify $\mathfrak{L}$ with the Lagrangian density, which in our case is

$$\mathfrak{L} = \tfrac{1}{2}(y_t^2 - y_x^2 - \tfrac{1}{3}\epsilon y_x^3) = \tfrac{1}{2}(v^2 - u^2 - \tfrac{1}{3}\epsilon u^3). \tag{4.6}$$

By "invariant under the family of transformation" we mean that the integral in (4.1) has the property

$$\int_{R^*} \mathfrak{L}\left(t^*, x^*, y^*, \frac{\partial y^*}{\partial t^*}, \frac{\partial y^*}{\partial x^*}\right) dt^*\, dx^* = \int_R \mathfrak{L}\left(t, x, y, \frac{\partial y}{\partial t}, \frac{\partial y}{\partial x}\right) dt\, dx. \tag{4.7}$$

We obtain an equation in conservation form, as (4.4), for each parameter η in the transformation group.

One can proceed in the opposite manner, namely, if we are given a conservation law then we can determine φ_0, φ_1, and ψ which characterize the near-identity transformation.

If we substitute (4.5) into (4.4) and expand, we obtain

$$(I_0)_t + (I_1)_x = 0, \tag{4.8}$$

where

$$\left.\begin{aligned} I_0 &= [\mathfrak{L}_v(\psi - u\varphi_1) + \varphi_0(\mathfrak{L} - v\mathfrak{L}_v)], \\ I_1 &= [\mathfrak{L}_u(\psi - v\varphi_0) + \varphi_1(\mathfrak{L} - u\mathfrak{L}_u)]. \end{aligned}\right\} \tag{4.9}$$

If we compare (4.8) with the equation in conservation form derived previously, namely (2.25),

$$(T_v)_t - (T_u)_x = 0, \tag{4.10}$$

we obtain two equations in the three unknowns ψ, φ_0, and φ_1,

$$I_0 = \mathfrak{L}_v(\psi - u\varphi_1) + \varphi_0(\mathfrak{L} - v\mathfrak{L}_v) = T_v, \tag{4.11}$$

$$I_1 = \mathfrak{L}_u(\psi - v\varphi_0) + \varphi_1(\mathfrak{L} - u\mathfrak{L}_u) = -T_u, \tag{4.12}$$

where we have ignored an arbitrary space-dependent function in (4.11) and an arbitrary time-dependent function in (4.12). Thus

$$\varphi_0 = \frac{T_v(\mathfrak{L} - u\mathfrak{L}_u) - uT_u\mathfrak{L}_v - \psi\mathfrak{L}\mathfrak{L}_v}{\mathfrak{L}(\mathfrak{L} - v\mathfrak{L}_v - u\mathfrak{L}_u)}, \tag{4.13}$$

$$\varphi_1 = \frac{-T_u(\mathfrak{L} - v\mathfrak{L}_v) + vT_v\mathfrak{L}_u - \psi\mathfrak{L}\mathfrak{L}_u}{\mathfrak{L}(\mathfrak{L} - v\mathfrak{L}_v - u\mathfrak{L}_u)}. \tag{4.14}$$

We see immediately that the energy and momentum conservation equations are consequences of the invariance of the action functional with respect to constant temporal and spatial displace-

[12] E. Noether, "Invariante Variationsprobleme," Kgl. Ges. Wiss. Nachr. Math.-Physik Kl. 2, 235 (1918).

[13] I. M. Gel'fand and S. V. Fomin, *Calculus of Variations*, translated by R. A. Silverman (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963). This well-written and superbly translated book contains a very lucid and rigorous account of Noether's theorems (Sec. 37.5) and their applicability to field theory problems (Secs. 37 and 38).

ments. If the Lagrangian of the system takes the form

$$\mathfrak{L} = \tfrac{1}{2}v^2 - F(u), \tag{4.15}$$

and if we set $\psi = \varphi_1 = 0$, $\varphi_0 =$ constant (temporal displacement), then (4.8) is the conservation equation for energy

$$0 = [\mathfrak{L} - v\mathfrak{L}_v]_t - [v\mathfrak{L}_u]_x = -\mathfrak{IC}_t + [vF_u]_x, \tag{4.16}$$

where $\mathfrak{IC} = \tfrac{1}{2}v^2 + F(u)$ is the energy density of the physical system. If we set $\psi = \varphi_0 = 0$ and $\varphi_1 =$ constant (spatial displacement), then (4.8) is the conservation equation for momentum

$$0 = -[u\mathfrak{L}_v]_t + [\mathfrak{L} - u\mathfrak{L}_u]_x = [-uv]_t + [\mathfrak{L} - u\mathfrak{L}_u]_x. \tag{4.17}$$

For Eq. (2.32) whose exact invariant densities are given by (3.10) and illustrated in Table I, we have

$$\mathfrak{L} = \tfrac{1}{2}(v^2 - u^2 - \tfrac{1}{3}\epsilon u^3), \quad \mathfrak{L}_v = v, \quad \mathfrak{L}_u = -(u + \tfrac{1}{2}\epsilon u^2), \tag{4.18}$$

and the denominator in (4.13) and (4.14) is

$$\begin{aligned} D &= \mathfrak{L}(\mathfrak{L} - v\mathfrak{L}_v - u\mathfrak{L}_u) \\ &= -\tfrac{1}{4}[v^4 - v^2(2u^2 + \epsilon u^3) + (u^2 + \tfrac{1}{3}\epsilon u^3)(u^2 + \tfrac{2}{3}\epsilon u^3)]. \end{aligned} \tag{4.19}$$

If T_{3a} is substituted into (4.13) and (4.14) and $\psi = 0$, then one easily finds that

$$\varphi_0 = -6 \quad \text{and} \quad \varphi_1 = 0, \tag{4.20}$$

which corresponds to the case of a constant temporal displacement. If T_{3b} is substituted into (4.13) and (4.14) and $\psi = 0$, then again one easily finds

$$\varphi_0 = 0 \quad \text{and} \quad \varphi_1 = -6, \tag{4.21}$$

which corresponds to the case of a constant spatial displacement. It is usually not possible to find a polynomial form for ψ which, when substituted into (4.13) and (4.14), results in the denominator being a factor of the numerator. Thus, φ_0 and φ_1 can be expressed only as rational functions of u and v.

5. ON THE RELATION BETWEEN THE EXACT INVARIANTS AND THE ADIABATIC INVARIANT

According to the general asymptotic theory presented by one of us,[9] there exists an *adiabatic invariant* for nearly recurrent Hamiltonian systems, expressible as a formal power series in ϵ. Now the solutions of the nonlinear partial differential equation (3.1) are nearly recurrent, so that even before we knew of the exact invariants, we sought an adiabatic invariant. To lowest order we found that the adiabatic invariant was the energy. In the next order, we obtained [after separating the contribution needed to complete the exact energy invariant—see (5.32)] a quantity which was almost invariant, namely the leading terms of T_{3b},

$$\int_{-1}^{1} dx(3v^2u + u^3).$$

We found that this could be made an exact invariant by addition of a term of order ϵ. With these results we were led to seek other invariants of the same general form and thereby came upon the procedure presented in previous sections of this paper.

In the previous work[9] [Eq. (B4)] the given physical system was described by a system of first-order ordinary differential equations $\mathbf{x}_s = \mathbf{f}(\mathbf{x})$, with all solutions nearly recurrent. The independent variable s represents the time and $\mathbf{f}$ is a vector function of the dependent variable $\mathbf{x}$. In the present work we apply some of the theory of Ref. 9 to (3.1) with $\alpha = 1$ and adapt the notation accordingly. The discrete vector index of Ref. 9 goes over into the continuous independent variable x of the present paper and summations correspondingly become integrations. The independent variable s becomes t and the dependent variable $\mathbf{x}$ (in a finite-dimensional space) becomes the pair of functions $y(x)$, $y_t(x) \equiv p(x)$ (in a function space). This theory leads to a special family of topological circles (called *rings*) in the function space, parametrized in a special way by an angle variable φ (here chosen to have period 2); each ring is a cyclic family of pairs of functions $y(x, \varphi)$, $p(x, \varphi)$.

The Hamiltonian for the present problem is

$$H = \int_{-1}^{1} d\xi\left(\frac{1}{2}p^2 + \frac{1}{2}y_\xi^2 + \frac{1}{6}\epsilon y_\xi^3\right), \tag{5.1}$$

where y_ξ is the Hamiltonian coordinate and $p(\xi)$ is its conjugate momentum. Hamilton's equations of motion are

$$p_t = -H_v = -\int_{-1}^{1} d\xi\left[y_\xi\,\partial_v y_\xi + \frac{1}{2}\epsilon y_\xi^2\,\partial_v y_\xi\right]$$

or

$$p_t = y_{\xi\xi} + \epsilon y_\xi y_{\xi\xi}, \tag{5.2}$$

where we have used

$$\partial_v y_\xi = \partial y_\xi(\xi)/\partial y(x) = -\delta'(x - \xi) \tag{5.3}$$

and then integrated by parts. Similarly,

$$y_t = H_p = \int_{-1}^{1} d\xi\, p\,\partial_p p = p. \tag{5.4}$$

The adiabatic invariant

$$J \equiv \oint d\varphi\, \mathbf{p}\cdot\mathbf{q}_\varphi$$

becomes

$$J \equiv \int_0^2 d\varphi \int_{-1}^{1} dx\, p(x;\varphi) y_\varphi(x;\varphi), \qquad (5.5)$$

where the φ integration is over a particular ring in the space of pairs of functions $y(x)$ and $p(x)$.

The Riemann invariants $r(x, t) \equiv r_+(x, t)$ and $s(x, t) \equiv r_-(x, t)$ introduced in Sec. 3, Eq. (3.7), are more convenient for analysis, as they permit one to replace the second-order equation by two symmetrical first-order partial differential equations. For the analysis to be followed we redefine them with a different additive constant as

$$\begin{aligned} r_\pm &= \tfrac{1}{2}\{\pm y_t + (2/3\epsilon)[(1+\epsilon u)^{\frac{3}{2}} - 1]\} \\ &= \tfrac{1}{2}\{\pm y_t + y_x + \tfrac{1}{4}\epsilon y_x^2 - \tfrac{1}{24}\epsilon^2 y_x^3 + O(\epsilon^3)\}, \end{aligned} \qquad (5.6)$$

and observe that they satisfy the coupled partial differential equations

$$\begin{aligned} (r_\pm)_t = \pm[1 &+ \tfrac{1}{2}\epsilon(r_+ + r_-) \\ &+ \tfrac{1}{4}\epsilon^2(r_+ + r_-)^2 + O(\epsilon^3)](r_\pm)_x. \end{aligned} \qquad (5.7)$$

We obtain a "more appropriate" set of dependent variables [denoted by $\mathbf{y}(\mathbf{x})$ and $v(\mathbf{x})$ in Sec. B5 of Ref. 9] by observing that we can eliminate the term $\pm(r_\pm)_x$ on the right side of each equation of (5.7) by transforming to an appropriate uniformly moving frame. This is readily accomplished by replacing the given variables $r_\pm(x, t)$ by the "more appropriate" variables

$$\rho_\pm(x, t) \equiv r_\pm(x \mp v, t) \quad \text{and} \quad v = t. \qquad (5.8)$$

We thereby obtain the equations

$$\begin{aligned} \rho_{\pm;t} &= \pm[\tfrac{1}{2}\epsilon\{\rho_\pm(x, t) + \rho_\mp(x \mp 2v, t)\} \\ &- \tfrac{1}{4}\epsilon^2\{\rho_\pm(x, t) + \rho_\mp(x \mp 2v, t)\}^2 + O(\epsilon^3)]\rho_{\pm;x}, \end{aligned} \qquad (5.9)$$

where we have used the relations

$$\begin{aligned} \rho_{\pm;t} &\equiv \partial_t r_\pm(x \mp v, t) = [\partial_t r_\pm(x, t) \mp \partial_x r_\pm(x, t)]_{x=x\mp v}, \\ \rho_{\pm;x} &\equiv \partial_x r_\pm(x \mp v, t) = [\partial_x r_\pm(x, t)]_{x=x\mp v}. \end{aligned} \qquad (5.10)$$

The subscript $x = x \mp v$ appended to the brackets indicates that the substitutions are made after the differentiations are performed. Equation (5.9) and $v_t = 1$ [corresponding to Eqs. (B13) of Ref. 9] show that the time derivative of $\boldsymbol{\rho} = (\rho_+, \rho_-)$ depends upon $\boldsymbol{\rho}$ and v.

We now seek "nice variables"

$$\mathbf{z} = \mathbf{Z}(\boldsymbol{\rho}, v) \quad \text{and} \quad \varphi = \Phi(\boldsymbol{\rho}, v),$$

(Sec. B6 of Ref. 9), namely those whose time derivatives can be expressed as functions of $\mathbf{z}$ *only*:

$$\mathbf{z}_t = \epsilon \mathbf{h}(\mathbf{z}), \qquad \varphi_t = \omega(\mathbf{z}). \qquad (5.11)$$

By definition a ring is generated by varying φ alone, so $\mathbf{z}$ is *constant* along a ring and hence labels it. That is, if $p(x, \varphi)$ and $y_\varphi(x, \varphi)$ in (5.5) are expressed in terms of $\mathbf{z}$, the φ integration is trivial. In the work that follows we assume that $r_\pm$, $\rho_\pm$, and $Z_\pm$ are *periodic functions* with zero mean over an interval of 2.

Now, if we expand the variables Φ, ω, $Z_\pm$, $h_\pm$ in power series in ϵ, for example,

$$Z_\pm = Z_\pm^{(0)} + \epsilon Z_\pm^{(1)} + \cdots,$$

and use the recursion relations given in Sec. B7 of Ref. 9, we can show that in lowest order

$$\varphi^{(0)} = \Phi^0 = \omega^{(0)} = v, \qquad (5.12)$$

$$h_\pm^{(0)} = \pm\tfrac{1}{2}\rho_\pm\rho_{\pm;x}, \qquad (5.13)$$

$$Z_\pm^{(0)} = \rho_\pm. \qquad (5.14)$$

Continuing to first order, we obtain

$$\varphi^{(1)} = \Phi^{(1)} = \omega^{(1)} = 0, \qquad (5.15)$$

$$Z_\pm^{(1)} = \mp\frac{1}{2}\left[\int_0^v \rho_\mp(x \mp 2v', t)\, dv'\right]\rho_{\pm;x}, \qquad (5.16)$$

or

$$Z_\pm^{(1)} = \tfrac{1}{4}[\bar{\rho}_\mp(x \mp 2v, t) - \bar{\rho}_\mp]\rho_{\pm;x}, \qquad (5.17)$$

where

$$\bar{f}(x) = \int_0^x f(x')\, dx' \quad \text{and} \quad \bar{f}(2) = 0. \qquad (5.18)$$

In Eqs. (5.13), (5.14), (5.16), and (5.17), the argument of $\rho_\pm$ and $Z_\pm$ is taken as (x, t) when not explicitly given.

Now, to evaluate the adiabatic invariant (5.5) we first express p and y_φ in terms of $Z_\pm$ and $\varphi = v$, perform the φ integration (where $Z_\pm$ is held constant), and then replace $Z_\pm$ by an equivalent expression in y_t and y_x. Using (5.17) and (5.14) we can write

$$\begin{aligned} \rho_\pm(x, t) = Z_\pm(x, t) &- \tfrac{1}{4}\epsilon[\bar{Z}_\mp(x \mp 2\varphi, t) - \bar{Z}_\mp(x, t)] \\ &\times Z_{\pm;x}(x, t) + O(\epsilon^2), \end{aligned} \qquad (5.19)$$

where we have set $v = \varphi$. Thus we obtain

$$\begin{aligned} y_t &= r_+(x) - r_-(x) = \rho_+(x + \varphi) - \rho_-(x - \varphi) \\ &= Z_+(x + \varphi) - Z_-(x - \varphi) \\ &- \tfrac{1}{4}\epsilon\{[\bar{Z}_-(x - \varphi) - \bar{Z}_-(x + \varphi)]Z_{+;x}(x + \varphi) \\ &- [\bar{Z}_+(x + \varphi) - \bar{Z}_+(x - \varphi)]Z_{-;x}(x - \varphi)\} + O(\epsilon^2). \end{aligned} \qquad (5.20)$$

In (5.20) we have omitted the second argument t and will continue to do so in the equations which follow. Similarly, to order ϵ we obtain

$$y_x = r_+(x) + r_-(x) - \tfrac{1}{4}\epsilon[r_+(x) + r_-(x)]^2 + O(\epsilon^2) \quad (5.21)$$

$$\begin{aligned} y_x = {} & Z_+(x+\varphi) + Z_-(x-\varphi) \\ & - \tfrac{1}{4}\epsilon[Z_+(x+\varphi) + Z_-(x-\varphi)]^2 \\ & - \tfrac{1}{4}\epsilon\{[\bar{Z}_-(x-\varphi) - \bar{Z}_-(x+\varphi)]Z_{+;x}(x+\varphi) \\ & + [\bar{Z}_+(x+\varphi) - \bar{Z}_+(x-\varphi)]Z_{-;x}(x-\varphi)\} \\ & + O(\epsilon^2). \end{aligned} \quad (5.22)$$

To obtain y_φ we integrate with respect to x and differentiate with respect to φ,

$$y_\varphi = \partial_\varphi\left[\int^x y_{x'}\, dx'\right].$$

We can ignore the constant in the x integration because it contributes merely an additive function of φ alone to y_φ and therefore nothing to (5.5), since p is a periodic function having zero mean. Thus the terms on the first two lines of (5.22) yield

$$\begin{aligned} & Z_+(x+\varphi) - Z_-(x-\varphi) - \frac{\epsilon}{4}[Z_+^2(x+\varphi) - Z_-^2(x-\varphi)] \\ & - \frac{\epsilon}{2}\partial_\varphi \int_0^x dx'\, Z_+(x'+\varphi)Z_-(x'-\varphi), \end{aligned} \quad (5.23)$$

and the terms within the braces yield

$$\begin{aligned} & -\tfrac{1}{4}\epsilon\,\partial_\varphi\Big\{[\bar{Z}_-(x-\varphi) - \bar{Z}_-(x+\varphi)]Z_+(x+\varphi) \\ & + [\bar{Z}_+(x+\varphi) - \bar{Z}_+(x-\varphi)]Z_-(x-\varphi) \\ & - \int_0^x dx'\,[Z_-(x'-\varphi) - Z_-(x'+\varphi)]Z_+(x'+\varphi) \\ & - \int_0^x dx'\,[Z_+(x'+\varphi) - Z_+(x'-\varphi)]Z_-(x'-\varphi)\Big\}, \end{aligned} \quad (5.24)$$

where we have integrated by parts. The last term of (5.23) cancels two of the integral terms of (5.24), and if we perform the φ differentiations, we obtain the result

$$\begin{aligned} y_\varphi = {} & Z_+(x+\varphi) - Z_-(x-\varphi) \\ & - \tfrac{1}{4}\epsilon[Z_+^2(x+\varphi) - Z_-^2(x-\varphi)] \\ & - \tfrac{1}{4}\epsilon\{[\bar{Z}_-(x-\varphi) - \bar{Z}_-(x+\varphi)]Z_{+;x}(x+\varphi) \\ & - [\bar{Z}_+(x+\varphi) - \bar{Z}_+(x-\varphi)]Z_{-;x}(x-\varphi)\} + O(\epsilon^2). \end{aligned} \quad (5.25)$$

Substituting (5.20) and (5.25) into (5.5) (and remembering that Z is constant when φ varies), we obtain the result

$$\begin{aligned} J = {} & 2\int_{-1}^{1} dx\,\{[Z_+^2(x) + Z_-^2(x)] \\ & - \tfrac{1}{4}\epsilon[Z_+^3(x) + Z_-^3(x) + Z_+^2Z_- + Z_-^2Z_+] + O(\epsilon^2)\}, \end{aligned} \quad (5.26)$$

where we have used the relations

$$\int_{-1}^{1} dx \int_0^2 d\varphi\, A(x+\varphi)B(x-\varphi) = 0,$$

$$\int_{-1}^{1} dx \int_0^2 d\varphi\, A(x \pm \varphi) = 2\int_{-1}^{1} dx\, A(x),$$

for any $A(x)$ and $B(x)$ periodic with zero mean over the interval of length 2.

We will now display the adiabatic invariant J in terms of partial derivatives of the dependent variable y. To obtain this result, we substitute for $Z_\pm$ the values given in (5.14) and (5.17), namely

$$Z_\pm = \rho_\pm + \tfrac{1}{4}\epsilon[\bar{\rho}_\mp(x \mp 2v) - \bar{\rho}_\mp(x)]\rho_{\pm;x} + O(\epsilon^2),$$

and obtain the integrand $j(x, t)$,

$$\begin{aligned} j(x,t) = {} & \rho_+^2 + \rho_-^2 - \tfrac{1}{4}\epsilon\{\rho_+^3 + \rho_-^3 + \rho_+^2\rho_- \\ & + \rho_-^2\rho_+ - 2\rho_+\rho_{+;x}[\bar{\rho}_-(x-2v,t) - \bar{\rho}_-] \\ & - 2\rho_-\rho_{-;x}[\bar{\rho}_+(x+2v,t) - \bar{\rho}_+]\} + O(\epsilon^2), \end{aligned} \quad (5.27)$$

where the argument of ρ_+ and ρ_- is (x, t) when not specified. We now express $\rho_\pm(x, t)$ in terms of $y_t(x, t)$ and $y_x(x, t)$ by means of (5.6),

$$\rho_\pm(x,t) = \tfrac{1}{2}[\pm y_t + y_x + \tfrac{1}{4}\epsilon y_x^2]_{x=x\mp v} + O(\epsilon^2). \quad (5.28)$$

When we substitute (5.28) into (5.27) and integrate from -1 to 1, the first four terms yield

$$\begin{aligned} & \int_{-1}^{1} dx\,[\rho_+^2 + \rho_-^2 - \tfrac{1}{4}\epsilon(\rho_+^3 + \rho_-^3)] \\ & = \int_{-1}^{1} dx\,[\tfrac{1}{2}(y_t^2 + y_x^2 + \tfrac{1}{2}\epsilon y_x^3) - \tfrac{1}{16}\epsilon(y_x^3 + 3y_t^2y_x)] \\ & + O(\epsilon^2), \end{aligned} \quad (5.29)$$

where the argument of y_t and y_x on the right side of (5.29) is (x, t). This follows because f is periodic, that is

$$\begin{aligned} \int_{-1}^{1} dx'\, f(x'+\alpha v, t) &= \int_{-1+\alpha v}^{1+\alpha v} dx\, f(x,t) \\ &= \int_{-1}^{1} dx\, f(x,t). \end{aligned}$$

In a similar manner, the last two terms of (5.27) contribute

$$\frac{1}{2}\epsilon\int_{-1}^{1} dx\ \{\rho_+\rho_{+;x}[\bar{\rho}_-(x-2v,t)-\bar{\rho}_-] + \rho_-\rho_{-;x}[\bar{\rho}_+(x+2v,t)-\bar{\rho}_+]\}$$
$$= \frac{\epsilon}{32}\int_{-1}^{1} dx\ \{[(y_t(x-v)+\bar{y}_x(x-v))^2]_x \times [-\bar{y}_t(x-v)+\bar{y}_x(x-v)+\bar{y}_t(x+v)-\bar{y}_x(x+v)] + [(-y_t(x+v)+y_x(x+v))^2]_x \times [\bar{y}_t(x+v)+\bar{y}_x(x+v)-\bar{y}_t(x-v)-\bar{y}_x(x-v)]\}. \tag{5.30}$$

If we integrate by parts, the terms with factors of two different arguments just cancel against the terms $\rho_+^2\rho_-$ and $\rho_-^2\rho_+$ in (5.27), and we are left with

$$-\frac{\epsilon}{16}\int_{-1}^{1} dx\ \{-y_xy_t^2+y_x^3\}. \tag{5.31}$$

If we combine (5.31) and (5.29) we obtain the final result

$$J = 2\int_{-1}^{1} dx\ \{[\tfrac{1}{2}(y_t^2+y_x^2+\tfrac{1}{3}\epsilon y_x^3)] - \tfrac{\epsilon}{24}[y_x^3+3y_xy_t^2]+O(\epsilon^2). \tag{5.32}$$

Thus the adiabatic invariant J gives no new information to order ϵ, since to this order it is expressible in terms of exact invariants. Indeed, this is historically how we discovered the exact invariants (beyond the trivial momentum and energy). It seems likely that J, up to any order, will be a combination of exact invariants.

6. SUMMARY AND CONCLUSIONS

A method has been given for deriving exact invariants of a class of nonlinear one-dimensional wave equations. Historically, we first calculated the *adiabatic invariant* for a special equation which has nearly periodic solutions and found that to order ϵ *it was an exact invariant.* This led us to search for higher-order forms of an exact invariant, and one of the main results of this paper followed, namely, that the *temporal invariant densities* were most easily derived as polynomial solutions of the *temporal hodograph equation.*

This hodograph equation was used previously by one of us to derive the exact solution to the same problem,[7] namely, (3.1) treated as an initial-value problem with periodic solutions in space. The exact solution of the hodograph equation and the perturbation solution previously developed by us[6] could only be expressed implicitly, that is, the independent variables were expressed in terms of the dependent variables. This implicit form is characteristic in the solution of nonlinear wave problems. Furthermore, in Sec. 4 we applied Noether's theorem and showed how to construct continuous transformation groups which leave the action functional invariant. These groups mix the dependent and independent variables. At present it is not clear how one uses these groups to determine properties of the solutions. It is interesting to observe that in the present work we have gone from a specific nonlinear interaction (or the equations of motion) to the exact invariants to the equations in conservation form. In practice, the reverse road is often travelled, namely invariance properties of a physical system are determined from an experiment and one seeks the nature of the interaction.

In some unpublished calculations we used the temporal invariants to construct "half" of the solution to an initial-value problem. That is, assuming that the basic period is separated into two regions by one maximum and one minimum, if we are given the solution in one region we can construct the solution in the other region from the values of the invariants.

Future work in this area should be directed to using properties of the temporal invariants and the continuous transformation groups they yield to construct properties of the solution to specific problems and in more than one spatial dimension. Furthermore, one should seek an algorithm for discretizing the continuous invariants such that they become the invariant or "almost" invariant quantities for the corresponding lattice or discrete chain of particles. To accomplish this it may be necessary to include higher spatial derivatives into the continuous equations before the correspondence between the discrete and continuous system can be made.

One-Dimensional Impenetrable Bosons in Thermal Equilibrium*

A. LENARD†
Institute for Advanced Study, Princeton, New Jersey
(Received 22 December 1965)

The simple relationship between the wavefunctions of a system of impenetrable one-dimensional bosons and impenetrable one-dimensional fermions is exploited to derive an expansion of the boson reduced density matrices in terms of the fermion reduced density matrices and vice versa. This expansion is independent of the statistical ensemble used, and of the interparticle potential (subject to the impenetrability condition). The special case of zero impenetrability radius with no other forces is treated in detail, using the grand canonical ensemble. This leads to a natural generalization of a formula previously found for zero temperature only, for the one-particle reduced density matrix.

IT has been shown that there is a close relationship between the quantum mechanical problem of impenetrable one-dimensional bosons and that of fermions interacting through the same forces.[1] Specifically, this correspondence is the following. To every wavefunction ψ^b of the boson problem there corresponds a wavefunction ψ^f of the fermion problem and vice versa:

$$\psi^b(x_1, x_2, \cdots, x_N) = A(x_1, x_2, \cdots, x_N)\psi^f(x_1, x_2, \cdots, x_N), \quad (1)$$

where the function A is defined by

$$A(x_1, x_2, \cdots, x_N) = \prod_{1\le i<j\le N} \operatorname{sign}(x_j - x_i), \quad (2)$$

$$\operatorname{sign}(x) = \begin{cases} +1 & x > 0 \\ 0 & x = 0. \\ -1 & x < 0 \end{cases}$$

$A = +1$ or -1 when the permutation P, defined by the positions on the real line of $x_1, x_2, \cdots, x_N$, is even or odd, respectively. For (1) to hold, the following conditions must be satisfied:

(i) Condition on the potential energy: It must include a hard core of radius $a \ge 0$, but may otherwise be arbitrary.

(ii) Boundary conditions: The same conditions hold for ψ^b and ψ^f in the cases of "rigid wall enclosure" ($\psi = 0$ at the walls) or periodicity if N is odd. If N is even, periodicity for ψ^b is associated with antiperiodicity for the corresponding ψ^f (or vice versa).

Further, if ψ^b is an energy eigenstate so is the corresponding ψ^f, the eigenvalues being identical. It follows from this correspondence that any properties depending only on the absolute value of the wavefunction in configuration representation or on the energy eigenvalues are the same for the bosons and fermions.

Other properties, in particular those describing distribution of momenta, are quite different, of course. In a previous paper[2] the author has investigated the one-particle density matrix (closely related to the momentum distribution) for the ground state of the boson problem in the case when the only forces present are the hard cores (with radius $a = 0$). This case is distinguished by a special simplicity because the corresponding fermion problem refers to *free* particles. For this reason a rather detailed analysis was possible, including a discussion of the behavior of the momentum distribution at the origin of momentum space.

Let us now consider the following problem. Suppose an arbitrary statistical mechanical ensemble of bosons is given. We describe this by the probabilities $P_{N\alpha} \ge 0$ attached to the occurrence of states $\psi^b_{N\alpha}$. Here N refers to the total number of particles, and the discrete index α enumerates the different states that may occur. We assume normalization in a "box" Ω (an interval of length L say)

$$\int_\Omega dx_1 \cdots \int_\Omega dx_N \, |\psi^b_{N\alpha}|^2 = 1. \quad (3)$$

Now we consider the (reduced) density matrices ρ^b_n ($n = 1, 2, \cdots$) defined as follows:

$$(x_1, \cdots, x_n \,|\rho^b_n|\, x'_1, \cdots, x'_n) = \sum_{N=n}^{\infty} \sum_\alpha P_{N\alpha} \frac{N!}{(N-n)!} \int_\Omega dx_{n+1} \cdots \int_\Omega dx_N \times \psi^b_{N\alpha}(x_1, \cdots, x_n, x_{n+1}, \cdots, x_N) \times \psi^{b*}_{N\alpha}(x'_1, \cdots, x'_n, x_{n+1}, \cdots, x_N). \quad (4)$$

* This work was supported by the National Science Foundation.

† On leave from the Plasma Physics Laboratory, Princeton University, Princeton, N. J.

[1] M. Girardeau, J. Math. Phys. 1, 516 (1960).

[2] A. Lenard, J. Math. Phys. 5, 930 (1964), hereafter referred to as I.

In a similar way we may define density matrices ρ_n^f for the *corresponding* fermion problem. This means that we consider the wavefunctions $\psi_{N\alpha}^f$ associated with $\psi_{N\alpha}^b$ via (1) and we average over a fermion ensemble using the *same* probabilities $P_{N\alpha}$. Now we raise the following question: How are the functions ρ_n^b and ρ_m^f $(n, m = 1, 2, \cdots)$ related?

When (1) is substituted into (4) then, in addition to the factors $\psi_{N\alpha}^f$ and $\psi_{N\alpha}^{f*}$ in the integrand, we also get the function

$$\begin{aligned} &A(x_1, \cdots, x_n, x_{n+1}, \cdots, x_N) \\ &\quad \times A(x_1', \cdots, x_n', x_{n+1}, \cdots, x_N) \\ &= A(x_1, \cdots, x_n)A(x_1', \cdots, x_n') \\ &\quad \times \prod_{i=n+1}^{N} \prod_{j=1}^{2n} \operatorname{sign}(x_i - y_j), \end{aligned} \tag{5}$$

where we denote by $y_1 < y_2 < \cdots < y_{2n}$ the numbers $x_1, \cdots, x_n, x_1', \cdots, x_n'$ arranged in order of increasing value. Let J denote the union of the disjoint intervals $(y_1, y_2), (y_3, y_4), \cdots, (y_{2n-1}, y_{2n})$, then the product

$$\prod_{j=1}^{2n} \operatorname{sign}(x - y_j) = \begin{cases} +1 & x \text{ not in } J, \\ -1 & x \text{ in } J. \end{cases} \tag{6}$$

Therefore the function (5) may be rewritten

$$A(x_1, \cdots, x_n)A(x_1', \cdots, x_n')(-1)^{\sigma(J)}, \tag{7}$$

where $\sigma(J)$ denotes the *number* of those integration variables among $x_{n+1}, x_{n+2}, \cdots, x_N$ which are in J. We get then, in a clear but abbreviated notation,

$$\begin{aligned} \rho_n^b = AA' \sum_{N=n}^{\infty} \sum_{\alpha} P_{N\alpha} \frac{N!}{(N-n)!} \\ \times \int_\Omega dx_{n+1} \cdots \int_\Omega dx_N (-1)^{\sigma(J)} \psi_{N\alpha}^f \psi_{N\alpha}^{f*}. \end{aligned} \tag{8}$$

Consider now any integral of the form

$$I = \int_\Omega dx_1 \cdots \int_\Omega dx_n \, (-1)^{\sigma(J)} f(x_1, \cdots, x_n), \tag{9}$$

where f is some symmetric function, J is a subdomain of Ω, and $\sigma(J)$ is the number of integration variables inside J. Obviously,

$$\begin{aligned} I = \sum_{m=0}^{n} \binom{n}{m} (-1)^m \int_J dx_1 \cdots \int_J dx_m \\ \times \int_{\Omega - J} dx_{m+1} \cdots \int_{\Omega - J} dx_n \, f, \end{aligned} \tag{10}$$

where $\Omega - J$ means the part of Ω outside J. Now we write

$$\int_{\Omega - J} dx_i = \int_\Omega dx_i - \int_J dx_i \quad (i = m+1, \cdots, n). \tag{11}$$

Therefore

$$\begin{aligned} I = \sum_{m=0}^{n} \binom{n}{m} (-1)^m \sum_{k=0}^{n-m} \binom{n-m}{k} (-1)^k \\ \times \int_J dx_1 \cdots \int_J dx_{m+k} \int_\Omega dx_{m+k+1} \cdots \int_\Omega dx_n \, f. \end{aligned} \tag{12}$$

The terms with fixed $m + k = j$ are summed first; this gives then

$$\begin{aligned} I = \sum_{j=0}^{n} \binom{n}{j} (-2)^j \int_J dx_1 \cdots \int_J dx_j \\ \times \int_\Omega dx_{j+1} \cdots \int_\Omega dx_n \, f. \end{aligned} \tag{13}$$

This formula will now be used to transform (8):

$$\begin{aligned} \rho_n^b = AA' \sum_{N=n}^{\infty} \sum_{\alpha} P_{N\alpha} \frac{N!}{(N-n)!} \sum_{j=0}^{N-n} \binom{N-n}{j} (-2)^j \\ \times \int_J dx_{n+1} \cdots \int_J dx_{n+j} \\ \times \int_\Omega dx_{n+j+1} \cdots \int_\Omega dx_N \, \psi_{N\alpha}^f \psi_{N\alpha}^{f*}. \end{aligned} \tag{14}$$

If now the order of summations is interchanged, it is recognized that the sum over N and α gives precisely the fermion density matrix ρ_{n+j}^f. We conclude then that

$$\begin{aligned} &(x_1, \cdots, x_n| \, \rho_n^b \, |x_1', \cdots, x_n') \\ &\quad = A(x_1, \cdots, x_n)A(x_1', \cdots, x_n') \\ &\quad \times \sum_{j=0}^{\infty} \frac{(-2)^j}{j!} \int_J dx_{n+1} \cdots \int_J dx_{n+j} \\ &\quad \times (x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+j}| \\ &\quad \times \rho_{n+j}^f \, |x_1', \cdots, x_n', x_{n+1}, \cdots, x_{n+j}). \end{aligned} \tag{15}$$

We repeat the definition of the region J: It is the union of the n *succeding alternate* intervals formed by the $2n$ points $x_1, \cdots, x_n, x_1', \cdots, x_n'$ on the real line.

Regarding this result we may make the following remarks:

(a) In deriving (15) we have made use only of the connection (1) between corresponding boson and fermion wavefunctions. Therefore it holds regardless of the nature of the interaction (so long as there are hard cores present).

(b) The nature of the ensemble over which the averaging is done is irrelevant. In particular, (15)

also holds if the density matrices are derived from a pure N-particle state (the ground state for instance). In such a case the sum contains only a finite number of nonvanishing terms, of course.

(c) There is no more reference in (15) to the interval Ω (the "box"). Hence, it holds also in the thermodynamic limit, interpreted in the usual way for an appropriate ensemble. The passage to the limit term-by-term can be justified under rather general conditions.

(d) The expansion (15) also holds in reverse, i.e., with the superscripts b and f interchanged. This follows from the same symmetry of (1).

In the following we make two special assumptions. First, we restrict attention to the case of "impenetrable but otherwise free" bosons. This means that no forces are present beside the hard cores. The hard-core radius is taken $a = 0$. In this case the corresponding fermion problem is that of completely free particles, the effect of the hard cores automatically being taken into account by the antisymmetry of wavefunctions.[1] Secondly, we assume a grand canonical ensemble over which the averaging (4) is defined.

A complete set of one-particle wavefunctions is taken to be

$$u_\nu(x) = \begin{cases} (2/L)^{\frac{1}{2}} \sin(\pi\nu x/L) & (\nu = 2, 4, \cdots), \\ (2/L)^{\frac{1}{2}} \cos(\pi\nu x/L) & (\nu = 1, 3, \cdots), \end{cases} \tag{16}$$

with the corresponding energy eigenvalues

$$\epsilon_\nu = \hbar^2\pi^2\nu^2/2mL^2 \qquad (\nu = 1, 2, 3, \cdots). \tag{17}$$

The wavefunctions satisfy the boundary conditions corresponding to "hard walls" $u_\nu(-\frac{1}{2}L) = u_\nu(\frac{1}{2}L) = 0$ and form an orthonormal system in the box $\Omega = (-\frac{1}{2}L, \frac{1}{2}L)$. The N-particle wavefunctions for the fermions are just Slater determinants

$$\psi^f_{N\alpha}(x_1, x_2, \cdots, x_n) = (N!)^{-\frac{1}{2}} \det_{1\le n,m\le N} u_{\nu_n}(x_m), \tag{18}$$

where the index α enumerates different sets $\{\nu_1, \nu_2, \cdots, \nu_N\}$ of distinct positive integers. The corresponding energy is

$$E_{N\alpha} = \frac{\hbar^2\pi^2}{2mL^2}(\nu_1^2 + \nu_2^2 + \cdots + \nu_N^2). \tag{19}$$

The grand canonical ensemble is defined by

$$P_{N\alpha} = \mathfrak{z}^{-1} z^N \exp(-E_{N\alpha}/kT), \tag{20}$$

where z is the fugacity, T the temperature, and

$$\mathfrak{z} = \mathfrak{z}(z, T, L) = \sum_{N=0}^{\infty} \sum_\alpha z^N \exp(-E_{N\alpha}/kT) \tag{21}$$

the grand partition function.

We now calculate ρ^f_n. The sum

$$z^N \sum_\alpha \exp(-E_{N\alpha}/kT) \times \psi^f_{N\alpha}(x_1, \cdots, x_N)\psi^{f*}_{N\alpha}(x'_1, \cdots, x'_N) \tag{22}$$

may be written as $(N!)^{-1}$ multiplied by a sum over all positive integers $\nu_1, \nu_2, \cdots, \nu_n$ *independently* of each other. If the definition of $\psi^f_{N\alpha}$ in the form of a determinant (18) is then substituted, one may interchange the sums over the ν_i with the sum over the $N!^2$ terms which arise by expanding the product of the two determinants. This results in the following expression for the quantity given in (22):

$$(N!)^{-1} \det_{1\le n,m\le N} f(x_n, x'_m), \tag{23}$$

where

$$f(x, y) \equiv z \sum_{\nu=1}^{\infty} \exp(-\epsilon_\nu/kT) u_\nu(x) u^*_\nu(y). \tag{24}$$

It is convenient to introduce a notation borrowed from the Fredholm theory of integral equations,

$$f\begin{pmatrix} x_1, x_2, \cdots, x_N \\ x'_1, x'_2, \cdots, x'_N \end{pmatrix} \equiv \det_{1\le n,m\le N} f(x_n, x'_m). \tag{25}$$

Using this notation, we have, for the grand partition function,

$$\begin{aligned} \mathfrak{z} &= \sum_{N=0}^{\infty} \sum_\alpha z^N \exp(-E_{N\alpha}/kT) \\ &= \sum_{N=0}^{\infty} z^N \sum_\alpha \exp(-E_{N\alpha}/kT) \\ &\quad \times \int_\Omega dx_1 \cdots \int_\Omega dx_N \, |\psi^f_{N\alpha}(x_1, \cdots, x_N)|^2 \\ &= \sum_{N=0}^{\infty} \frac{1}{N!} \int_\Omega dx_1 \cdots \int_\Omega dx_N \, f\begin{pmatrix} x_1, \cdots, x_N \\ x_1, \cdots, x_N \end{pmatrix} \equiv \mathfrak{D}_F. \end{aligned} \tag{26}$$

In the last step, we used again a notation from integral equation theory: $\mathfrak{D}_F$ is the Fredholm determinant belonging to the kernel $f(x, y)$ on the interval Ω, the subscript calling attention to the dependence on the kernel. In a similar way, we have

$$\begin{aligned} &\sum_{N=n}^{\infty} z^N \sum_\alpha \exp(-E_{N\alpha}/kT) \frac{N!}{(N-n)!} \\ &\times \int_\Omega dx_{n+1} \cdots \int_\Omega dx_N \, \psi^f_{N\alpha}(x_1, \cdots, x_n, x_{n+1}, \cdots, x_N) \\ &\times \psi^{f*}_{N\alpha}(x'_1, \cdots, x'_n, x_{n+1}, \cdots, x_N) = \sum_{N=n}^{\infty} \frac{1}{(N-n)!} \\ &\times \int_\Omega dx_{n+1} \cdots \int_\Omega dx_N \, f\begin{pmatrix} x_1, \cdots, x_n, x_{n+1}, \cdots, x_N \\ x'_1, \cdots, x'_n, x_{n+1}, \cdots, x_N \end{pmatrix} \\ &\equiv \mathfrak{D}_F\begin{pmatrix} x_1, \cdots, x_n \\ x'_1, \cdots, x'_n \end{pmatrix}. \end{aligned} \tag{27}$$

This is the Fredholm minor of order n belonging to the kernel $f(x, y)$. We have now

$$(x_1, \cdots, x_n|\ \rho_n^{\mathrm{f}}\ |x_1', \cdots, x_n') = \mathfrak{D}_{\mathrm{F}}\begin{bmatrix} x_1, \cdots, x_n \\ x_1', \cdots, x_n' \end{bmatrix} \Big/ \mathfrak{D}_{\mathrm{F}}. \tag{28}$$

This expression is the ratio of two infinite series. It may be further simplified by means of an identity from the theory of integral equations. Suppose we let $F(x, y)$ stand for the resolvent kernel belonging to the integral equation whose kernel is $f(x, y)$, i.e., $F(x, y)$ is the solution of

$$F(x, y) + \int_\Omega f(x, t)F(t, y)\, dt = f(x, y). \tag{29}$$

Then the identity[3] asserts that

$$\mathfrak{D}_{\mathrm{F}}\begin{bmatrix} x_1, \cdots, x_n \\ x_1', \cdots, x_n' \end{bmatrix} \Big/ \mathfrak{D}_{\mathrm{F}} = F\begin{bmatrix} x_1, \cdots, x_n \\ x_1', \cdots, x_n' \end{bmatrix}, \tag{30}$$

where the right side is interpreted in accordance with the notation (25).

To complete the calculation of ρ_n^{f} it remains then to determine $F(x, y)$. This is easily done by expanding $F(x, y)$ in the orthonormal system $u_\nu(x)u_{\nu'}(y)$ and solving for the coefficients from (29) and (24). One obtains

$$F(x, y) = \sum_{\nu=1}^{\infty} \frac{1}{1 + z^{-1} \exp(\epsilon_\nu/kT)}\, u_\nu(x)u_\nu(y). \tag{31}$$

Making use of the explicit form (16) for the eigenfunctions and (17) for the eigenvalues we can carry out the limit $L \to \infty$ when the sum turns into an integral

$$\lim_{L\to\infty} F(x, y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{ds\, \cos s(x - y)}{1 + z^{-1} \exp[(\hbar^2 s^2)/(2mkT)]}. \tag{32}$$

This is, of course, the Fourier transform of the Fermi momentum distribution for free particles, corresponding to the general relation between the single-particle density matrix and the momentum distribution. The identity following from (28) and (30),

$$(x_1, \cdots, x_n|\ \rho_n^{\mathrm{f}}\ |x_1', \cdots, x_n') = \det_{1\le i,j\le n} (x_i|\ \rho_1^{\mathrm{f}}\ |x_j') \tag{33}$$

is well known; we have presented the above derivation only because it is different from the usual ones found in the textbooks and brings out the connection to the identity (30) which, to the author's knowledge, has not been noted before.

Utilizing the general theorem expressed by (15), we have now shown that

$$\begin{aligned}(x_1, \cdots, x_n|\ \rho_n^{\mathrm{b}}\ |x_1', \cdots, x_n') &= A(x_1, \cdots, x_n)A(x_1', \cdots, x_n') \\ &\times \sum_{j=0}^{\infty} \frac{(-2)^j}{j!} \int_J dx_{n+1} \cdots \int_J dx_{n+j} \\ &\times F\begin{bmatrix} x_1, \cdots, x_n, x_{n+1}, \cdots, x_{n+j} \\ x_1', \cdots, x_n', x_{n+1}, \cdots, x_{n+j} \end{bmatrix},\end{aligned} \tag{34}$$

where the kernel $F(x, y)$ is just the single-particle density matrix for free fermions given by (32). We shall now consider the dependence on the parameters defining our ensemble, the temperature T, and fugacity z.

We notice, first of all, that the temperature T enters in a quite trivial way. Specifically, for fixed z but varying T, the functions ρ_n^{f} and therefore also ρ_n^{b} remain the same if distance is measured in an appropriate (T-dependent) unit. Therefore the only nontrivial dependence is on the parameter z. In this regard, the situation is quite like that for the free Fermi gas. We may call the limit of small z the case of no degeneracy, and the limit of large z the case of degeneracy.

The simpler of the two limiting cases is the one of no degeneracy. In this case we can expand the function F, given in (32), in a power series in z

$$\begin{aligned}F(x, y) &= \frac{z}{2\pi}\int_{-\infty}^{\infty} ds\, \cos s(x - y) \\ &\times \exp[-(\hbar^2 s^2)/(2mkT)] + O(z^2) \\ &= (z/\lambda) \exp(-\pi(x - y)^2/\lambda^2) + O(z^2),\end{aligned} \tag{35}$$

where by λ we denote the thermal de Broglie wavelength $(2\pi\hbar^2/mkT)^{\frac{1}{2}}$. This way the series (34) also generates a power series in z, the jth term beginning with $O(z^{n+j})$. Therefore, the dominant term is just the first one $j = 0$, and so

$$\begin{aligned}\rho_n^{\mathrm{b}} = AA'\rho_n^{\mathrm{f}} &= AA' \det_{1\le i,j\le n} \frac{z}{\lambda} \\ &\times \exp\{-[\pi(x_i - x_j')^2/\lambda^2]\}.\end{aligned} \tag{36}$$

In the limit of no degeneracy, the boson and fermion density matrices become identical (apart from the sign factor AA' demanded by the proper symmetry). The function (35) is of course just the Fourier transform of the "classical" or Maxwell–Boltzmann velocity distribution.

The opposite limit of extreme degeneracy is

[3] For a proof see W. A. Hurwitz, Bull. Am Math. Soc. **20**, 406 (1914).

handled in a well-known manner. There it is convenient to set $z = \exp(\mu/kT)$ where μ is the chemical potential per particle, and let $T \to 0$. In this limit

$$F(x, y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{ds\, \cos s(x-y)}{1 + \exp\{[(\hbar^2 s^2/2m) - \mu]/kT\}}$$
$$\to \frac{1}{2\pi}\int_{-(2m\mu)^{\frac{1}{2}}/\hbar}^{(2m\mu)^{\frac{1}{2}}/\hbar} ds\, \cos s(x-y)$$
$$= [1/\pi(x-y)] \sin\{[(2m\mu)^{\frac{1}{2}}/\hbar](x-y)\}. \quad (37)$$

The number density of particles is given by the diagonal element of the one-particle density matrix which is, in this limit $T \to 0$,

$$(0|\, \rho_1^b\, |0) = (0|\, \rho_1^f\, |0) = F(0,0) = (2m\mu)^{\frac{1}{2}}/\pi\hbar. \quad (38)$$

Thus, if we choose μ so that $(2m\mu)^{\frac{1}{2}} = \pi\hbar$, the number density is unity, and with

$$F(x, y) = \sin \pi(x-y)/\pi(x-y) \quad (39)$$

and

$$(0|\, \rho_1^b\, |\xi) = \sum_{j=0}^{\infty} \frac{(-2)^j}{j!}$$
$$\times \int_0^{\xi} dx_1 \cdots \int_0^{\xi} dx_j\, F\begin{bmatrix} 0 & x_1 & \cdots & x_j \\ \xi & x_1 & \cdots & x_j \end{bmatrix}, \quad (40)$$

we recover exactly the formula (88) of I.

We wish to make two remarks on this contact of the present paper with I. The first concerns the proper order of limiting processes when the "zero-temperature case" of a macroscopic system is under investigation from the standpoint of statistical mechanics. The general rule is always to let the system become infinitely large first (thermodynamic limit) and only afterwards should thermodynamic parameters tend to special limiting values if desired. Therefore, strictly speaking, the procedure of I was illegitimate inasmuch as there we have taken $T = 0$ (equivalent to looking at the pure ground state) before going to the thermodynamic limit. The correct procedure is contained in this paper where, for instance, (40) arises as a limit $L \to \infty$ first and then $T \to 0$. From this point of view it is gratifying that the results actually do not differ.

Our second remark concerns the discussion contained in Sec. 8 of I. The final formula (88) derived there was based on an extraneous limiting process and derived in the final analysis from the spin analogy of Lieb, Schultz, and Mattis. That derivation appears rather artificial and contrasts with the derivation in the present paper which is quite straightforward. In addition, our point of view brings out the reason for the appearance of the kernel (39): It is just the zero-temperature fermion density matrix which is needed in accordance with the general theorem (15). In our previous work, however, the really natural representation was the one given in Theorem 3. It was this representation which was derived from the Toeplitz determinant of Theorem 1. Now, the Toeplitz determinant representation was important because it allowed the use of Szegö's theorem for deriving a bound on the behavior of $(x|\, \rho_1^b\, |x')$ as $|x - x'| \to \infty$. The natural question is therefore whether a similar representation might not hold for $T > 0$ also. Quite possibly the answer is affirmative, but the author has been unable to derive such a representation. For this reason we do not have a good bound for ρ_1^b as $|x - x'| \to \infty$ when $T > 0$, although on physical ground there cannot be much doubt that the decrease is faster than in the ground state. In fact, it is likely that the decrease is exponential which implies that the momentum distribution is an analytic function, even in the neighborhood of zero momentum. If this is true, then the momentum dependence suggested in (78) of I arises only in the limit $T \to 0$.

ACKNOWLEDGMENTS

It is a pleasant duty to acknowledge financial support of the National Science Foundation, and the hospitality of the Institute for Advanced Study during the time this research was done.

Classical Field Theory and Gravitation in a de Sitter World*

P. ROMAN AND J. J. AGHASSI
Boston University, Department of Physics, Boston, Massachusetts
(Received 1 November 1965)

The algebra and the calculus pertaining to a certain class of representations of the de Sitter group is developed. This permits us to formulate covariant field equations in de Sitter space, and, in particular, to construct quantum mechanical equations of motion associated with particles of given spin. The gauge principle is invoked and a spin-two field emerges, which we identify with the gravitational field. Its coupling to sources is discussed and conservation laws are derived. The emerging nongeometric theory of gravitation is compared with both the Einstein–Riemann type and other previously proposed nongeometrical theories.

1. INTRODUCTION

HISTORICALLY speaking, theories of gravitation can be grouped into two main categories. In theories of the first category, gravitation is regarded as a field of force generated by and acting upon the material content of space–time. Space–time, in turn, is looked upon as a rigid and absolute superstructure. Thus, the motions, but not the metrical properties, of material objects are affected by gravitation. Furthermore, *all* fields are treated on equal footing, and differ only in their laws of propagation and in their interactions with their respective generating currents.

The second category of theories is epitomized by Einstein's magnificent theory of 1916. In this framework, gravitation is the manifestation of space–time's dynamical structure. The geometry is of the Riemannian type and is fully determined by the material content of the universe. In turn, both the motions and the metrical properties of material objects are determined by the curvature of space-time. Thus, gravitation assumes a *unique* role among all forces of nature.

In spite of its stunning beauty and a number of experimental verifications, there are certain difficulties inherent in the geometrical interpretation of gravitation. First of all, in these theories the boundary conditions for the field equations cannot be uniquely specified without an *a priori* knowledge of the topological properties of the geometry. However, in many problems and *par excellence* in cosmology, such knowledge is not available prior to solving the field equations themselves. Secondly, the lack of an obvious asymptotic symmetry leads to ambiguities in the definition of the energy-momentum tensor. Thirdly, because of its interaction with matter fields, gravitation itself is subject to quantization, and the fluctuations thereby introduced will be transferred to the metric structure of space-time, resulting in an ambiguous geometry in the small. Finally, the unique status of gravitation is bound to cause a feeling of uneasiness. In fact, with the now known large number of matter fields, a unified geometrical field theory appears hopeless.

In view of these difficulties connected with the geometrical interpretation of gravity, we feel that it may be worthwhile to reconsider the theories where gravitation appears as a field of force in a rigid space-time background. Of course, difficulties are abundant in this class of theories, too. To start with, we have to discard theories which give only an approximate description of gravity, such as Newton's theory, Nordström's Lorentz-scalar theory, and the linearized tensor theory. On the other hand, nonlinear theories in a Lorentz background and formally *equivalent* to Einstein's have been proposed with some success by Gupta[1] and Thirring.[2] Here, the nonlinearity arises from the coupling of the tensor field with its own energy-momentum tensor. The Lorentz metric is unobservable, and rods and clocks do attain Riemannian metric properties as a result of the formal combination of the Lorentz metric and the field tensor. However, the boundary conditions are specified with respect to the Lorentz background; and, in general, this prescription may become incompatible with the actual resulting Riemannian nature of the solutions.

Another class of proposals in this category (e.g., Weyl,[3] Utiyama,[4] Schwinger,[5] Sciama[6]) is based on the powerful non-Abelian gauge field concept of

* This research was sponsored by the U. S. Air Force under Grant No. AF-AFOSR-385-65.

[1] S. Gupta, Rev. Mod. Phys. **29**, 334 (1957).
[2] W. E. Thirring, Ann. Phys. **16**, 96 (1961).
[3] H. Weyl, Z. Physik **56**, 330 (1929).
[4] R. Utiyama, Phys. Rev. **101**, 1597 (1956).
[5] J. Schwinger, Phys. Rev. **130**, 1253 (1963).
[6] D. W. Sciama, Rev. Mod. Phys. **36**, 463 (1964).

Yang and Mills.[7] Here, along with the metric tensor, the affine connections assume a fundamental role. They are coupled to the angular momentum tensor, while the metric is coupled to the energy-momentum tensor. These theories go beyond the Riemannian theory, for the affine connections reduce to the Christoffel symbols only when we deal with coherent, spinless matter. The gauge theories suffer from the same shortcoming as the previously discussed class: The metric is known only after a complete solution is found.

The main purpose of the present paper is to propose a novel, nongeometric theory of gravitation, leaning heavily on group theoretical ideas. While this theory combines the appealing features of the nonlinear self-coupled and gauge-type theories (mentioned in the preceding two paragraphs), it avoids their basic pitfalls.[8]

Our starting point consists in taking a rigid space–time background for which, however, we choose not the flat Lorentz world but rather the de Sitter world with positive curvature. We now wish to motivate this choice of ours. It is well known that the only nonflat world structures, which are homogeneous and isotropic both in space and in time, are the Riemann spaces of constant curvature, i.e., de Sitter spaces. Cosmological observations favor positive curvature. An outstanding feature of the de Sitter space is that it is the only nonflat space which admits a maximal group of motions. This can be seen by embedding the de Sitter space in a five-dimensional flat space with signature $(+\,+\,+\,-\,+)$. The maximal symmetry group is the de Sitter group $\mathfrak{L}_5$ whose Lie algebra is isomorphic with that of the five-dimensional proper rotation group R_5. Moreover, in the local limit and when the curvature tends to zero, the symmetry operations of $\mathfrak{L}_5$ reduce to those of the inhomogeneous extended Lorentz group $I\mathfrak{L}_4'$. Thus, in a sense, the well-established local validity of the Poincaré group can be looked upon as a consequence of Mach's principle: The local isotropy and homogeneity expressed by the Poincaré group is a reflection of a similar mass distribution in the large. Moreover, the presence of an asymptotic limit will allow an unambiguous definition of the energy-momentum tensor and other conserved quantities with reference to the curved space. A final ingredient in our reasoning is the presently highly topical endeavor to merge the internal and space-time symmetries of elementary particles. If such attempts are to be taken seriously, the space-time symmetry must encompass the true (as opposed to only the local) structure of the universe. While the covariant merger of $I\mathfrak{L}_4'$ and the internal symmetry (say, SU_3) meets with serious difficulties, we have shown recently that a unification comes about quite naturally and leads to excellent physical predictions[9] if the space-time symmetry is that of the de Sitter world.

The first step in our program is the development of general field theory in a de Sitter background. While such attempts have already a rather long history,[10] we endeavored to formulate a more general and comprehensive framework, and a substantial part of this paper is concerned with this topic. Special attention is paid to the gauge principle in general and to the interpretation of $\mathfrak{L}_5$ as a gauge group. The postulate that the total Lagrangian be invariant under a general $\mathfrak{L}_5$ transformation with parameters which are arbitrary functions of the coordinates leads, in the usual manner, to the existence of a conserved current $\mathcal{J}$ and a compensating gauge field φ. The latter is related to the gravitational field, the existence of which thus attains a *raison d'être*. The conserved current contains not only the matter fields but also φ itself. Hence, φ obeys nonlinear equations. The transformation properties of φ are, of course, different from those of the metric tensor, so that gravitational and metrical effects are separated. Thereby the crucial difficulty of the aforementioned nongeometrical theories is eliminated. Actually, the boundary value problem is that of nonlinear field equations, with respect to a fixed and rigid de Sitter background. As a further consequence, the quantization of gravitation reduces to a conceptually simpler problem, no more difficult than the problem of quantization of the Yang–Mills field.

It is shown that one can split off from φ a coordinate dependent affinity. It is, in fact, this remainder Σ which is coupled to the current. Both Σ and $\mathcal{J}$ turn out to belong to the 35-dimensional representation of the de Sitter group. When going

[7] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954). See also J. J. Sakurai, Ann. Phys. **11**, 1 (1960) and P. Roman, Nuovo Cimento **21**, 747 (1961).

[8] Some of the ideas developed below were first expressed by us in an essay written for the 1965 Gravity Research Foundation Contest. We are indebted to the Foundation for awarding Honorable Mention to this essay.

[9] P. Roman and J. J. Aghassi, Phys. Letters **14**, 68 (1965); Nuovo Cimento **36**, 1062 (1965); **37**, 354 (1965); **38**, 1092 (1965).

[10] See, for example, P. A. M. Dirac, Ann. Math. **36**, 657 (1935) and Max Planck Festschrift, p. 339 (Berlin, 1958); E. Schrödinger, Proc. Roy. Irish Acad. **A46**, 25 (1940); F. Gürsey and T. D. Lee, Proc. Natl. Acad. Sci. U. S. **49**, 179 (1963); C. Fronsdal, Rev. Mod. Phys. **37**, 221 (1965); and the literature quoted in these papers.

to the local frame and taking the limit of vanishing curvature, the de Sitter tensors decompose in such a way that the locally significant components of $\mathcal{J}$ will be identified with the components of the symmetric angular momentum tensor (including gravitation) and with the components of the angular momentum tensor, including spin. Similarly, Σ decomposes in such a way that one part of it simulates the Einstein gravitational potential tensor, another part the total affinity (Christoffel tensor plus Sciama's spin-induced gauge field), and a third part simulates the curvature tensor.

To conclude this introductory survey, we would like to state clearly our motivation concerning the adoption of the gauge principle as an essential ingredient in our theory. While the use of this device is certainly prompted by personal preference (and perhaps even preoccupation) we feel that at the present stage of development it is of immense heuristic value as a guiding principle. The recent history of particle and field theory has taught us that we are in dire need of devices which can serve as limitations when forms of interaction must be selected out of a vast class of open possibilities. In the present case, the gauge principle serves, first of all, the purpose of leading unambiguously to the tensor Σ as the carrier of gravitational effects. Thus, it provides the framework which brings about the dissociation of metric and gravitation. Furthermore, it leads to a concrete form of the currents which are automatically conserved. Finally, the use of this principle casts our theory of gravitation into a form which is fully analogous to the theory of electromagnetism and, at the same time, has an essential nonlinearity built in it in a natural way.

2. THE DE SITTER GROUP

In this section we collect the basic mathematical tools needed to set up a field theory in de Sitter space.

The de Sitter space will be embedded in a five-dimensional flat pseudo-Euclidean space $\mathcal{S}_5$ with *real* coordinates x_a and with the metric tensor[11]

$$g_{11} = g_{22} = g_{33} = -g_{44} = g_{55} = 1,$$
$$g_{ab} = 0 \quad (a \neq b). \tag{1}$$

We define g^{ab} by

$$g^{ab}g_{ac} = \delta^b_c, \quad \text{i.e.,} \quad g^{ab} = g_{ab}, \tag{1a}$$

and we use the metric tensor to raise and lower indices.

[11] Lower case latin indices $a, b, \ldots$ from the beginning of the alphabet run from 1 to 5.

The locus of the de Sitter space in $\mathcal{S}_5$ is the hypersphere

$$x_a x^a = x_1^2 + x_2^2 + x_3^2 - x_4^2 + x_5^2 = R^2, \tag{2}$$

where R is the inverse curvature of the universe. At any point Q of this sphere a local normal unit vector $n(Q)$ is defined by $n_a(Q)\, dx^a = 0$. This implies

$$n_a(Q) = (1/R)x_a(Q). \tag{3}$$

Next, we choose[12] four tangential unit vectors $t^{(\mu)}(Q)$, orthogonal to $n(Q)$,

$$n_a t^{a(\mu)} = 0, \tag{4a}$$

and mutually orthogonal,

$$t^{(\mu)}_a t^{a(\nu)} = g^{\mu\nu}. \tag{4b}$$

These $t^{(\mu)}(Q)$ are unique, up to a Lorentz transformation; $t^{(4)}$ is timelike and the other three are spacelike. We define the *local frame* in the physical Lorentz space $\mathcal{S}_4$ at Q by the vectors $t^{(\mu)}$.

The de Sitter group $\mathcal{L}_5$ is the set of real linear homogeneous transformations which leave the form (2) invariant. It is a 10-parameter noncompact group of rank two whose Lie algebra is locally isomorphic to that of the five-dimensional real rotation group R_5. The generators are denoted[13] by $M_{[ab]}$ and satisfy the algebra

$$[M_{[ab]}, M_{[cd]}] = g_{bc}M_{[ad]} + g_{ad}M_{[bc]} - g_{bd}M_{[ac]} - g_{ac}M_{[bd]}. \tag{5}$$

A realization of these in the function space defined over $\mathcal{S}_5$ is obtained by setting

$$L_{[ab]} \equiv x_a\partial_b - x_b\partial_a. \tag{6}$$

At any given point Q, the subset $L_{[\mu\nu]}$ provides the six generators of the local homogeneous Lorentz group $\mathcal{L}_4$. Furthermore, for a local observer at Q, we have $x_5 = R$, hence $\partial_5 \sim 0$, so that $L_{[5\mu]} \sim R\partial_\mu$, i.e., we get the four generators of local translations in $\mathcal{S}_4$. Thus, the Poincaré group $I\mathcal{L}_4$ is obtained from $\mathcal{L}_5$ by contraction,[14] in the limit $R \to \infty$.

To facilitate the interpretation of our subsequent work, we also point out that

$$\partial^T_b \equiv (1/R)n_a L_{[ab]} \tag{7}$$

can be considered as the local tangential derivative, since $n_b\, \partial^T_b = 0$.

[12] Greek indices $\mu, \nu, \ldots$ from the end of the alphabet run from 1 to 4.

[13] Square brackets around indices denote antisymmetrization.

[14] E. İnönü and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. 39, 510 (1953); 40, 119 (1954).

We now construct a class of representation of $\mathfrak{L}_5$ which is to be used in our formulation of classical (i.e., not second-quantized) field theory in de Sitter space. The idea is to utilize the local isomorphism of $\mathfrak{L}_5$ and R_5, and to construct the representations $J_{[ab]}$ of the generators $M_{[ab]}$ as the sum of an orbital and an intrinsic spin term. The finite dimensional representations so obtained will not be unitary in Hilbert space, but as long as we wish only to set up $\mathfrak{L}_5$-covariant quantum mechanical equations of motion, this does not entail any difficulty. For the purpose of second quantization, an indefinite metric will then be needed in Hilbert space. The details of this procedure go beyond our present framework and will be discussed elsewhere.

We first recall that the vectors spanning an irreducible representation of R_5 are eigenvectors of the two Casimir invariants[15]

$$\mathfrak{N}_2 = \tfrac{1}{2}J_{[ab]}J^{[ab]} \tag{8}$$

and

$$\mathfrak{N}_4 = \tfrac{1}{2}W_aW^a, \tag{9}$$

where, if ϵ is the Levi–Civita tensor in S_5,

$$W_a = \epsilon_{[abcde]}J^{[bc]}J^{[de]}. \tag{9a}$$

An alternative characterization of a representation is given by specifying the "coordinates" λ_1 and λ_2 of the highest weight in the weight diagram. Here $\lambda_1 \geq \lambda_2 \geq 0$, both being simultaneously integers or half-integers.[16] We then have

$$\mathfrak{N}_2 = \tfrac{1}{2}[\lambda_1(\lambda_1 + 3) + \lambda_2(\lambda_2 + 1)], \tag{10}$$

$$\mathfrak{N}_4 = \tfrac{1}{2}(\lambda_1 + 2)(\lambda_1 + 1)(\lambda_2 + 1)\lambda_2. \tag{11}$$

The dimensionality of the representation $\mathfrak{D}^d_{\lambda_1\lambda_2}$ is

$$d = \tfrac{2}{3}(\lambda_1 + \tfrac{3}{2})(\lambda_2 + \tfrac{1}{2})[(\lambda_1 + \tfrac{3}{2})^2 - (\lambda_2 + \tfrac{1}{2})^2]. \tag{12}$$

For a totally symmetric representation $\lambda_2 = 0$, λ_1 arbitrary integer, and we can choose for the representative of $J_{[ab]}$ the differential operator $L_{[ab]}$ given by (6). Apart from a numerical factor, this is nothing else but the *orbital angular momentum* in S_5 space. The eigenfunctions spanning the weight space of $\mathfrak{D}^d_{\lambda 0}$ can be constructed from the normal vectors n_a defined by (3). We set, symbolically,

$$Y^\lambda_{klm}(x) = n_{(a_1}n_{a_2}\cdots n_{a_\lambda)} - \tfrac{1}{5}g_{(a_1a_2}n_{a_3}n_{a_4}\cdots n_{a_\lambda)}. \tag{13}$$

These are traceless symmetric functions of the x_a's. Introducing polar coordinates by

$$\begin{aligned} x_1 &= R\sinh\alpha\sinh\beta\sin\theta\cos\varphi,\\ x_2 &= R\sinh\alpha\sinh\beta\sin\theta\sin\varphi,\\ x_3 &= R\sinh\alpha\sinh\beta\cos\theta,\\ x_4 &= R\sinh\alpha\cosh\beta,\\ x_5 &= R\cosh\alpha, \end{aligned} \tag{14}$$

it is easy to show that

$$Y^\lambda \sim P^{\lambda+1}_{\kappa+1}(\cosh\alpha)P^{\kappa+\frac{1}{2}}_{l+\frac{1}{2}}(\cosh\beta)P^l_m(\cos\theta)e^{im\varphi} \times [\sinh\alpha]^{-1}[\sinh\beta]^{-\frac{1}{2}}, \tag{15}$$

and that this is an eigenfunction of the angular part of the 5-d'Alambertian $\partial_a\partial^a$, i.e.,

$$\begin{aligned} &\Big\{(\sinh\alpha)^{-3}\,\partial_\alpha\sinh^3\alpha\,\partial_\alpha - (\sinh\alpha)^{-2}\\ &\quad\times\Big[(\sinh\beta)^{-2}\,\partial_\beta\sinh^2\beta\,\partial_\beta + (\sinh\beta)^{-2}\\ &\quad\times\Big(\frac{1}{\sin\theta}\,\partial_\theta\sin\theta\,\partial_\theta + \frac{1}{\sin^2\theta}\,\partial^2_\varphi\Big)\Big]\Big\}Y^\lambda\\ &= -\lambda(\lambda+3)Y^\lambda. \end{aligned} \tag{16}$$

Here λ, κ, l, and m are the labels of the representations of $\mathfrak{L}_5 \subset \mathfrak{L}_4 \subset R_3 \subset R_2$, respectively.[17]

In a local frame, the representations of $\mathfrak{L}_5$ reduce[18] into direct sums of irreducible representations of $\mathfrak{L}_4$. The reason for this is that, in the local frame, $x_5 \sim R$ becomes an invariant component in $\mathfrak{L}_4$, causing a reduction in the order of $Y^\lambda(x)$ according to how many times the indices $a_1, \cdots, a_\lambda$ assume the value 5.

To simplify further work, at this point we introduce the single index $A = 1, 2, \cdots, 10$ in place of the antisymmetric pair-indices $[ab]$, by means of the following transcription rule[19]:

[15] The correct definition of raising and lowering general indices (and in particular the case of doing that with the indices belonging to the adjoint representation) are given in the Appendix.

[16] For tensorial representations, λ_1 and λ_2 are integers and can be thought of as the number of nodes in the first and second row of the Young diagram pertaining to the symmetric group S_5.

[17] In general, the Lorentz subgroup $\mathfrak{L}_4$ has the representations labeled by $\kappa_1 \leq \lambda_1$, $\kappa_2 \leq \lambda_2$ contained in $\mathfrak{D}_{\lambda_1\lambda_2}$. Thus, in the present case, $0 \leq \kappa \leq \lambda$, $0 \leq l \leq \kappa$, $-l \leq m \leq l$.

[18] More about the reduction is said at the beginning of Sec. 3.

[19] In general, upper case Latin indices $A, B, \cdots$ from the beginning of the alphabet run from 0 to 10, and are used to identify components of the adjoint representation $\mathfrak{D}^{10}_{11}$.

A	1	2	3	4	5	6	7	8	9	10
$[ab]$	[23]	[31]	[12]	[14]	[24]	[34]	[15]	[25]	[35]	[45]

We can now proceed to construct the representatives J_A for arbitrary representations. We set

$$J_A = L_A + \sigma_A, \tag{17}$$

where the σ_A are finite dimensional matrix operators, obeying the same algebra as L_A [i.e., the relations (5)]. They are the familiar "spin matrices" of R_5. If a representation $\mathfrak{D}^d_{\lambda_1\lambda_2}$ of R_5 is spanned by the vectors φ^s [$s = 1, 2, \cdots, d$, where d is given by (12)] then we denote[20] the matrix elements of σ_A by $\sigma_A{}^r{}_s$. The total set of commutation relations can be summarized as

$$[L_B, L_A] = \gamma_B{}^C{}_A L_C, \tag{18}$$

$$[\sigma_B, \sigma_A]^r{}_s = \gamma_B{}^C{}_A \sigma_C{}^r{}_s, \tag{19}$$

$$[L_A, \sigma_B] = 0. \tag{20}$$

Here $\gamma_B{}^C{}_A$ are the structure constants of the group. We note that, in particular, the 10×10 matrices $\sigma_B{}^C{}_A$, giving the representation of the spin generators for the adjoint representation, are related to the structure constants as

$$2\sigma_B{}^C{}_A = \gamma_B{}^C{}_A. \tag{21}$$

Hence, they are totally antisymmetric, and in particular

$$\sigma_B{}^C{}_A = -\sigma_A{}^C{}_B. \tag{21a}$$

We now introduce the concept of *comma derivation*. Let $\psi^r(x)$ span the space[20] of a representation $\mathfrak{D}^d_{\lambda_1\lambda_2}$ of $\mathfrak{L}_5$. We define the comma derivative by putting

$$\psi^r{}_{,A} \equiv J_A\psi^r(x) = L_A\psi^r(x) + \sigma_A{}^r{}_s\psi^s(x). \tag{22}$$

Comma derivation is commutative,

$$\psi^r{}_{,A,B} = \psi^r{}_{,B,A}. \tag{23}$$

The proof is as follows. We note that J_A transforms according to the adjoint representation $\mathfrak{D}^{10}_{11}$, so that $\psi^s{}_{,C}$ transforms according to $\mathfrak{D}^d_{\lambda_1\lambda_2} \otimes \mathfrak{D}^{10}_{11}$. Then we obtain, using (21a),

$$\psi^r{}_{,A,B} = L_BL_A\psi^r + \sigma_A{}^r{}_sL_B\psi^s + (\sigma_B{}^r{}_s\,\delta^C_A - \sigma_B{}^C{}_A\,\delta^r_s)\psi^s{}_{,C}.$$

A similar expression, with A and B interchanged, is obtained for $\psi^r{}_{,B,A}$. Thus, with (22) and (21a), we easily get

$$\psi^r{}_{,B,A} - \psi^r{}_{,A,B} = [L_A, L_B]\psi^r + [\sigma_A, \sigma_B]^r{}_s\psi^s + 2\sigma_B{}^C{}_A\psi^r{}_{,C}.$$

Using (18) and (19), recombining terms via (22), we find

$$\psi^r{}_{,B,A} - \psi^r{}_{,A,B} = \gamma_A{}^C{}_B\psi^r{}_{,C} + 2\sigma_B{}^C{}_A\psi^r{}_{,C}.$$

Taking note of (21), (21a) we see that the right-hand side vanishes, Q.E.D.

It is clear that comma derivation is the natural generalization for $\mathfrak{L}_5$-covariant theories of the ordinary partial derivation in the $\mathfrak{L}_4$-covariant theories. This is to be understood *in the group theoretic* sense, so that what we mean is that the role played by ∂_μ as a displacement operator in $\mathfrak{S}_4$, is taken over by J_A in the group structure of $\mathfrak{L}_5$ defined over $\mathfrak{S}_5$.

We also note that the metric g^{ab}, being an invariant, satisfies the rule

$$g^{ab}{}_{,A} = 0. \tag{24}$$

Finally, we point out that contravariant comma derivation is easily and consistently defined by putting[15]

$$\psi^{s,A} = g^{AB}\psi^s{}_{,A}. \tag{25}$$

3. FIELDS IN DE SITTER SPACE

The theory of fields and of quantum mechanical equations of motion can now be easily transcribed into an $\mathfrak{S}_5$ framework. In close analogy to the standard $\mathfrak{L}_4$-covariant equations of motion, we propose to take the field equations to be of the form

$$\mathfrak{N}_2\psi_{\lambda_1\lambda_2}(x) = \alpha(\lambda_1, \lambda_2)\psi_{\lambda_1\lambda_2}(x), \tag{26}$$

where α, the eigenvalue of $\mathfrak{N}_2$ for the representation spanned by the components of $\psi_{\lambda_1\lambda_2}$, is given by Eq. (11). We also allow for possible subsidiary conditions involving $\mathfrak{N}_4$.

Next, we note that, in the local frame and in the limit of very small curvature,[21] the invariants of $\mathfrak{L}_5$ reduce to

$$\mathfrak{N}_2 \sim \tfrac{1}{2}R^2\Box^2 \tag{27}$$

and

$$\mathfrak{N}_4 \sim \tfrac{1}{4}R^2v_\mu v^\mu. \tag{28}$$

Here $\Box^2 = \partial_\mu\partial^\mu \equiv -P_\mu P^\mu$ and

$$v_\mu = \epsilon_{[\mu\nu\rho\tau]}\sigma_{[\nu\rho]}\,\partial_\tau. \tag{28a}$$

[20] In general, Latin lower case indices r, s, ... from the end of the alphabet denote components of an arbitrary representation, and run from 1 to d.

[21] That is, keeping only the leading term in R.

The eigenvalue[22] of $\Box^2$ we denote by m^2, and call m the physical mass associated with the representation $\mathfrak{D}^d_{\lambda_1\lambda_2}$. The eigenvalue of $\sigma_{[\nu\rho]}\sigma^{[\nu\rho]} \equiv -S^2$ is $s(s+1)$, and we call s the spin associated with the $\mathfrak{D}^d_{\lambda_1\lambda_2}$. Thus

$$\mathfrak{N}_2 \sim \tfrac{1}{2}R^2m^2, \tag{29}$$

$$\mathfrak{N}_4 \sim -\tfrac{1}{2}R^2m^2s(s+1) - R^2\sigma_{[\mu\nu]}\sigma^{[\mu\rho]}\,\partial_\rho\,\partial^\nu. \tag{30}$$

If we go to the local *rest* frame, then the right-hand side of (30) simplifies to $\frac{1}{2}R^2m^2s(s+1)$. In view of (11) we then can *define* the spin of the particle associated with the representation to be given by λ_2, i.e., we set

$$s = \lambda_2. \tag{31}$$

This implies that the squared spin operator is defined as being proportional to $\mathfrak{N}_4$ when this is taken in the local rest frame.

The other label λ_1 we then associate, according to (29) and (10), with the mass. This implies that Rm and λ_1 are linearly related.

The wavefunctions spanning a representation of $\mathfrak{L}_5$ can always be written in the form

$$\psi_{\lambda_1\lambda_2}(x) = Y^\lambda(x)\varphi_{\lambda_1'\lambda_2'}, \tag{32}$$

where the coordinate independent φ carries the intrinsic spin. For finite m, the λ_1 will be very large (because R is very large), so that the orbital angular momentum $\lambda(\lambda+1)$ associated with the pseudospherical harmonic Y^λ is, in general, very large. The decomposition (32) cannot be made unique, since for a given set λ, λ_1', λ_2' various values of λ_1 and λ_2 are possible. But this seeming lack of uniqueness is already present in the case of relativistic quantum mechanical equations of motion in $\mathfrak{S}_4$. In that case, the various covariant realizations of fields characterized by a given mass and spin can be made equivalent by a suitable choice of the subsidiary condition on the field components required to eliminate the unwanted spin components. In the rest frame, all these equivalent realizations reduce to a unique Foldy representation. Pursey[23] has clarified this point and shown the unique status of the subsidiary relation in the form $v_\mu v^\mu = m^2s(s+1)$. An "arbitrary" covariant representation is related to the canonical Foldy representation by a generalized Foldy–Wouthuysen transformation. In our case of $\mathfrak{S}_5$, the situation is complicated by the fact that we do not have in $\mathfrak{L}_5$ an invariant Abelian subgroup. Nevertheless, *quantum equations of motion which involve only $\mathfrak{N}_2$ and $\mathfrak{N}_4$ give unique canonical representations* in the *local* rest frame. In particular, the subsidiary condition (31) fixing the spin serves the purpose of selecting a particular representation contained in the outer product $Y^\lambda \otimes \varphi_{\lambda_1'\lambda_2'}$. This still does not fix λ_1, but then, the mass (associated with λ_1) cannot be "quantized" in the classical theory and is an arbitrary parameter.

Suppose now that we consider the limiting case of vanishing orbital angular momentum, i.e., $\lambda = 0$ so that $\lambda_1 = \lambda_1'$ and $\lambda_2 = \lambda_2'$. Then, by (29) and (10)

$$m^2 = (1/R^2)[\lambda_1(\lambda_1+3) + \lambda_2(\lambda_2+1)]. \tag{33}$$

Thus, the mass can never be zero in the $\mathfrak{L}_5$ covariant theory.[24] For example, the photon would correspond to the adjoint representation with no orbital momentum, i.e., $\lambda = 0$, $\lambda_2 = 1$, $\lambda_1 = 1$, giving $m_\gamma = 6/R$, and the neutrino would have $\lambda = 0$, $\lambda_2 = \frac{1}{2}$, $\lambda_1 = \frac{1}{2}$, giving $m_\nu = 5/2R$. Taking the radius of the universe to be $R \sim 10^{28}$ cm, the numerical values are of the order 10^{-65} g, way below the experimental lower limits.

Fields associated with $\lambda = 0$ and thus having a mass of the order $m \sim 1/R$ we call *minimal mass fields*. In the limit of vanishing curvature, they correspond to the zero-mass fields of the $\mathfrak{L}_4$ covariant theories. Such fields require a special treatment. In the $R = \infty$ limit of the corresponding $\mathfrak{L}_4$ theory, massless fields are subject to (a) subsidiary conditions to eliminate timelike polarizations, and (b) gauge transformations of the second kind. The latter, in conjunction with (a), prevent longitudinal polarizations from becoming physically observable, i.e., from becoming coupled to observable currents in the Lagrangian. For example, in case of a massless symmetric tensor field $\psi_{\mu\nu}$, whose source is the current $j_{\mu\nu}$, we have the field equations

$$\Box^2\psi_{\mu\nu} = j_{\mu\nu},$$

the subsidiary condition

$$\partial_\mu\,\psi_{\mu\nu} = 0,$$

and the gauge transformation

$$\psi'_{\mu\nu} = \psi_{\mu\nu} + \partial_\mu\,\Lambda_\nu + \partial_\nu\,\Lambda_\mu.$$

The gauge is restricted by the constraint condition

$$\Box^2\Lambda_\mu = 0,$$

so that the subsidiary condition is consistent with the gauge transformation and we also have

$$\Box^2\psi'_{\mu\nu} = j_{\mu\nu}.$$

[22] Units: $\hbar = c = 1$.

[23] D. L. Pursey, Ann. Phys. **32**, 157 (1965).

[24] The only exception is $\lambda = \lambda_1 = \lambda_2 = 0$, the trivial one-dimensional representation. However, no physical particle corresponds to this, because in view of (32), the field associated with this case is a constant.

Conservation of the current, $\partial_\nu j_{\mu\nu} = 0$ is guaranteed by the subsidiary condition.

However, if the mass does not vanish (as is the case in our framework), we are faced with the incompatibility of the subsidiary condition and the gauge transformation. In the usual theory, gauge transformation is then abandoned, and this results in a coupling between the longitudinal and transversal components of the field. On the other hand, gauge transformations are supposedly the *raison d'être* for certain fundamental fields in nature,[25] and we feel forced to adopt an alternative procedure. That is, we keep the gauge transformation and give up the subsidiary condition as an exact relation. The procedure is explained within the limiting $\mathfrak{L}_4$ framework, taking the above example of the tensor field. We now have

$$(\Box^2 - m^2)\psi_{\mu\nu} = j_{\mu\nu}.$$

(i) Define now, as before, the gauge transformed field

$$\psi'_{\mu\nu} = \psi_{\mu\nu} + \partial_\mu \Lambda_\nu + \partial_\nu \Lambda_\mu.$$

(ii) *Require* that

$$(\Box^2 - m^2)\psi'_{\mu\nu} = j_{\mu\nu}.$$

(iii) For consistency, restrict Λ_μ by

$$(\Box^2 - m^2)\Lambda_\mu = 0.$$

(iv) The expression $\partial_\mu\psi_{\mu\nu}$ is not gauge invariant, since

$$\partial_\mu \psi'_{\mu\nu} = \partial_\mu \psi_{\mu\nu} + m^2\Lambda_\nu + \partial_\nu \partial_\mu \Lambda_\mu.$$

Hence, we cannot postulate $\partial_\mu\psi_{\mu\nu} = 0$ for all gauges.

(v) However, due to (iv) and (iii), the expression $(\Box^2 - m^2)\,\partial_\mu\psi_{\mu\nu}$ *is* gauge invariant, and *can* be given a unique value.

(vi) In fact, we *must* set

$$(\Box^2 - m^2)\,\partial_\mu\,\psi_{\mu\nu} = 0$$

so as to obtain a conserved current for the source of the *physical* field $\psi_{\mu\nu}$.

In a sense, (vi) plays the role of a subsidiary condition. In this scheme, decoupling of the longitudinal components is again only approximate and holds only locally. But the timelike components are eliminated by the gauge transformation itself. This can be seen by considering the Fourier transform of the field and gauge ($\tilde\psi_{\mu\nu}$ and $\tilde\Lambda_\mu$) and noting that in the local rest frame $\tilde\psi_{\mu\nu}$ and $\tilde\psi_{\mu\nu} + k_\mu\tilde\Lambda_\nu + k_\nu\tilde\Lambda_\mu$ are physically equivalent. But k_μ is timelike, hence $k_\mu\tilde\Lambda_\nu$ and $k_\nu\tilde\Lambda_\mu$ are additions to the timelike components of $\tilde\psi_{\mu\nu}$, making these unaccessible to physical observation.

The argument, as presented above, carries over to the case of the $\mathfrak{L}_5$ theory, by simply making the replacements

$$\partial_\mu \to J_A \quad \text{(comma derivation)},$$
$$\Box^2 \to (2/R^2)\mathfrak{N}_2.$$

The procedure is then applicable to *all* physical fields, with minimal mass or otherwise, and arbitrary spin as defined by (31), and is indeed adopted in all cases.

4. THE GAUGE PRINCIPLE

Let us consider an $\mathfrak{L}_4$ covariant theory and assume that the Lagrangian density is invariant under a set of linear homogeneous transformations of the field components, with *constant* coefficients:

$$(\psi_i)' \equiv \psi_{i'} = \Lambda^i_{i'}\psi_i. \tag{34a}$$

We assume that the contragradient transformation is

$$(\psi^i)' \equiv \psi^{i'} = \Lambda^{i'}_i\psi^i, \tag{34b}$$

so that

$$\Lambda^{k'}_i\Lambda^i_{j'} = \delta^{k'}_{j'}. \tag{35}$$

We then say that the system possesses a *phase symmetry group*. For an infinitesimal transformation

$$\Lambda^i_j = \delta^i_j + \epsilon\alpha^i_j, \qquad (|\epsilon| \ll 1) \tag{36}$$

the resulting variation in the Lagrangian density has the form $\sim\epsilon\alpha^i_j\,\partial_\mu\mathcal{J}^{\mu j}_i$. The assumed invariance gives the conservation law

$$\partial_\mu\,\mathcal{J}^{\mu j}_i = 0. \tag{37}$$

If we allow the Λ's to be space–time functions, we call the group a *gauge group*. As is well known, the Lagrangian density will cease to be invariant unless an additional field is introduced and coupled to the ψ field components in a well-prescribed manner. To see this we note that now

$$\partial_\mu\,(\psi_i)' = \Lambda^i_{i'}\,\partial_\mu\,\psi_i + \psi_i\,\partial_\mu\,\Lambda^i_{i'},$$

and the second term breaks the invariance. Thus, the compensating field $\varphi_{\mu i}{}^k$ must be entered into the Lagrangian so that $\partial_\mu\psi_i$ is replaced everywhere by

$$\partial_\mu\psi_i \to D_\mu\psi_i \equiv \partial_\mu\psi_i - \varphi_{\mu i}{}^k\psi_k, \tag{38a}$$

$$\partial_\mu\psi^l \to D_\mu\psi^l \equiv \partial_\mu\psi^l + \varphi_{\mu k}{}^l\psi^k. \tag{38b}$$

[25] See Ref. 7. Actually, we derive the gravitational field from a gauge principle.

If we subject φ to the accompanying transformation

$$(\varphi_{\mu i}{}^{k})' = \Lambda_l^{k'}\Lambda_{i'}^{i}\varphi_{\mu i}{}^{l} + \Lambda_l^{k'}(\partial_\mu\Lambda_{i'}^{l}), \tag{39}$$

then the invariance is restored, because then[26]

$$\begin{aligned} D_\mu(\psi_i)' &= \partial_\mu(\psi_i)' - (\varphi_{\mu i}{}^{k})'(\psi_k)' \\ &= \Lambda_{i'}^{i}(\partial_\mu\psi_i - \varphi_{\mu i}{}^{l}\psi_l) = \Lambda_{i'}^{i}\, D_\mu\psi_i. \end{aligned} \tag{40}$$

It is convenient to set

$$\Lambda_i^{i} = \exp(-i\Phi_i^{i}) = \exp(-i\theta_m X_{mi}{}^{i}), \tag{41}$$

where X_m are the Hermitian generators of the symmetry group which gives rise to at least one additively conserved quantity and renders the system phase invariant. If the parameters θ_m are made functions of space-time, we obtain a representation of the gauge group. The summation index m runs from 1 to p, where p is the dimensionality of the adjoint representation. Equation (39) now becomes

$$\begin{aligned} (\varphi_{\mu i}{}^{k})' &= \Lambda_l^{k'}\Lambda_{i'}^{i}(\varphi_{\mu i}{}^{l} - i\,\partial_\mu\Phi_i^{l}) \\ &= \Lambda_l^{k'}\Lambda_{i'}^{i}[\varphi_{\mu i}{}^{l} - i(\partial_\mu\theta_m)X_{mi}{}^{l}]. \end{aligned} \tag{42}$$

When the gauge group is not Abelian, i.e., the X_m do not commute, the φ field will carry the charge generated by the group, i.e., adds to the current its own contribution so that the total charge is conserved. The φ field obeys a nonlinear field equation and current conservation assumes the form[27]

$$\partial_\mu\mathcal{J}^{\mu i}{}_{l} + \varphi_{\mu i}{}^{i}\mathcal{J}^{\mu i}{}_{l} - \varphi_{\mu l}{}^{k}\mathcal{J}^{\mu i}{}_{k} = 0. \tag{43}$$

After this review of the gauge principle,[28] we now consider its application to our $\mathfrak{L}_5$ framework. Clearly, all considerations can be carried over by simply changing the ∂_μ derivatives to the appropriate comma derivatives. Thus, in case of a gauge group, the comma derivative must be replaced in the Lagrangian by what we call the *bar derivative*:

$$\psi_{r,A} \to \psi_{r|A} \equiv \psi_{r,A} - \varphi_{Ar}{}^{t}\psi_t, \tag{44a}$$

$$\psi^{s}{}_{,A} \to \psi^{s}{}_{|A} \equiv \psi^{s}{}_{,A} + \varphi_{At}{}^{s}\psi^{t}. \tag{44b}$$

These are the analogs of (38a, b) and the transcriptions of the subsequent formulas are also obvious.

Of particular interest to us is the case when the gauge group is generated by the space–time symmetry itself, i.e., by the de Sitter group $\mathfrak{L}_5$. Then the indices r, s, etc. carry tensor (or spinor) character and the Λ's are representations of $\mathfrak{L}_5$. In addition, because the index A belongs in the gauge group, a factor Λ_B^{A} appears in the transformation of $\psi^{r}{}_{|B}$. That is, it suffers a homogeneous gauge transformation of the first kind,

$$(\psi^{r}{}_{|B})' = \Lambda_s^{r'}\Lambda_{B'}^{A}\psi^{s}{}_{|A}. \tag{45}$$

The compensating φ field is now subject to a gauge transformation of the second kind,

$$(\varphi_{Br}{}^{s})' = \Lambda_B^{A}\Lambda_t^{s'}(\Lambda_{r'}^{p}\varphi_{Ap}{}^{t} + \Lambda_{r',A}^{t}). \tag{46}$$

Using the representation of type (41) for the Λ, we have alternatively

$$\begin{aligned} (\varphi_{Br}{}^{s})' &= \Lambda_B^{A}\Lambda_t^{s'}\Lambda_{r'}^{p}[\varphi_{Ap}{}^{t} - i(\Phi_p^{t})_{,A}] \\ &= \Lambda_B^{A}\Lambda_t^{s'}\Lambda_{r'}^{p}[\varphi_{Ap}{}^{t} + \theta_{G,A}J_{Gi}{}^{t}]. \end{aligned} \tag{47}$$

Here, θ_G are the 10 parameters of the $\mathfrak{L}_5$ transformation, J_G the anti-Hermitian generators. Note that we have $\theta_{G,A}$ rather than $\theta_{G|A}$: the θ are unobservable gauge functions; hence they do not couple to the gauge field and do not contribute to the gauge current.

The field $\varphi_{Ar}{}^{s}$ is associated with a current $\mathcal{J}^{Ar}{}_{s}$ which is conserved in the sense that

$$\mathcal{J}^{Ar}{}_{s|A} \equiv \mathcal{J}^{Ar}{}_{s,A} + \varphi_A{}^{r}{}_{t}\mathcal{J}^{At}{}_{s} - \varphi_A{}^{t}{}_{s}\mathcal{J}^{Ar}{}_{t} = 0. \tag{48}$$

5. THEORY OF GRAVITATION

As we point out toward the end of Sec. 4, the group $\mathfrak{L}_5$ of general de Sitter transformations gives rise to a gauge group and the compensating gauge field $\varphi_{Ar}{}^{s}$ serves as an affinity to define a *covariant derivation* which we identify with the pertinent case of bar differentiation,

$$\psi_{r|A} = \psi_{r,A} - \varphi_{Ar}{}^{s}\psi_s. \tag{49}$$

This provides a prescription to *couple the φ field to all other fields* that carry $\mathfrak{L}_5$ transformation properties, i.e., span representations of $\mathfrak{L}_5$. However, the gauge is not completely arbitrary: a preferred coordinate system exists which helps to fix the gauge. This preferred system is, in fact, a Cartesian coordinate system in the $\mathcal{S}_5$ space, defined up to an arbitrary $\mathfrak{L}_5$ transformation, i.e., up to a phase transformation. To show that such a system exists, it is sufficient to show that in it the metric g_{ab} is not affected by the group. Now, in a Cartesian system we indeed find, using (49) and (24),

---

[26] Note that in consequence of (35),

$$\Lambda_l^{k'}\,\partial_\mu\Lambda_{i'}^{l} = -\Lambda_{i'}^{l}\,\partial_\mu\Lambda_l^{k'}.$$

[27] J. Schwinger, Phys. Rev. **125**, 1043 (1962).

[28] Although (apart from some points of presentation) the preceding paragraphs are standard knowledge, we found it worthwhile to put them in context. More details can be found, for example, in an article by B. S. DeWitt, in *Relativity, Groups, and Topology* (Gordon and Breach Science Publishers, Inc., New York, 1964), p. 585.

$$g_{ab|A} = g_{ab,A} - \varphi_{Aa}{}^{c} g_{cb} - \varphi_{Ab}{}^{c} g_{ac}$$
$$= -\varphi_{Aab} - \varphi_{Aba} = 0,$$

where the last step follows from the fact that φ_A belongs to the adjoint representation which is antisymmetric.

Thus, in general, we can split off a coordinate dependent affinity from $\varphi_{Aa}{}^{b}$. This affinity will be, in effect, the Christoffel symbol related to the particular preferred coordinate system. We can then write

$$\psi_{a|A} = \psi_{a,A} - \{_{Aa}{}^{b}\}\psi_b - \Sigma_{Aa}{}^{b}\psi_b. \tag{50}$$

If we define *semicolon derivation* by putting

$$\psi_{a;A} \equiv \psi_{a,A} - \{_{Aa}{}^{b}\}\psi_b, \tag{51}$$

then (50) assumes the form

$$\psi_{a|A} = \psi_{a;A} - \Sigma_{Aa}{}^{b}\psi_b. \tag{52}$$

Here $\Sigma_{Aa}{}^{b}$ is a tensor, not an affinity, because it is the difference of two affinities,

$$\Sigma_{Aa}{}^{b} = \varphi_{Aa}{}^{b} - \{_{Aa}{}^{b}\}. \tag{53}$$

Actually, $\Sigma_{Aab} \equiv g_{bc}\Sigma_{Aa}{}^{c}$ is antisymmetric in the indices ab, because Σ_A belongs to the adjoint representation.[29] Since the index A, when expressed by vector indices c, d is given as $[cd]$ according to the transliteration table of Sec. 2, we can write

$$\Sigma_{Aab} \equiv \Sigma_{[cd][ab]}. \tag{54}$$

Thus, Σ *belongs to the* 35*-dimensional representation* $\lambda_1 = 2$, $\lambda_2 = 2$ of $\mathfrak{L}_5$. This then also implies that

$$\Sigma_{[ab][cd]} = \Sigma_{[cd][ab]}. \tag{55}$$

In view of the role of Σ as discussed above, *we now identify this object with the gravitational field.* Gravitons, clearly, will have spin two.[30]

We now show that the field $\Sigma_{Aa}{}^{b}$ is associated with a conserved current $\mathcal{J}^{Aa}{}_{b}$. Let us choose the Lagrangian density of the coupled matter and gravitational fields to be[31]

$$L = \tfrac{1}{2}\psi^{s}{}_{|G}\psi_{s}{}^{|G} + \tfrac{1}{2}m^2R^2\psi^{s}\psi_{s}$$
$$+ \tfrac{1}{2}\Sigma_{Aa}{}^{b}{}_{|G}\Sigma^{Aa}{}_{b}{}^{|G} + \tfrac{1}{2}\mu^2\Sigma_{Aa}{}^{b}\Sigma^{Aa}{}_{b}. \tag{56}$$

This is easily seen to be gauge invariant and we obtain the field equations:

$$\psi^{s}{}_{|G}{}^{|G} - m^2R^2\psi^{s} = 0, \tag{57}$$

$$\Sigma^{Aa}{}_{b|G}{}^{|G} - \mu^2R^2\Sigma^{Aa}{}_{b} = \mathcal{J}^{Aa}{}_{b}. \tag{58}$$

Here m and μ are the masses of the ψ and Σ field, respectively,[30] and the current is of the form

$$\mathcal{J}^{Aa}{}_{b} = \gamma_{G}{}^{r}{}_{s}\gamma^{Ga}{}_{b}(\psi^{s}\psi_{r}{}^{|A} - \psi_{r}\psi^{s|A})$$
$$+ \gamma^{Ga}{}_{b}\gamma_{G}{}^{F}{}_{H}(\Sigma^{Hc}{}_{d}\Sigma_{F}{}^{d}{}_{c}{}^{|A} - \Sigma^{Hc}{}_{d}{}^{|A}\Sigma_{F}{}^{d}{}_{c}). \tag{59}$$

The γ are the Clebsch–Gordan coefficients which couple to the G component of the adjoint representation. Note that $\mathcal{J}$ contains the Σ field, both explicitly (in the second term) and also implicitly through the bar derivatives (in the first term). Hence, *the gravitational field equation* (58) *is nonlinear*, as expected.

From the form (59) of the current and from (57), one then obtains, through a somewhat lengthy calculation—using also the commutation relations of the Clebsch–Gordan coefficients,

$$\mathcal{J}^{Aa}{}_{b;A} = 0. \tag{60}$$

This expresses in a neat form the *conservation of the total current.*

We now integrate this equation over a 3-volume in $\mathcal{S}_5$. The volume element is

$$d^3V_{[de]} = \epsilon_{[defgh]}\,dx^{f}\,dx^{g}\,dx^{h}. \tag{61}$$

Now, the integral of (60) gives, by Gauss' theorem (in the preferred frame),

$$\int dV_{[de]}\mathcal{J}^{[de]a}{}_{b} = \text{const}, \tag{62}$$

which contains contributions from both the matter field ψ *and from gravitation.* Equation (62) is the integral conservation law.

Incidentally, the total current $\mathcal{J}_{Aab}$ coupled to Σ_{Aab} obviously carries the same symmetry as Σ_{Aab}, i.e., belongs to the $\lambda_1 = \lambda_2 = 2$ representation. Thus,

$$\mathcal{J}_{Aab} = \mathcal{J}_{[cd][ab]} = \mathcal{J}_{[ab][cd]}. \tag{63}$$

It is instructive to consider the $\mathfrak{L}_4$ transformation properties of the various components of the total current in the local frame. We set

$$\mathcal{K}_{[ab]} \equiv \int d^3V_{[cd]}\mathcal{J}_{[ab]}{}^{[cd]}. \tag{64}$$

Now, locally, only the $d^3V_{[\mu 5]}$ components of the volume element (62) are large. In particular, $d^3V_{[45]}$ is timelike, i.e., involves integration over spacelike local surfaces. Thus, we see that $\mathcal{K}_{[\mu 5]}$ is essentially the conventional energy-momentum vector and $\mathcal{K}_{[\mu\nu]}$ the angular momentum tensor. Hence, $\mathcal{J}_{[\mu 5][\nu 5]}$

[29] We can say that $\Sigma_{A[ab]}$ is the antisymmetric part of φ_{Aab} whereas $\{_{Aab}\} \equiv \frac{1}{2}g_{ab,A}$ is its symmetric part. Note that in a non-Cartesian system tensors of $\mathfrak{L}_5$ do not have definite symmetry.

[30] Since we take the gravitational field to be minimal mass field, Eq. (33) gives for the mass of the graviton $m_G = 4/R$.

[31] From the manipulations explained in the Appendix, it follows that $\psi_{s}{}^{|G} = g^{GA}\psi_{s|A}$.

can be identified with the conventional symmetric energy-momentum tensor density, and $\mathcal{J}_{[\mu 5][\nu\sigma]}$ with the angular-momentum tensor density. It includes spin. Finally, $\mathcal{J}_{[\mu\nu][\sigma\tau]}$ does not couple to the large components of the volume element; hence it has little local physical significance.

In a similar way[32] we can convince ourselves that, in the local weak limit, $\Sigma_{[\mu 5][\nu 5]}$ simulates the Einstein gravitational potential $(g_{\mu\nu} - \eta_{\mu\nu})$, where $\eta_{\mu\nu}$ is the Lorentz metric. Further, $\Sigma_{[\mu 5][\nu\sigma]}$ simulates $\Gamma_{\mu\nu\sigma}$, the total affinity made up from the Christoffel connection plus Sciama's spin-induced gauge field[6] $S_{\mu\nu\sigma}$. Finally, $\Sigma_{[\mu\nu][\sigma\tau]}$ simulates the curvature tensor $R_{\mu\nu\sigma\tau}$. Thus, in the local frame *and* in the limiting case of weak gravitational fields our theory is equivalent to Einstein's theory as modified by Sciama.

6. CONCLUDING COMMENTS

We believe that the advantage of our proposed theory lies mainly in the fact that the symmetry properties of the Σ field prevent it from becoming coupled to the metric (i.e., unlike the Einstein potential). The theory does not possess a geometrical interpretation. The boundary value problem is well defined.

In the large, our theory is clearly not equivalent to a Riemannian theory. There exists a preferred frame, or equivalently, φ can be split into Σ and the Christoffel affinity. This, in turn, allows an unambiguous separation of the inertial effects (due to the choice of the coordinate system) and the truly gravitational effects (due to the Σ field). In an observer's local frame the two effects are temporarily tied together to simulate an apparent equivalence principle. However, if the observer performs long-range observations, say on the cosmological scale (such as a galactic red-shift measurement), he can immediately discover the existence of the preferred frame which is attached to the de Sitter substratum. This then allows him to discriminate between inertial and gravitational effects. The equivalence principle seems to appear only in case of observations performed on a local level, such as in the "elevator experiment." This, however, does not constitute a "complete" observation since, by limiting the domain of observation, the nature of the global (asymptotic) symmetry is lost.

The principle of general relativity will hold only in the restricted sense of general covariance with respect to arbitrary coordinate transformations belonging to the de Sitter space group. On the other hand, the difference between the presently proposed and the standard theory consist only in effects of the order of $1/R$, at least for the linearized theory. As for truly nonlinear effects, a check such as the computation of the perihelion motion will have to be performed.

We note that the existence of the preferred frame also allows an unambiguous definition and computation of the current related to the Σ field. In this context, we point out that the matter fields ψ are coupled to Σ and to the Christoffel symbol in an identical way, so that the equality of gravitational and inertial mass is guaranteed.

Finally, we observe that, although local gravitational effects are unrelated to geometry, global geometrical effects of gravitation are encompassed, insofar as they are absorbed into the observed curvature in the large. This is so because the numerical value of R will be modified by the nonlinear effects of coupling to the sources. The cosmology is that of a de Sitter world, without being tied down in detail to the material contents of the universe. Nevertheless, by relating the curvature to some average of local interactions, Mach's principle is, in a sense, satisfied. We feel that our scheme contains the seeds of a unified theory of the macro- and micro-world.

APPENDIX

The calculus of $\mathfrak{L}_5$ is greatly facilitated by making use of the Clebsch–Gordan coefficients. Let us denote the Clebsch–Gordan coefficient which couples the r, s component of a representation $\mathfrak{D}^{d'}_{\lambda_1'\lambda_2'}$ to the outer product of the p and t components of the representation $\mathfrak{D}^{d}_{\lambda_1\lambda_2}$ by the symbol $\gamma_{rs}{}^{pt}$. Here $r, s = 1, 2, \cdots, d'$ and $p, t = 1, 2, \cdots, d$. In particular, if $\mathfrak{D}_{\lambda_1'\lambda_2'}$ is the one-dimensional representation, i.e., if $\lambda_1' = \lambda_2' = 0$, we shall use the symbol $\gamma_0{}^{pt}$. For short, we also write

$$g^{pt} \equiv \gamma_0{}^{pt}. \tag{A1}$$

We define the inverse g_{pt} by demanding

$$g^{pt} g_{tu} = \delta^p_u. \tag{A2}$$

The g_{pt} is symmetric (antisymmetric) in its indices according to whether the representation $\mathfrak{D}^{d}_{\lambda_1\lambda_2}$ is a tensor (spinor) representation.

The raising of an index of a component of a vector belonging to an arbitrary representation can now be consistently written by setting

$$\psi^r = g^{rs}\psi_s. \tag{A3}$$

[32] Going back, say, to Eqs. (58) and (52).

Similarly,

$$\psi_r = g_{rs}\psi^s. \tag{A4}$$

The invariance of $\psi_r\psi^r$ is guaranteed by the orthogonality of the Clebsch–Gordan coefficients.

In particular, we can write

$$g_{AB} = \gamma_A{}^{[ab]}\gamma_B{}^{[cd]}g_{[ab][cd]}, \tag{A5}$$

where

$$g_{[ab][cd]} \equiv \tfrac{1}{2}(g_{ac}g_{bd} - g_{ad}g_{bc}). \tag{A6}$$

Hence,

$$g_{AB} = \gamma_A{}^{[ab]}\gamma_{B[ab]}. \tag{A7}$$

As a special consequence, we have

$$L_A = \gamma_A{}^{[ab]}L_{[ab]} \tag{A8}$$

or

$$L_{[ab]} = \gamma^A{}_{[ab]}L_A. \tag{A9}$$

This is equivalent to the transliteration table given in Sec. 2.

It is worth mentioning the connection between the spin generators and the Clebsch–Gordan coefficients. We have

$$2\sigma_A{}^{rs} = \gamma_A{}^{rs}, \tag{A10}$$

provided the product $\mathfrak{D}^d_{\lambda_1\lambda_2} \otimes \mathfrak{D}^d_{\lambda_1\lambda_2}$ contains the adjoint representation only once. If, however, $\mathfrak{D}^{10}_{11}$ is contained in the product twice (i.e., when $\lambda_1 \neq \lambda_2 \neq 0$), then

$$2\sigma_A{}^{rs} = \gamma_A{}^{[rs]} \tag{A11}$$

or

$$2\sigma_A{}^{rs} = \gamma_A{}^{(rs)}, \tag{A12}$$

according to whether $\mathfrak{D}^d_{\lambda_1\lambda_2}$ is a tensor or spinor representation, respectively.[33] Equation (21) is a special case of (A10).

Finally, we note that the covariant derivation given by Eq. (52) can be extended to arbitrary representations by writing

$$\psi_{s|A} = \psi_{s;A} - \Sigma_{As}{}^r\psi_r, \tag{A13}$$

where

$$\Sigma_{As}{}^r \equiv \gamma_{[cd]s}{}^r\Sigma_A{}^{[cd]}. \tag{A14}$$

Note added in proof: While the manuscript of this paper was in press, Prof. D. D. Ivanenko of Moscow kindly called our attention to a large body of work by Soviet authors which is closely related to various ideas expressed in our paper. Some results of these authors can be found, for example, in the following papers: D. Ivanenko, and G. H. A. Sokolik, Zh. Eksperim, i Teor. Fiz. **41**, 10 (1961); A. M. Brodskii, D. Ivanenko, and G. A. Sokolik, Zh. Eksperim, i Teor. Fig. **41**, 1307 (1961) [English Transl.: Soviet Phys.—JETP **4**, 930 (1962)]. D. Ivanenko, Progr. Theoret, Phys. (Kyoto) Suppl., p. 161 (1965); G. A. Sokolik and N. P. Konopleva, Nucl. Phys. **72**, 667 (1965); A. M. Brodskii, D. Ivanenko, Compt. Rend. Acad. Bulgare Sci. **17**, 801 (1964); B. N. Frolov, Vestn. Mosk. Univ. **6**, 48 (1963). See also various articles in the "Theses of the 2nd Soviet Gravit. Conf." (Tbilissi University, 1965). We also wish to call attention to a related paper by Y. Tanikawa, Progr. Theoret. Phys. (Kyoto) Suppl., p. 609 (1965).

[33] (rs) means symmetrization.

Unitary Irreducible Representations of $SL(3, R)$

GERALD ROSEN*
Southwest Research Institute, San Antonio, Texas
(Received 24 December 1965)

It is shown that to each finite-dimensional single-valued irreducible representation of $SL(3, R)$ there corresponds an infinite-dimensional representation which is unitary on any member of a certain one-parameter family of Hilbert spaces. We set up an eigenfunction problem for the members of a three-parameter family of Hilbert subspaces on which such a unitary representation is irreducible. The relatively simple but especially important three-dimensional case is worked out completely. Unitary irreducible representations for the unimodular real linear groups $SL(N, R)$ with $N > 3$ and their subgroups can be obtained by generalizing the formalism described here.

1. INTRODUCTION

IT has been conjectured that $SL(3, R)$, the noncompact group of real unimodular linear transformations in three dimensions, may be of significance for the over-all classification of one-particle hadron states[1] and may also govern the approximate symmetry between physical states associated with strong quantum-gravitational fields.[2] Such applications to physics require appropriate unitary irreducible representations of $SL(3, R)$, or equivalently, appropriate Hermitian irreducible representations for the generators of the $SL(3, R)$ Lie algebra. Thus, a problem of current interest in mathematical physics is to find solutions of the Lie equations

$$[G_A, G_B] = iC_{AB}{}^{D}G_D \tag{1.1}$$

for sets of eight $SL(3, R)$ generators $G_A = G_A^{\dagger}$, Hermitian on some suitably prescribed (separable) Hilbert space. In the case of $SL(3, R)$, the structure constants in (1.1),

$$C_{AB}{}^{D} \equiv C_{ABC}\Omega^{CD}, \tag{1.2}$$

$$\Omega_{AB} \equiv \tfrac{1}{12}C_{AE}{}^{F}C_{BF}{}^{E} = \text{diag}\,[-1, -1, -1, 1, 1, 1, 1, 1] \equiv \Omega^{AB}, \tag{1.3}$$

follow from the totally antisymmetric C_{ABC} with the independent nonvanishing components[3]

$$C_{123} = C_{146} = C_{247} = C_{178} = C_{286} = C_{367} = 1, \quad C_{165} = C_{257} = \sqrt{3}, \quad C_{384} = 2. \tag{1.4}$$

Because $SL(3, R)$ is semisimple and noncompact,[4] it has no finite-dimensional unitary representation[5] and the Cartan–Weyl theory for obtaining unitary irreducible representations of compact semisimple Lie groups does not apply. A complete unitary ir-

* Present address: Drexel Institute of Technology, Philadelphia, Pennsylvania.

[1] Y. Dothan, M. Gell-mann, and Y. Ne'eman, Phys. Letters **17**, 148 (1965).

[2] G. Rosen, Nuovo Cimento **42**, 797 (1966).

[3] Nonzero structure constants have an odd number of indices equal to 1, 2, or 3, so with the Ω "metric" used to raise enumerator indices, we have $C_{ABC} \equiv -C^{ABC}$. It should be remarked that our indices 1, 2, 3 label the generators for the $O(3)$ subgroup of $SL(3, R)$, the maximal compact subgroup contained in $SL(3, R)$.

[4] The (unique) compact complex extension of $SL(3, R)$ is $SU(3)$ (of order eight, rank two, and semisimple), Lie equations for the Hermitian generators of $SU(3)$ being obtained formally from Eqs. (1) by making the correspondence $G_A \to G_A$ [$A = 1, 2, 3$], $G_A \to iG_A$ [$A = 4, 5, 6, 7, 8$] and letting some of the nonzero structure constants $C_{AB}{}^{D}$ (namely, those with $A, B > 3 \geq D$) absorb a minus sign. Thus our Hermitian representation problem for the $SL(3, R)$ Lie algebra can be viewed as an unusual $SU(3)$ representation problem, one with Hermitian operators required for the three generators of an $O(3)$ subgroup but anti-Hermitian operators required for the other five generators of the $SU(3)$ Lie algebra. Although we do not have general recourse to a compact complex extension point of view in order to derive our unitary representations of $SL(3, R)$, the compact complex extension point of view is often very useful.[6-7] [In this connection, it is interesting that the homogeneous Lorentz group was analyzed via its compact complex extension $O(4)$ in the original representation theory by P. A. M. Dirac, Proc. Roy. Soc. (London) **A155**, 447 (1936).]

[5] This is true with the usual meaning of "unitary," namely, unitary on a bonafide Hilbert space endowed with a positive-definite metric. However, if the physical application were to admit a pseudo-Hilbert space with an indefinite metric, then finite-dimensional "unitary" irreducible representations of $SL(3, R)$ could readily be constructed from the finite-dimensional unitary irreducible representations of $SU(3)$. One simply takes the eight Hermitian generators $\mathring{G}_A = \mathring{G}_A^{\dagger}$ for a representation of the $SU(3)$ Lie algebra and sets

$$G_A \equiv \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \times \mathring{G}_A & [A = 1, 2, 3] \\ \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \times \mathring{G}_A & [A = 4, 5, 6, 7, 8]. \end{cases}$$

It follows[4] that these generators satisfy the $SL(3, R)$ Lie equations as a consequence of the $SU(3)$ Lie equations and are "Hermitian" with respect to the indefinite metric

$$\sum \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \times \mathbb{1}, \qquad G_A \sum = \sum G_A^{\dagger},$$

as a consequence of hermicity of the $\mathring{G}_A$.

reducible representation theory for $SL(2, R)$, the universal covering group for proper homogeneous Lorentz transformations in three-dimensional $(2 + 1)$ Minkowski space, has been given by Bargmann in a very beautiful and comprehensive paper.[6] Despite the fact that the latter work is an ideal prototype for noncompact Lie group unitary representation theory, such complete results as those obtained by Bargmann for rank one $SL(2, R)$ are not practical for rank two $SL(3, R)$; in addition to much greater technical complications, a comprehensive theory for $SL(3, R)$ would have to feature a classification of unitary irreducible representations based on all pairs of admissible values for the two functionally independent (quadratic and cubic) Casimir invariants, bringing in many complicated representations which are manifestly academic in character and unlikely to be of physical interest. Our objective in the present paper is to study the rather obvious and most natural unitary representations of $SL(3, R)$, to obtain a formalism which is probably sufficient for physical applications although in no sense exhaustive from the mathematical point of view.

In general, for a unitary representation the Hermitian generators G_A satisfying (1) must be described either by differential operators or by equivalent infinite matrices, derived by taking the differential operators between a complete set of functions in the Hilbert space. A large class of unitary representations of $SL(3, R)$, each representation in the form of a set of eight Hermitian differential operators which act on a certain properly defined Hilbert space of infinitely differentiable functions, is presented in Sec. 2. Then in Sec. 3 we formulate the general eigenfunction problem for obtaining irreducible constituents from these unitary representations of $SL(3, R)$. Finally, the unitary irreducible representations which are most likely to be of physical interest are analyzed in greater detail in Sec. 4.

2. GENERAL FORM OF THE UNITARY REPRESENTATIONS

All finite-dimensional (necessarily nonunitary) single-valued irreducible representations of $SL(3, R)$ are well known.[7] The eight generators of these finite-dimensional single-valued irreducible representations are $n \times n$ real traceless matrices $M_{Ak}{}^{l} = (M_{Ak}{}^{l})^*$, $M_{Ai}{}^{i} = 0$, which satisfy real Lie equations, differing formally from Eqs. (1) only by a trivial factor i,

$$M_{Ak}{}^{i}M_{Bi}{}^{l} - M_{Bk}{}^{i}M_{Ai}{}^{l} = C_{AB}{}^{D}M_{Dk}{}^{l}. \qquad (2.1)$$

As the simplest faithful single-valued irreducible representations of $SL(3, R)$, we have the *fundamental* representation with $n = 3$ (employing a conjunctive expression for the eight matrices, the γ's merely being arbitrary numerical parameters),

$$(M_{Ak}{}^{l}\gamma^{A}) = \begin{bmatrix} \gamma^4 + \frac{1}{\sqrt{3}}\gamma^5 & \gamma^3 + \gamma^8 & -\gamma^2 + \gamma^7 \\ -\gamma^3 + \gamma^8 & -\gamma^4 + \frac{1}{\sqrt{3}}\gamma^5 & \gamma^1 + \gamma^6 \\ \gamma^2 + \gamma^7 & -\gamma^1 + \gamma^6 & -\frac{2}{\sqrt{3}}\gamma^5 \end{bmatrix}; \qquad (2.2)$$

[6] V. Bargmann, Ann. Math. **48**, 568 (1947). This paper was originally intended to be Part I of an even more detailed work. On page 571 Bargmann remarks: "In an appendix to Part II, Dirac's *expansor representations* are analyzed...," referring to the paper by P. A. M. Dirac, Proc. Roy. Soc. (London) **A183**, 284 (1945). Although Bargmann's Part II was never published, a Dirac expansor treatment of $SL(2, R)$ has been given recently by A. O. Barut and C. Fronsdal, Proc. Roy. Soc. (London) **A287**, 532 (1965). The latter work does not exhibit the unitary irreducible representations in explicit form, employing abstract Hilbert space notation throughout and disregarding questions of Hilbert space definition or actual representation realizability, but has the virtue of being readily accessible to a physicist. A formal Dirac expansor treatment of $S\tilde{U}(3)$, a noncompact complex extension of $SU(3)$ differing essentially from $SL(3, R)$ (see remarks at the end of Sec. 2), has also been worked out along the same lines by C. Fronsdal, Proc. Roy. Soc. (London) **A288**, 98 (1965). In contrast to the method of Bargmann, which focuses analysis on the definition of proper Hilbert spaces and the eigenfunctions of the generalized Casimir invariants, the sufficiency of a unitary irreducible representation theory based on purely formal Dirac expansors is always open to question, especially for the more complicated noncompact groups [e.g., $SL(3, R)$].

[7] For example, see: M. Hamermesh, *"Group Theory"* (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1964), pp. 378–391. A basis tensor for a finite-dimensional single-valued irreducible representation of $SL(3, R)$ has N_1 totally symmetric covariant (upper) indices, N_2 totally symmetric contravariant (lower) indices, and is totally traceless with respect to the contraction of any pair of covariant and contravariant indices. Expressed in terms of N_1 and N_2, the dimension of the irreducible representation n and the two independent Casimir constants $q_{(II)}$ and $q_{(III)}$ [see our Eqs. (3.4) and (3.5)] are given by

$$n = \tfrac{1}{2}(N_1 + 1)(N_2 + 1)(N_1 + N_2 + 2),$$

$$q_{(II)} = \tfrac{4}{3}(N_1^2 + N_1N_2 + N_2^2) + 4(N_1 + N_2),$$

$$q_{(III)} = \tfrac{16}{9}(N_1^3 - N_2^3) + \tfrac{8}{3}(N_1^2N_2 - N_1N_2^2) + 8(N_1 - N_2)(N_1 + N_2 + 1).$$

These formulas are identical to those for $SU(3)$.[4] We note that $q_{(III)} = 0$ (implying that $N_1 = N_2$) is a necessary and sufficient condition for a representation to be self-conjugate.

the associated cogradient representation with $n = 6$,

$$(M_{Ak}{}^{l}\gamma^{A}) = \begin{bmatrix} 2\gamma^4 + \frac{2}{\sqrt{3}}\gamma^5 & 0 & 0 & 2\gamma^3 + 2\gamma^8 & 0 & -2\gamma^2 + 2\gamma^7 \\ 0 & -2\gamma^4 + \frac{2}{\sqrt{3}}\gamma^5 & 0 & -2\gamma^3 + 2\gamma^8 & 2\gamma^1 + 2\gamma^6 & 0 \\ 0 & 0 & -\frac{4}{\sqrt{3}}\gamma^5 & 0 & -2\gamma^1 + 2\gamma^6 & 2\gamma^2 + 2\gamma^7 \\ -\gamma^3 + \gamma^8 & \gamma^3 + \gamma^8 & 0 & \frac{2}{\sqrt{3}}\gamma^5 & -\gamma^2 + \gamma^7 & \gamma^1 + \gamma^6 \\ 0 & -\gamma^1 + \gamma^6 & \gamma^1 + \gamma^6 & \gamma^2 + \gamma^7 & -\gamma^4 - \frac{1}{\sqrt{3}}\gamma^5 & -\gamma^3 + \gamma^8 \\ \gamma^2 + \gamma^7 & 0 & -\gamma^2 + \gamma^7 & -\gamma^1 + \gamma^6 & \gamma^3 + \gamma^8 & \gamma^4 - \frac{1}{\sqrt{3}}\gamma^5 \end{bmatrix}; \quad (2.3)$$

and the *adjoint* representation with $n = 8$, $M_{Ak}{}^{l} \equiv -C_{Ak}{}^{l}$. A generalized *contragradient* or *conjugate* representation with the generators

$$\tilde{M}_{Ak}{}^{l} \equiv -\Lambda_{jk}\Lambda^{il}M_{Ai}{}^{j} \quad (2.4)$$

is associated with each finite-dimensional representation, where $\Lambda_{ij} = \Lambda_{ji}$ is a nonsingular $n \times n$ matrix of real constants ($\Lambda^{ij} = \Lambda^{ji}$, the inverse matrix), the conjugate representation (2.4) also satisfying Eqs. (2.1) for any nonsingular symmetrical Λ matrix. A representation is *self-conjugate* (again in a general sense) if there exists a particular Λ matrix for which $\tilde{M}_{Ak}{}^{l} \equiv M_{Ak}{}^{l}$; no such Λ matrix exists for either representation (2.2) or (2.3), but the adjoint representation $M_{Ak}{}^{l} = -C_{Ak}{}^{l}$ is self-conjugate with $\Lambda_{ij} = \Omega_{ij}$.

Let $\xi = (\xi^1, \cdots, \xi^n)$ denote a real n-tuple [$n \geq 3$] of coordinates which label the points of an n-dimensional simply connected compact real manifold $\mathfrak{M}_n$ having the topology of a hollow n ball. Let $\mathfrak{M}_n$ be described analytically in terms of a positive-definite real homogeneous function of the second degree in ξ, $\phi(\xi) \equiv \lambda^{-2}\phi(\lambda\xi)$, for all real $\lambda > 0$, by $\mathfrak{M}_n \equiv \{\xi: a \leq \phi(\xi) \leq b\}$, where a and b are fixed real constants such that $0 < a < b < \infty$. Thus, all compact hypersurfaces $\phi(\xi) = \phi_c \equiv$ const are contained in $\mathfrak{M}_n$ for $a \leq \phi_c \leq b$ and the boundary of $\mathfrak{M}_n$ is given by

$$\partial\mathfrak{M}_n \equiv \{\xi: \phi(\xi) = b\} - \{\xi: \phi(\xi) = a\}, \quad (2.5)$$

the two bounding hypersurfaces being related homeomorphically by the scale dilatation $\xi \to (b/a)^{\frac{1}{2}}\xi$. Defined over $\mathfrak{M}_n$, the eight differential operators associated with a single-valued n-dimensional irreducible representation of $SL(3, R)$

$$G_A \equiv i\xi^k M_{Ak}{}^{l}(\partial/\partial\xi^l) \quad (2.6)$$

satisfy (1.1) as a consequence of (2.1) and therefore constitute an infinite-dimensional representation of $SL(3, R)$.

We seek Hilbert function spaces on which the $SL(3, R)$ generators (2.6) act as Hermitian operators. Let $C^\infty/\mathfrak{M}_n$ denote the space of infinitely differentiable complex-valued functions of ξ over $\mathfrak{M}_n$. Let $\mathfrak{N} = \mathfrak{N}(n; \epsilon_1)$ be the Hilbert space in $C^\infty/\mathfrak{M}_n$ composed of homogeneous functions of degree $-\frac{1}{2}(n + i\epsilon_1)$ in ξ with the parameter ϵ_1 free to range continuously over the real line,

$$\mathfrak{N} = \mathfrak{N}(n; \epsilon_1) \equiv \{f = f(\xi) \text{ in } C^\infty/\mathfrak{M}_n; f(\lambda\xi) = \lambda^{-\frac{1}{2}(n+i\epsilon_1)}f(\xi) \text{ for all real } \lambda > 0\}, \quad (2.7)$$

and with an inner product for any of the Hilbert spaces defined by

$$(f, g) \equiv \int \cdots \int_{\mathfrak{M}_n} f^* g \, d\xi^1 \cdots d\xi^n \quad [f \text{ and } g \text{ in } \mathfrak{N}]. \quad (2.8)$$

Then the generators (2.6) have the closure property on any member of the one-parameter family of Hilbert spaces (2.7), $\{G_A f\}$ in $\mathfrak{N}$ for all f in $\mathfrak{N}$, and are Hermitian on $\mathfrak{N}$ with the inner product (2.8), $(f, G_A g) = (G_A f, g)$, because we have

$$\begin{aligned}(f, G_A g) - (G_A f, g) &\equiv iM_{Ak}{}^{l} \int \cdots \int_{\mathfrak{M}_n} \xi^k \left(f^* \frac{\partial g}{\partial \xi^l} + \frac{\partial f^*}{\partial \xi^l} g \right) d\xi^1 \cdots d\xi^n \\ &= iM_{Ak}{}^{l} \int \cdots \int_{\partial\mathfrak{M}_n} \xi^k f^* g \, d\sigma_l(\xi), \end{aligned} \quad (2.9)$$

where the reality and traceless properties of the matrices $M_{Ak}{}^{l}$ are used together with the divergence theorem; the final integral in (2.9) vanishes as a consequence of (2.5) with f and g in $\mathfrak{R}$, for then the differential integrand $\xi^k f^* g \, d\sigma_l(\xi)$ is invariant with respect to the simple scale dilatation $\xi \to (b/a)^{\frac{1}{2}}\xi$.

Hence, to each finite-dimensional single-valued irreducible representation of $SL(3, R)$ there corresponds a family of infinite-dimensional representations, parameterized above by the real constant ϵ_1. Although members of the family of Hilbert spaces (2.7) that are associated with different values of ϵ_1 are "equivalent" in the sense of a unitary phase transformation, matrix elements of the Hermitian generators (2.6) depend essentially on the parameter ϵ_1, and thus different values of ϵ_1 characterize inequivalent unitary representations. These unitary representations of $SL(3, R)$ are generally reducible with respect to the Hilbert spaces (2.7), but unitary representations with the generators (2.6) are irreducible on certain suitably prescribed Hilbert subspaces $\hat{\mathfrak{R}}$ contained in $\mathfrak{R}$. With respect to such restricted subspaces $\hat{\mathfrak{R}}$, the representations are most useful for physical applications, and in Sec. 3 we formulate appropriate conditions for the functions that compose an $\hat{\mathfrak{R}}$. Solution of the latter conditions provides a general and systematic analytical procedure for extracting irreducible constituent representations from the reducible unitary representations defined above.

It is important to note that no eigenfunction of any of the five generators G_4, G_5, G_6, G_7, G_8 can be contained in the Hilbert space $\mathfrak{R}$ (or in any subspace of $\mathfrak{R}$). For example, if f in $\mathfrak{R}$ were an eigenfunction of the Hermitian generator G_4, $G_4 f = \lambda f$ with λ real, then the commutation relation

$$[G_4, (G_1 + G_6)] = -i(G_1 + G_6) \qquad (2.10)$$

obtained from the Lie equations (1.1) implies that $g \equiv (G_1 + G_6)f$ would also be an eigenfunction of the Hermitian generator G_4, $G_4 g = (\lambda - i)g$ but with a complex eigenvalue, and this is impossible. A similar argument shows that no eigenfunction of the other generators G_A with $A \geq 4$ can be contained in the Hilbert space $\mathfrak{R}$. Thus, essentially unlike a Hilbert space unitary representation for compact $SU(3)$ or a noncompact $\widetilde{SU}(3)$[6], eigenfunctions of two commuting generators (for example, G_4 and G_5, generators that would be diagonalized in a formal treatment of the Lie algebra along the lines suggested by Cartan's theory) are not contained in the representation space for a unitary representation of $SL(3, R)$, and so a basis for a unitary irreducible representation of $SL(3, R)$ cannot be defined in terms of the simultaneous eigenfunctions associated with two commuting generators. Rather, to set up a basis for an $SL(3, R)$ unitary irreducible representation, it is necessary to work exclusively with the generators G_1, G_2, G_3 of the maximal compact subgroup. Functions which compose the basis of a unitary irreducible representation are most conveniently taken to be eigenfunctions of, say, G_3 and $(G_1^2 + G_2^2 + G_3^2)$, the simplest operator function of G_1, G_2, G_3 that commutes with G_3, and are labeled accordingly, as shown by example in Sec. 4.

3. GENERAL FORMULATION OF THE EIGENFUNCTION PROBLEM FOR UNITARY IRREDUCIBLE REPRESENTATIONS

Schur's lemma holds good for infinite-dimensional representations of a Lie group, and thus any operator which commutes with the eight generators of $SL(3, R)$ must act like a constant on the basis of an irreducible representation. Conversely, a representation of $SL(3, R)$ is irreducible if all operators which commute with the eight generators behave like constants over the entire Hilbert space of the representation. It follows that all simultaneous eigenfunctions of a maximal set of operators which commute with the eight generators, eigenfunctions associated with a fixed set of eigenvalues, would constitute a restricted Hilbert space $\hat{\mathfrak{R}}$ contained in $\mathfrak{R}$ for an irreducible representation. Our program in the following is to formulate the appropriate eigenfunction problem which serves to define such restricted Hilbert spaces $\hat{\mathfrak{R}}$.

There are two distinct types of operators that commute with the eight generators of $SL(3, R)$. First, we have the generalized Casimir invariants, operator-valued functions of the generators themselves. As shown in Appendix A, the quadratic and cubic Casimir invariants [denoted by (II) and (III), respectively] suffice as a complete set of functionally independent Casimir invariants for $SL(3, R)$. In addition to the generalized Casimir invariants, we also have the dilatation point transformations $\xi \to \lambda\xi$ with λ a real positive parameter and the parity point transformation $\xi \to -\xi$, linear transformations of $\mathfrak{M}_n$ which leave our representations with the form (2.6) invariant. No other linear point transformation of $\mathfrak{M}_n$ leaves (2.6) invariant as this is guaranteed by the irreducibility of the n-dimensional generators $M_{Ak}{}^{l}$ in (2.6) and Schur's lemma. Invariance of the function space with respect to dilatation transformations is already a property of (2.7) and therefore of any subspace, but the discrete

parity transformation $\xi \to -\xi$ is effective in the decomposition of $\mathfrak{R}$ for irreducible representations. Hence a unitary representation of $SL(3, R)$ with the generators (2.6) is irreducible with respect to any member of a three-parameter family of restricted Hilbert spaces

$$\begin{aligned}\hat{\mathfrak{R}} &= \hat{\mathfrak{R}}(n; \epsilon_1, \epsilon_2, \epsilon_3; \pm) \\ &\equiv \{f = f(\xi) \text{ in } \mathfrak{R}(n; \epsilon_1); (\mathbf{II})f = \epsilon_2 f, \\ &\quad (\mathbf{III})f = \epsilon_3 f; f(-\xi) = \pm f(\xi)\},\end{aligned} \tag{3.1}$$

where the ϵ's are real parameters which characterize a unitary irreducible representation.

Let us work out explicit operator expressions for the quadratic and cubic Casimir invariants, (A3) and (A4) in Appendix A. By putting (2.6) into the latter formulas, we obtain the representations

$$(\mathbf{II}) = \xi^i M_{Ai}{}^j M^A{}_j{}^k \frac{\partial}{\partial \xi^k} + \xi^i \xi^j M_{Ai}{}^k M^A{}_j{}^l \frac{\partial^2}{\partial \xi^k \partial \xi^l}, \tag{3.2}$$

$$\begin{aligned}(\mathbf{III}) = i\Omega^{ABC}\Big(&\xi^i M_{Ai}{}^j M_{Bj}{}^k M_{Ck}{}^l \frac{\partial}{\partial \xi^l} \\ &+ 3\xi^h \xi^i M_{Ah}{}^j M_{Bj}{}^k M_{Ci}{}^l \frac{\partial^2}{\partial \xi^k \partial \xi^l} \\ &+ \xi^\sigma \xi^h \xi^i M_{A\sigma}{}^j M_{Bh}{}^k M_{Ci}{}^l \frac{\partial^3}{\partial \xi^j \partial \xi^k \partial \xi^l}\Big),\end{aligned} \tag{3.3}$$

where terms in (3.3) have been combined by exploiting the total symmetry of Ω^{ABC}. These differential operator representations of the Casimir invariants are simplified somewhat by making use of the relations

$$M_{Ai}{}^j M^A{}_j{}^k = q_{(II)} \delta_i^k, \tag{3.4}$$

$$\Omega^{ABC} M_{Ai}{}^j M_{Bj}{}^k M_{Ck}{}^l = q_{(III)} \delta_i^l, \tag{3.5}$$

with $q_{(II)}$ and $q_{(III)}$ the two independent Casimir constants characteristic of the n-dimensional irreducible representation.[7] By symmetrizing, splitting off the diagonal part of each tensorial coefficient, and evoking (3.4) and (3.5), Eqs. (3.2) and (3.3) become

$$(\mathbf{II}) = (n+1)^{-1} q_{(II)} (X^2 + nX) + \xi^i \xi^j Q_{ij}{}^{kl} \frac{\partial^2}{\partial \xi^k \partial \xi^l}, \tag{3.6}$$

$$\begin{aligned}(\mathbf{III}) = {}& i(n+1)^{-1}(n+2)^{-1} q_{(III)} (2X^3 + 3nX^2 + n^2 X) \\ &+ i(n+4)^{-1}(2X + n)\xi^i \xi^j R_{ij}{}^{kl} \frac{\partial^2}{\partial \xi^k \partial \xi^l} \\ &+ i\xi^\sigma \xi^h \xi^i W_{\sigma hi}{}^{jkl} \frac{\partial^3}{\partial \xi^j \partial \xi^k \partial \xi^l},\end{aligned} \tag{3.7}$$

where

$$X \equiv \xi^i (\partial / \partial \xi^i) \tag{3.8}$$

and with the introduction of the irreducible (totally symmetric and traceless) tensorial quantities

$$\begin{aligned}Q_{ij}{}^{kl} \equiv {}& \tfrac{1}{2}(M_{Ai}{}^k M^A{}_j{}^l + M_{Ai}{}^l M^A{}_j{}^k) \\ &- \tfrac{1}{2}(n+1)^{-1} q_{(II)} (\delta_i^k \delta_j^l + \delta_i^l \delta_j^k),\end{aligned} \tag{3.9}$$

$$\begin{aligned}R_{ij}{}^{kl} \equiv {}& \tfrac{3}{4}\Omega^{ABC}(M_{Ai}{}^k M_{Bh}{}^k M_{Cj}{}^l + M_{Ai}{}^k M_{Bh}{}^l M_{Cj}{}^k \\ &+ M_{Aj}{}^k M_{Bh}{}^k M_{Ci}{}^l + M_{Aj}{}^k M_{Bh}{}^l M_{Ci}{}^k) \\ &- \tfrac{3}{2}(n+1)^{-1} q_{(III)} (\delta_i^k \delta_j^l + \delta_i^l \delta_j^k),\end{aligned} \tag{3.10}$$

$$\begin{aligned}W_{\sigma hi}{}^{jkl} \equiv {}& \tfrac{1}{6} \sum_{\substack{\text{Perm} \\ \sigma, h, i}} \Omega^{ABC} M_{A\sigma}{}^j M_{Bh}{}^k M_{Ci}{}^l \\ &- \tfrac{1}{18}(n+4)^{-1} \sum_{\substack{\text{Perm} \\ \sigma, h, i}} \sum_{\substack{\text{Perm} \\ j, k, l}} \delta_\sigma^j R_{hi}{}^{kl} \\ &- \tfrac{1}{3}(n+2)^{-1}(n+1)^{-1} q_{(III)} \sum_{\substack{\text{Perm} \\ \sigma, h, i}} \delta_\sigma^j \delta_h^k \delta_i^l.\end{aligned} \tag{3.11}$$

The tensor character of (3.9) is expressed by

$$M_{Ai}{}^h Q_{hj}{}^{kl} + M_{Aj}{}^h Q_{ih}{}^{kl} - M_{Ah}{}^k Q_{ij}{}^{hl} - M_{Ah}{}^l Q_{ij}{}^{kh} = 0, \tag{3.12}$$

similar equations expressing the tensor character of (3.10) and (3.11). That the latter quantities are traceless follows from the orthonormality conditions

$$M_{Ak}{}^l M_{Bl}{}^k = \tfrac{1}{8} n q_{(II)} \Omega_{AB} \quad \text{and} \quad \Omega^{ABC} \Omega_{BC} = 0.$$

Returning to the eigenfunction problem for restricted Hilbert spaces (3.1) with the Casimir invariants represented by (3.6) and (3.7), we see that for any f in $\mathfrak{R}(n; \epsilon_1)$,

$$Xf = -\tfrac{1}{2}(n + i\epsilon_1) f, \tag{3.13}$$

and hence with (3.6) and (3.7) the eigenfunction conditions in (3.1) reduce to

$$\xi^i \xi^j Q_{ij}{}^{kl} \frac{\partial^2 f}{\partial \xi^k \partial \xi^l} = [\epsilon_2 + \tfrac{1}{4}(n^2 + \epsilon_1^2)(n+1)^{-1} q_{(II)}] f, \tag{3.14}$$

$$\begin{aligned}&i\xi^\sigma \xi^h \xi^i W_{\sigma hi}{}^{jkl} \frac{\partial^3 f}{\partial \xi^j \partial \xi^k \partial \xi^l} \\ &\quad + \epsilon_1 (n+2)^{-1} \xi^i \xi^j R_{ij}{}^{kl} \frac{\partial^2 f}{\partial \xi^k \partial \xi^l} \\ &\quad = [\epsilon_3 + \tfrac{1}{4}\epsilon_1(n^2 + \epsilon_1^2)(n+1)^{-1}(n+2)^{-1} q_{(III)}] f.\end{aligned} \tag{3.15}$$

Equations (3.14) and (3.15) serve to define the restricted Hilbert spaces (3.1) for irreducible rep-

resentations, admissible values of the real parameters ϵ_2 and ϵ_3 appearing as eigenvalues with f in $\mathfrak{N}(n; \epsilon_1)$ and with f of definite parity. By virtue of the high symmetry featured by the irreducible tensorial coefficients $Q_{ij}{}^{kl}$, $R_{ij}{}^{kl}$, and $W_{ghi}{}^{jkl}$, explicit solution of this eigenfunction problem is feasible for the lower-dimensional cases. We illustrate this in Sec. 4 by working out the relatively simple but especially important three-dimensional case in complete detail.

4. UNITARY IRREDUCIBLE REPRESENTATIONS ASSOCIATED WITH THE THREE-DIMENSIONAL CASE

For the case $n = 3$ with the $M_{Ak}{}^{l}$ given by (2.2), straightforward calculation shows that the tensorial quantities (3.9), (3.10), and (3.11) all vanish,

$$Q_{ij}{}^{kl} = 0, \qquad R_{ij}{}^{kl} = 0, \qquad W_{ghi}{}^{jkl} = 0. \tag{4.1}$$

Equations (4.1) express the major simplifying and distinguishing feature of the three-dimensional case, for which the conditions (3.14) and (3.15) reduce to

$$\epsilon_2 = -\tfrac{1}{16}(9 + \epsilon_1^2)q_{(II)}, \tag{4.2}$$

$$\epsilon_3 = -\tfrac{1}{80}\epsilon_1(9 + \epsilon_1^2)q_{(III)}, \tag{4.3}$$

with f in $\mathfrak{N}(n; \epsilon_1)$ but otherwise unrestricted. We have[7] $q_{(II)} = 16/3$ and $q_{(III)} = 160/9$ for $n = 3$, and thus the family of restricted Hilbert spaces (3.1) becomes

$$\hat{\mathfrak{N}} = \hat{\mathfrak{N}}[3; \epsilon_1, -3(1 + \tfrac{1}{9}\epsilon_1^2), -2\epsilon_1(1 + \tfrac{1}{9}\epsilon_1^2); \pm], \tag{4.4}$$

a member determined by the single real parameter ϵ_1, $[-\infty < \epsilon_1 < +\infty]$, and the parity of the constituent functions. These unitary irreducible representations correspond qualitatively to the continuous Bargmann[6] series $C_q^{\frac{1}{2}}$ for $SL(2, R)$; the signature of the quadratic Casimir invariant defined by (A3) is -2 [opposite in sign to Bargmann's definition for $SL(2, R)$], so that admissible values of the quadratic Casimir invariant run continuously toward $-\infty$ with (4.4). By putting (2.2) into (2.6), we obtain the following representation for the Hermitian generators:

$$G_1 = i\left(\xi^2 \frac{\partial}{\partial \xi^3} - \xi^3 \frac{\partial}{\partial \xi^2}\right),$$

$$G_2 = i\left(\xi^3 \frac{\partial}{\partial \xi^1} - \xi^1 \frac{\partial}{\partial \xi^3}\right),$$

$$G_3 = i\left(\xi^1 \frac{\partial}{\partial \xi^2} - \xi^2 \frac{\partial}{\partial \xi^1}\right),$$

$$G_4 = i\left(\xi^1 \frac{\partial}{\partial \xi^1} - \xi^2 \frac{\partial}{\partial \xi^2}\right),$$

$$G_5 = \frac{i}{\sqrt{3}}\left(\xi^1 \frac{\partial}{\partial \xi^1} + \xi^2 \frac{\partial}{\partial \xi^2} - 2\xi^3 \frac{\partial}{\partial \xi^3}\right),$$

$$G_6 = i\left(\xi^2 \frac{\partial}{\partial \xi^3} + \xi^3 \frac{\partial}{\partial \xi^2}\right),$$

$$G_7 = i\left(\xi^1 \frac{\partial}{\partial \xi^3} + \xi^3 \frac{\partial}{\partial \xi^1}\right),$$

$$G_8 = i\left(\xi^1 \frac{\partial}{\partial \xi^2} + \xi^2 \frac{\partial}{\partial \xi^1}\right), \tag{4.5}$$

while the constituent functions of the family of restricted Hilbert spaces (4.4) can be expressed generically in terms of an infinite series in spherical harmonics

$$f = f(\xi) = \sum_l \sum_{m=-l}^{l} f_m^l y_l^m, \tag{4.6}$$

where the f_m^l's are complex constants,

$$y_l^m = y_l^m(\rho, \theta, \omega) \equiv \left[\frac{(l + \frac{1}{2})(l - m)!}{\pi \ln (b/a)(l + m)!}\right]^{\frac{1}{2}} \times \rho^{-\frac{1}{2}(3 + i\epsilon_1)} P_l^m(\cos\theta) e^{-im\omega}, \tag{4.7}$$

and

$$\xi^1 \equiv \rho \sin\theta \cos\omega, \qquad \xi^2 \equiv \rho \sin\theta \sin\omega, \qquad \xi^3 \equiv \rho \cos\theta. \tag{4.8}$$

Summation with regard to the index l in (4.6) is extended over all even nonnegative integers or over all odd positive integers, depending on the parity of the Hilbert space (4.4). In view of the spherical coordinates (4.8), the most convenient bounding hypersurfaces for $\mathfrak{M}_3$ [compatible with the conditions stated above Eq. (2.5)] are obtained by setting $\phi(\xi) \equiv \rho^2$, and then it follows that the functions (4.7) comprise an orthonormal set with respect to the inner product (2.8),

$$(y_{l'}^{m'}, y_l^m) = \delta_{m'm}\delta_{l'l}. \tag{4.9}$$

Individual spherical harmonics in (4.6) are eigenfunctions for the main commuting operators associated with the $O(3)$ subgroup,

$$(G_1^2 + G_2^2 + G_3^2)y_l^m = l(l + 1)y_l^m, \tag{4.10}$$

$$G_3 y_l^m = m y_l^m, \tag{4.11}$$

the hypersurfaces $\rho = \text{const}$ being invariant varieties for the $O(3)$ subgroup. However, the other generators G_4, G_5, G_6, G_7, G_8 in (4.5) have nonvanishing

off-diagonal matrix elements, as well as diagonal matrix elements, with regard to the l index in the basis (4.7). Computation of the matrix elements for the latter generators in (4.5) is facilitated by transforming to differential operators in spherical coordinates and making use of well-known connection formulas for the spherical harmonics, such as the formulas

$$\sin\theta\frac{\partial y_l^m}{\partial\theta} = \left[\frac{(l+1)^2-m^2}{4(l+1)^2-1}\right]^{\frac{1}{2}} l y_{l+1}^m - \left[\frac{l^2-m^2}{4l^2-1}\right]^{\frac{1}{2}}(l+1)y_{l-1}^m \tag{4.12}$$

and

$$(\cos\theta)y_l^m = \left[\frac{(l+1)^2-m^2}{4(l+1)^2-1}\right]^{\frac{1}{2}} y_{l+1}^m + \left[\frac{l^2-m^2}{4l^2-1}\right]^{\frac{1}{2}} y_{l-1}^m. \tag{4.13}$$

We present the results of the matrix element computation for

$$G_5 = \frac{i}{\sqrt{3}}\left[(1-3\cos^2\theta)\rho\frac{\partial}{\partial\rho} + 3\cos\theta\sin\theta\frac{\partial}{\partial\theta}\right] \tag{4.14}$$

with

$$\begin{aligned}(y_{l'}^{m'}, G_5 y_l^m) = \Big\{&\frac{\epsilon_1}{\sqrt{3}}\left[\left(\frac{(l+1)^2-m^2}{4(l+1)^2-1}\right)l - \left(\frac{l^2-m^2}{4l^2-1}\right)(l+1)\right]\delta_{ll'} \\ &- \frac{\sqrt{3}}{2}[\epsilon_1 + i(2l-1)]\left[\frac{l^2-m^2}{4l^2-1}\right]^{\frac{1}{2}} \\ &\times\left[\frac{(l-1)^2-m^2}{4(l-1)^2-1}\right]^{\frac{1}{2}}\delta_{l-1,l'+1} \\ &- \frac{\sqrt{3}}{2}[\epsilon_1 - i(2l'-1)]\left[\frac{l'^2-m^2}{4l'^2-1}\right]^{\frac{1}{2}} \\ &\times\left[\frac{(l'-1)^2-m^2}{4(l'-1)^2-1}\right]^{\frac{1}{2}}\delta_{l'-1,l+1}\Big\}\delta_{mm'}.\end{aligned} \tag{4.15}$$

More generally, the matrix element $(y_{l'}^{m'}, G_A y_l^m)$ may be nonzero for any of the generators with $A = 1, 2, 3$ only if $l = l'$ and $(m - m') = 0$ or ± 1, while for any of the generators with $A = 4, 5, 6, 7, 8$ only if $(l - l') = 0$ or ± 2 and $(m - m') = 0$ or ± 2. Finally, we note that the set of linear combinations of the generators in (A1) takes a particularly simple form for the three-dimensional case, namely,

$$S_b^a = \xi^a\frac{\partial}{\partial\xi^b} - \tfrac{1}{3}\delta_b^a\xi^c\frac{\partial}{\partial\xi^c}. \tag{4.16}$$

5. CONCLUDING REMARKS

Alluded to in the literature[1] as "ladder representations," for the three-dimensional case discussed in the preceding section the generators G_1, G_2, G_3 take a form equivalent to the direct sum of an infinite sequence of unitary irreducible $O(3)$ representations, with l either even or odd for all members in the sequence, while the generators G_4, G_5, G_6, G_7, G_8 interrelate adjacent $O(3)$ representations with $\Delta l = 2$ in the sequence. For the n-dimensional cases with $n \geq 6$ formulated in Sec. 3, the generators G_1, G_2, G_3 also take a form equivalent to the direct sum of an infinite sequence of unitary irreducible $O(3)$ representations, but the generators G_4, G_5, G_6, G_7, G_8 act on the basis of the representations in a considerably more complicated manner with no "ladderlike" relationship being in evidence. Likewise, the eigenvalue spectra of the Casimir invariants (II) and (III) on $\mathfrak{R}(n; \epsilon_1)$, defining the irreducible representations, are very rich and complicated spectra, as exemplified by the $n = 8$ case of the adjoint representation (with $q_{(II)} = 12$, $q_{(III)} = 0$, and $R_{ij}{}^{kl} = 0$), for which the conditions (3.14) and (3.15) reduce to

$$\xi^i\xi^j Q_{ij}{}^{kl}\frac{\partial^2 f}{\partial\xi^k\,\partial\xi^l} = [\epsilon_2 + \tfrac{1}{3}(64+\epsilon_1^2)]f, \tag{5.1}$$

$$i\xi^g\xi^h\xi^i W_{ghi}{}^{jkl}\frac{\partial^3 f}{\partial\xi^j\,\partial\xi^k\,\partial\xi^l} = \epsilon_3 f, \tag{5.2}$$

where

$$\begin{aligned}Q_{ij}{}^{kl} &= \tfrac{1}{2}(C_i{}^{kA}C_{Aj}{}^{l} + C_i{}^{lA}C_{Aj}{}^{k}) - \tfrac{2}{3}(\delta_i^k\delta_j^l + \delta_i^l\delta_j^k) \\ &= \tfrac{1}{2}(\Omega_i{}^{kA}\Omega_{Aj}{}^{l} + \Omega_i{}^{lA}\Omega_{Aj}{}^{k}) - \Omega_{ij}{}^{A}\Omega_A{}^{kl} \\ &\quad - \tfrac{8}{3}\Omega_{ij}\Omega^{kl} + \tfrac{2}{3}(\delta_i^k\delta_j^l + \delta_i^l\delta_j^k)\end{aligned} \tag{5.3}$$

and

$$W_{ghi}{}^{jkl} = -\tfrac{1}{6}\sum_{\substack{Perm\\ g,h,i}}\Omega^{ABC}C_{Ag}{}^{j}C_{Bh}{}^{k}C_{Ci}{}^{l}, \tag{5.4}$$

with the tensorial quantities (5.3) and (5.4) having many nonzero components. From general considerations one would expect all sufficiently *large negative* values to be admissible for the eigenvalue ϵ_2 of the quadratic Casimir invariant for every family of unitary irreducible representations associated with the n-dimensional case.

Even more complicated unitary irreducible representations can be obtained by forming the direct sum of a number of representations (2.6), each defined over a distinct manifold $\mathfrak{M}_n$, and regarding these direct sum representations as defined in the natural (and thus unitary) way on the associated direct-product manifold. The eigenfunction problem for extracting new irreducible constituent representations from these manifestly reducible unitary representations is made formidable from the standpoint of explicit solution by cross terms of differential operators which appear in the generalized Casimir invariants. This is illustrated by the $(3 \oplus 3)$-dimensional case with the generators represented by

$$G_A = iM_{Ak}{}^l\left(\xi^k \frac{\partial}{\partial \xi^l} + \eta^k \frac{\partial}{\partial \eta^l}\right), \tag{5.5}$$

where the $M_{Ak}{}^l$ are defined by (2.2); the quadratic Casimir invariant (A3) is thus

$$\begin{aligned}(\mathrm{II})_3 = {} & \frac{4}{3}\left(\xi^k \frac{\partial}{\partial \xi^k}\right)^2 + 4\xi^k \frac{\partial}{\partial \xi^k} \\ & + 4\xi^k \eta^l \left(\frac{\partial^2}{\partial \xi^l\, \partial \eta^k} - \frac{1}{3}\frac{\partial^2}{\partial \xi^k\, \partial \eta^l}\right) \\ & + \frac{4}{3}\left(\eta^k \frac{\partial}{\partial \eta^k}\right)^2 + 4\eta^k \frac{\partial}{\partial \eta^k}\end{aligned} \tag{5.6}$$

with manifestly involved eigenfunctions in the Hilbert space

$$\begin{aligned}&\mathfrak{H}(3 \oplus 3;\, \epsilon_1', \epsilon_1'') \equiv \{f = f(\xi, \mathfrak{n}) \text{ in } C^\infty/\mathfrak{M}_{3\oplus 3};\\ &f(\xi, \mathfrak{n}) = \lambda^{\frac{1}{2}(3+i\epsilon_1')} f(\lambda\xi, \mathfrak{n}) = \lambda^{\frac{1}{2}(3+i\epsilon_1'')} f(\xi, \lambda\mathfrak{n})\\ &\qquad\qquad \text{for all real } \lambda > 0\}\end{aligned} \tag{5.7}$$

owing to the cross term in (5.6).

In the Appendixes we treat some mathematical topics which are closely related to the representations discussed in the preceding sections. Finally, it should be noted that unitary irreducible representations for the unimodular real linear groups $SL(N, R)$ with $N > 3$ and their subgroups can be obtained by generalizing the formalism described here.

ACKNOWLEDGMENTS

The author would like to thank Professor George W. Mackey of Harvard University and Dr. Morton Hamermesh of the Argonne National Laboratory for early discussions which confirmed the nonexistence of previously reported results and the need for research on the $SL(3, R)$ unitary representation problem. Thanks are also due Professor Leon Ehrenpreis of Harvard University for a helpful private communication.

APPENDIX A. GENERALIZED CASIMIR INVARIANTS

Certain operator-valued functions of the eight generators for $SL(3, R)$ satisfying (1.1), the so-called "generalized Casimir invariants," commute with all of the generators and thus play a key role in the reduction theory discussed in Sec. 3. In order to facilitate a systematic construction of the generalized Casimir invariants for $SL(3, R)$, we introduce the set of linear combinations of the generators

$$\begin{aligned}
S_1^1 &\equiv \frac{1}{2i}\left(G_4 + \frac{1}{\sqrt{3}} G_5\right),\\
S_1^2 &\equiv \frac{1}{2i}(G_3 + G_8),\\
S_1^3 &\equiv \frac{1}{2i}(-G_2 + G_7),\\
S_2^1 &\equiv \frac{1}{2i}(-G_3 + G_8),\\
S_2^2 &\equiv \frac{1}{2i}\left(-G_4 + \frac{1}{\sqrt{3}} G_5\right),\\
S_2^3 &\equiv \frac{1}{2i}(G_1 + G_6),\\
S_3^1 &\equiv \frac{1}{2i}(G_2 + G_7),\\
S_3^2 &\equiv \frac{1}{2i}(-G_1 + G_6),\\
S_3^3 &\equiv -\frac{1}{\sqrt{3}\, i} G_8.
\end{aligned} \tag{A1}$$

Since we have $S_a^a \equiv S_1^1 + S_2^2 + S_3^3 \equiv 0$, only eight of the operators (A1) are linearly independent, but as a consequence of (1.1) and (1.4) the latter operators satisfy the symmetrical set of commutation relations

$$[S_c^a, S_d^b] = \delta_c^b S_d^a - \delta_d^a S_c^b, \tag{A2}$$

which show that S_b^a has the transformation character of a mixed tensor with respect to real unimodular transformations in three dimensions. It follows immediately from (A2) that any *invariant* operator function of S_b^a, formed by contraction with no free tensor indices, commutes with all the S_b^a. By making use of (A1) and bearing in mind that $S_a^a \equiv 0$, the

simplest nontrivial Casimir invariants are obtained as[8]

$$2S_b^a S_a^b = -\Omega^{AB}G_A G_B \equiv -G_A G^A \equiv (\mathrm{II}), \tag{A3}$$

$$4i\epsilon_{abc}\epsilon^{def}S_d^a S_e^b S_f^c = 4i(S_b^a S_c^b S_a^c + S_a^b S_b^c S_c^a)$$
$$= -\Omega^{ABC}G_A G_B G_C \equiv (\mathrm{III}), \tag{A4}$$

where the totally symmetric $\Omega^{ABC} \equiv \Omega_{ABC}$ has the independent nonvanishing components

$$\Omega_{114} = \Omega_{128} = \Omega_{236} = \Omega_{317} = \Omega_{477} = \Omega_{678} = 1,$$
$$\Omega_{115} = \Omega_{225} = \frac{1}{\sqrt{3}}, \qquad \Omega_{445} = \Omega_{588} = \frac{2}{\sqrt{3}}, \tag{A5}$$
$$\Omega_{224} = \Omega_{466} = -1, \qquad \Omega_{566} = \Omega_{577} = -\frac{1}{\sqrt{3}},$$
$$\Omega_{335} = \Omega_{555} = -\frac{2}{\sqrt{3}},$$

and satisfies the equations

$$\Omega_{ABC}\Omega^{BC} \equiv \Omega_{AB}{}^{B} = 0, \quad \Omega_A{}^{EF}\Omega_{EF}{}^{B} = \tfrac{20}{3}\delta_A^B, \tag{A6}$$

$$\Omega_{ABE}C_{FC}{}^{E} + \Omega_{BCE}C_{FA}{}^{E} + \Omega_{CAE}C_{FB}{}^{E} = 0. \tag{A7}$$

The other linear combination of invariants cubic in the S_b^a,

$$S_b^a S_c^b S_a^c - S_a^b S_b^c S_c^a = 3S_b^a S_a^b, \tag{A8}$$

is not functionally independent of the quantity (A3). Furthermore, all higher-order invariants, containing more than three S_b^a in a totally contracted term, depend functionally on (A3) and (A4) by virtue of the Cayley operator equations[9]

$$S_e^a S_f^e S_b^f = 3S_e^a S_b^e + (\tfrac{1}{2}S_f^e S_e^f - 2)S_b^a + (\tfrac{1}{3}S_e^a S_f^e S_c^f - S_f^e S_e^f)\delta_b^a, \tag{A9}$$

$$S_a^e S_e^f S_b^f = -3S_a^e S_e^b + (\tfrac{1}{2}S_f^e S_e^f - 2)S_a^b + (\tfrac{1}{3}S_c^e S_e^f S_f^c + S_f^e S_e^f)\delta_a^b, \tag{A10}$$

implied by (A2). Hence the quadratic and cubic Casimir invariants (A3) and (A4) suffice as a complete set of functionally independent operators which commute with all eight of the $SL(3, R)$ generators, in the abstract algebraic sense and thus also for any representation.

APPENDIX B. THE $SL(3, R)$ GROUPS OF MOTIONS

Let $\mathcal{S}_n$ denote the unbounded n-dimensional real manifold with Euclidean topology, and let the points of $\mathcal{S}_n$ be labeled by the real n-tuple of coordinates $\xi = (\xi^1, \cdots, \xi^n)$, $[-\infty < \xi^k < \infty]$; $\mathfrak{M}_n$, the real manifold defined in Sec. 2, is a certain compact subspace contained in $\mathcal{S}_n$. Associated with each unitary representation of $SL(3, R)$ given by (2.6) with (2.7) and (2.8) is an $SL(3, R)$ group of linear and homogeneous point transformations of $\mathcal{S}_n$,

$$\xi^l \to \bar{\xi}^l = T_k{}^l\xi^k \equiv (\exp ix^A G_A)\xi^l$$
$$\equiv (\exp - x^A \mathbf{M}_A)_k{}^l\xi^k$$
$$\equiv (\delta_k^l - x^A M_{Ak}{}^l + \tfrac{1}{2}x^A x^B M_{Ak}{}^i M_{Bi}{}^l + \cdots)\xi^k, \tag{B1}$$

in which the real $n \times n$ unimodular transformation matrix $T_k{}^l$ depends on the eight group parameters x^A but not on the manifold points ξ. In the following we analyze the geometrical possibilities for $\mathcal{S}_n$ and thus for $\mathfrak{M}_n$ admitted by the groups of point transformations (B1). The nature of such admissible *Riemannian geometries* for $\mathfrak{M}_n$ with $n \geq 8$ is of practical interest in connection with the reduction theory for irreducible representations outlined in Sec. 3 and of academic interest in connection with the Bargmann[6] treatment of $SL(2, R)$, where all representations are interpreted fundamentally as homeomorphic point transformations on a certain manifold.

For values of $n \geq 8$ we introduce a Riemannian metric into $\mathcal{S}_n$ with the line element $ds^2 = g_{ij}d\xi^i d\xi^j$. Then the point transformations (B1) can be viewed as a group of motions mapping $\mathcal{S}_n$ homeomorphically into itself with preservation of metrical structure provided that the metric tensor $g_{ij} = g_{ji}$ satisfies the Killing equations

$$\xi^k M_{Ak}{}^l \frac{\partial g_{ij}}{\partial \xi^l} + M_{Ai}{}^l g_{lj} + M_{Aj}{}^l g_{li} = 0. \tag{B2}$$

---

[8] Our expression (A4) for the cubic Casimir invariant agrees with the less explicit result obtained by L. C. Biedenharn, J. Math. Phys. **4**, 436 (1963). The latter reference, an exposition of Racah's construction method for obtaining bonafide and independent invariants for $SU(N)$, applies equally well to the noncompact $SL(N, R)$. Unfortunately, some of the recent literature on semisimple Lie groups [for example, A. Salam, in *Theoretical Physics* (International Atomic Energy Agency, Vienna, 1963), p. 178] still suggests the quantity $C^{AD}{}_E C^{BE}{}_F C^{CF}{}_D G_A G_B G_C$ as a useful cubic Casimir invariant, supposed to be functionally independent of (A3) for a semisimple Lie group of rank two. However, the latter quantity is actually proportional to (A3), for we have

$$C^{AD}{}_E C^{BE}{}_F C^{CF}{}_D G_A G_B G_C$$
$$= -\tfrac{1}{2}C^{ADE}(C^B{}_{EF}C^{FC}{}_D + C_D{}^B{}_F C^{FC}{}_E)G_A G_B G_C$$
$$= \tfrac{1}{2}C^{ADE}C_{EDF}C^{FCB}G_A G_B G_C$$
$$= -6C^{ABC}G_A G_B G_C = 36iG_A G^A,$$

where we have used the quadratic Lie identities, the definition part of (1.3), and a relation derived from (1.1) and (1.3), $C^{ABC}G_B G_C = -i\,6\,G^A$.

[9] S. Okubo, Progr. Theor. Phys. (Kyoto) **27**, 949 (1962), in particular, pp. 961–965. An interesting but less detailed derivation is given by H. Goldberg and Y. Lehrer-Ilamed, J. Math. Phys. **4**, 501 (1963).

The integrability conditions for these Killing equations are satisfied identically,[10] and thus the system of Eqs. (B2) is completely integrable. It follows from the general theory of linear partial differential equations that solutions of (B2) exist for all $n \geq 8$ and that the metric tensor for a solution of (B2) can be prescribed arbitrarily on a certain suitable $(n - m)$-dimensional subspace $\hat{\mathcal{S}}_{n-m}$ contained in $\mathcal{S}_n$, where m denotes the generic rank of the Killing vector array $\xi^i M_{Ai}{}^k$. (We have $m = 7$ for the $n = 8$ adjoint representation and m less than or equal to 8 for all representations with $n > 8$.) Furthermore, as a consequence of m being less than n for all representations with $n \geq 8$, the groups of motions are intransitive, there being certain invariant varieties (submanifolds in $\mathcal{S}_n$) which are mapped homeomorphically into themselves with preservation of their metrical structure under the point transformations (B1). Since the group is noncompact, the invariant varieties are manifestly open in $\mathcal{S}_n$, and no invariant variety can be wholly contained in the compact subspace $\mathfrak{M}_n$. From (B2) we obtain

$$\xi^k M_{Ak}{}^l \frac{\partial \psi}{\partial \xi^l} \equiv -iG_A \psi = 0, \tag{B3}$$

where

$$\psi = \psi(\xi) \equiv g_{ij}\xi^i\xi^j, \tag{B4}$$

and hence the hypersurfaces $\psi = \text{const}$ are generally invariant varieties. The additional conditions that the metric tensor be positive-definite and such that

$$g_{ij}(\lambda\xi) = g_{ij}(\xi), \qquad \det(g_{ij}) \equiv 1, \tag{B5}$$

are compatible with (B2) and do not alter the general integrability of these equations, in effect being restrictive conditions for the metric tensor prescribed (otherwise arbitrarily) in the $(n - m)$-dimensional subspace $\hat{\mathcal{S}}_{n-m}$.

To obtain an admissible positive-definite metric tensor which satisfies the conditions (B5), we first write the solution to (B2) in the integrated form

$$g_{ij}(\xi) = T_i{}^k T_j{}^l g_{kl}(\hat{\xi}), \tag{B6}$$

where $\hat{\xi}$ is related to ξ by the transformation matrix $T_k{}^l$ according to (B1). Now let the $(n - m)$-dimensional subspace $\hat{\mathcal{S}}_{n-m}$ be such that $(\lambda\hat{\xi})$ is contained in $\hat{\mathcal{S}}_{n-m}$ for all real $\lambda > 0$ if $\hat{\xi}$ is contained in $\hat{\mathcal{S}}_{n-m}$, and let the metric tensor take on the values

$$g_{kl}(\hat{\xi}) = \delta_{kl} \quad \text{for} \quad \hat{\xi} \text{ in } \hat{\mathcal{S}}_{n-m}. \tag{B7}$$

Then (B6) becomes

$$g_{ij}(\xi) = \sum_{k=1}^{n} T_i{}^k T_j{}^k \quad \text{for} \quad (T_k{}^l\xi^k) = \hat{\xi} \text{ in } \hat{\mathcal{S}}_{n-m} \tag{B8}$$

in which the point ξ is associated with a transformation matrix $T_k{}^l$ that maps ξ into a $\hat{\xi}$ in the "conical" subspace $\hat{\mathcal{S}}_{n-m}$. Let us introduce a smooth system of coordinates $(\omega_1, \cdots, \omega_{n-m})$ that covers $\hat{\mathcal{S}}_{n-m}$, the points in $\hat{\mathcal{S}}_{n-m}$ being prescribed by n differentiable functions $\hat{\xi}^l = \hat{\xi}^l(\omega_1, \cdots, \omega_{n-m})$ such that the rank of the matrix $(\partial\hat{\xi}^l/\partial\omega_i)$ equals $(n - m)$. By recalling (B1), we can then rewrite the metric tensor (B8) in an explicit parametric form

$$g_{ij} = \sum_{k=1}^{n} (\exp - x^A \mathbf{M}_A)_i{}^k (\exp - x^B \mathbf{M}_B)_j{}^k \tag{B9}$$

at the point

$$\xi^k = (\exp + x^A \mathbf{M}_A)_l{}^k \hat{\xi}^l(\omega_1, \cdots, \omega_{n-m}). \tag{B10}$$

Parametric redundancy in the latter equation with the n ξ's expressed in terms of $(n - m + 8)$ x's and ω's can be eliminated by a suitable definition for the domain of the x's which would involve $(8 - m)$ analytical functional constraints between them. With (B9) and (B10) the (positive-definite) function (B4) reduces to

$$\psi(\xi) = \sum_{k=1}^{n} [\hat{\xi}^k(\omega_1, \cdots, \omega_{n-m})]^2, \tag{B11}$$

where ξ is given by (B10). It is readily seen that the invariant variety $\psi = 0$ bounds the domain of definition for the positive-definite metric tensor (B9) and (B10), so that all points in $\mathcal{S}_n$ are not accessible to the form (B10), irrespective of how suitably one prescribes the $(n - m)$-dimensional conical subspace $\hat{\mathcal{S}}_{n-m}$. Generally, it is impossible to ascribe a positive-definite metric tensor globally throughout all of $\mathcal{S}_n$, or even globally throughout all of $\mathfrak{M}_n$, but it is still interesting that one can write down a positive-definite metric tensor (B9) for an open and unbounded subspace of $\mathcal{S}_n$, notwithstanding the noncompact character of $SL(3, R)$.

For the self-conjugate representations (necessarily with $n \geq 8$) we have

$$M_{Ak}{}^l = -\Lambda_{jk}\Lambda^{il} M_{Ai}{}^j, \tag{B12}$$

where $\Lambda_{ij} = \Lambda_{ji}$ is a nonsingular $n \times n$ matrix of real constants ($\Lambda^{ij} = \Lambda^{ji}$, the inverse matrix) and an immediate solution of (B2) is $g_{ij} = \Lambda_{ij}$. Such metric tensors Λ_{ij} are indefinite, and conversely, no positive-

[10] The computation specializes the proof of a theorem by L. Bianchi, *Lezioni sulla Teoria dei Gruppi Continui Finiti di Transformazioni* (Spoerri, Pisa, 1918), pp. 522–524. Formally, one applies the operator equations

$$\left[\xi^k M_{Ak}{}^h \frac{\partial}{\partial \xi^h}, \xi^q M_{Bq}{}^l \frac{\partial}{\partial \xi^l}\right] = C_{AB}{}^D \xi^k M_{Dk}{}^l \frac{\partial}{\partial \xi^l}$$

to g_{ij} and invokes (B2) to eliminate the partial derivatives of the metric tensor; the resulting algebraic equations in g_{ij} are satisfied identically.

definite metric tensor which satisfies (B2) and (B5) can be identically constant over $\mathfrak{M}_n$ for any of the $n \geq 8$ representations, self-conjugate or otherwise. As in the general case, an invariant variety associated with these admissible indefinite Riemannian geometries cannot be wholly contained in the compact subspace $\mathfrak{M}_n$; not only does the $SL(3, R)$ group of point transformations rule out the possibility of any compact invariant variety, but the noncompact character of the group also rules out the possibility of a one-parameter family of positive-definite and homogeneous invariant varieties [like (B11)] which pass through every point in $\mathfrak{M}_n$.

APPENDIX C: ALGEBRAIC MAKEUP OF THE QUADRATIC CASIMIR INVARIANT WITH RESPECT TO A LINEAR DECOMPOSITION

Let us consider the Casimir invariant (A3) in the form

$$(\mathrm{II}) = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4 + \Gamma_5, \tag{C1}$$

where the Γ's are the Hermitian operators

$$\begin{aligned} \Gamma_1 &\equiv G_1^2 - G_6^2, \qquad \Gamma_2 \equiv G_2^2 - G_7^2, \\ \Gamma_3 &\equiv G_3^2 - G_8^2, \qquad \Gamma_4 \equiv -G_4^2, \\ \Gamma_5 &\equiv -G_5^2. \end{aligned} \tag{C2}$$

The remarkable feature of the decomposition (C1) is that commutators of the Γ's involve only one independent operator, for by using Eqs. (1.1) and (1.4), we obtain

$$i[\Gamma_\alpha, \Gamma_\beta] = k_{\alpha\beta}\Delta, \tag{C3}$$

in which

$$(k_{\alpha\beta}) \equiv \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{C4}$$

and

$$\Delta \equiv 4(G_1G_7G_8 + G_2G_6G_8 + G_3G_6G_7 + G_3G_2G_1), \tag{C5}$$

the latter operator being Hermitian as a consequence of (C3) and the hermicity of the generators G_A. This Δ operator (C5) takes an interesting symmetrical form in terms of the linear combinations of generators (A1)

$$\Delta = \frac{8i}{3}(S_3^1\{S_1^2S_2^3\} + S_1^2\{S_2^3S_3^1\} + S_2^3\{S_3^1S_1^2\} - S_2^1\{S_1^3S_3^2\} - S_3^2\{S_2^1S_1^3\} - S_1^3\{S_3^2S_2^1\}), \tag{C6}$$

which is equivalent to

$$\Delta = \frac{8i}{3}(S_3^1\{S_1^aS_a^3\} + S_1^2\{S_2^aS_a^1\} + S_2^3\{S_3^aS_a^2\} - S_2^1\{S_1^aS_a^2\} - S_3^2\{S_2^aS_a^3\} - S_1^3\{S_3^aS_a^1\}). \tag{C7}$$

Equation (C7) shows that Δ is something of a "skew counterpart" to the cubic Casimir invariant (A4), even though Δ does not have the latter operator's marked property of commuting with all the generators. By using Eqs. (A2) to achieve a maximum reduction of (C6), we find

$$\Delta = 16i(S_1^2S_2^3S_3^1 - S_1^3S_3^2S_2^1). \tag{C8}$$

It follows immediately from (C8) that Δ vanishes for the unitary representation (2.6) with $n = 3$ and thus with S_b^a given by (4.16). More generally however, Δ is finite and although the Γ-commutators (C3) are strikingly simple compared to the G-commutators (1.1), higher-order commutators involving Γ_1, Γ_2, Γ_3, and Δ are considerably more complicated, with no closure property being evident for a finite number of operators generated by commutation.

Small-Angle X-Ray Scattering from Rods and Platelets*

PAUL W. SCHMIDT
Physics Department, University of Missouri, Columbia, Missouri
(Received 10 September 1965)

For highly elongated rods or flat platelets, there is a range of scattering angles in which the intensity of small-angle x-ray scattering cannot be conveniently approximated either by techniques normally used at small scattering angles or by the asymptotic expansion which is applicable in the outer part of the scattering curve. Expressions for the scattered intensity from rods and platelets have therefore been developed which can be used both at these intermediate scattering angles and also in the outer portion of the small-angle x-ray scattering curve.

INTRODUCTION

CALCULATIONS of the predicted scattered intensity are often useful for analyzing small-angle x-ray scattering data.

In a recent investigation of the small-angle x-ray scattering from an assembly of identical, independent, randomly oriented right cylinders with arbitrary cross section and uniform electron density,[1] an expression was developed which approximated the scattered intensity in the outer part of the scattering curve. When all dimensions of the cylinder are nearly equal in magnitude, this expression will be applicable at scattering angles which satisfy the condition $hL_{\min} \gg 1$, where $h = 4\pi\lambda^{-1} \sin(\phi/2)$, λ is the x-ray wavelength, ϕ is the scattering angle, and $L_{\min}$ is the smallest length characterizing the cylinder.

For highly elongated cylinders, which will be called rods, the asymptotic expansion of Ref. 1 is found to be useful only for h values which satisfy the more restrictive condition $h(L_{\min})^2/(vD) \ll 1$, where vD is the length of the cylinder and D is the maximum diameter of the cross section—that is, the length of the longest line that can be contained in the cross section. Since for a rod $L_{\min}/(vD) \ll 1$, the range of applicability of the asymptotic expansion of Ref. 1 is considerably reduced for rods.

A similar effect is found for very thin generalized cylinders, which will be referred to as platelets. In this case, the asymptotic expansion of Ref. 1 can be applied only when $hv^2D \gg 1$.

For greatly elongated or flattened cylinders, there is a range of h values such that for one or more dimensions L_i of the cylinders, $hL_i \gg 1$, even though the general asymptotic expansion of Ref. 1 cannot be applied. These values of h ordinarily are so large that expansions in powers of h are not practical. Special expressions for the scattered intensity in this angular range are therefore desirable.

Approximations for the scattered intensity have previously been developed for elongated rods[2,3] for the range of angles for which $hvD \gg 1$ and $hD < 1$. (Both these conditions can be satisfied only when $v \gg 1$.) While these relations are useful within the angular region in which they apply, the error cannot always be easily estimated. Also, it would be preferable to have expressions for the intensity which are useful both in this range of h and also for all larger values of h.

For thin platelets, only the limiting form of the expression for the scattering from generalized cylinders has been evaluated.[4]

In the calculations described below, expressions for the scattered intensity are developed which can be applied to rods and platelets at all h values for which, in the above notation, $hL_i \gg 1$ for certain lengths L_i, regardless of the magnitude of $hL_{\min}$. By use of expansions which are valid for all angles for which $hL_i \gg 1$, regardless of the magnitude of $hL_{\min}$, the analysis of experimental scattering data can often be simplified. Also, these approximate expressions for the scattered intensity often show general properties of the scattering process which could not be deduced by examination of tables or graphs of the calculated scattering for a few particle shapes.

PLATELETS

By rearrangement of Eq. 2 of Ref. 3, the average intensity $I(h)$ scattered by a randomly oriented platelet with thickness vD and maximum cross section, diameter D can be expressed

$$I(h) = \frac{2}{vD}\int_0^{vD} dx \left(1 - \frac{x}{vD}\right) M(h, x), \qquad (1)$$

where

$$M(h, x) = \frac{2\pi}{A}\int_x^{(x^2+D^2)^{\frac{1}{2}}} dr \frac{\sin hr}{h} \beta[(r^2 - x^2)^{\frac{1}{2}}],$$

* This work was supported by the National Science Foundation.

[1] P. W. Schmidt, J. Math. Phys. 6, 424 (1965).

[2] A. R. Stokes, Proc. Phys. Soc. (London) B70, 379 (1957).

[3] A. Miller and P. W. Schmidt, J. Math. Phys. 3, 92 (1962).

[4] G. Porod, Acta Phys. Austriaca 2, 270, 278 (1948–1949).

and where $\beta(r)$ and A are the characteristic function and the area of the cross section.

In order to obtain an expression for the scattered intensity, an approximation for $M(h, x)$ must first be developed. This approximation is found to depend on the behavior of $\beta(r)$ in the neighborhood of $r = 0$ and of all points $r = a_i$ at which derivatives of $\beta(r)$ are discontinuous. The available information about the properties of $\beta(r)$ suggests that most discontinuities in derivatives of $\beta(r)$ occur at r values which are called type-I and type-II points.[1] The a_i corresponding to type-I and type-II points are designated by a_{2i+1} and a_{2i}, respectively. Thus, odd and even indices on the a_i refer to type-I and type-II points, respectively.

In the neighborhood of a type-I point $r = a_{2i+1}$, $\beta(r)$ is assumed to be expressible in the form

$$\beta(r) = G_{2i+1}(r) + K_{2i+1}(r), \qquad r \geq a_{2i+1},$$
$$\beta(r) = G_{2i+1}(r), \qquad r \leq a_{2i+1},$$

where all derivatives of the function $G_{2i+1}(r)$ are continuous at $r = a_{2i+1}$, and

$$K_{2i+1}(r) = \sum_{n=0}^{\infty} A_n^{2i+1}\left[\left(\frac{r}{a_{2i+1}}\right)^2 - 1\right]^{n+\alpha_{2i+1}}.$$

At a type-II point $r = a_{2i}$, $\beta(r)$ is written

$$\beta(r) = G_{2i}(r) + K_{2i}(r), \qquad r \leq a_{2i},$$
$$\beta(r) = G_{2i}(r), \qquad r \geq a_{2i},$$

where all derivatives of $G_{2i}(r)$ are continuous at $r = a_{2i}$, and

$$K_{2i}(r) = \sum_{n=0}^{\infty} A_n^{2i}\left[1 - \left(\frac{r}{a_{2i}}\right)^2\right]^{n+\alpha_{2i}}.$$

The point $r = 0$, at which all derivatives of $\beta(r)$ ordinarily exist, is called a_0. Then for $i > 0$, none of the a_i are zero. The point $r = D$, which is always a type-II point, is called a_{2I}. With this notation, i ranges from 0 to $2I$.

Equation (1) requires an expression for $M(h, x)$ valid for $ha_i \gg 1$ for $i \geq 1$, regardless of whether hx is large or small. Unless $hx \gg 1$, however, the usual techniques for asymptotic expansion of Fourier integrals[5] are not applicable. A modified approximation technique is therefore necessary.

As is shown in the Appendix,

$$M(h, x) = W_L(h, x) + \sum_{i=1}^{2I} M_i(h, x) + E(h, x), \quad (2)$$

where

$$W_L(h, x) = \sum_{l=0}^{2L-2} W_l \frac{(hx)^{(l+1)/2}}{h^{l+2}}\left[\cos\left(\frac{l+1}{2}\pi\right)J_{(l+1)/2}(hx) - \sin\left(\frac{l+1}{2}\pi\right)N_{(l+1)/2}(hx)\right],$$

$$W_l = (2\pi/A)(\pi/2)^{\frac{1}{2}}(\beta^{(l)}(0)/l!)2^{l/2}\Gamma(l/2 + 1),$$

$$\beta^{(l)}(r) = d^l\beta/dr^l$$

$$M_i(h, x) = \sum_{n=0}^{N-1} \frac{d_{ni}[(1 + x^2/a_i^2)^{\frac{1}{2}}]}{(ha_i)^{\alpha_i+n+2}} \cos\left[h(x^2 + a_i^2)^{\frac{1}{2}} + \gamma_i + \frac{n\pi}{2}\right],$$

$$d_{ni}(z) = \sum_{j=0}^{n} D_{nij} z^{2j+\alpha_i-n},$$

$$D_{nij} = \frac{2\pi a_i^2(-1)^{(i+1)(j+1)}2^{2j+\alpha_i-n}\Gamma(n + 1 + \alpha_i)\Gamma(j + 1 + \alpha_i)}{A(n - j)!\Gamma(2j + \alpha_i + 1 - n)} A_j^i,$$

$$\gamma_i = (-1)^{i+1}\alpha_i\pi/2,$$

and where $J_l(x)$ and $N_l(x)$ are the Bessel functions of order l and of the first and second kinds, respectively. The A_j^i and the α_i are obtained from the series expansions of the $K_i(r)$ defined in the discussion of type-I and type-II points.

When (2) is useful, the quantity $E(h, x)$, which is discussed in the Appendix, is negligible. The integers N, L, and m are chosen to make $E(h, x)$ as small as possible, with the restrictions that $N \geq L + 1$ and $m > L - 1$.

When (2) is substituted into (1), $I(h)$ can be approximately expressed by the relation

$$I(h) = \sum_{i=0}^{2I} I_i(h), \quad (3)$$

where

$$I_0(h) = \frac{2\pi}{A} \sum_{l=0}^{2L-2} \frac{\beta^{(l)}(0)\, Y_l(hvD)}{h^{l+2}},$$

[5] A. Erdélyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956), p. 49.

$$Y_{2l}(x) = \frac{2(-1)^l}{x^2}\left[(2l+1) + \left(\frac{\pi}{2}\right)^{\frac{1}{2}} \frac{\Gamma(\frac{1}{2})\Gamma(l+1)}{\Gamma(l+\frac{1}{2})} \sum_{K=0}^{l} \frac{x^{K+\frac{1}{2}} N_{K+\frac{1}{2}}(x)}{2^K K!}\right],$$

$$Y_{2l+1}(x) = \frac{\pi(-1)^l \Gamma(l+\frac{3}{2})}{xl!\Gamma(\frac{3}{2})} \times \left[J_1(x) - \int_0^x dt\, J_0(t) + \sum_{K=0}^{l} \frac{\Gamma(\frac{3}{2})x^K}{\Gamma(K+\frac{3}{2})} \frac{J_{K+1}(x)}{2^K}\right],$$

and where, for $i \geq 1$,

$$I_i(h) = \sum_{n=0}^{N-1} (ha_i)^{-(n+\alpha_i+2)} I_{ni}(h),$$

$$I_{ni} = \frac{2}{vD} \int_0^{vD} dx \left(1 - \frac{x}{vD}\right) d_{ni}\left[\left(1 + \frac{x^2}{a_i^2}\right)^{\frac{1}{2}}\right] \times \cos\left(h(x^2 + a_i^2)^{\frac{1}{2}} + \gamma_i + \frac{n\pi}{2}\right).$$

In (3), the error will not exceed the maximum value of $|E(h, x)|$ in the interval $0 \leq x \leq vD$.

For $i \geq 1$, the $I_i(h)$ can be expanded in the series

$$I_i(h) = \sum_{n=0}^{N-1} \sum_{l=0}^{\infty} \left(\frac{hv^2 D^2}{2a_i}\right)^l X_{nl} \times \frac{\cos[ha_i + \gamma_i + \frac{1}{2}(n+l)\pi]}{(ha_i)^{n+\alpha_i+2}}, \quad (4)$$

where

$$X_{nl} = \frac{2^{l+1}}{l!} \int_0^1 dt\,(1-t) \frac{t^{2l}\, d_{ni}\,[(1 + t^2 v^2 D^2/a_i^2)^{\frac{1}{2}}]}{[1 + (1 + t^2 v^2 D^2/a_i^2)^{\frac{1}{2}}]^l}.$$

Equation (4) is convenient to use when the quantity

$$hv^2 D^2/2a_i$$

is not large. When only the $l = 0$ term in the sum in (4) is appreciable, the $I_i(h)$ represent the terms in the asymptotic expansion of the intensity scattered by a platelet with negligible thickness. When $hv^2D^2/(2a_i) \gg 1$, the scattering can be calculated from the asymptotic expansion developed in Ref. 1 or by asymptotic expansion of the integrals in (4). Information about this expansion is available from the author.

RODS

For an elongated rod, $v \gg 1$, and, as shown in Ref. 3, p. 95, for $hvD \gg 1$ the scattered intensity $I(h)$ can be conveniently expressed

$$I(h) = \pi(hvD)^{-1} I_a(h) - 2(hvD)^{-2} I_b(h) + I_c(h), \quad (5)$$

where

$$I_a(h) = \frac{2\pi}{A} \int_0^D x\, dx\, \beta(x) J_0(hx),$$

$$I_b(h) = \frac{2\pi}{A} \int_0^D dx\, x\beta(x) \cos hx,$$

$$I_c(h) = \frac{4\pi}{A(hvD)} \int_0^D x\, dx\, \beta(x) \times \left[\frac{\cos h(x^2 + v^2 D^2)^{\frac{1}{2}}}{hv D} - \int_{(v^2D^2+x^2)^{\frac{1}{2}}}^{\infty} \frac{dr \sin hr}{(r^2 - x^2)^{\frac{1}{2}}}\right].$$

When $hvD \gg 1$, the inner integral in $I_c(h)$ can be evaluated by an asymptotic expansion. Then by use of the relation

$$P_n(t) = C_n^{\frac{1}{2}}(t)$$

and Eq. (13) of the Appendix, $I_c(h)$ can be approximated by the expression

$$I_c(h) = \sum_{n=0}^{N-2} \frac{2(n+1)!(-1)^{n+1} T_n(h)}{(hv D)^{n+3}}, \quad (6)$$

where $P_n(t)$ is the Legendre polynomial of order n, and

$$T_n(h) = \frac{2\pi}{A} \int_0^D x\, dx\, \beta(x) P_{n+1}\left[\left(1 + \frac{x^2}{v^2 D^2}\right)^{\frac{1}{2}}\right] \times \sin\left[h(x^2 + v^2 D^2)^{\frac{1}{2}} + \frac{n\pi}{2}\right].$$

The $T_n(h)$ can be expanded in the convergent series

$$T_n(h) = \sum_{l=0}^{\infty} T_{nl} \left(\frac{hD}{2v}\right)^l \sin\left(hv D + \frac{n+l}{2}\pi\right), \quad (7)$$

where

$$T_{nl} = \frac{2\pi}{A} \frac{1}{l!} \int_0^D dx\, x P_{n+1}\left[\left(1 + \frac{x^2}{v^2 D^2}\right)^{\frac{1}{2}}\right] \times \beta(x) \left(\frac{x}{D}\right)^{2l} \frac{2^l}{[1 + (1 + x^2/v^2 D^2)^{\frac{1}{2}}]^l}.$$

If $hD/(2v)$ is so small that only the term for $l = 0$ is appreciable, then when h is set equal to 0 in $I_a(h)$ and $I_b(h)$ in (5), the asymptotic expansion for a rod with completely negligible cross section is obtained. Higher approximations can be calculated by including terms for larger l values. The resulting expressions are generalizations of the relations obtained by Stokes[2] and by Miller and Schmidt.[3]

When $ha_i^2/(2vD) \gg 1$ for all the a_i, the asymptotic expansion for a generalized cylinder[1] can be used instead of (5), or the $T_n(h)$ in (6) can be approximated by asymptotic expansions.

Let p be the largest value of any of the $|P_n(x)|$ for $n = 1, 2, 3, \cdots N$ and for $1 \leq x \leq (1 + v^{-2})^{\frac{1}{2}}$. Then

$$|T_n(h)| \leq p \tag{8}$$

since[3]

$$\frac{2\pi}{A}\int_0^D x\,dx\,\beta(x) = 1.$$

Inequality (8) is useful in setting bounds on the error in (6).

DISCUSSION

Equation (3) gives the intensity scattered by a platelet whenever $hD \gg 1$, regardless of whether hvD is large or small. The asymptotic expansions in Ref. 1, on the other hand, can be used only when $hv^2D^2/(2a_i) \gg 1$. For small hvD, (3) approaches the asymptotic expansion of the intensity scattered by a platelet with a completely negligible thickness. As hvD becomes larger, the thickness can no longer be considered negligible. For sufficiently large h, (3) is equivalent to the asymptotic expansion of Ref. 1.

Similarly, (5) gives the scattered intensity for a rod for all values of hD when $hvD \gg 1$. When hD is small, (5) approximates the asymptotic expansion for the scattering from a rod with negligible cross section. As hD becomes larger, the effects of the cross section become appreciable, and for sufficiently large hD, (5) should give the same results as the asymptotic expansion of Ref. 1.

In most cases in which (3) and (5) are useful for analysis of experimental scattering curves, only a few terms are necessary in each sum. When more terms are required, another type of approximation is usually preferable. The higher-order terms in (3) and (5), however, can be useful in predicting general properties of the scattering. An advantage of (3) and (5) in numerical calculations is that methods are provided for setting limits on the error in these approximate expressions for the scattered intensity. Consideration of the error is usually very important in numerical calculations of the intensity of small-angle x-ray scattering because of the relatively complex form of the expression for the intensity.

While (3) in principle gives the scattered intensity for any platelet, the expression is useful only for a thin platelet, for which there is a range of h values for which $hD \gg 1$ even though hvD is not large. Such a range of h values occurs only when $v \ll 1$. For larger v, the expansion of Ref. 1 is ordinarily preferable. Similarly, for rods, (5) is useful primarily when $v \gg 1$.

If a platelet has a cross section such that the smallest nonzero r value at which a derivative of $\beta(r)$ is discontinuous is a type-II point instead of a type-I point, the value of a_i in (3) can be any convenient r value such that $a_1 < a_2$. In this case, in (3), $I_1(h)$ is set equal to 0.

When hvD approaches zero, all of the $Y_{2l}(hvD)$ which appear in I_0 in (3) approach finite nonzero limiting values. On the other hand, for small hvD, the $Y_{2l+1}(hvD)$ are proportional to $hvD^{(2l+2)}$. Recent calculations by Schmidt[6] and by Kirste and Porod[7] suggest that for cross-section boundaries which are smooth, without corners, $\beta_{2l}(0) = 0$ for $l > 0$. If this result holds, then if the cross section has a smooth boundary, I_0 in (3) will contain only $Y_0(hvD)$ and the Y_{2l+1}. The latter functions are proportional to hvD^{2l+2} for small hvD. Thus, when the platelet thickness is small but not negligible, one can expect that for a given value of hvD, I_0 can be better approximated by only the $l = 0$ term for a smooth cross-section boundary than for a boundary with corners.

APPENDIX

The function $M(h, x)$, which is defined in (1), can be written

$$M(h, x) = M_0(h, x) + U_1(h, x) + U_2(h, x), \tag{9}$$

where

$$M_0(h, x) = \frac{2\pi}{Ah}\sum_{l=0}^{2L-2}\frac{x^{l+1}\beta^{(l)}(0)}{l!}\,Q_l\left(hx, \frac{a_1}{x}\right),$$

$$U_1(h, x) = \frac{2\pi}{Ah}\int_{(x^2+a_1^2)^{\frac{1}{2}}}^{(x^2+D^2)^{\frac{1}{2}}} dr \sin hr\beta[(r^2 - x^2)^{\frac{1}{2}}],$$

$$U_2(h, x) = \frac{2\pi}{Ah}\int_x^{(x^2+a_1^2)^{\frac{1}{2}}} dr \sin hr S_{L-1}[(r^2 - x^2)^{\frac{1}{2}}],$$

$$Q_l(y, x) = \int_1^{(1+x^2)^{\frac{1}{2}}} dt\,(t^2 - 1)^{l/2} \sin yt,$$

$$S_L(x) = \beta(x) - \sum_{l=0}^{2L}\frac{\beta^{(l)}(0)y^l}{l!},$$

$$\beta^{(l)}(x) = d^l\beta/dx^l.$$

By application of the usual techniques for asymptotic expansion of Fourier integrals,[5] $U_1(h, x)$ can be expressed

$$U_1(h, x) = E_1(h, x) + B_1(h, x) + \sum_{i=1}^{2I} M_i(h, x), \tag{10}$$

[6] P. W. Schmidt, *Proceedings of the Conference on Small Angle X-Ray Scattering, Syracuse, New York, June, 1965*, edited by H. Brumberger (Gordon and Breach Science Publishers, Inc., New York, in press.)

[7] R. Kirste and G. Porod, Kolloid Zeits. **184**, 1 (1962).

where $E_1(h, x)$ is the error involved in the asymptotic expansion of $U_1(h, x)$ and

$$B_1(h, x) = \frac{2\pi}{A}\sum_{n=0}^{N-1}\frac{\cos\left[h(x^2+a_1^2)^{\frac{1}{2}}+\frac{1}{2}(n\pi)\right]}{h^{n+2}} \times \frac{d^n G_1[(r^2-x^2)^{\frac{1}{2}}]}{dr^n}\bigg|_{r=(x^2+a_1^2)^{\frac{1}{2}}}.$$

The $M_i(h, x)$ are defined in (2).

For $l = -1$, $Q_l(h, x)$ can be written[8]

$$Q_l(y, z) = \left(\frac{\pi}{2}\right)^{\frac{1}{2}} 2^{l/2}\Gamma\left(1+\frac{l}{2}\right) y^{-(l+1)/2} J_{-(l+1)/2}(y) - R_{ml}[y, (1+z^2)^{\frac{1}{2}}], \quad (11)$$

where $J_k(x)$ is the Bessel function of order k and of the first kind, and

$$R_{ml}(y, z) = \sum_{n=0}^{m}\frac{F_{l/2}^{(n)}(z)\cos(yz+n\pi/2)}{y^{n+1}} + \frac{1}{y^{m+1}}\int_z^\infty dt\, F^{(m+1)}(t)\sin\left(yt+\frac{m+1}{2}\pi\right),$$

$$F_{l/2}^{(m)}(z) = d^m(z^2-1)^{l/2}/dz^m.$$

By use of an integral representation of Bessel functions,[8] (11) is found to be true for $l = 0$ and $m \geq 0$. By differentiation, (11) then can be shown to hold whenever $l \geq 0$, provided that $m \geq l$.

According to Eq. (11), $M_0(h, x)$ can be written

$$M_0(h, x) = W_2(h, x) - \frac{2\pi}{hA}\sum_{l=0}^{2L-2}\frac{\beta^{(l)}(0)}{l!}x^{l+1}R_{ml}\left[hx, \left(1+\frac{a_1^2}{x^2}\right)^{\frac{1}{2}}\right].$$

The $W_L(h, x)$ are defined in Eq. (2).

The first $L-1$ derivatives of the integrand of $U_2(h, x)$ in (9) will be continuous throughout the entire interval of integration, since by hypothesis a_1 is a type-I point. Therefore, $U_2(h, x)$ can be integrated L times by parts, giving

$$U_2(h, x) = U_3(h, x) - \frac{2\pi}{A}\sum_{l=0}^{L-1}\frac{\cos\left[h(x^2+a_1^2)^{\frac{1}{2}}+l\pi/2\right]}{h^{l+2}} \times \frac{d^l S_{L-1}[(r^2-x^2)^{\frac{1}{2}}]}{d_r^l}\bigg|_{r=(x^2+a_1^2)^{\frac{1}{2}}}$$

where

$$U_3(h, x) = \frac{2\pi}{Ah^{L+1}}\int_x^{(x^2+a_1^2)^{\frac{1}{2}}} dr\,\frac{d^L S_{L-1}[(r^2-x^2)^{\frac{1}{2}}]}{dr^L} \times \sin\left(hr+\frac{L\pi}{2}\right).$$

Thus, assuming that $N-1 \geq L$,

$$M(h, x) = W_L(h, x) + \sum_{i=1}^{2I} M_i(h, x) + E(h, x), \quad (12)$$

where

$$E(h, x) = E_1(h, x) + E_2(h, x) + E_3(h, x) + E_4(h, x)$$

with $E_1(H, x)$ being given by (10) and with

$$E_2(h, x) = \frac{2\pi}{A}\sum_{n=L}^{N-1}\frac{\cos\left[h(x^2+a_1^2)^{\frac{1}{2}}+n\pi/2\right]}{h^{n+2}} \times \frac{d^n G_1[(r^2-x^2)^{\frac{1}{2}}]}{dr^n}\bigg|_{r=(x^2+a_1^2)^{\frac{1}{2}}},$$

$$E_3(h, x) = -\frac{2\pi}{A}\sum_{l=0}^{2L-2}\frac{\beta^{(l)}(0)}{l!} \times \sum_{n=L}^{m}\frac{d^n[(r^2-x^2)^{\frac{1}{2}l}]}{dr^n}\bigg|_{r=(x^2+a_1^2)^{\frac{1}{2}}} \times \frac{\cos\left[h(x^2+a_1^2)^{\frac{1}{2}}+n\pi/2\right]}{h^{n+2}},$$

$$E_4(h, x) = U_3(h, x) - \frac{2\pi}{A}\sum_{l=0}^{2L-2}\frac{\beta^{(l)}(0)}{l!h^{m+2}} \times \int_{(x^2+a_1^2)^{\frac{1}{2}}}^\infty dr\sin\left(hr+\frac{m+1}{2}\pi\right)\frac{d^{m+1}(r^2-x^2)^{\frac{1}{2}l}}{dr^{m+1}}.$$

The value of h will be assumed to be large enough that $E(h, x)$ is negligible. From the form of $E(h, x)$, one can show that

$$|E(h, x)| < Bh^{-L-\frac{1}{2}},$$

where B is a constant.

The integers N, L, and m can be chosen to give the smallest value of $E(h, x)$, subject to the restrictions that $N \geq L+1$ and $m > L-1$. If $N = L$, $E_2(h, x) = 0$. However, $E_3(h, x)$ involves derivatives of $(r^2-x^2)^{\frac{1}{2}l}$ of order as high as $2L-2$. Some of these derivatives may be so large that the net error $E(h, x)$ may be smaller if $N > L$ than if $N = L$, since if $N > L$, some of the terms in $E_2(h, x)$ and $E_3(h, x)$ may tend to cancel each other. In numerical calculation, $E(h, x)$ should be examined in detail, to find the optimum choices of N, L, and M.

The magnitudes of the terms in $E_3(h, x)$ can often be conveniently estimated by writing r^2-x^2 in the form

$$r^2-x^2 = a_1^2\left\{1+2\left[\frac{r}{a_1}-\left(1+\frac{x^2}{a_1^2}\right)^{\frac{1}{2}}\right] \times \left(1+\frac{x^2}{a_1^2}\right)^{\frac{1}{2}}+\left[\frac{r}{a_1}-\left(1+\frac{x^2}{a_1^2}\right)^{\frac{1}{2}}\right]^2\right\}.$$

[8] W. Magnus and F. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics* (Chelsea Publishing Company, New York, 1954), p. 27.

Then[9]

$$\left.\frac{d^n (r^2 - x^2)^{l/2}}{dr^n}\right|_{r=(x^2+a_1^2)^{\frac{1}{2}}} = (-1)^n n! a_1^{l-n} C_n^{-l/2}\left[\left(1 + \frac{x^2}{a_1^2}\right)^{\frac{1}{2}}\right], \quad (13)$$

[9] Reference 8, p. 76.

where the $C_n^r(t)$ are the Gegenbauer functions. For platelets, $x \ll a_1$, and so

$$C_n^{-l/2}[(1 + x^2/a_1^2)^{\frac{1}{2}}] \approx C_n^{-l/2}(1)$$

for $n > l$, $C_n^{-l/2}(1) = 0$, while for $n \leq l$,

$$C_n^{-l/2}(1) = (-1)^n (l)![n!(l-n)!]^{-1}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 7 JULY 1966

On the Chakrabarti Transformation

J. SESMA
Escuela de Física y Matemáticas. Universidad Central de Venezuela. Caracas, Venezuela
(Received 19 October 1965)

In this note, we intend to present an appropriate Hamiltonian formulation for the extension of the Chakrabarti transformation [A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963)] to spin-0 and spin-1 particles. The suggested formulation is derived from that of Duffin–Kemmer and has been used previously [L. M. Garrido and P. Pascual, Nuovo Cimento **12**, 181 (1959); L. M. Garrido and J. Sesma, Am. J. Phys. **30**, 887 (1962); J. Sesma, J. Biel, and L. M. Garrido, Am. J. Phys. **32**, 559 (1964)] to achieve a generalization of the Foldy–Wouthuysen transformation.

INTRODUCTION

The relativistic spin-$\frac{1}{2}$ free-particle equation

$$H\varphi \equiv (\alpha \mathbf{p} + \beta\kappa)\varphi = i\,\partial\varphi/\partial t, \quad (1.1)$$

known as the Dirac equation, takes on the canonical form

$$\beta E \varphi_{\mathrm{F-W}} = i\,\partial\varphi_{\mathrm{F-W}}/\partial t \quad (1.2)$$

when the Foldy–Wouthuysen transformation[1] is performed. Instead, Chakrabarti proposes[2] another transformation which brings the explicitly covariant equation

$$(\gamma \cdot p - \kappa)\varphi = 0 \quad (1.3)$$

to the covariant canonical form

$$(\gamma^0 m - \kappa)\varphi_Q = 0. \quad (1.4)$$

This transformation is general, valid for arbitrary spin, and is the most appropriate for definition of relativistic polarization and position operators.[2,3] However, in the absence of a general Hamiltonian formulation, Chakrabarti does not show the transformation in the case of spin 1, and discusses the expression of particular operators only in the case of spin-$\frac{1}{2}$ particles. In this paper we treat the Chakrabarti transformation by using a general formulation for particles of spin 0, $\frac{1}{2}$, and 1, that permits a very natural extension of the Foldy–Wouthuysen transformation to spin-0 and -1 particles.[4,5]

[1] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
[2] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).
[3] A. Chakrabarti, J. Math. Phys. **4**, 1223 (1963).

2. GENERAL FORMULATION

The general relativistic wave equation is[6]

$$(\beta_\mu\,\partial_\mu + \kappa)\varphi = 0, \quad (2.1)$$

where the β_μ are 5×5 and 10×10 matrices for spin 0 and 1, respectively, which obey the Duffin–Kemmer algebra, while for spin $\frac{1}{2}$, they are 4×4 matrices obeying the Dirac algebra. By using the notation

$$\gamma^0 = \beta_4 \equiv \beta, \qquad \gamma_k = i\beta_k, \quad (2.2)$$

and the relations

$$p^0 = i\partial^0 = -\partial_4, \qquad p_k = -i\partial_k, \quad (2.3)$$

[4] L. M. Garrido and P. Pascual, Nuovo Cimento **12**, 181 (1959).
[5] L. M. Garrido and J. Sesma, Am. J. Phys. **30**, 887 (1962).
[6] H. Umezawa, *Quantum Field Theory* (North-Holland Publishing Company, Amsterdam, 1956), Chap. V.

Then[9]

$$\left.\frac{d^n (r^2 - x^2)^{l/2}}{dr^n}\right|_{r=(x^2+a_1^2)^{\frac{1}{2}}} = (-1)^n n! a_1^{l-n} C_n^{-l/2}\left[\left(1 + \frac{x^2}{a_1^2}\right)^{\frac{1}{2}}\right], \quad (13)$$

[9] Reference 8, p. 76.

where the $C_n^r(t)$ are the Gegenbauer functions. For platelets, $x \ll a_1$, and so

$$C_n^{-l/2}[(1 + x^2/a_1^2)^{\frac{1}{2}}] \approx C_n^{-l/2}(1)$$

for $n > l$, $C_n^{-l/2}(1) = 0$, while for $n \leq l$,

$$C_n^{-l/2}(1) = (-1)^n (l)![n!(l-n)!]^{-1}.$$

JOURNAL OF MATHEMATICAL PHYSICS VOLUME 7, NUMBER 7 JULY 1966

On the Chakrabarti Transformation

J. SESMA
Escuela de Física y Matemáticas. Universidad Central de Venezuela. Caracas, Venezuela
(Received 19 October 1965)

In this note, we intend to present an appropriate Hamiltonian formulation for the extension of the Chakrabarti transformation [A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963)] to spin-0 and spin-1 particles. The suggested formulation is derived from that of Duffin–Kemmer and has been used previously [L. M. Garrido and P. Pascual, Nuovo Cimento **12**, 181 (1959); L. M. Garrido and J. Sesma, Am. J. Phys. **30**, 887 (1962); J. Sesma, J. Biel, and L. M. Garrido, Am. J. Phys. **32**, 559 (1964)] to achieve a generalization of the Foldy–Wouthuysen transformation.

INTRODUCTION

The relativistic spin-$\frac{1}{2}$ free-particle equation

$$H\varphi \equiv (\alpha \mathbf{p} + \beta\kappa)\varphi = i\,\partial\varphi/\partial t, \quad (1.1)$$

known as the Dirac equation, takes on the canonical form

$$\beta E \varphi_{\mathrm{F-W}} = i\,\partial\varphi_{\mathrm{F-W}}/\partial t \quad (1.2)$$

when the Foldy–Wouthuysen transformation[1] is performed. Instead, Chakrabarti proposes[2] another transformation which brings the explicitly covariant equation

$$(\gamma \cdot p - \kappa)\varphi = 0 \quad (1.3)$$

to the covariant canonical form

$$(\gamma^0 m - \kappa)\varphi_Q = 0. \quad (1.4)$$

This transformation is general, valid for arbitrary spin, and is the most appropriate for definition of relativistic polarization and position operators.[2,3] However, in the absence of a general Hamiltonian formulation, Chakrabarti does not show the transformation in the case of spin 1, and discusses the expression of particular operators only in the case of spin-$\frac{1}{2}$ particles. In this paper we treat the Chakrabarti transformation by using a general formulation for particles of spin 0, $\frac{1}{2}$, and 1, that permits a very natural extension of the Foldy–Wouthuysen transformation to spin-0 and -1 particles.[4,5]

[1] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950).
[2] A. Chakrabarti, J. Math. Phys. **4**, 1215 (1963).
[3] A. Chakrabarti, J. Math. Phys. **4**, 1223 (1963).

2. GENERAL FORMULATION

The general relativistic wave equation is[6]

$$(\beta_\mu\,\partial_\mu + \kappa)\varphi = 0, \quad (2.1)$$

where the β_μ are 5×5 and 10×10 matrices for spin 0 and 1, respectively, which obey the Duffin–Kemmer algebra, while for spin $\frac{1}{2}$, they are 4×4 matrices obeying the Dirac algebra. By using the notation

$$\gamma^0 = \beta_4 \equiv \beta, \qquad \gamma_k = i\beta_k, \quad (2.2)$$

and the relations

$$p^0 = i\partial^0 = -\partial_4, \qquad p_k = -i\partial_k, \quad (2.3)$$

[4] L. M. Garrido and P. Pascual, Nuovo Cimento **12**, 181 (1959).
[5] L. M. Garrido and J. Sesma, Am. J. Phys. **30**, 887 (1962).
[6] H. Umezawa, *Quantum Field Theory* (North-Holland Publishing Company, Amsterdam, 1956), Chap. V.

the general equation can be given the explicitly covariant form

$$(\gamma \cdot p - \kappa)\varphi = 0, \tag{2.4}$$

as proposed by Chakrabarti.[2] This equation also admits a Hamiltonian form which for spin-$\frac{1}{2}$ particles is the well-known Dirac equation, and for spin 0 and 1 the equation obtained by Case.[7] In both cases we have the same expression,

$$(\alpha \mathbf{p} + \beta\kappa)\phi = E\phi, \tag{2.5}$$

for the Hamiltonian eigenvalue equation.

The operator that achieves the transformation of Eq. (1.3) into (1.4) is

$$Q = \exp\{-iS_{k4}p_k[\tanh^{-1}(p/m)]/p\}, \tag{2.6}$$

where $m = [p^\nu p_\nu]^{\frac{1}{2}}$ is the mass operator, and p does not represent the 4-vector momentum, but

$$p = [\mathbf{pp}]^{\frac{1}{2}}, \tag{2.7}$$

The $S_{\mu\nu}$ are the generators of infinitesimal Lorentz transformations, as defined in Ref. 5. In fact,

$$S_{k4} = -f\alpha_k, \tag{2.8}$$

where f is a numerical factor:

$$f = \tfrac{1}{2}, \quad \text{for spin } \tfrac{1}{2},$$
$$f = 1, \quad \text{for spin 0 and 1.} \tag{2.9}$$

In the case of Duffin–Kemmer particles, the operator (2.6) can be written as

$${}_{0,1}Q = 1 - \alpha\mathbf{p}/m + (\alpha\mathbf{p})^2/m(p^0 + m), \tag{2.10}$$

and in the Dirac case, (2.6) reduces to

$${}_{\frac{1}{2}}Q = (p^0 + m - \alpha\mathbf{p})/[2m(p^0 + m)]^{\frac{1}{2}}, \tag{2.11}$$

an expression that can easily be brought to the form proposed by Chakrabarti.

By applying the operator Q, we pass to a new representation, which is referred to as the Chakrabarti representation. In it, the general covariant equation assumes the form

$$(\gamma^0 m - \kappa)\varphi_Q = 0, \tag{2.12}$$

where $\varphi_Q = Q\varphi$ is the wavefunction in the new representation.

3. OPERATORS IN THE CHAKRABARTI REPRESENTATION

We can now give the explicit form of the operators discussed in Refs. 2 and 3. The expressions that we present are simultaneously valid for spin-0, -$\frac{1}{2}$, and -1 particles. The single differences is in the algebra obeyed by the matrices that appear. As we have mentioned, Dirac algebra must be used in the case of spin $\frac{1}{2}$, and Duffin–Kemmer algebra for spin-0 and -1 particles.

The infinitesimal operators of the Poincaré group

$$p = p, \tag{3.1a}$$
$$\mathbf{M} = \mathbf{x} \times \mathbf{p} + f\boldsymbol{\sigma}, \tag{3.1b}$$
$$\mathbf{N} = (x^0\mathbf{p} - \mathbf{x}p^0) + if\boldsymbol{\alpha}, \tag{3.1c}$$

become in the new representation

$$p_Q = p, \tag{3.2a}$$
$$\mathbf{M}_Q = \mathbf{x} \times \mathbf{p} + f\boldsymbol{\sigma} = \mathbf{M}, \tag{3.2b}$$
$$\mathbf{N}_Q = (x^0\mathbf{p} - \mathbf{x}p^0) + f(\boldsymbol{\sigma} \times \mathbf{p})/(p^0 + m). \tag{3.2c}$$

A mean spin operator in the usual representation, $\boldsymbol{\Sigma}$, can be defined as that which becomes $\boldsymbol{\sigma}$ in the Chakrabarti representation; i.e.,

$$\boldsymbol{\Sigma} = Q^{-1}\boldsymbol{\sigma}Q$$
$$= (p^0/m)\boldsymbol{\sigma} + (i/m)\boldsymbol{\alpha} \times \mathbf{p} - [1/m(p^0 + m)](\boldsymbol{\sigma}\mathbf{p})\mathbf{p}. \tag{3.3}$$

This operator, valid for spin 0, $\frac{1}{2}$, and 1, commutes with the Hamiltonian and, as shown in Ref. 2, becomes the most appropriate for the definition of a relativistic polarization operator.

It is easy to prove that the Chakrabarti transformation in its general form does not mix opposite chirality states. In fact, the operator

$$\gamma^5 = -(1/4!)e^{\mu\nu\lambda\rho}\gamma_\mu\gamma_\nu\gamma_\lambda\gamma_\rho, \tag{3.4}$$

which enables one to construct the projectors, $\frac{1}{2}(1 \pm i\gamma^5)$ on the positive and negative chirality states, is invariant under the transformation. [In the Dirac case, (3.4) reduces to $\gamma^5 = -\gamma^0\gamma^1\gamma^2\gamma^3$.]

Finally, in a manner analogous to the definition of mean spin, a mean position operator can be considered. Its expression and properties for spin $\frac{1}{2}$ are discussed in Ref. 3. This operator,

$$X = Q^{-1}xQ, \tag{3.5}$$

has two different expressions:

$$X^0 = x^0 - i(f/m^2)\alpha\mathbf{p}, \tag{3.6a}$$
$$\mathbf{X} = \mathbf{x} - i(f/m)\boldsymbol{\alpha} + [f/m(p^0 + m)] \times \{\boldsymbol{\sigma} \times \mathbf{p} - (i/m)(\alpha\mathbf{p})\mathbf{p}\}, \tag{3.6b}$$

when Q is expressed in terms of

$$p^0 = i\,\partial/\partial x^0, \qquad m = \pm[p^\nu p_\nu]^{\frac{1}{2}}, \tag{3.7}$$

[7] K. M. Case, Phys. Rev. 100, 1513 (1955).

and

$$X''^0 = x^0, \tag{3.8a}$$

$$\mathbf{X}'' = \mathbf{x} - i(f/m)\boldsymbol{\alpha} + [f/m(p^0 + m)] \times \{\boldsymbol{\sigma}\times\mathbf{p} + (i/p^0)(\boldsymbol{\alpha}\mathbf{p})\mathbf{p}\}, \tag{3.8b}$$

if in (2.6) one takes

$$p^0 = \pm E = \pm[\mathbf{p}^2 + \kappa^2]^{\frac{1}{2}}, \qquad m = \pm\kappa. \tag{3.9}$$

The velocities associated with these operators,

$$\langle\dot{\mathbf{X}}\rangle = \langle\beta\mathbf{p}/\kappa\rangle, \tag{3.10a}$$

$$\langle\dot{\mathbf{X}}''\rangle = \langle\mathbf{p}/p^0\rangle, \tag{3.10b}$$

have the same form in both the Dirac and the Duffin–Kemmer cases. Their physical properties are, equally, the same as discussed in Ref. 3.

ACKNOWLEDGMENTS

I wish to thank Professor Louis Michel for hospitality at the Centre de Physique Théorique de l'Ecole Polytechnique, Paris, where I became interested in the Chakrabarti transformation.

This note has been presented at the Fifteenth Convention of ASOVAC (Asociación Venezolana para el Avance de la Ciencia) in Caracas, May 1965.

APPENDIX

We here give some of the relations used in obtaining the preceding expressions.

(a) *Dirac Algebra.* Besides the more familiar properties of the α and γ matrices, we have made use of

$$(\boldsymbol{\alpha}\mathbf{p})\boldsymbol{\alpha} = \mathbf{p} + i(\boldsymbol{\sigma}\times\mathbf{p}), \tag{A1}$$

$$[\boldsymbol{\sigma}, (\boldsymbol{\alpha}\mathbf{p})] = -2i(\boldsymbol{\alpha}\times\mathbf{p}), \tag{A2}$$

$$(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\alpha}\times\mathbf{p}) = i\{(\boldsymbol{\sigma}\mathbf{p})\mathbf{p} - \mathbf{p}^2\boldsymbol{\sigma}\}, \tag{A3}$$

$$[\boldsymbol{\alpha}, (\boldsymbol{\alpha}\mathbf{p})] = -2i(\boldsymbol{\sigma}\times\mathbf{p}), \tag{A4}$$

(b) *Duffin–Kemmer Algebra.* The relations used are some of those mentioned in Ref. 4 and the following:

$$[\beta, (\boldsymbol{\alpha}\mathbf{p})] = i(\boldsymbol{\beta}\mathbf{p}), \tag{A5}$$

$$[\beta, (\boldsymbol{\alpha}\mathbf{p})^2] = 2(\boldsymbol{\beta}\mathbf{p})^2\beta - \mathbf{p}^2\beta, \tag{A6}$$

$$[\boldsymbol{\beta}, (\boldsymbol{\alpha}\mathbf{p})] = -i\beta\mathbf{p}, \tag{A7}$$

$$[\boldsymbol{\beta}, (\boldsymbol{\alpha}\mathbf{p})^2] = (\boldsymbol{\beta}\mathbf{p})\{1 - 2\beta^2\}\mathbf{p}, \tag{A8}$$

$$(\boldsymbol{\alpha}\mathbf{p})\beta = -i(\boldsymbol{\beta}\mathbf{p})\beta^2, \tag{A9}$$

$$(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\beta}\mathbf{p}) = i\{\mathbf{p}^2 - (\boldsymbol{\beta}\mathbf{p})^2\}\beta, \tag{A10}$$

$$(\boldsymbol{\alpha}\mathbf{p})^2\beta = \{\mathbf{p}^2 - (\boldsymbol{\beta}\mathbf{p})^2\}\beta \tag{A11}$$

$$(\boldsymbol{\alpha}\mathbf{p})^2(\boldsymbol{\beta}\mathbf{p}) = \mathbf{p}^2(\boldsymbol{\beta}\mathbf{p})\beta^2, \tag{A12}$$

$$\beta(\boldsymbol{\alpha}\mathbf{p}) = i(\boldsymbol{\beta}\mathbf{p})\{1 - \beta^2\}, \tag{A13}$$

$$(\boldsymbol{\alpha}\mathbf{p})\beta(\boldsymbol{\alpha}\mathbf{p}) = 0, \tag{A14}$$

$$(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\beta}\mathbf{p})(\boldsymbol{\alpha}\mathbf{p}) = 0, \tag{A15}$$

$$[\boldsymbol{\alpha}, (\boldsymbol{\alpha}\mathbf{p})] = i(\mathbf{p}\times\boldsymbol{\sigma}), \tag{A16}$$

$$[\boldsymbol{\alpha}, (\boldsymbol{\alpha}\mathbf{p})^2] = i\{(\mathbf{p}\times\boldsymbol{\sigma})(\boldsymbol{\alpha}\mathbf{p}) + (\boldsymbol{\alpha}\mathbf{p})(\mathbf{p}\times\boldsymbol{\sigma})\}, \tag{A17}$$

$$[\boldsymbol{\sigma}, (\boldsymbol{\alpha}\mathbf{p})] = -i(\boldsymbol{\alpha}\times\mathbf{p}), \tag{A18}$$

$$[\boldsymbol{\sigma}, (\boldsymbol{\alpha}\mathbf{p})^2] = -i\{(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\alpha}\times\mathbf{p}) + (\boldsymbol{\alpha}\times\mathbf{p})(\boldsymbol{\alpha}\mathbf{p})\}, \tag{A19}$$

$$[(\boldsymbol{\alpha}\times\mathbf{p}), (\boldsymbol{\alpha}\mathbf{p})] = i\{\mathbf{p}^2\boldsymbol{\sigma} - (\boldsymbol{\sigma}\mathbf{p})\mathbf{p}\}, \tag{A20}$$

$$(\boldsymbol{\alpha}\mathbf{p})\boldsymbol{\alpha}(\boldsymbol{\alpha}\mathbf{p}) = (\boldsymbol{\alpha}\mathbf{p})\mathbf{p}, \tag{A21}$$

$$(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\alpha}\times\mathbf{p})(\boldsymbol{\alpha}\mathbf{p}) = 0, \tag{A22}$$

$$(\boldsymbol{\alpha}\mathbf{p})\boldsymbol{\sigma}(\boldsymbol{\alpha}\mathbf{p}) = (\boldsymbol{\alpha}\mathbf{p})^2\boldsymbol{\sigma} - i(\boldsymbol{\alpha}\mathbf{p})(\boldsymbol{\alpha}\times\mathbf{p}), \tag{A23}$$

$$\beta(\boldsymbol{\alpha}\mathbf{p})^2 + (\boldsymbol{\alpha}\mathbf{p})^2\beta = \mathbf{p}^2\beta, \tag{A24}$$

$$(\boldsymbol{\beta}\mathbf{p})(\boldsymbol{\alpha}\mathbf{p})^2 + (\boldsymbol{\alpha}\mathbf{p})^2(\boldsymbol{\beta}\mathbf{p}) = \mathbf{p}^2(\boldsymbol{\beta}\mathbf{p}), \tag{A25}$$

$$(\boldsymbol{\alpha}\times\mathbf{p})\boldsymbol{\beta} = -i(\boldsymbol{\beta}\times\mathbf{p})\beta^2, \tag{A26}$$

Global Covariant Conservation Laws in Riemannian Spaces. I

BOHDAN SHEPELAVEY
General Electric Company, HMED, Syracuse, New York
(Received 1 July 1965)

It is demonstrated that in the theory of general relativity integral conservation laws can be obtained from a vanishing covariant divergence of a tensor in a completely covariant form. This is achieved by introducing into tensor calculus a new operation, namely, *tensor integration* which is defined as an inverse of tensor differentiation. In particular, a one-dimensional absolute integration of a tensor along a curve is defined as an inverse operation to the same type of differentiation. A representation of absolute integration is developed by a perturbation method as an infinite series where each term consists of ordinary integrations only. As an example of absolute integration, a vector field is obtained whose components can be employed as a coordinate frame having a very close resemblance to the Riemannian coordinates. Covariant integration is then introduced as an inverse of covariant differentiation; however, its usefulness is severely limited by the conditions of integrability which have to be satisfied to make covariant integration possible. A set of unspecified covariantly constant base vectors is used to explain the idea of a covariantly constant tensor and to express symbolically absolute integration in terms of the ordinary integration. The one-dimensional absolute integration is then extended to higher dimensions in such a way that it is independent of the order of integrations. Finally, Gauss' theorem is proved for the absolute integration which enables one to convert a volume integral of a covariant divergence of a tensor into a corresponding surface integral of the same tensor.

1. INTRODUCTION

IN generally covariant field theories[1] the requirement of general covariance has been successfully imposed on the local level. Thus, both the field equations and the vanishing divergence of the energy-momentum tensor can always be stated in a covariant language. However, attempts to advance these theories to the global level presented many difficulties which could be overcome in some measure by sacrificing the requirement of general covariance. In the theory of general relativity, noncovariant coordinate constraints[2] are sometimes admitted in order to be able to integrate the equations of motion. A better example, and of primary interest in this work, is the case of global conservation laws. The vanishing covariant divergence of the energy-momentum tensor contains other terms besides the ordinary divergence. They prevent the use of Gauss' theorem which otherwise would allow to convert the four-dimensional integral into the three-dimensional one with the latter yielding the global energy-momentum conservation law. In order to obtain such laws, what is done is that the requirement of general covariance is dropped, and instead conservation laws of generally covariant field theories are derived from Noether's theorems.[3-5] The form of such laws is

$$\partial_\mu S^\mu = 0,$$

where S^μ is a function of the field variables of a number of arbitrary functions. In general, S^μ is not a tensor under general coordinate transformations. Whenever it is a tensor it still depends on an arbitrary vector field. Barring the case when S^μ is a vector, the conserved quantities S^μ lead to three distinct difficulties if one attempts either to interpret what S^μ conserves locally or to formulate global (integral) conservation laws.

First, if S^μ is not a tensor then it has no invariant or rather covariant local or global meaning. Secondly, if it is a tensor its volume integral over a finite domain generally is not a tensor so that it has no global covariant meaning. And, probably the most serious difficulty is that S^μ is not unique, but of an infinite variety.[6]

It is easy to see that waiving of the general covariance requirement with regard to the global conservation laws has not helped either to formulate them in a unique manner or to make an extensive use of them (as is done in Lorentz covariant theories). For some of these reasons it has been claimed[7] that the concepts of energy and momentum are not of great importance in the general relativity theory. However, it may be presumed that importance of these concepts would be regained should it be possible to construct a global energy–momentum tensor unambiguously and covariantly. The primary

[1] P. G. Bergmann, Phys. Rev. 75, 680 (1949).
[2] J. Plebanski and J. Ryten, J. Math. Phys. 2, 677 (1961).
[3] J. N. Goldberg, Phys. Rev. 89, 263 (1953).
[4] P. G. Bergmann, Phys. Rev. 122, 287 (1958).
[5] *Gravitation, an Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), Chap. 5.
[6] J. N. Goldberg, Phys. Rev. 111, 315 (1958).
[7] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1953), p. 193.

objective of this work is to demonstrate how global covariant conservation laws can be constructed within the context of the Riemannian geometry. The main assumption underlying this demonstration is that all conservation laws have to be manifestly covariant. Among other things this implies that it is meaningful to speak of a constant global tensor only when it is covariantly constant. The last concept can be easily defined in mathematical terms; nonetheless, one lacks intuition, pictorial representation, and physical content of this idea. Although an attempt is made to develop both intuition and pictorial description of this concept, the approach, in general, is not slanted in that direction.

2. TENSOR INTEGRATION

2.1 Absolute Integration in One Dimension

When one tries to integrate a vanishing covariant divergence of a tensor other than a vector the difficulties encountered are not only the additional terms besides the ordinary divergence. Another distinct difficulty, even in the absence of such additional terms, is the fact that integration of a tensor destroys its tensor character. Therefore, before one can hope to obtain global or integrated covariant conservation laws, it is necessary, first of all, to modify the ordinary integration so as to make it a tensor operation. The choice of modification is not as wide as one may think if it is recalled that integration and differentiation are inverse operations. Covariant differentiation and absolute differentiation[8] are well defined and widely used tensor operations. However, operations which are inverse to them are conspicuously absent in tensor calculus.

Thus in tensor calculus it is customary to pose the following questions: Given a tensor p defined along the curve u, is it possible to generate other tensors from p? The answer, of course, is "yes", because by means of absolute differentiation an infinite number of derived tensors can be readily gotten. We would like to pose the same question in the reverse order. Given T, an absolute derivative of a tensor p, defined along the curve u, is it possible to obtain the original tensor p? This question can be stated in mathematical terms as follows. A tensor p with N components is related to its absolute derivative along u, denoted as the tensor T, by

$$\frac{\delta p^n}{\delta u} \equiv \frac{dp^n}{du} + \Lambda^n_k p^k = T^n. \tag{1}$$

[8] J. L. Synge and A. Schild, *Tensor Calculus* (University of Toronto Press, Toronto, 1952).

For convenience all terms in the absolute derivative of p which are proportional to p are lumped together. The tensor indices $\alpha_1 \cdots \alpha_N$ may be either covariant, contravariant, or mixed, but are denoted by one Latin letter. It is also clear that the function Λ stands for a linear expression in $\Gamma^\alpha_{\beta\gamma}(dx^\gamma/du)$ as is required by the tensor character of p.

Now, the question posed can be rephrased to read: does the system of equations in (1) possess solutions for p^n when T^n and Λ^n_k are known functions of the parameter u. Whenever such solutions exist the tensor p is said to be an absolute integral of T along u and is written as

$$p^n = \int T^n \, \delta u + p^n_{(0)}; \qquad \frac{\delta p^n_{(0)}}{\delta u} = 0, \tag{2}$$

where $p^n_{(0)}$ is the set of integration "constants" whose absolute derivative is zero.

Absolute integration is then defined as integration of the system of differential equations in (1). Being an inverse operation to absolute differentiation it is by definition a tensor operation.

When the tensor p^n is described by N independent components, none of which vanish identically, then Eq. (1) represents N independent, linear, first-order, inhomogeneous differential equations which, in general, are also coupled. By successive differentiations and eliminations of all but one tensor component, this system of equations can be reduced to a Nth-order, linear, inhomogeneous, differential equation whose coefficients are Λ, T, and their derivatives.

Proofs and conditions under which such equations possess solutions can be found in any standard book on differential equations, such as Ince[9] or Forsyth.[10] Here it is assumed that such solutions exist as they do for ordinary points of Λ and T. In special types of Riemannian spaces these solutions can be obtained in a finite number of elementary calculus operations. Thus Eq. (1) can be integrated when Λ's are such that all equations can be successively decoupled, or when all Λ's are constant, that is independent of u.

If no restrictions are put on the Riemannian space, Eq. (1) can be solved only when an infinite number of elementary operations are applied to the function T and Λ, assuming, of course, that there are at least two independent equations. A perturbation solution of Eq. (1) falls in this category.

[9] E. L. Ince, *Ordinary Differential Equations* (Dover Publications, Inc., New York, 1956), p. 72.

[10] A. R. Forsyth, *Theory of Differential Equations* (Dover Publications, Inc., New York), Vol. 1–6.

It may be obtained by introducing an auxiliary equation

$$\frac{dp^n}{du} + \lambda(\Lambda^n_k - \Lambda^n_{k0})p^k + \Lambda^n_{k0}p^k = T^n, \qquad (3)$$

where λ does not depend on u and Λ^n_{k0} are arbitrary functions of u. Here, p^n and T^n denote the nth component of the tensors p and T. For $\lambda = 1$, Eq. (3) reduces exactly to Eq. (1). Now, it is assumed that p^n can be expressed in a power series of λ,

$$p^n = \sum_{\alpha=0}^{\infty} \lambda^\alpha p^n_\alpha. \qquad (4)$$

Substituting the power series for p^n in Eq. (3) and considering the latter an identity in λ, an infinite set of equations is obtained:

$$\begin{aligned} \frac{dp^n_0}{du} + \Lambda^n_{k0}p^k_0 &= T^n, \\ \frac{dp^n_\alpha}{du} + \Lambda^n_{k0}p^k_\alpha &= (\Lambda^n_{k0} - \Lambda^n_k)p^k_{\alpha-1}. \end{aligned} \qquad (5)$$

The functions Λ^n_{k0} may be chosen so that all equations in (5) can be integrated. Among such choices are

$$\text{(a)}\quad \Lambda^n_{k0} = 0, \qquad \text{(b)}\quad \Lambda^n_{k0} = \text{const},$$
$$\text{(c)}\quad \Lambda^n_{k0} = \text{triangular matrix}, \qquad (6)$$

where in a triangular matrix all elements either below or above the main diagonal are zero. The rate at which the perturbation solution in Eq. (4) converges can be controlled to some extent through an appropriate selection of Λ^n_{k0}. In a neighborhood of a point on u, it can be entirely controlled by choosing constant Λ^n_{k0} in (6b) to be equal to Λ^n_k at the point.

Absolute integration in Eq. (2) can be represented in many different ways in terms of an infinite number of ordinary integrations. One such representation is obtained with the choice (6a); then

$$\begin{aligned} p^n &= \sum_{\alpha=0}^{\infty} p^n_\alpha; \qquad p^n_0 = \int T^n\,du + c^n; \\ -p^n_\alpha &= \int \Lambda^n_k p^k_{\alpha-1}\,du, \end{aligned} \qquad (7)$$

where c^n are ordinary constants of integration. The solution in Eq. (7) indeed splits up into two parts. One of them, proportional to c^n, which in Eq. (2) was denoted by $p^n_{(0)}$, is the solution of the homogeneous version of Eq. (1). The other part being the particular solution of Eq. (1) is linear in the source term T^n and in Eq. (2) was denoted as the absolute integral of T^n.

2.2 A Vector Field

As an application of absolute integration one may attempt to obtain a position vector in a general Riemannian space. It is well known that the coordinate components x^μ which define a point in such a space do not transform as a vector unless the space is flat. The latter is a trivial case from the point of view of general relativity and is not considered here.

The first vector that is closely related to the coordinate components x^μ at a given point is the vector tangent to a curve u which passes through that point. Thus we may postulate that there exists a vector X^μ whose absolute derivative along the curve u is equal to the tangent vector dx^μ/du, that is,

$$\frac{\delta X^\mu}{\delta u} \equiv \frac{dX^\mu}{du} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{du}X^\beta = \frac{dx^\mu}{du}. \qquad (8)$$

In the notation of the previous section, the solution of Eq. (8) can be written as

$$X^\mu = \int \frac{dx^\mu}{du}\,\delta u + X^\mu_0 = \int_{u_1}^{u_2} \frac{dx^\mu}{du}\,\delta u, \qquad (9)$$

where in the last integral the "constants" of integration have been absorbed in the integration limits. From the last expression, it is seen that in a flat space, where absolute integration becomes just ordinary integration, the vector X^μ is a difference of two position vectors defined by the two points u_1 and u_2 and is independent of the curve u. In curved spaces the latter is no longer true, for the integration in Eq. (9) is explicitly dependent on the curve u so that, with the two points u_1 and u_2, it is possible to associate as many vectors X^μ as the number of curves can be passed through these two points. Unless one can single out from among these a unique vector to go with each pair of points, the concept of a vector field X^μ introduced in Eq. (8) will remain, at best, of questionable value.

Within the class of curves which pass through two given points there exists a curve of particular interest and significance. It is the straightest curve possible in a general Riemannian space, that is a geodesic. The equations which determine a geodesic $u : x^\mu = x^\mu(u)$, are solutions of

$$\text{(a)}\quad \frac{\delta}{\delta u}\frac{dx^\mu}{du} = 0; \quad g_{\mu\nu}\frac{dx^\mu}{du}\frac{dx^\nu}{du} = \epsilon$$
$$\text{for a nonnull geodesic,} \qquad (10)$$
$$\text{(b)}\quad \frac{\delta}{\delta u}\frac{dx^\mu}{du} = 0; \quad g_{\mu\nu}\frac{dx^\mu}{du}\frac{dx^\nu}{du} = 0$$
$$\text{for a null geodesic,}$$

where ϵ is the indicator.[8,11] Although the same symbol u is used for the arc length of a null and a nonnull geodesic, it should be noted that they are different. For example, in the space of general relativity the first is the arc length in a three-dimensional space,[11] whereas the second is the arc length or separation[11] in the four-dimensional space; moreover, it vanishes on the null cone.

If the integration in Eq. (9) is performed along a geodesic, then X^μ becomes

$$X^\mu = u \left.\frac{dx^\mu}{du}\right|_{u_1}^{u_2}, \tag{11}$$

where $(dx^\mu/du)\ \delta u$ was written as $\delta[u(dx^\mu/du)]$ in view of Eq. (10) and thus could be immediately integrated.

The interesting properties of the vector field X^μ are as follows. It is a two-point function

$$X^\mu = X^\mu(u_2, u_1) = -X^\mu(u_1, u_2).$$

Furthermore, it transforms as a vector with regard to either of the two points u_2 or u_1.

Since the vector X^μ is defined in terms of a pair of points in a Riemannian space, it is *a propos* to inquire about the type of correspondence that there exists between such a pair of points and the vector X^μ. This is entirely equivalent to the question of how many distinct geodesics can be passed through two given points. For a Riemannian space which is topologically the same as the flat space or when the two points are limited to a small enough, although not necessarily infinitesimal, neighborhood, it is clear that there exists only one distinct geodesic for any two points.[12] However, it is just as clear that, in multiply-connected spaces or in closed spaces, more than one distinct geodesic can be passed through two given points. In general, the correspondence between the vector X^μ and a pair of points will be many-to-one. What is significant, however, is that the multivaluedness of the vector X^μ with regard to a pair of points reflects the space topology. In this sense the vector X^μ enables one either to incorporate the topological aspects of space implicitly wherever it is used at its face value, or to deal with them directly if a one-to-one correspondence is set up between the multivaluedness of X^μ and the space topology.

Affinity of the vector X^μ to the Riemannian coordinates introduced in Ref. 8 and normal coordinates of Ref. 13 is rather evident. In fact, they are of the same form and differ only in that the tangent vector dx^μ/du in Eq. (11) is evaluated at both limits in the same way as the arc length u, whereas in the Riemannian coordinates dx^μ/du is evaluated at the origin (u_1) and is held fixed at this value at other points. The vector character of this coordinate system is limited to just one point, its origin u_1 where it coincides with the vector X^μ.

Other interesting properties of the vector field X^μ stem from the assumption that is usually made in the general theory of relativity, namely, that physical action propagates along geodesics. Usefulness of the vector field X^μ when one attempts to integrate the field equations of this theory is strongly suspected, but this line of thought is not pursued here.

2.3 Covariant Integration

In absolute differentiation or integration it is sufficient to specify the considered tensor as a function of the curve along which the operation is performed. Often the tensors of interest are either known or are considered known not only along a given curve but throughout the entire Riemannian space. In such cases it is meaningful to speak of a covariant derivative of a tensor which generates at a point an infinity of absolute derivatives corresponding to all possible directions of a curve passing through that point.

Consequently, in analogy to absolute integration, we define covariant integration as an inverse operation to that of covariant differentiation, that is, as a solution to a set of partial differential equations

$$\frac{\delta p^n}{\delta x^\mu} \equiv \frac{\partial p^n}{\partial x^\mu} + \Lambda^n_{\mu k} p^k = T^n_\mu, \tag{12}$$

where T and Λ are known point functions throughout the space of interest, and where the tensor p^n is the unknown. Unlike the set of ordinary differential equations in (1) which always possess solutions at ordinary points of Λ and T, the set of partial differential equations in (12) possess solutions only under very special conditions. These arise from the fact that there may or may not exist a required number of functions, the integrating factors, which upon multiplication convert Eq. (12) into a set of perfect differentials. When a required number of such integrating factors exists, Eq. (12) is said to be integrable. Methods of determining inte-

[11] *Relativity Groups and Topology, Les Houches 1963*, C. DeWitt and B. S. DeWitt, Eds. (Gordon and Breach Science Publishers, Inc., New York, 1964). Lectures by J. L. Synge.

[12] L. P. Eisenhart, *An Introduction to Differential Geometry* (Princeton University Press, Princeton, New Jersey, 1947), p. 173.

[13] L. P. Eisenhart, *Non-Riemannian Geometry* (American Mathematical Society, New York, 1927), Chap. II.

grating factors as well as conditions under which these exist can be found in Vol. 1 of Forsyth.[10]

The conditions of integrability are

$$\frac{\delta}{\delta x^\sigma} T^n_\mu - \frac{\delta}{\delta x^\mu} T^n_\sigma = \left(\frac{\delta}{\delta x^\sigma}\frac{\delta}{\delta x^\mu} - \frac{\delta}{\delta x^\mu}\frac{\delta}{\delta x^\sigma}\right)p^n. \qquad (13)$$

It is seen that they can be satisfied if the tensor T^n_μ is a covariant derivative of a tensor Q^n. Then the solutions of Eq. (12) can be written down as

$$p^n = Q^n + Q^n_0; \qquad \delta Q^n_0/\delta x^\mu = 0, \qquad (14)$$

where Q^n_0 is a covariantly constant tensor. Conversely, if Eq. (12) can be integrated, then the tensor T^n_μ can be expressed as a covariant derivative of another tensor. This is a generalization of a well-known statement in vector analysis that a vanishing curl of a vector implies that the vector can be written as a gradient of a scalar.

The operations of absolute and covariant differentiation and absolute integration can be applied to a given tensor any number of times to generate other tensors. This is not true of covariant integration, since it can be applied only to a specia class of tensors which satisfy the integrability condition in Eq. (13); moreover, repeated applications are not automatic but must be inspected anew for the existence of integrating factors.

Due to these limitations it is evident that covariant integration will not be as useful as the other three tensor operations.

2.4 Covariantly Constant Base Vectors

The concepts of absolute and covariant differentiation or integration become much clearer and intuitively acceptable if they are defined in terms of covariantly constant base vectors.[14-16]

Thus, suppose that with each point of the Riemannian space we associate a set of independent base vectors e_μ which transform as vectors under general coordinate transformations. The relative magnitudes and orientations of these vectors are assumed to be given by the metric tensor $g_{\mu\nu}$, that is,

$$e_\mu \cdot e_\nu = g_{\mu\nu}; \qquad e^\mu = g^{\mu\nu}e_\nu. \qquad (15)$$

Here the dot indicates a scalar product or projection of one vector on the other. Next it is required that the base vectors be covariantly constant, that is, that they satisfy

$$(\delta/\delta x^\sigma)e_\mu = (\partial e_\mu/\partial x^\sigma) - \Gamma^\nu_{\sigma\mu}e_\nu = 0. \qquad (16)$$

The last requirement is a special case of Eq. (12) with T^n_μ equal zero. The conditions of integrability of Eq. (16) reduce to

$$R_{\mu\nu\sigma}{}^\alpha e_\alpha = 0, \qquad (17)$$

where $R_{\mu\nu\sigma}{}^\alpha$ is the Riemann curvature tensor. In flat spaces, where the curvature tensor vanishes identically, these are satisfied so that it is possible to integrate Eq. (16). However, in unrestricted Riemannian spaces, (17) cannot be satisfied and the set of base vectors with the property (16) does not exist.

Since the base vectors e_μ cannot, in general, be determined as point functions over the entire space they are treated henceforth as unspecified entities possessing the properties in Eqs. (15) and (16). For this reason it was not necessary to introduce more explicit notation for the dot product in (15).

By means of the base vectors e_μ, each tensor equation can be converted into an invariant form if its free indices are contracted with the base vectors. Thus there is a one-to-one correspondence between the tensor $A^{\mu\nu}$ and the invariant form $A^{\mu\nu}e_\mu e_\nu$. Moreover, due to Eq. (16), the covariant derivative of $A^{\mu\nu}$ can be written as an ordinary partial derivative as follows:

$$e_\mu e_\nu(\delta A^{\mu\nu}/\delta x^\sigma) = (\partial/\partial x^\sigma)(A^{\mu\nu}e_\mu e_\nu). \qquad (18)$$

Appearance of the additional terms in a tensor derivative besides the ordinary derivative can now be explained as the corrective terms which are required in order to compensate for the curvature in the coordinates, that is, variations of the base vectors from point to point, regardless of whether this curvature arises from the choice of coordinates or from the curvature of the space. Equation (18) shows very lucidly what is meant by a covariantly constant tensor. First of all, such a tensor can be converted into an invariant form. The latter can be thought of to consist of a bundle of arrows (or vectors) emanating from each point of space where the tensor is defined. The number of arrows corresponds to the tensor rank and the magnitude and direction of each arrow is determined by the tensor components at the point. In a covariantly constant tensor, both the magnitudes and directions of all its arrows remain the same throughout the region where its covariant derivative vanishes. In a tensor which is covariantly constant along a curve u, the magnitudes and directions of all its arrows remain constant on u as long as its absolute derivative along u vanishes.

[14] T. J. Willmore, *Introduction to Differential Geometry* (Oxford University Press, New York, 1959).
[15] L. Brand, *Vector and Tensor Analysis* (John Wiley & Sons, Inc., New York, 1953), 4th ed., p. 366.
[16] Reference 11, Lectures by C. Misner.

It is seen that a covariantly constant tensor conforms to all our preconceived notions of constancy.

Absolute integration can be expressed very simply as ordinary integration with the use of the base vectors. Thus converting Eq. (1) into an invariant form one obtains

$$(d/du)(p^{\alpha_1\cdots\alpha_k}e_{\alpha_1}\cdots e_{\alpha_k}) = T^{\alpha_1\cdots\alpha_k}e_{\alpha_1}\cdots e_{\alpha_k}, \quad (19)$$

where the tensor indices of p and T are written explicitly. Equation (19) can be immediately integrated formally to read

$$p^{\alpha_1\cdots\alpha_k} = e^{\alpha_1}\cdots e^{\alpha_k}\vdots\int T^{\beta_1\cdots\beta_k}e_{\beta_1}\cdots e_{\beta_k}\,du, \quad (20)$$

where the vertical dots imply a k-fold dot product. However, the integration in Eq. (20) cannot be performed straightforwardly because e_μ are not specified. It is possible to integrate (20) by parts and, if Eq. (16) is used to eliminate the derivatives of the base vectors, the procedure can be repeated indefinitely. This yields the same representation of absolute integration as was obtained by the perturbation method in Eq. (7).

2.5 Integration in Higher Dimensions

Absolute integration can be applied more than once to a given tensor along the same curve. If the tensor T is defined on more than one curve, say $u_1, u_2 \cdots u_n$ multiple integrals of T can be formulated along various curves such as

$$\iint T\,\delta u_1\,\delta u_2; \qquad \iiint T\,\delta u_1\,\delta u_2\,\delta u_3; \quad \cdots. \quad (21)$$

Of particular interest is the case when the families of curves $u_1, u_2; u_1, u_2, u_3; \cdots$ define a two-dimensional, three-dimensional, etc. subspaces, such that when one travels along u_k other u remain constant. Then the integrals in (21) become absolute integrals of T over a 2-surface, 3-surface, and higher-dimensional subspaces.

However, absolute integrations along different paths do not commute just like absolute differentiations. In view of this, it is possible to associate not one but $k!$ absolute integrals of T over a k-dimensional subspace. The number $k!$ represents all the permutations of the order of integrations among k distinct integrations.

The multiplicity of such integrals and especially their differences prompts one to ask the following question. Is it possible to define an absolute integral of a tensor over a k-dimensional (sub)space which is both unique and independent of the order of integrations? The key to the answer lies in the last requirement. It can be fulfilled if such an integral is completely symmetrized with regard to the order of integrations. Since, among the $k!$ integrals there exists only one combination which is completely symmetric, its uniqueness is automatically established.

A k-dimensional symmetrized "volume" element δV_k is a direct sum of all permutations of k integrations

$$\delta V_k = \frac{1}{k!}\epsilon^{ij\cdots n}\,\delta u_i\,\delta u_j\cdots\delta u_n; \qquad i, j, \cdots n = 1, \cdots k, \quad (22)$$

where $\epsilon^{ij\cdots n} = 1$ if none of its indices are repeated and $\epsilon^{ij\cdots n} = 0$ otherwise. In (22) the sum is implied over the indices $i, j \cdots n$. Examples of δV_k for $k = 2, 3$ are

$$\begin{aligned}\delta V_2 &= \tfrac{1}{2}(\delta u_1\,\delta u_2 + \delta u_2\,\delta u_1),\\ \delta V_3 &= \tfrac{1}{6}(\delta u_1\,\delta u_2\,\delta u_3 + \delta u_1\,\delta u_3\,\delta u_2 + \delta u_2\,\delta u_3\,\delta u_1\\ &\quad + \delta u_2\,\delta u_1\,\delta u_3 + \delta u_3\,\delta u_1\,\delta u_2 + \delta u_3\,\delta u_2\,\delta u_1).\end{aligned} \quad (23)$$

By its construction δV_s is independent of the order of integrations since interchanging the kth and the mth integrations in (22) only rearranges the terms in the sum.

2.6 Gauss' Theorem for Absolute Integration

In the preceding sections, absolute integration has been sufficiently well defined to enable one to integrate a covariantly vanishing divergence of a tensor. The latter can be written as

$$(\delta/\delta x^\mu)\mathfrak{J}^\mu = 0, \quad (24)$$

where the free indices on the tensor density $\mathfrak{J}^\mu$ are suppressed for clarity and convenience. In a four-dimensional space of general relativity, one wishes to integrate Eq. (24) over a 4-volume V_4 bounded by a 3-surface V_3. If the 4-volume V_4 is a part of a multiple-connected space, the bounding surface may consist of several disjoint sections such as two concentric spherical surfaces in a three-dimensional case.

The absolute volume integral of Eq. (24) is

$$\int_{V_4}\frac{\delta\mathfrak{J}^\mu}{\delta x^\mu}\,\delta V_4 = \frac{1}{4!}\sum_{(ijkl)}\iiiint\left(\frac{\delta\mathfrak{J}^0}{\delta x^0} + \frac{\delta\mathfrak{J}^1}{\delta x^1} + \frac{\delta\mathfrak{J}^2}{\delta x^2} + \frac{\delta\mathfrak{J}^3}{\delta x^3}\right)\delta x^i\,\delta x^j\,\delta x^k\,\delta x^l, \quad (25)$$

where $\sum_{(ijkl)}$ is the sum of all permutations of 0123, and where the coordinates themselves are taken as the curves along which absolute integration is performed.

The right-hand side of Eq. (25) can be divided into four classes of integrals as follows:

$$\int_{V_4} \frac{\delta \mathfrak{J}^\mu}{\delta x^\mu}\,\delta V_4$$

$$= \frac{1}{4!}\sum_{n=0}^{3}\sum_{(ijk)}\Big[\iiint \frac{\delta \mathfrak{J}^n}{\delta x^n}\,\delta x^n\,\delta x^i\,\delta x^j\,\delta x^k \qquad (26a)$$

$$+ \iiint \frac{\delta \mathfrak{J}^n}{\delta x^n}\,\delta x^i\,\delta x^n\,\delta x^j\,\delta x^k \qquad (26b)$$

$$+ \iiint \frac{\delta \mathfrak{J}^n}{\delta x^n}\,\delta x^i\,\delta x^j\,\delta x^n\,\delta x^k \qquad (26c)$$

$$+ \iiint \frac{\delta \mathfrak{J}^n}{\delta x^n}\,\delta x^i\,\delta x^j\,\delta x^k\,\delta x^n\Big], \qquad (26d)$$

with $\sum_{(ijk)}$ being the sum of all permutations of ijk not containing n.

It is obvious that only the first integral (26a) can be integrated once to yield immediately a 3-surface integral, whereas the others cannot because absolute integrations do not commute. However, the sum of all four integrals is invariant when any two or more integrations are interchanged. Next, it is shown that the integral (26a) remains invariant when the order of integration is altered. First of all, it is obvious that (26a) is unchanged if any of the ijk indices are relabeled—in fact this is true for each fixed n in (26a) not only for the sum over n. Also (26a) or any integral with a fixed n in (26a) can be written in five different ways,

$$\iiint \frac{\delta \mathfrak{J}^n}{\delta x^n}\,\delta x^n\,\delta x^i\,\delta x^j\,\delta x^k$$

$$= \iiint \frac{\delta}{\delta x^n}\int \mathfrak{J}^n\,\delta x^i\,\delta x^n\,\delta x^j\,\delta x^k$$

$$= \iint \frac{\delta}{\delta x^n}\iint \mathfrak{J}^n\,\delta x^i\,\delta x^j\,\delta x^n\,\delta x^k$$

$$= \int \frac{\delta}{\delta x^n}\iiint \mathfrak{J}^n\,\delta x^i\,\delta x^j\,\delta x^k\,\delta x^n$$

$$= \iiint \mathfrak{J}^n\,\delta x^i\,\delta x^j\,\delta x^k. \qquad (27)$$

Now, suppose that the first and second integrations are to be interchanged in (26a); then choose that integral for (26a) in (27), where δx^n is either in the third or fourth place. The interchange would correspond then to relabeling i by j and j by i, but this does not affect the value of the integral (26a). Similar choices can always be made if not all four integrations are involved in the reordering.

Using the above invariance of (26a) and the complete symmetry of (26) with regard to the order of any two integrations, we switch the first and the second integrations in (26), whereupon (26b) reduces to (26a). Again, switching the first with the third, and the fourth in succession, (26c) and (26d) are reduced to (26a). Consequently, (25) can be written as

$$\int_{V_4} \frac{\delta \mathfrak{J}^\mu}{\delta x^\mu}\,\delta V_4 = \int_{V_3} \mathfrak{J}^\mu\,\delta V_\mu$$

$$\equiv \frac{1}{3!}\sum_{(ijk)}\iiint \mathfrak{J}^n\,\delta x^i\,\delta x^j\,\delta x^k. \qquad (28)$$

The result (28) shows that Gauss' theorem holds true also for absolute integration. Although it was proven only for a four-dimensional space, the theorem can be readily generalized to an n-dimensional Riemannian space.

From (24) and (28) it follows that

$$\int_{V_3} \mathfrak{J}^\mu\,\delta V_\mu = 0 = \frac{1}{3!}\Big[\sum_{(123)}\iiint \mathfrak{J}^0\,\delta x^i\,\delta x^j\,\delta x^k$$

$$+ \sum_{(230)}\iiint \mathfrak{J}^1\,\delta x^i\,\delta x^j\,\delta x^k$$

$$+ \sum_{(301)}\iiint \mathfrak{J}^2\,\delta x^i\,\delta x^j\,\delta x^k$$

$$+ \sum_{(012)}\iiint \mathfrak{J}^3\,\delta x^i\,\delta x^j\,\delta x^k\Big]. \qquad (29)$$

It is of interest to take the absolute time derivative of (29). If one applies the arguments used in derivation of (28) to the last three integrals in (29), the time derivative of (29) yields

$$\frac{\delta}{\delta x^0}\Big(\frac{1}{3!}\sum_{(123)}\iiint \mathfrak{J}^0\,\delta x^i\,\delta x^j\,\delta x^k\Big)$$

$$+ \tfrac{1}{2}\sum_{\text{cyclic}(123)}\iint \mathfrak{J}^i(\delta x^j\,\delta x^k + \delta x^k\,\delta x^j) = 0. \qquad (30)$$

It is seen that the absolute time derivative of the global tensor $\int_{V_3} \mathfrak{J}^0\,\delta V_0$ is equal to the absolute surface integral of $\mathfrak{J}^i$ over V_2, where V_2 encloses the spatial volume V_3. Although the absolute surface integral is not the same as the ordinary surface integral, the two have one important point in common, that is, the evaluation of either of them requires $\mathfrak{J}^i$ to be specified only on the surface V_2.

Any tensor density $\mathfrak{J}^\mu$ satisfying (24) and whose surface integral vanishes gives rise to a conserved global tensor p, where

$$p = \frac{1}{3!}\sum_{(123)}\iiint \mathfrak{J}^0\,\delta x^i\,\delta x^j\,\delta x^k;\quad \frac{\delta}{\delta x^0}p = 0. \qquad (31)$$

Nontrivial tensor densities $\mathfrak{J}^\mu$ satisfying Eq. (24) are known to exist in the theory of general relativity. They are investigated in Part II of this report.

Global Covariant Conservation Laws in Riemannian Spaces. II

BOHDAN SHEPELAVEY
General Electric Company, HMED, Syracuse, New York
(Received 9 November 1965)

Using the idea of tensor integration, the vector field developed in Part I of this report, and the full Bianchi identities, it is shown that in a general Riemannian space there are four global covariantly-conserved tensors. The ranks of these tensors are three, four, five, and six. The traces of the first two of these tensors yield the generally covariant equivalent of the familiar linear and angular momentum. The remaining four traceless tensors describe, residually, the gravitational field. With each covariantly conserved tensor one can associate a number of independent invariants. Such invariants are conserved in the ordinary sense. Among these are two types of rest energies and two types of angular momentum magnitudes obtained from the trace and traceless tensors. Examples of global, conserved tensors are derived for a Schwarzschild metric with the electron mass, and a metric of a point electron. It is shown that the rest energy of the Coulomb field diverges as $\ln(1/r)$ at the origin and the second rest energy, that is, the rest energy of the gravitational field diverges as $\ln r$ as r approaches infinity. When cutoffs are introduced at the Schwarzschild radius r_0, at the classical electron radius r_1, and at the radius of the visible universe r_2, the rest energy of the gravitational field contained in the shell of thickness $r_1 - r_0$ is approximately 100 times that of the electron rest energy. It is twice this value in the entire visible universe. Since the gravitational field is described by the traceless tensors and the former forms a heavy, compact cloud around the point particle, it is conjectured that the traceless tensors represent the internal degrees of freedom of the elementary particles.

1. INTRODUCTION

THE objective of this report is to demonstrate, by an actual construction and a specific application, the existence of a finite and unique set of global, covariantly conserved tensors in a general Riemannian space. The requisite mathematical tools for this task have been developed in Part I of this report and they are readily recognized to be a key factor in Part II, although no emphasis is being put on them here.

The local covariant conservation laws, which go beyond the commonly accepted laws, are formulated in Sec. 2. They are derived from the full Bianchi identities and the vector field of Part I in a form of a fourth-rank tensor and its first three moments, all of them with an identically vanishing covariant divergence. The use of a fourth-rank tensor in a similar capacity is suggested by Trautman,[1] but not in a generally covariant context.

In Sec. 3, tensor integration and the Gauss theorem for such integration is used to convert the locally conserved tensors into the corresponding global tensors whose ranks, due to integration, are lowered by one unit. These global tensors are also covariantly conserved provided that certain surface integrals vanish. It is shown that, in spaces where this provision is satisfied, the algebraic structure of the four global tensors admits of a decomposition into two trace tensors of ranks one and two and four traceless tensors of ranks three, four, five, and six. All of these tensors are separately conserved. The trace tensors are the linear and angular momenta of the matter fields. The traceless tensors are the zeroth, first, second, and third moments of the fourth-rank energy–momentum tensor of the gravitational field plus the matter field.

In Sec. 5 the third-rank global tensor is obtained for the Riemannian spaces of the Schwarzschild metric and the metric of a point electron. This enables one, for the first time, to calculate in a manifestly covariant manner the total rest energy of the gravitational field.

The final section concludes this report with an interpretation of the traceless tensors, besides the linear and angular momentum, as the additional degrees of freedom of a generally covariant dynamical system.

2. LOCAL CONSERVATION LAWS

2.1 Bianchi Identities

The field equations of the theory of general relativity are

$$R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R = (8\pi k/c^4)T^{\mu\nu}, \tag{1}$$

where $R^{\mu\nu}$ is the Ricci tensor, R is the scalar curvature, and $T^{\mu\nu}$ is the energy–momentum tensor of the matter. In the absence of matter, that is, when $T^{\mu\nu}$ is identically zero, Eq. (1) describes gravitational radiation. It is then typical of gravitational radiation

[1] A. Trautman, in *Gravitation, An Introduction to Current Research*, L. Witten, Ed. (John Wiley & Sons, Inc., New York, 1962), Chap. 5, pp. 183–188.

that, if it carries any energy and momentum, a covariant quantity describing the energy and momentum of gravitational radiation must assume a form which is different from the terms of Eq. (1) (that is it cannot be a second-rank symmetric tensor). This is significant in the fact that, in Lorentz covariant theories,[2] the matter tensor $T^{\mu\nu}$ describes completely and adequately the energy and the momentum state of a dynamical system. In such theories the symmetric tensor $T^{\mu\nu}$ satisfies

$$(\partial/\partial x^{\mu})T^{\mu\nu} = 0. \tag{2}$$

From Eq. (2) and the symmetry of $T^{\mu\nu}$ it follows that

$$(\partial/\partial x^{\mu})(T^{\mu\nu}x^{\sigma} - T^{\mu\sigma}x^{\nu}) = 0. \tag{3}$$

Equations (2) and (3) define ten local conservation laws which, in Lorentz covariant theories, can be easily integrated.

In the theory of general relativity one could proceed to obtain ten conservation laws in exactly the same way. It is known that Eq. (1) satisfies a generally covariant equivalent of Eq. (2)

$$\frac{\delta}{\delta x^{\mu}}(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R) = \frac{8\pi k}{c^4}\frac{\delta}{\delta x^{\mu}}T^{\mu\nu} = 0. \tag{4}$$

Similarly, it is shown later that, using the symmetry of $T^{\mu\nu}$ and the vector field X^{μ} introduced in Part I, a generally covariant equivalent of Eq. (3) also holds. From these generally covariant local conservation laws, ten global conservation laws can be obtained in complete analogy to the Lorentz case by means of the tensor integration developed in Part I.

Although the initial objective here, as well as in many other investigations,[3-5] was to exhibit the ten conserved quantities of the Lorentz group in the theory of general relativity, it is obvious that having done this the case of the generally covariant conservation laws cannot be considered closed. There still remains the question of energy and momentum transfer by gravitational radiation. If it is assumed that the energy and momentum of a dynamical system are solely expressed by and derived from the tensor $T^{\mu\nu}$, one has to conclude that gravitational radiation cannot possess energy and momentum. At present the experimental evidence is inconclusive and can neither deny nor confirm this assumption. However, unless there is experimental evidence to the contrary, one would like to believe that gravitational radiation, just like other types of radiation, is a carrier of energy and momentum. When this point of view is adopted, it is clear that there ought to be another conserved tensor besides $T^{\mu\nu}$. The same question put in a more formal way is whether in the Riemannian geometry there are other tensors besides $R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R$ with a vanishing covariant divergence.

In order to look for such tensors, it is best to discard the tensor $T^{\mu\nu}$ which is not an object of the Riemannian geometry and to consider the tensor $(c^4/8\pi k)(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R)$ instead. Its vanishing divergence expressed by Eq. (4) is a direct consequence of the contracted Bianchi identies. At this point it is natural to go back to the general statement of the Bianchi identities and see if they yield a vanishing divergence of a tensor less restricted than $R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R$.

If the Riemann curvature tensor and the Ricci tensor are defined as

$$\begin{gathered} R_{\mu\nu\rho}{}^{\sigma} = \partial_{\mu}\Gamma^{\sigma}_{\nu\rho} - \partial_{\nu}\Gamma^{\sigma}_{\mu\rho} + \Gamma^{\tau}_{\nu\rho}\Gamma^{\sigma}_{\mu\tau} - \Gamma^{\tau}_{\mu\rho}\Gamma^{\sigma}_{\nu\tau}, \\ R_{\nu\rho} = R_{\mu\nu\rho}{}^{\mu}; \qquad \partial_{\mu} \equiv \partial/\partial x^{\mu}, \end{gathered} \tag{5}$$

then the Bianchi identities take the form

$$D_{\tau}R_{\mu\nu\rho}{}^{\sigma} + D_{\mu}R_{\nu\tau\rho}{}^{\sigma} + D_{\nu}R_{\tau\mu\rho}{}^{\sigma} = 0; \quad D_{\mu} \equiv \delta/\delta x^{\mu}. \tag{6}$$

It is seen by inspection that, when Eq. (6) is contracted on τ and σ, one obtains a vanishing divergence of a linear combination of the Riemann curvature tensor, namely

$$D_{\sigma}(R_{\mu\nu\rho}{}^{\sigma} - \delta^{\sigma}_{\mu}R_{\nu\rho} + \delta^{\sigma}_{\nu}R_{\mu\rho}) = 0. \tag{7}$$

From the first term in (7) one concludes that this linear combination should possess all the symmetries of the Riemann curvature tensor. When the index σ is lowered, however, the Ricci tensor terms do not exhibit the required symmetry in the pair of $\sigma\rho$ indices. This situation can be easily corrected if one considers two contracted versions of Eq. (7),

$$D_{\sigma}(R^{\sigma}_{\nu} - \tfrac{1}{2}\delta^{\sigma}_{\nu}R)g_{\mu\rho} - D_{\sigma}(R^{\sigma}_{\mu} - \tfrac{1}{2}\delta^{\sigma}_{\mu}R)g_{\nu\rho} = 0. \tag{8}$$

Adding Eqs. (7) and (8) together one obtains

$$D_{\sigma}T_{\mu\nu\rho}{}^{\sigma} = 0, \tag{9}$$

---

[2] L. Landau and E. Lifshitz, *The Classical Theory of Fields*, translated from the Russian by M. Hamermesh (Addison-Wesley Publishing Corporation, Inc., Reading, Massachusetts, 1951), p. 80.

[3] Ref. 2, pp. 316–323.

[4] J. Rayski, "Conservation Laws in General Relativity," Bull. Polish Acad. Sci. 9, 33 (1961).

[5] C. Møller, Tetrad Fields and Conservation Laws in General Relativity, in Proc. Intern. School Phys. "Enrico Fermi," June-July 1961.

where the tensor $T_{\mu\nu\rho}{}^{\sigma}$ is defined as

$$T_{\mu\nu\rho}{}^{\sigma} \equiv (-c^4/8\pi k)\{R_{\mu\nu\rho}{}^{\sigma} + g_{\mu\rho}R_{\nu}^{\sigma} - \delta_{\mu}^{\sigma}R_{\nu\rho} + \delta_{\nu}^{\sigma}R_{\mu\rho} - g_{\nu\rho}R_{\mu}^{\sigma} + \tfrac{1}{2}R(g_{\nu\rho}\delta_{\mu}^{\sigma} - g_{\mu\rho}\delta_{\nu}^{\sigma})\}. \quad (10)$$

It can be verified directly that the sum of the Ricci tensor terms in Eq. (10) possess the same symmetries as the Riemann curvature tensor; consequently

$$T^{\mu\nu\rho\sigma} + T^{\nu\mu\rho\sigma} = 0, \quad (11)$$

$$T^{\mu\nu\rho\sigma} + T^{\mu\nu\sigma\rho} = 0, \quad (12)$$

$$T^{\mu\nu\rho\sigma} - T^{\rho\sigma\mu\nu} = 0, \quad (13)$$

$$T_{\mu\nu\rho}{}^{\sigma} + T_{\nu\rho\mu}{}^{\sigma} + T_{\rho\mu\nu}{}^{\sigma} = 0. \quad (14)$$

The gravitational constant in the definition of $T_{\mu\nu\rho}{}^{\sigma}$ in Eq. (10) was introduced in order to make $T_{\mu\nu\rho}{}^{\sigma}$ assume the dimensions of energy. One observes that the contracted $T_{\mu\nu\rho}{}^{\sigma}$ coincides with the Einstein tensor $R_{\nu\rho} - \frac{1}{2}g_{\nu\rho}R$ times the gravitational constant, and, modulo the field equations in Eq. (1), it is equal to the energy–momentum tensor of the matter.

$$T_{\sigma\nu\rho}{}^{\sigma} = T_{\nu\rho} = (c^4/8\pi k)(R_{\nu\rho} - \tfrac{1}{2}g_{\nu\rho}R). \quad (15)$$

Thus, the matter tensor is a trace of a higher-rank tensor $T_{\mu\nu\rho}{}^{\sigma}$. The former's vanishing divergence can be thought of as following from the local conservation law satisfied by $T_{\mu\nu\rho}{}^{\sigma}$ as expressed by Eq. (9). The tensor $T_{\mu\nu\rho}{}^{\sigma}$ can be invariantly decomposed into its traceless and trace tensors just like the Riemann tensor,[6]

$$T_{\mu\nu\rho}{}^{\sigma} = (-c^4/8\pi k)\{C_{\mu\nu\rho}{}^{\sigma} + P_{\sigma[\nu}g_{\mu]\rho} + P_{\rho[\mu}g_{\nu]\sigma} + \tfrac{1}{12}R(g_{\nu\rho}g_{\mu\sigma} - g_{\mu\rho}g_{\nu\sigma})\}, \quad (16)$$

where $C_{\mu\nu\rho}{}^{\sigma}$ is the Weyl conform tensor, $P_{\mu\nu}$ is the traceless part of the Ricci tensor $R_{\mu\nu} - \frac{1}{4}g_{\mu\nu}R$, and the square brackets around the indices indicate the antisymmetric part $[\mu\nu] = \frac{1}{2}(\mu\nu - \nu\mu)$. Gravitational radiation is characterized by a vanishing of all parts in $T_{\mu\nu\rho}{}^{\sigma}$ except for the Weyl tensor $C_{\mu\nu\rho}{}^{\sigma}$, so that the latter may be interpreted as the energy–momentum tensor of the gravitational field (radiation). It is covariantly conserved only when nothing else but gravitational field is present. When matter is present then the sum of $C_{\mu\nu\rho}{}^{\sigma}$ and the matter tensor in the form prescribed by Eq. (16) is covariantly conserved together.

Having found a more general conservation law than that of Eq. (4), it is of interest to ascertain whether it is unique. Before this is undertaken one has to state the acceptability criteria for such a conservation law. In generally covariant field theories, it seems reasonable to require that a conserved quantity be

(1) a covariant quantity,
(2) with a vanishing covariant divergence,
(3) containing quadratic terms in the first derivatives of the field variables.

To these requirements may be added the traditional one that the conserved quantity should not contain higher than the first derivatives of the field variables. For the gravitational field $g_{\mu\nu}$ it cannot be reconciled with the first two requirements which, here, are considered more important, therefore exemption from higher derivatives is dropped.[7]

Moreover, it should be noted that some conserved quantities, homogeneously linear in higher derivatives of the field variables, have recently been discovered by Lipkin,[8] but since they do not seem to be of practical importance[9] they are excluded here by the requirement (3).

Within the Riemannian geometry there is only one tensor that satisfies requirements (1) and (3), namely, the Riemann curvature tensor. But the tensor $T_{\mu\nu\rho}{}^{\sigma}$, being a linear combination of the latter and possessing its symmetries, satisfies (1) and (3) as a matter of course. In addition it satisfies (2) in view of the Bianchi identities. Thus the uniqueness of $T_{\mu\nu\rho}{}^{\sigma}$ is established.

2.2 Angular and Higher Moments

In Lorentz-covariant theories the symmetry of the matter tensor and its vanishing divergence led to a conservation law for the angular momentum in Eq. (3). This approach can be utilized in generally covariant theories. A new element that is needed for this purpose is the position vector. In Riemannian geometry no such vector exists; however, in Part I a vector field X^{μ} was derived which can be used in the capacity of a position vector. It is defined by the differential equation

$$\frac{\delta X^{\mu}}{\delta u} = \frac{dX^{\mu}}{du} + \Gamma_{\alpha\beta}^{\mu}\frac{dx^{\alpha}}{du}X^{\beta} = \frac{dx^{\mu}}{du}. \quad (17)$$

[6] J. Ehlers and W. Kundt, Ref. 1, Chap. 2.

[7] If another affine connection, e.g., that of a flat space $\bar{\Gamma}_{\beta\gamma}^{\alpha}$, is admitted into the Riemannian geometry, then the tensor $\phi_{\beta\gamma}^{\alpha} = \Gamma_{\beta\gamma}^{\alpha} - \bar{\Gamma}_{\beta\gamma}^{\alpha}$ offers a possibility to eliminate higher than the first derivatives. However, attempts have so far failed to prove the existence or nonexistence of a quadratic expression in ϕ with a vanishing divergence.

[8] D. M. Lipkin, J. Math. Phys. 5, 696 (1964).

[9] T. W. B. Kibble, J. Math. Phys. 6, 1022 (1965).

If the curve u is a geodesic $x^\mu = x^\mu(u)$ determined by the equations

$$\frac{d^2x^\mu}{du^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{du}\frac{dx^\beta}{du} = 0; \quad g_{\alpha\beta}\frac{dx^\alpha}{du}\frac{dx^\beta}{du} = \epsilon, \tag{18}$$

where $\epsilon = -1, 0, +1$ if dx^μ/du is timelike, null, or spacelike, then the vector X^μ is

$$X^\mu = u\frac{dx^\mu}{du}\Big|^{u_2}_{u_1}. \tag{19}$$

The vector X^μ depends on two points u_1 and u_2 and a geodesic u that passes through these two points. If u_1 is made to coincide with the origin of the coordinate system and the point u_2 is allowed to wander over the entire domain, then with each point of the domain one can associate a vector X^μ. Next, it is necessary to determine the covariant derivative of X^μ at any point of the domain if it is to be used in the role of the position vector.

For this purpose, consider the solution of Eq. (18) which assumes the form

$$x^\mu = x^\mu(x^\mu_0, p^\mu, u), \tag{20}$$

where x^μ_0 and p^μ are constants of integration. More explicitly, x^μ can be expanded as a power series[10] in u,

$$x^\mu = x^\mu_0 + up^\mu + \frac{1}{2!}A^\mu_{\alpha\beta}u^2p^\alpha p^\beta + \frac{1}{3!}A^\mu_{\alpha\beta\gamma}u^3p^\alpha p^\beta p^\gamma + \cdots. \tag{21}$$

Here, A^μ are the Γ's and their derivatives evaluated at $x^\mu = x^\mu_0$. They are obtained by repeated differentiation of Eq. (18). The constants of integration are sufficient to pass the geodesic curve through any two desired points. Since one of the points is to be the origin, x^μ_0 must be set equal to zero ($x^\mu_0 = 0$). The other constants p^μ describe the direction of the geodesic at the origin, that is

$$p^\mu = \frac{dx^\mu}{du}\Big|_{u=0}.$$

Any point may be specified by prescribing either its coordinates x^μ or alternatively by stating the corresponding values of u and p^μ. Consequently, x^μ may be considered as functions of x^μ_0, p^μ, and u [as is shown in Eq. (20)], but in our case x^μ_0 are fixed so that x^μ depend only on p^μ and u.

It follows from Eq. (21) by direct calculation [and therefore must also be true of Eq. (20)] that

$$p^\gamma\frac{\partial x^\beta}{\partial p^\gamma} = u\frac{\partial x^\beta}{\partial u} \quad \text{or} \quad \frac{\partial x^\beta}{\partial p^\gamma} = u\frac{\partial x^\beta}{\partial u}p_\gamma, \tag{22a}$$

$$p^\sigma\frac{\partial^2x^\mu}{\partial p^\gamma\,\partial p^\sigma} = u^2\frac{\partial^2x^\mu}{\partial u^2}p_\gamma. \tag{22b}$$

Here $p_\gamma = \partial u/\partial x^\gamma|_{u=0}$ and, since the right-hand sides of Eq. (22) are obtained at $p^\alpha = \text{const}$, it follows that

$$p^\mu p_\alpha = \frac{dx^\mu}{du}\frac{\partial u}{\partial x^\alpha}\Big|_{u=0} = \delta^\mu_\alpha.$$

In view of Eq. (20) the covariant derivative of X^μ may be written as

$$\frac{\delta X^\mu}{\delta x^\alpha} = \frac{\partial u}{\partial x^\alpha}\frac{\delta X^\mu}{\delta u} + \frac{\partial p^\gamma}{\partial x^\alpha}\frac{\delta X^\mu}{\delta p^\gamma}. \tag{23}$$

Each term in Eq. (23) can be evaluated by substitution of $u(\partial x^\mu/\partial u)$ for X^μ, where dx^μ/du is now written as $\partial x^\mu/\partial u$ due to the fact that x^μ is considered also a function of the integration constants p^μ as is indicated in Eq. (20). The covariant derivatives in Eq. (23) are

$$\frac{\delta X^\mu}{\delta u} = \frac{\partial x^\mu}{\partial u} + u\left(\frac{\partial^2x^\mu}{\partial u^2} + \Gamma^\mu_{\alpha\beta}\frac{\partial x^\alpha}{\partial u}\frac{\partial x^\beta}{\partial u}\right), \tag{24}$$

$$\frac{\delta X^\mu}{\delta p^\gamma} = \frac{\partial}{\partial p^\gamma}\left(u\frac{\partial x^\mu}{\partial u}\right) + \Gamma^\mu_{\alpha\beta}u\frac{\partial x^\alpha}{\partial u}\frac{\partial x^\beta}{\partial p^\gamma}. \tag{25}$$

With the repeated use of Eqs. (22), the last expression may be converted to

$$\frac{\delta X^\mu}{\delta p^\gamma} = \frac{\partial x^\mu}{\partial p^\gamma} + u^2p_\alpha\frac{\partial p^\alpha}{\partial x^\gamma}\left(\frac{\partial^2x^\mu}{\partial u^2} + \Gamma^\mu_{\sigma\rho}\frac{\partial x^\sigma}{\partial u}\frac{\partial x^\rho}{\partial u}\right). \tag{26}$$

Due to Eq. (18) the last terms in Eqs. (24) and (26) vanish so that the final result is

$$\frac{\delta X^\mu}{\delta x^\alpha} = \frac{\partial u}{\partial x^\alpha}\frac{\partial x^\mu}{\partial u} + \frac{\partial p^\gamma}{\partial x^\alpha}\frac{\partial x^\mu}{\partial p^\gamma} = \frac{dx^\mu}{dx^\alpha} = \delta^\mu_\alpha. \tag{27}$$

This result suffices to make the intended use of the vector X^μ. In analogy to Eq. (3), the angular momentum of the energy–momentum tensor $T^{\mu\nu\rho\sigma}$ can be formulated as

$$D_\mu(T^{\mu\nu\rho\sigma}X^\tau - T^{\mu\sigma\tau\nu}X^\rho) = 0. \tag{28}$$

It is satisfied, in view of Eqs. (9), (13), (27), and the distributive character of the covariant derivative indicated by the symbol D_μ. It should be noted that the trace of Eq. (28) reduces to the generally covariant equivalent of Eq. (3),

$$D_\mu(T^{\mu\tau}X^\rho - T^{\mu\rho}X^\tau) = 0. \tag{29}$$

[10] J. L. Synge and A. Schild, *Tensor Calculus* (The University of Toronto Press, Toronto, 1952), p. 60.

The other three symmetries of $T^{\mu\nu\rho\sigma}$ in Eqs. (11), (12), and (14) can be utilized to write down similar expressions to that in Eq. (28); however, they are nothing more than linear combinations of Eq. (28) so that they need not be considered.

Since the divergence of $T^{\mu\nu\rho\sigma}$ in Eq. (9) contains three free indices, it is possible to formulate higher moments of $T^{\mu\nu\rho\sigma}$ with a vanishing divergence. One can easily verify that

$$D_k \sum_{\mu\nu\tau} [(T^{k\mu\rho\nu}X^\sigma - T^{k\nu\sigma\mu}X^\rho)X^\tau] = 0, \qquad (30)$$

where $\sum_{\mu\nu\tau}$ stands for a sum of three terms in which $\mu\nu\tau$ are cyclicly permuted. Since the divergence of the parentheses vanishes by Eq. (28), it is sufficient to show that

$$\sum_{\mu\nu\tau} (T^{\tau\mu\rho\nu}X^\sigma - T^{\tau\nu\sigma\mu}X^\rho) = 0$$

in order to prove Eq. (30). But the last expression is zero in view of Eq. (14). Finally the divergence of the third moment of $T^{\mu\nu\rho\sigma}$ also vanishes if it is defined by

$$D_k \sum_{\rho\sigma\epsilon} \sum_{\mu\nu\tau} [(T^{k\mu\rho\nu}X^\sigma - T^{k\nu\sigma\mu}X^\rho)X^\tau X^\epsilon] = 0. \qquad (31)$$

Again one needs only prove that

$$\sum_{\rho\sigma\epsilon} \sum_{\mu\nu\tau} (T^{\epsilon\mu\rho\nu}X^\sigma - T^{\epsilon\nu\sigma\mu}X^\rho)X^\tau = 0.$$

Summing this expression first on $\rho\sigma\epsilon$, we get

$$\sum_{\mu\nu\tau} \{(T^{\epsilon\mu\rho\nu}X^\sigma - T^{\epsilon\nu\sigma\mu}X^\rho)X^\tau + (T^{\rho\mu\sigma\nu}X^\epsilon - T^{\rho\nu\epsilon\mu}X^\sigma)X^\tau + (T^{\sigma\mu\epsilon\nu}X^\rho - T^{\sigma\nu\rho\mu}X^\epsilon)X^\tau\} = 0.$$

Thus it is zero because terms cancel in pairs due to Eq. (13). It is not possible to formulate higher moments of $T^{\mu\nu\rho\sigma}$ than the third, because in Eq. (31) there are no more free indices left in the tensor $T^{\mu\nu\rho\sigma}$ for mixing with the index of the vector X^μ. This is also true of the tensor $T^{\mu\nu}$ in Eq. (29), consequently the higher moments conservation laws corresponding to Eqs. (30) and (31) do not exist in the Lorentz-covariant theories.

3. CONSERVED GLOBAL TENSORS

Any one of the locally conserved quantities in Eqs. (9), (28), (30), or (31) can be converted into a global tensor if it is integrated over some volume of the Riemannian space. Such integration requires in the integrand a factor of the Jacobian of transformation $g^{\frac{1}{2}}$, where g is the absolute value of the metric tensor determinant. This factor, being covariantly constant ($D_\mu g^{\frac{1}{2}} = 0$), can be pulled inside the divergence to make the tensor in question a tensor density. It was shown in Part I that the integral of the divergenceless tensor density $\mathfrak{I}^\mu$ (free indices suppressed on $\mathfrak{I}^\mu$) can be written as

$$\frac{\delta}{\delta x^0}\left(\frac{1}{3!}\sum_{123}\iiint_{V_3} \mathfrak{I}^0\, \delta x^i\, \delta x^j\, \delta x^k\right) + \tfrac{1}{2}\sum_{\text{cycl.}123}\iint_{V_2} \mathfrak{I}^i(\delta x^j\, \delta x^k + \delta x^k\, \delta x^j) = 0, \qquad (32)$$

where ijk refer to the spatial components. The same result can be expressed more compactly,

$$D_\mu p^\mu = \frac{\delta p^0}{\delta x^0} + \sum_{i=1}^{3}\frac{\delta p^i}{\delta x^i} = 0, \qquad (33)$$

where

$$p^\mu = \frac{1}{3!}\sum_{123}\iiint_{V_3} \mathfrak{I}^\mu\, \delta x^i\, \delta x^j\, \delta x^k.$$

Equation (32) [or (33)] is the global conservation law, or rather the global equation of continuity, which states that the covariant rate at which the amount of the quantity $\mathfrak{I}$ changes in the spatial volume V_3 is equal to the flow of that quantity through the surface V_2 bounding the volume V_3.

If the flow through some surface is zero for each component of the tensor, then the amount of the tensor within the corresponding volume remains covariantly constant in time.

$$\frac{\delta p}{\delta x^0} = 0; \quad p = p^0 = \frac{1}{3!}\sum_{123}\iiint \mathfrak{I}^0\, \delta x^i\, \delta x^j\, \delta x^k. \qquad (34)$$

The above statement implies that associated with this volume there are n quantities which are conserved in the ordinary sense. Here n is the number of independent invariants that can be formed from the tensor p. Clearly, from Eq. (34), it follows that

$$\frac{\delta}{\delta x^0} I_i(p) = \frac{\partial I_i}{\partial p}\frac{\delta p}{\delta x^0} = \frac{\partial I_i}{\partial x^0} = 0, \quad i = 1, \cdots, n, \qquad (35)$$

$I_i(p)$ being the independent invariants of p.

From now on only those spaces, in which the surface integrals at the spatial infinity vanish for the tensors of Eqs. (9), (28), (30), and (31), are considered.

In the stipulated spaces there are four covariantly conserved tensors which exceed the corresponding tensors of the Lorentz group in rank and number. Since only the latter are well understood, many questions relating to the algebraic properties, physical meaning and importance of the four tensors remain to be answered. Although no thorough investigation of them has been undertaken so far,

it is possible to comment on the more obvious algebraic structure as well as to make some inferences about the importance of these conservation laws.

Thus, when matter is completely absent ($T^{\mu\nu} = 0$), the Ricci tensor vanishes in view of the field equations (1), and the only surviving part in the energy tensor $T^{\mu\nu\rho\sigma}$ is the Weyl tensor $C^{\mu\nu\rho\sigma}$. In such a space filled with the gravitational radiation (field) only, none of the conserved tensors vanish identically. Consequently, the residual part of each conserved tensor pertains to the gravitational field. When matter is present, it is possible to split the conserved tensors invariantly into the matter part consisting of the same type of linear or angular momentum as in the Lorentz-covariant theory and the new, higher-rank traceless tensor which consists of the residual gravitational field plus those contributions of the matter fields which interact with the gravitational radiation.

This is well examplified by the global tensors derived from Eqs. (9) and (28). The first of these, $p^{\mu\nu\rho}$, is a third-rank tensor antisymmetric in $\nu\rho$,

$$p^{\mu\nu\rho} = \frac{1}{3!}\sum_{123}\iiint \sqrt{g}\, T^{0\mu\nu\rho}\, \delta x^1\, \delta x^2\, \delta x^3. \tag{36}$$

There are six components in $T^{\mu\nu\rho\sigma}$ without a single zero index, so that $p^{\mu\nu\rho}$ consists of 14 components.

It can be split into its traceless and trace parts as follows:

$$p^{\mu\nu\rho} = q^{\mu\nu\rho} + \tfrac{1}{3}(g^{\mu\nu}p^{\rho} - g^{\mu\rho}p^{\nu}), \tag{37}$$

where $p^{\rho} = g_{\mu\nu}p^{\mu\nu\rho}$ is the linear energy–momentum vector of the matter fields. From the definition of $q^{\mu\nu\rho}$ in Eq. (37), it follows that $q^{\mu\nu\rho}$ is a sum of two terms

$$q^{\mu\nu\rho} = c^{\mu\nu\rho} + m^{\mu\nu\rho}. \tag{38}$$

Here $c^{\mu\nu\rho}$ is the integral of the conformal tensor density $g^{\frac{1}{2}}C^{0\mu\nu\rho}$. It represents the gravitational field contribution to the energy and momentum of the system. $m^{\mu\nu\rho}$ is the integral of the tensor density entirely defined by the Ricci tensor so that it represents the contribution of the matter fields to the traceless energy–momentum tensor $q^{\mu\nu\rho}$. The decomposition of $p^{\mu\nu\rho}$ into two mutually orthogonal tensors in Eq. (37) (that is, a 10 component tensor $q^{\mu\nu\rho}$ and a four-component vector p^{ρ}) is invariant with regard to the general coordinate transformations. A number of significant consequences can be drawn from this invariance. First, the tensors $q^{\mu\nu\rho}$ and p^{ρ} are conserved separately,

$$\frac{\delta}{\delta x^0}p^{\rho} = 0; \qquad \frac{\delta}{\delta x^0}q^{\mu\nu\rho} = 0. \tag{39}$$

Secondly, the magnitude of p^{ρ}, $p = (|p^{\rho}p_{\rho}|)^{\frac{1}{2}}$, and the magnitude of $q^{\mu\nu\rho}$, $q = (|q^{\mu\nu\rho}q_{\mu\nu\rho}|)^{\frac{1}{2}}$, being two independent invariants of $p^{\mu\nu\rho}$, are both constants of the motion

$$\frac{\partial}{\partial x^0}p = 0; \qquad \frac{\partial}{\partial x^0}q = 0. \tag{40}$$

In analogy to p, which is the rest energy (or rest mass) of the dynamical system consisting of the matter fields, q can be interpreted as the rest energy of the gravitational field and those parts of the matter fields which interact with it. *Thus, in generally covariant theories, each dynamical system, is characterized not by one but by two rest masses.*

Thirdly, there can be no exchange of the linear energy–momentum (p^{ρ}) between the gravitational field and any of the matter fields due to the fact that the energy and momentum of the gravitational field is always expressed by the traceless tensor $q^{\mu\nu\rho}$. Exchange of energy and momentum between the matter fields and the gravitational field is allowed by means of the tensors $c^{\mu\nu\rho}$ and $m^{\mu\nu\rho}$, for neither of them is individually conserved although their sum is.

The latter part of the third conclusion has been known in various forms,[11] namely, that the lowest observable interaction mode between the gravitational radiation and a test particle is through quadrupole oscillations.

The global tensor derived from Eq. (28) expressing the conservation of angular momentum is a fourth-rank tensor

$$p^{\tau\nu\rho\sigma} = \frac{1}{3!}\sum_{123}\iiint_{V_3}\sqrt{g}\,(T^{0\nu\rho\sigma}X^{\tau} - T^{0\sigma\tau\nu}X^{\rho})\,\delta x^1\,\delta x^2\,\delta x^3. \tag{41}$$

It is antisymmetric with regard to the transposition of the first and the second pair of indices,

$$p^{\tau\nu\rho\sigma} = -p^{\rho\sigma\tau\nu}. \tag{42}$$

The maximum number of independent components in $p^{\tau\nu\rho\sigma}$ cannot exceed the product of the components of $T^{0\nu\rho\sigma}$ and X^{τ} or $14 \times 4 = 56$. Among these components, however, nine are identically zero,

$$p^{1010} = p^{1020} = p^{1030} = p^{2020} = p^{2030} = 0,$$
$$p^{2121} = p^{3030} = p^{3131} = p^{3232} = 0,$$

and five differ only by a sign,

$$p^{2120} = -p^{2021}, \quad p^{2312} = -p^{1232}, \quad p^{3130} = -p^{3031},$$
$$p^{3231} = -p^{3132}, \quad p^{3230} = -p^{3032}.$$

[11] B. DeWitt, in Ref. 1, p. 340.

Consequently, $p^{\tau\nu\rho\sigma}$ consists of 42 independent components. When it is contracted on $\nu\sigma$, a generally covariant equivalent of the angular momentum in the Lorentz theory is obtained.

$$p^{\tau\rho} = g_{\nu\sigma}p^{\tau\nu\rho\sigma}$$

$$= \frac{-1}{3!}\sum_{123}\iiint g^{\frac{1}{2}}(T^{0\rho}X^{\sigma} - T^{0\sigma}X^{\rho})\,\delta x^{i}\,\delta x^{j}\,\delta x^{k}. \quad (43)$$

Thus the trace of the tensor $p^{\tau\nu\rho\sigma}$ is the angular momentum of the matter fields. The other trace of $p^{\tau\nu\rho\sigma}$, $g_{\tau\rho}p^{\tau\nu\rho\sigma}$, consists of $p^{\nu\sigma}$ and another part dependent on the Weyl conform tensor C,

$$g_{\tau\rho}p^{\tau\nu\rho\sigma} = -p^{\nu\sigma} + f^{\nu\sigma}(c). \quad (44)$$

In analogy to the previously considered tensor, $p^{\tau\nu\rho\sigma}$ can also be decomposed into its traceless and trace parts

$$p^{\tau\nu\rho\sigma} = q^{\tau\nu\rho\sigma} + \tfrac{1}{4}g^{\nu\sigma}p^{\tau\rho}, \quad (45)$$

where $q^{\tau\nu\rho\sigma}$ is the traceless tensor with regard to the index pair $\nu\sigma$. Again, the tensor $q^{\tau\nu\rho\sigma}$ is a sum of two different terms, the residual term that is the integral of the conform tensor density $g^{\frac{1}{2}}(C^{0\nu\rho\sigma}X^{\tau} - C^{0\sigma\tau\nu}X^{\rho})$ which may be interpreted as the angular momentum of the gravitational field, and the matter term—the integral, whose density consists only of the Ricci tensor,

$$q^{\tau\nu\rho\sigma} = c^{\tau\nu\rho\sigma} + m^{\tau\nu\rho\sigma}. \quad (46)$$

From Eq. (45) it follows that the six-component tensor $p^{\tau\rho}$ and the 36-component tensor $q^{\tau\nu\rho\sigma}$ are conserved separately

$$\frac{\delta}{\delta x^{0}}p^{\tau\rho} = 0; \qquad \frac{\delta}{\delta x^{0}}q^{\tau\nu\rho\sigma} = 0. \quad (47)$$

Their magnitudes, being two independent invariants of $p^{\tau\nu\rho\sigma}$, are constants of the motion

$$\frac{\partial}{\partial x^{0}}(|p^{\tau\rho}p_{\tau\rho}|)^{\frac{1}{2}} = 0; \qquad \frac{\partial}{\partial x^{0}}(|q^{\tau\nu\rho\sigma}q_{\tau\nu\rho\sigma}|)^{\frac{1}{2}} = 0. \quad (48)$$

Conclusions drawn about the energy and momentum tensor $p^{\mu\nu\rho}$ are equally valid for the angular momentum tensor $p^{\tau\nu\rho\sigma}$ when appropriate terms referring to momentum are substituted with terms that refer to angular momentum.

In a generally covariant dynamical system, the angular momentum of the matter fields is described by the familiar six-component, antisymmetric, covariantly conserved tensor $p^{\tau\rho}$. The angular momentum of the gravitational field and of the matter fields which interact with the former is described by a new traceless, fourth-rank, covariantly conserved tensor $q^{\tau\nu\rho\sigma}$.

There can be no exchange of the angular momentum $p^{\tau\rho}$ between the gravitational field and the matter fields due to the traceless character of $q^{\tau\nu\rho\sigma}$. The exchange of the angular momentum between the gravitational radiation and the matter fields is allowed by means of the tensors $c^{\tau\nu\rho\sigma}$ and $m^{\tau\nu\rho\sigma}$, since neither of them is conserved.

Therefore, *with a generally covariant dynamical system, one associates not only two rest energies or masses but also two types of angular momentum.* Magnitudes of these momenta are two independent constants of the system. The remaining two conserved tensors derived from Eqs. (30) and (31) obviously cannot be decomposed into the traceless tensors and the lower-rank trace tensors of the Lorentz group. In the Lorentz-covariant theories, there are no known conserved tensors expressing either the second or the third moments of the energy and momentum to fulfill the role of traces. It is concluded, then, that both these tensors are of the same character as the previously discussed q tensors, that is they govern the exchange of the second and the third energy–momentum moments between the gravitational field and the matter fields. This is also borne out by the fact that traces of rank one and two of these tensors do not reduce to any tensors of the Lorentz group but vanish identically.

The second moment tensor of Eq. (30) is of the fifth rank. It will be written as

$$q^{\tau\sigma\mu\rho\nu} = \frac{1}{3!}\sum_{123}\iiint_{V_3} g^{\frac{1}{2}}\sum_{\mu\nu\tau}(T^{0\mu\rho\nu}X^{\sigma} - T^{0\nu\sigma\mu}X^{\rho})X^{\tau}\,\delta x^{i}\,\delta x^{j}\,\delta x^{k}, \quad (49)$$

where, it is recalled, $\sum_{\mu\nu\tau}$ stands for a sum of three terms in which $\mu\nu\tau$ are cyclicly permuted. Using the symmetry arguments below one can show that

$$g_{\mu\nu}g_{\sigma\rho}q^{\tau\sigma\mu\rho\nu} = 0 \quad (50)$$

so that $q^{\tau\sigma\mu\rho\nu}$ is indeed traceless. A number of symmetries of $q^{\tau\sigma\mu\rho\nu}$ follow directly from its definition in Eq. (49).

$$q^{\tau\sigma\mu\rho\nu} = q^{\mu\sigma\nu\rho\tau} = q^{\nu\sigma\tau\rho\mu}, \quad (51)$$

$$q^{\tau\sigma\mu\rho\nu} = -q^{\mu\rho\tau\sigma\nu} = -q^{\tau\rho\nu\sigma\mu}. \quad (52)$$

They can be easily checked by writing out the sum $\sum_{\mu\nu\tau}$ explicitly.

The number of independent components of $q^{\tau\sigma\mu\rho\nu}$ cannot exceed the product of the components of $T^{0\mu\rho\nu}$ and $X^{\sigma}X^{\tau}$ or $14 \times 10 = 140$. However, in view of the above symmetries, many components may be linearly related or may vanish identically.

The last conserved tensor describing the third moment of the energy and momentum is a six-rank tensor $q^{\tau\sigma\mu\rho\nu\epsilon}$, where

$$q^{\tau\sigma\mu\rho\nu\epsilon} = \frac{1}{3!}\sum_{123}\iiint_{V_3} g^{\frac{1}{2}}\sum_{\mu\nu\tau}\sum_{\rho\sigma\epsilon}(T^{0\mu\rho\nu}X^{\sigma} - T^{0\nu\sigma\mu}X^{\rho})X^{\tau}X^{\epsilon}\,\delta x^{i}\,\delta x^{j}\,\delta x^{k}. \quad (53)$$

In view of the sums $\sum_{\mu\nu\tau}$ and $\sum_{\rho\sigma\epsilon}$ in the definition of $q^{\tau\sigma\mu\rho\nu\epsilon}$, it is obvious that $q^{\tau\sigma\mu\rho\nu\epsilon}$ does not change when either $\mu\nu\tau$ or $\rho\sigma\epsilon$ are cyclicly permuted as in Eq. (51). When the sums are written out explicitly, one recognizes by inspection the following symmetries: $q^{\tau\sigma\mu\rho\nu\epsilon}$ is completely antisymmetric in the three indices $\mu\nu\tau$ and completely symmetric in the remaining three indices $\rho\sigma\epsilon$. Obviously, all traces of the antisymmetric index pairs $\mu\nu$, $\mu\rho$, $\nu\rho$ vanish. The symmetric index triple $\rho\sigma\epsilon$ generates 20 distinct components and the antisymmetric one only four, consequently, the number of independent components of $q^{\tau\sigma\mu\rho\nu\epsilon}$ should not exceed the product of these, that is 80 components.

4. SUMMARY

In generally covariant theories, a dynamical system is characterized by four global tensors which express conservation of the system's energy and momentum, and their first three moments. The first two tensors are reducible into traceless and trace tensors. The latter constitute the generally covariant equivalents of the conserved tensors of the Lorentz theory; however, contrary to common expectations, they say nothing about the energy or momentum of the gravitational field but describe exclusively the matter fields. The gravitational field energy and momentum as well as its first three moments are contained in the four traceless tensors of the third, fourth, fifth, and sixth rank, respectively. The same tensors contain also the contribution from the matter fields via the Ricci tensor terms. Each traceless tensor is covariantly conserved, but the gravitational field and the matter field parts in it are not conserved separately. This fact allows for interaction of the gravitational field and matter fields with an exchange of energy and momentum or any of its first three moments between them.

Conservation of the second and third moments of the energy and momentum arise in consequence of the high rank of the energy–momentum tensor. The latter is required by the nature of the gravitational field in the general relativity theory. One can easily see by considering quadrupole or higher-order multipole radiation that momentum transferred to a test particle by such radiation is not confined to one direction, as is the case with the linear momentum, but is distributed in different directions simultaneously. Such distribution can be described only by a tensor of higher rank. Moreover, the net linear momentum in any one direction imparted by the quadrupole or higher-order multipole radiation to the test particle is zero. This explains the traceless character of the four q tensors.

5. EXAMPLES OF GLOBAL CONSERVED TENSORS

Implications and usefulness of the global conserved tensors are easily shown by obtaining examples of some of these tensors in specific Riemannian spaces. This also serves other purposes. First, the method of tensor integration is demonstrated. Secondly, it is possible to show that some of the assumptions made, such as the one about vanishing of certain surface integrals at the spatial infinity, do not lead to a trivial class of conserved tensors. Finally, the numerical results have an important bearing on the physical interpretation of spaces under consideration. Also, they may shed some light not only on the relative magnitudes of the matter tensors and the q tensors, but, hopefully, indicate their relative importance in general.

Two considered spaces are that of a neutral mass m or the Schwarzschild metric and that of a mass m with an electric charge e. The metric and the curvature tensor need be specified only for the latter since they become identical to those of the neutral space when the charge is set equal to zero ($e = 0$). This metric and the corresponding nonzero Γ's are[12–14]

$$g_{\mu\nu}\,dx^{\mu}\,dx^{\nu} = -\left(1 - \frac{r_0}{r} + \frac{r_0 r_1}{r^2}\right)(c\,dt)^2 + \frac{dr^2}{1 - (r_0/r) + (r_0 r_1/r^2)} + r^2(d^2\theta + \sin^2\theta\, d^2\varphi),$$

$$\Gamma^0_{01} = \Gamma^0_{10} = \tfrac{1}{2}g^{00}\,\partial_1 g_{00};\quad \Gamma^1_{00} = -\tfrac{1}{2}g^{11}\,\partial_1 g_{00};$$

$$\Gamma^1_{11} = \tfrac{1}{2}g^{11}\,\partial_1 g_{11},\quad \Gamma^1_{22} = \sin^{-2}\theta\,\Gamma^1_{33} = r g_{00};$$

$$\Gamma^2_{12} = \Gamma^2_{21} = \Gamma^3_{13} = \Gamma^3_{31} = 1/r, \quad (54)$$

$$\Gamma^2_{33} = -\sin\theta\cos\theta;\quad \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta.$$

Here ct, r, θ, ϕ are the coordinates. The two constants

[12] P. G. Bergmann, *Introduction to the Theory of Relativity* (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1953).
[13] R. C. Tolman, *Relativity, Thermodynamics and Cosmology* (Clarendon Press, Oxford, England, 1934).
[14] H. Weyl, *Space-Time-Matter* (Dover Publications, Inc, New York, 1950).

r_0 and r_1 are the Schwarzschild radius and the charge radius, respectively.

$$r_0 = 2km/c^2; \qquad r_1 = e^2/2mc^2, \tag{55}$$

where k is the Newtonian gravitational constant which already appeared in Eq. (1). The six nonvanishing components of the curvature tensor are

$$\begin{aligned} R_{011}{}^0 &= \left(\frac{r_0}{r^3} - \frac{3r_0r_1}{r^4}\right)\left(1 - \frac{r_0}{r} + \frac{r_0r_1}{r^2}\right)^{-1}; \\ R_{022}{}^0 &= R_{122}{}^1 = -\frac{1}{2}\left(\frac{r_0}{r} - \frac{2r_0r_1}{r^2}\right), \\ R_{033}{}^0 &= -\frac{1}{2}\left(\frac{r_0}{r} - \frac{2r_0r_1}{r^2}\right)\sin^2\theta = R_{133}{}^1; \\ R_{233}{}^2 &= \left(\frac{r_0}{r} - \frac{r_0r_1}{r^2}\right)\sin^2\theta. \end{aligned} \tag{56}$$

The nonvanishing components of $T^{\mu\nu\rho\sigma}$ containing at least one zero index are

$$\begin{aligned} T_{01}{}^{10} &= \frac{-c^4}{8\pi k}\left(\frac{r_0}{r^3} - \frac{r_0r_1}{r^4}\right); \\ T_{02}{}^{20} &= T_{03}{}^{30} = \frac{-c^4}{8\pi k}\left(-\frac{r_0}{2r^3} + \frac{r_0r_1}{r^4}\right). \end{aligned} \tag{57}$$

All components of the form $T_{ij}{}^{i0}$ $i \neq 0$ vanish identically so that the surface integrals of Eq. (32) are zero for all four tensor densities and for any arbitrary choice of the spatial surface V_2.

In view of Eq. (57) the energy–momentum tensor $p^{\mu\nu\rho}$ consists only of three components $p^i{}_{i0}$, $i = 1, 2, 3$ which are given by

$$\begin{aligned} p^i{}_{i0} = \frac{1}{6}\iiint r^2 \sin\theta T^{0i}{}_{i0}\{&\delta r(\delta\theta\ \delta\varphi + \delta\varphi\ \delta\theta) \\ &+ \delta\theta(\delta r\ \delta\varphi + \delta\varphi\ \delta r) + \delta\varphi(\delta r\ \delta\theta + \delta\theta\ \delta r)\}. \end{aligned} \tag{58}$$

If any one of the six integrals in Eq. (58) is denoted by $p_n^{i\,i0}$ then it can be shown that

$$\begin{aligned} &(D_iD_j - D_jD_i)p_n^{s\,s0} = (D_iD_j - D_jD_i)D_kp_n^{s\,s0} = 0, \\ &i, j, k, s = 1, 2, 3 \quad i \neq j \neq k, \quad n = 1, \cdots, 6, \end{aligned} \tag{59}$$

so that all tensor integrations δr, $\delta\theta$, $\delta\phi$ commute in this space. Consequently, $p^i{}_{i0}$ are

$$\begin{aligned} p^i{}_{i0} &= \iiint r^2 \sin\theta T^{0i}{}_{i0}\ \delta r\ \delta\theta\ \delta\varphi \\ &= 4\pi\int r^2 T^{0i}{}_{i0}\ \delta r. \end{aligned} \tag{60}$$

In the last integral the integrations over ϕ and θ have been performed over the entire space. The result is 4π, since they happen to coincide with the ordinary integration. The r integration does not, but it can be performed by taking the covariant derivative of Eq. (60) with regard to r and solving the resulting differential equation for $p^i{}_{i0}$. This yields

$$\begin{aligned} p^i{}_{i0} &= 4\pi(-g_{00})^{\frac{1}{2}}\int r^2(-g_{00})^{-\frac{1}{2}}T^{0i}{}_{i0}\,dr; \\ -g_{00} &= 1 - \frac{r_0}{r} + \frac{r_0r_1}{r^2}. \end{aligned} \tag{61}$$

Substituting the expressions for $T^{0i}{}_{i0}$ from Eq. (57), the three components are

$$\begin{aligned} p^1{}_{10} &= -mc^2(-g_{00})^{\frac{1}{2}}(F_1 - F_2), \\ p^2{}_{20} = p^3{}_{30} &= -mc^2(-g_{00})^{\frac{1}{2}}(-\tfrac{1}{2}F_1 + F_2). \end{aligned} \tag{62}$$

The functions F_i are two integrals

$$\begin{aligned} F_1 &\equiv \ln\left[\frac{r}{r_0}(-g_{00})^{\frac{1}{2}} + \frac{r}{r_0} - \frac{1}{2}\right] + c_1, \\ F_2 &\equiv \left(\frac{r_1}{r_0}\right)^{\frac{1}{2}}\left\{\ln\left[(-g_{00})^{\frac{1}{2}} + \frac{(r_0r_1)^{\frac{1}{2}}}{r} - \frac{1}{2}\left(\frac{r_0}{r_1}\right)^{\frac{1}{2}}\right] + c_2\right\}, \end{aligned} \tag{63}$$

where c_1 and c_2 are the constants of integration which can be identified with the lower limit of the integral in Eq. (61) if it is definite. The trace of $p^i{}_{i0}$ and its traceless tensor are

$$\begin{aligned} p_0 &= \sum_{i=1}^{3} p^i{}_{i0} = -mc^2(-g_{00})^{\frac{1}{2}}F_2, \\ q^1{}_{10} &= -mc^2(-g_{00})^{\frac{1}{2}}(F_1 - \tfrac{4}{3}F_2); \\ q^2{}_{20} = q^3{}_{30} &= -mc^2(-g_{00})^{\frac{1}{2}}(-\tfrac{1}{2}F_1 + \tfrac{2}{3}F_2). \end{aligned} \tag{64}$$

The magnitudes or rest energies of p_0 and $q^i{}_{i0}$ are

$$p = mc^2\,|F_2|; \qquad q = mc^2\,|(\tfrac{3}{2})^{\frac{1}{2}}F_1 - (\tfrac{8}{3})^{\frac{1}{2}}F_2|. \tag{65}$$

In the Riemannian space of the Schwarzschild metric r_1 is zero so that F_2 vanishes. But the rest energy of the gravitational field contained in the spherical shell of thickness $r - r_L$ does not vanish, and according to Eq. (65) is equal to

$$q_{\text{Sch}} = mc^2(\tfrac{3}{2})^{\frac{1}{2}}\ln\frac{r(1 - r_0/r)^{\frac{1}{2}} + r - \frac{1}{2}r_0}{r_L(1 - r_0/r_L)^{\frac{1}{2}} + r_L - \frac{1}{2}r_0}. \tag{66}$$

Here c_1 in F_1 is chosen so as to coincide with the lower limit r_L of the integral in Eq. (61). The lowest value that r_L can assume is the Schwarzschild radius r_0. The upper limit r may be made to approach infinity, in which case the rest energy of the gravitational field diverges logarithmically. In order to obtain a finite result, it is necessary either to intro-

duce a cutoff or to inquire about the energy contained in shells of finite thickness.

First, the energy between r_0 and r_1 is calculated, where it is assumed that r_0 in Eq. (55) is produced by the smallest known mass m, that is, the mass of the electron and r_1 is of the order of nuclear size, or more accurately, the classical electron radius of Eq. (55) with e, m being the electron charge and mass. For this case, one obtains approximately

$$q_{\text{Sch}}(r_0, r_1) = 119mc^2. \tag{67}$$

The energy of the gravitational field within the volume of the size of a nuclear particle is two orders of magnitude greater than the rest energy of the particle which produces the field.

In the second calculation, let the upper limit r be extended up to the radius of the visible universe, that is, $r = r_2 = 10^{28}$ cm. In fact, it may be argued that this is the maximum that the upper limit should assume, since the regions beyond this point are not causally connected with the field-producing particle. The rest energy of the gravitational field within the volume of the size of the visible universe is only about twice of the value in Eq. (67), that is,

$$q_{\text{Sch}}(r_0, r_2) = 234mc^2. \tag{68}$$

One concludes from this that the gravitational field energy is concentrated in the immediate neighborhood of the Schwarzschild radius r_0.

In the Riemannian space of the charged particle with the metric in Eq. (54) the field is characterized by two rest energies, that of the matter field which in this case is the Coulomb field and that of the gravitational field and matter field. The rest energy of the Coulomb field contained in a spherical shell of thickness $r_u - r$ is

$$p = mc^2\left(\frac{r_1}{r_0}\right)^{\frac{1}{2}} \times \ln \frac{\left(1 - \frac{r_0}{r} + \frac{r_0 r_1}{r^2}\right)^{\frac{1}{2}} + \frac{(r_0 r_1)^{\frac{1}{2}}}{r} - \frac{1}{2}\left(\frac{r_0}{r_1}\right)^{\frac{1}{2}}}{\left(1 - \frac{r_0}{r_u} + \frac{r_0 r_1}{r_u^2}\right)^{\frac{1}{2}} + \left(\frac{r_0 r_1}{r_u}\right)^{\frac{1}{2}} - \frac{1}{2}\left(\frac{r_0}{r_1}\right)^{\frac{1}{2}}}. \tag{69}$$

The upper limit r_u can be extended to infinity with no ill consequences. However, the lower limit r, on approaching zero, yields a divergent result. This is not surprising since the Coulomb self-energy is known to diverge as $1/r$. What is new here is that the general covariance removes one degree of divergence so that the Coulomb energy p in Eq. (69) diverges only logarithmically.

To get a finite result for p a cutoff has to be introduced. Since the classical electron radius r_1 is indicative of the size of charge distribution, the lower limit may be chosen to be on the order of r_1, say r_1/n, where n is close to unity. With $r_u = \infty$ one obtains

$$p = mc^2\left(\frac{r_1}{r_0}\right)^{\frac{1}{2}} \ln \frac{1 + (n - \frac{1}{2})\left(\frac{r_0}{r_1}\right)^{\frac{1}{2}}}{1 - \frac{1}{2}\left(\frac{r_0}{r_1}\right)^{\frac{1}{2}}} = nmc^2. \tag{70}$$

In the calculation of the second rest energy q it is necessary to introduce also the upper cutoff $r_u = r_2$. With these limits on the integral in Eq. (61) q becomes

$$q\left(\frac{r_1}{n}, r_2\right) = (105.75 + 1.67n + 1.22 \ln n)mc^2. \tag{71}$$

In the above calculation the cutoffs were imposed on both integrals F_1 and F_2 of Eq. (64), however, each function F requires only one cutoff. If the function F_1 were to be calculated with the upper cutoff only, then q in Eq. (71) should be augmented by

$$q\left(0, \frac{r_1}{n}\right) = 60mc^2. \tag{72}$$

The removal of the upper cutoff on F_2 changes its value only infinitesimally since F_2 converges at $r \to \infty$.

Although the Riemannian space of the charged particle is considerably different from the space of a neutral particle, e.g., the charge removes the Schwarzschild interior region of $r < r_0$, the qualitative features of both these spaces are the same. First, there is the field-producing particle represented by a singularity in the interior of the Schwarzschild radius or in the interior of the charge distribution. The latter is assumed to be a point charge due to the lack of a more satisfactory theoretical or experimental charge model. In view of its singular nature, this particle is not governed by the field equations. The singular particle is surrounded by two types of fields, the matter field—provided the particle carries a matter "charge"—and the gravitational field. The rest energy of the matter field is on the order of the rest energy of the singularity, whereas the rest energy of the gravitational field is two orders of magnitude greater than either of the other two. The "heavy" gravitational cloud surrounding the particle is mostly contained in a volume of the size of a nuclear particle.

6. INTERPRETATION

Appearance of the q tensors on the scene raises some questions in regard to their importance relative to the trace tensors, delineation of domains in which they are of primary significance, and their physical meaning.

Answers to some of these questions are interrelated and have to be discussed jointly to a degree.

It is shown in Sec. 3 that the class of four conserved tensors divides into two subclasses, one containing tensors of the Lorentz theory, and the other consisting of high rank, traceless q tensors. Since tensors in each subclass are conserved separately any measurement or knowledge of a tensor in one subclass does not extend to, or say anything about, a similar tensor in the other subclass. Thus the q tensors can be interpreted as those degrees of freedom which are necessary to specify a generally covariant dynamical system in addition to the familiar linear and angular momentum.

Although it is not possible to ascribe different levels of importance to various degrees of freedom, the q tensors do seem to be more fundamental in the following sense. In a Riemannian space none of the q tensors need be zero when all matter fields vanish. At the same time none of the q tensors can be made to vanish in the presence of any nonzero matter field without violating the field equations.

Usefulness of the q tensors in those theories, where the gravitational field has to be dealt with explicitly, is rather evident and need not be elaborated on here. What is most interesting, and at the same time least certain, is the speculation that the q tensors are the internal degrees of freedom of the elementary particles. This is strongly suggested by the picture of the particle which emerged from the specific examples of the conserved tensors considered in the previous section. Thus, can the second rest energy q account for heaviness of some elementary particles and is its conservation synonymous with the conservation of heavy particles? Is there any connection between the higher moments q tensors and the various particle spins? However, any such specific identifications are premature at this time. The conjecture that there is a connection between the q tensors and the internal degrees of freedom of the elementary particles can be proved or disproved only after these tensors are exhaustively studied and analyzed for their formal structure and symmetries, after they are applied to more realistic Riemannian spaces where cutoffs need not be introduced, and after one develops a set of observables of these tensors. The correspondence, or a lack thereof, between the q tensors and the internal degrees of freedom of the particles will then be easily recognized.

ACKNOWLEDGMENTS

The author is indebted to Professor Rohrlich and others of Syracuse University for reading the first draft of the material appearing in Parts I and II and for their valuable criticisms and suggestions.

Cylindrical Gravitational News*

JOHN J. STACHEL
Department of Physics, Boston University, Boston, Massachusetts
(Received 16 December 1965)

The concepts of news function and mass aspect are generalized to a class of cylindrically symmetric metrics containing both degrees of freedom of the gravitational field. It is proved that the mass/unit length always decreases if there is any cylindrical news. The asymptotic behavior of the Riemann tensor in the cylindrical case is analyzed and a peeling theorem proved for this case. An example is given to show that asymptotic conditions on the metric or the Riemann tensor which are analogous to the conditions used in the asymptotically spherical case do not exclude certain infinite incoming radiation trains in the cylindrical case. Pure incoming and outgoing solutions are defined for the cylindrical case, and their generalization to the asymptotically spherical case is suggested. An exactly conserved quantity is shown to exist which may be the cylindrical analog of the ten exactly conserved quantities recently discovered by Newman and Penrose.

I. INTRODUCTION

IT is well known that, according to the general theory of relativity, the gravitational field is capable of supporting radiation with two degrees of freedom.[1] In recent years a great deal of progress has been made in the analysis of the asymptotic gravitational radiation field of bounded sources by carrying out the analysis on null hypersurfaces.[2] Yet, so far no exact solution has been found representing the exterior field around a bounded radiating source, much less a solution representing the source itself together with its radiation field.[3] In the absence of such solutions, the linearized theory has been used recently to investigate the relation between the asymptotic field and its sources; and a definition of the multipole structure of sources in the exact theory has been proposed on the basis of this analysis.[4]

Although the study of bounded sources and their fields is undoubtedly the best model for realistic physical situations (where one might hope that gravitational radiation plays some role), there is much to be said for the study of any exact solutions available which allow some aspects of the problem to be studied, even if they are physically unrealistic. They can serve to demonstrate how well the new methods accomplish their aims in a simplified model. And, more importantly, they can serve as starting points for the approach to problems too complex to be solved in the more realistic but mathematically more intractable cases.

Cylindrical gravitational waves offer such a model where both degrees of freedom of the gravitational field may be studied in a relatively simple mathematical context with the hope of relating them exactly to their sources. This paper applies some of the recently developed methods of analysis to this case. Section II summarizes the nature of the class of cylindrical radiation metrics now known, and gives a new solution to field equations of this type. In Sec. III a news function is defined for the cylindrical case, and the subclass of solutions for which this news function exists is discussed. Section IV discusses the asymptotic structure of the Riemann tensor in the cylindrical case. Section V is a summary, with indications of further problems remaining to be solved.

II. CYLINDRICAL GRAVITATIONAL WAVES

A class of time-dependent cylindrically symmetric solutions to the Einstein field equations seems to have been first discovered by Beck,[5] who found them by making the complex transformation $z \to it$, $t \to iz$ on the Weyl–Levi-Civita class of axially symmetric static metrics. They were rediscovered ten years later by Einstein and Rosen[6] in an explicit search for time-dependent cylindrically symmetric solutions to the field equations. This class of metrics contains

* This research was partially supported by the National Science Foundation while the author was at the University of Pittsburgh; and by the Aerospace Research Laboratories of the Office of Aerospace Research, United States Air Force, during a visit to Temple University.

[1] See, for example, R. K. Sachs, *Relativity, Groups and Topology*, C. DeWitt and B. DeWitt, Eds. (Gordon and Breach, Science Publishers, Inc. New York, 1964), p. 523.

[2] H. Bondi, M. van der Burg, and A. Metzner, Proc. Roy. Soc. (London) **A269**, 21 (1962); R. K. Sachs, *ibid.* (London) **A270**, 103 (1962); E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[3] A solution representing asymptotically spherical waves has been given by I. Robinson and A. Trautman, Proc. Roy. Soc. (London) **A265**, 463 (1962), but it does not represent radiation from a bounded source.

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

[5] G. Beck, Z. Physik **33**, 713 (1925).

[6] A. Einstein and N. Rosen, J. Franklin Inst. **223**, 43 (1937).

only one degree of freedom of the radiation field. The second degree of freedom was implicitly excluded by Einstein and Rosen's demand for reflection symmetry of their metric with respect to the transformation $z \to -z$. A class of solutions containing the second degree of freedom was discovered by Ehlers,[7] using a method similar to Beck's. He discovered a generalization of the Weyl–Levi-Civita metrics and performed a complex transformation on them. Kompaneets[8] independently discovered this class by a generalization of the Einstein–Rosen method, dropping the demand for reflection symmetry.

The metric, in the form given by Jordan and Ehlers, is

$$ds^2 = e^{2(\gamma-\psi)}(dt^2 - d\rho^2) - e^{2\psi}\,dz^2 - 2\chi e^{2\psi}\,dz\,d\phi - (\chi^2 e^{2\psi} + \rho^2 e^{-2\psi})\,d\phi^2, \quad (2.1)$$

where $\psi(\rho, t)$ and $\chi(\rho, t)$, representing the two degrees of freedom of the gravitational field, are determined by the two coupled nonlinear partial differential equations

$$(1/\rho)(\rho\psi_{,\rho})_{,\rho} - \psi_{,tt} = (1/2\rho^2)e^{4\psi}(\chi_{,\rho}^2 - \chi_{,t}^2), \quad (2.2)$$

$$(1/\rho)(\rho\chi_{,\rho})_{,\rho} - \chi_{,tt} = (2/\rho)\chi_{,\rho} + 4(\chi_{,t}\psi_{,t} - \chi_{,\rho}\psi_{,\rho}). \quad (2.3)$$

The function $\gamma(\rho, t)$ is determined by the equations

$$\gamma_{,\rho} = \rho(\psi_{,\rho}^2 + \psi_{,t}^2) + (1/4\rho)e^{4\psi}(\chi_{,t}^2 + \chi_{,\rho}^2),$$
$$\gamma_{,t} = 2\rho\psi_{,\rho}\psi_{,t} + (1/2\rho)e^{4\psi}\chi_{,\rho}\chi_{,t}, \quad (2.4)$$

once ψ and χ are given. If χ equals zero, the metric reduces to the Einstein–Rosen metric, which has only one degree of freedom ψ determined by Eq. (2.2), which then reduces to the scalar cylindrical wave equation. This case has been extensively discussed by Rosen,[9] Weber and Wheeler,[10] Marder,[11] and Thorne.[12] It has been shown that the static solution $\psi = a \ln \rho + b$ represents the field exterior to a massive infinite cylinder.[11] The values $a = 0$ and $a = 1$ correspond to flat space–time; but $a = 1$ corresponds to an interchange of the roles of z and ϕ. So it is more convenient to use the value $a = 0$ for flat space with this labeling of coordinates. Outgoing and incoming radiation solutions representing finite pulses of radiation and infinite wave trains have been constructed,[9] and the outgoing solutions shown to carry off energy and momentum as measured by the Einstein pseudotensor calculated in a suitable way.[13] A linear equation governs ψ, so any arbitrary combination of incoming and outgoing solutions may be superposed.

Since the case of one degree of freedom, the Einstein–Rosen case, leads to a simple wave equation for ψ, one is led to investigate the possibilities of simple results for the other degree of freedom. However, if we try to solve the coupled set of Eq. (2.2) and (2.3) with $\psi = 0$, we easily see that no solution is possible except $\chi = \text{const.}$, which can be eliminated by the coordinate transformation $z' = z + c\phi$. The next simplest possibility is to take the static solution $\psi = a \ln \rho + b$, and see if any solution for χ is compatible with it. It is found that only for $a = \frac{1}{4}$ can there be any nonconstant solution, and in this case χ may be any function of $(t - \rho)$ or of $(t + \rho)$ (but not their sum, because of the nonlinearity of the equations).[14] So we do indeed find a class of solutions depending on an arbitrary function of a particularly simple type for the second degree of freedom when $a = \frac{1}{4}$. Unfortunately, this solution is of even less physical significance than other cylindrical metrics. Examination of the behavior of a distant test particle in the field of the static solution $\psi = a \ln \rho + b$ shows that, for $0 < a < 1$, the test particle is repelled from the cylinder, indicating that it has negative mass. Application of the general relativistic analog of Gauss' theorem for static solutions[15] also shows that the gravitational mass/unit length of the cylinder is proportional to $a(a - 1)$; so that, for $0 < a < 1$, the enclosed gravitational mass is seen to be negative by this method as well. However, this solution does not seem to be untypical of the behavior of a class of solutions for χ corresponding to values of a which make the central cylinder attractive, as a study of weak linearized solutions to Eqs. (2.2) and (2.3) or the equivalent Eqs. (A12) and (A13) shows. This solution will be of some interest when we study the

[7] First reported by P. Jordan, "Research on the Theory of Relativity," U. S. Air Force Tech. Note A. R. L. WCIJ TN 88-1.
[8] A. Kompaneets, Soviet Phys.—JETP **34**, 659 (1958).
[9] N. Rosen, Bull. Res. Council Israel **3**, 328 (1954).
[10] J. Weber and J. Wheeler, Rev. Mod. Phys. **29**, 509 (1957).
[11] L. Marder, Proc. Roy. Soc. (London) **A244**, 524 (1958).
[12] K. Thorne, Phys. Rev. **138**, B251 (1965).
[13] J. Boardman and P. Bergmann, Phys. Rev. **115**, 1318 (1959) shows this for the Einstein–Rosen case. The author showed that the same result holds [namely, the loss of energy is proportional to $\partial/\partial t\ (e^{2\gamma}-1)$] when both states of polarization are present, in an independent calculation: "Energy Flow in Cylindrical Gravitational Waves" (Master's Thesis, Stevens Institute of Technology, 1959).
[14] J. Stachel, Bull. Am. Phys. Soc. **6**, 305 (1961).
[15] E. T. Whittaker, Proc. Roy. Soc. (London) **A149**, 384 (1935).

asymptotic behavior of solutions and the news function.

III. GRAVITATIONAL NEWS IN CYLINDRICAL CASE

A. Coordinates and Tetrads

In order to see when a news function type of analysis is possible for the class of cylindrical wave metrics discussed in the previous section, it is necessary to introduce a null coordinate so that we may examine the behavior of the solutions on a family of null hypersurfaces. Considerations of symmetry immediately suggest using $u = t - \rho$ as the null coodinate in this case. In the analysis of the asymptotically spherical case,[2] one may then proceed to use either the affine parameter, the luminosity distance, or the parallax distance along the radial null geodesics on the null hypersurface $u = \text{const}$, because the ratio of all three distances approaches unity sufficiently rapidly as we go to infinity (see Sachs, Ref. 2 for definitions and discussion of this point). But this is no longer true in the cylindrical case, for, unless $a = 0$ or 1 in the static part of the solution for ψ, the space does not get asymptotically flat rapidly enough as we go toward infinity along a null hypersurface, and the ratio of the three distances does not approach one. So we must decide which is the most convenient coordinate in this case. For a cylindrically symmetric metric it seems reasonable to define the luminosity distance so that the brightness falls off linearly with the distance from a uniform infinite line source. With this definition, ρ is the luminosity distance, and it is convenient to continue it as one of our coordinates. Sachs[2] notes in the spherical case that "the ratio of the parallax to luminosity distance... appears as a kind of correction term for the specifically nonlinear effects of the field in many places...," and we shall find this to be true in the cylindrical case as well. Since it does not approach unity (except for $a = 0$ or 1), it appears in the final results that, in many places, it drops out of the spherical results. Its value in our case is $e^{2(\gamma-\psi)}$.

It is also necessary to choose an orthonormal tetrad and a null tetrad (or Sachsbein), with respect to which we later take the physical components of the Riemann tensor. Considerations of symmetry uniquely fix the directions of the timelike and radial spacelike unit vectors of the orthonormal tetrad, and thus the directions of the two real null vectors of the Sachsbein. However, the normalizations of the vectors are not thereby fixed, and two different choices prove useful. For taking physical components of the Riemann tensor, it is more useful to use the sum of the timelike and radial spacelike vectors for the future outward-directed null vector l^μ.[16] For the study of the field equations and the news function in this section, it proves more convenient to use a slightly different normalization, since l^μ, as defined above, is not the gradient of u. Multiplying l^μ by $e^{(\psi-\gamma)}$ (and correspondingly renormalizing the other null vector—this amounts to a Lorentz transformation in the plane of the two null vectors), we make the new $\bar{l}^\mu$ the gradient of u. An element of arbitrariness remains in the choice of the other two spacelike tetrad vectors (and thus of the complex null Sachsbein vector). If χ were zero, however, they would be uniquely defined by symmetry considerations, and we partially remove the arbitrariness by choosing them to take advantage of the additional symmetry when χ is zero.

Orthonormal tetrad	*Null tetrads*
$e^\mu_0 = e^{\psi-\gamma}\,\delta^\mu_t,$	$l^\mu = e^\mu_0 + e^\mu_1, \quad \bar{l}^\mu = e^{(\psi-\gamma)} l^\mu,$
$e^\mu_1 = e^{\psi-\gamma}\,\delta^\mu_\rho,$	$n^\mu = \tfrac{1}{2}(e^\mu_0 - e^\mu_1), \quad \bar{n}^\mu = e^{(\gamma-\psi)} n^\mu,$
$e^\mu_2 = e^{-\psi}\,\delta^\mu_z,$	$m^\mu = (1/\sqrt{2})(e^\mu_2 + ie^\mu_3),$
$e^\mu_3 = (\delta^\mu_\phi + \chi\,\delta^\mu_z)(e^\psi/\rho).$	$\bar{m}^\mu = (1/\sqrt{2})(e^\mu_2 - ie^\mu_3).$

(3.1)

In this section we shall refer exclusively to the null tetrad $(\bar{l}^\mu, \bar{n}^\mu, m^\mu, \bar{m}^\mu)$. Straightforward computation shows that the divergence and the shear of the null congruence $\bar{l}^\mu$ are given by

divergence:

$$\theta = \nabla_\mu \bar{l}^\mu = e^{2(\psi-\gamma)}/\rho; \tag{3.2}$$

shear:

$$|\sigma|^2 = \tfrac{1}{2}(\nabla_\nu \bar{l}^\mu)(\nabla^\nu \bar{l}_\mu) - \tfrac{1}{4}\theta^2$$
$$= e^{4(\psi-\gamma)}[\psi^2_{,\rho} - \psi_{,\rho}/\rho + 1/4\rho^2 + 5/4(\chi^2_{,\rho}/\rho^2)e^{4\psi}]. \tag{3.3}$$

For static solutions $\psi = a \ln \rho + b$, the value of the shear is $[e^{2(\psi-\gamma)}(\tfrac{1}{2} - a)]/\rho$; and we see that the value of a strongly affects the shear, as well as the divergence. For $a = \tfrac{1}{2}$ the congruence is shear-free, and as we should expect from the Goldberg–Sachs theorem,[17] the Riemann tensor proves to be al-

[16] We use extensively the concepts and try to use the notations of Bondi, Sachs, Newman, and Penrose in Ref. 2, wherever possible, but the notations sometimes conflict.

[17] J. Goldberg and R. Sachs, Acta Phys. Polon. **22**, 13 (1962).

gebraically special in this case (see Sec. IV). For cylindrical null hypersurfaces in Minkowski space, $|\sigma|^2 = \frac{1}{4}(\theta^2)$, as can be seen from the geometrical interpretation of the shear and divergence of a null congruence, or simply by putting $\psi = \chi = \gamma = 0$ in Eqs. (3.2) and (3.3); so we would expect these relationships to hold asymptotically for our cylindrically symmetric null hypersurfaces if the space becomes asymptotically flat fast enough at infinity. If we compute the asymptotic values of the divergence and shear for outgoing ψ and χ waves superimposed on a static solution $\psi = a \ln \rho + b$, we indeed find that the cylindrical relations between shear and divergence are verified asymptotically if and only if $a = 0$ or 1—the only values for which the space would be flat except for the radiation field. For other values of a, although the curvature tensor asymptotically vanishes as $\rho \to \infty$, the properties of the radiation field are more strongly influenced by the presence of the static field which makes many quantities fall off asymptotically more slowly as $\rho \to \infty$ than the time-dependent part of the field does. We see another example of this behavior in Sec. IV. Thus, the word "cylindrical", when used in connection with this class of metrics, implies only cylindrical symmetry; only if $a = 0$ or 1 does the more strict criterion hold asymptotically.

B. Field Equations

It will be recalled that Bondi *et al.*[2] and Sachs[2] show how the field equations break up into four groups when their components with respect to a null tetrad are taken: the hypersurface equations, standard equations, trivial equation, and supplementary conditions. The most important are the standard equations, which determine the evolution of the two degrees of freedom of the field and which lead to the definition of the news functions (we shall speak of two real instead of one complex news function); and the supplementary conditions, which lead to the result that any news decreases the mass of the system (defined as the integral of the mass aspect over the two-sphere at infinity).

Straightforward computation gives the breakup of the field equations with respect to the null tetrad given above. The hypersurface equation, standard equations, trivial equation, and supplementary condition are given in the Appendix. Let us look at the standard equations, (A8) and (A9). They are just equivalent to Eqs. (2.2) and (2.3) written in null coordinates. In the asymptotically spherical case, only the asymptotic form of these equations can be written down, and even these involve other parts of the metric besides the two functions defining the two degrees of freedom of the field, so that these degrees of freedom are only implicitly given by the standard equations. In the cylindrical case, the exact equations involve only ψ and χ, so that their evolution and interaction is explicitly isolated from the rest of the metric. (Of course, we pay for this mathematical simplicity at the expense of physical realism in the model.) It proves most useful to replace χ by $\bar{\chi} = e^{2\psi}x/2\rho$ in future work, since ψ and $\bar{\chi}$ enter more symmetrically into the standard equations; and (more importantly) they then both have the same type of asymptotic behavior at infinity, as we show. Our choice of ψ and $\bar{\chi}$ to represent the two degrees of freedom of the cylindrical gravitational field differs from that of Sachs.[2] Roughly speaking, and disregarding the nonlinearities which make these statements only asymptotically true, his choice represents a symmetric choice of the two functions about the axis $\theta = 0$ in his coordinate system, so that for axially symmetric solutions he takes his two degrees of freedom equal. Our choice represents the "sum" and "difference" of his functions, so that for the axially symmetric solutions one of our functions vanishes ($\bar{\chi} = 0$).

The standard equations in terms of ψ and $\bar{\chi}$, Eqs. (A12) and (A13), may be put into the form

$$X_{,\rho} = fY + gX + h, \tag{3.4}$$

$$Y_{,\rho} = uX + vY + w, \tag{3.5}$$

where $X = \rho^{\frac{1}{2}}\psi_{,u}$, $Y = \rho^{\frac{1}{2}}\bar{\chi}_u$, and u, v, w and f, g, h are functions of ψ, $\bar{\chi}$ and their derivatives with respect to ρ. Thus, if we are given the values of X and Y at some point $\rho = \rho_0$, and the values of ψ and $\bar{\chi}$ over the rest of the null surface $u =$ const, these equations may be integrated to give the value of X and Y over the rest of the null hypersurface. If we are given the values of ψ and $\bar{\chi}$ on some initial null hypersurface $u = u_0$, and the values of X and Y on the timelike tube $\rho = \rho_0$ (or any other timelike tube for that matter) we can then determine the evolution of ψ and $\bar{\chi}$ off the initial hypersurface $u = u_0$. This of course is just a particular example of a null hypersurface initial-value problem for a system of hyperbolic partial differential equations, and illustrates the general result that initial data on the null hypersurface never suffice to determine a solution, but must be supplemented by data off the hypersurface.[18]

[18] See article by Sachs in Ref. 1 for detailed discussion and references.

So far, our analysis would apply to any solution of the standard equations, representing any mixture of incoming or outgoing waves. If the timelike tube bearing the "news" (not contained in the hypersurface initial data) is placed at any finite value of ρ_0, any solution can be described in this way.

C. Definition of Outgoing and Incoming Radiation

In an attempt to limit the class of solutions to those characterized by outgoing radiation only, Bondi *et al.*[2] and Sachs[2] set boundary conditions on the metric analogous to the well-known Sommerfeld radiation condition; while Newman and Penrose[2] set conditions on the asymptotic behavior of the Riemann tensor. Both approaches proved essentially equivalent. In this section, we follow the Bondi–Sachs method of examining the metric, leaving discussion of the Riemann tensor for the next section. The Bondi–Sachs boundary conditions proved strong enough to enable them to define a news function at infinity. Specification of certain data on the initial null hypersurface, plus the news function on the two-sphere at infinity for a range of values of u (what we shall call the timelike tube at infinity) as well as certain data on the initial two-sphere at infinity serve to determine a unique solution to the field equations. It was soon realized that this class of solutions would have to include at least some with finite pulses of incoming radiation. Since one was basically working by analogy with linear theories, and had no independent definition of incoming or outgoing radiation in the exact theory, nor any exact solutions on which to test these concepts, it seems of interest to see how these ideas work out in our case, where we can give an independent meaning to incoming and outgoing radiation solutions, and even write down certain exact solutions.

The form of the metric Eq. (2.1) is canonical, in the sense (among others) that the only coordinate transformations of ρ and t which do not change this form are linear: $\bar{\rho} = k\rho$, $\bar{t} = kt + c$.[19] So the null hypersurfaces $u = t - \rho = \text{const.}$, $v = t + \rho = \text{const.}$ are uniquely defined. We can thus define incoming and outgoing solutions to the field equations as those solutions that are generated respectively by incoming and outgoing solutions to the equations for ψ and χ, Eqs. (2.2) and (2.3); or equivalently to those for ψ and $\bar{\chi}$, Eqs. (3.4) and (3.5). In the case where $\bar{\chi} = 0$ (the Einstein–Rosen case), this means incoming and outgoing solutions to the scalar wave equation for ψ, or any linear combination of them, generate corresponding solutions to the field equations. To give a precise meaning to the concept of incoming and outgoing waves in a nonlinear theory, we adopt the following definition.[20] A time-dependent disturbance propagating itself into a region where all the field variables are time-independent must have its wavefront on a null hypersurface. If the null hypersurface joining the time-independent and time-dependent regions is a forward null hypersurface $u = u_0$, we call the disturbance an outgoing wave. If the null hypersurface joining the two regions is a backward null hypersurface $v = v_0$, we call the disturbance an incoming wave. Notice that this definition does not restrict us to wave pulses, since the disturbance need never cease once it has started, but does exclude disturbances infinite in both time directions. For linear theories it is easy to extend the concept of outgoing (or incoming) waves to cover this case by defining any superposition of outgoing (or incoming) waves as an outgoing (or incoming) wave; but it is not so easy to see how to define outgoing and incoming waves infinite in both directions in the nonlinear case. At any rate, we do not need such waves for our purposes, and leave open the question of such a definition.[21]

On this definition of incoming and outgoing waves, it is clear that the solution $\chi = g(u)$ or $\bar{\chi} = g(u)/2\rho^{\frac{1}{2}}$ with $\psi = \frac{1}{4}\ln\rho$, is an outgoing solution if $g(u) \equiv 0$ for u less than some u_0; while $\chi = g(v)$ or $\bar{\chi} = g(v)/2\rho^{\frac{1}{2}}$ is an incoming solution if $g(v) = 0$ for v less than some v_0.

D. News Function

We define the news function for this class of metrics by giving the "news" off the null hypersurface on the timelike tube at infinity. More explicitly, we define two functions dc_1/du and dc_2/du:

$$dc_1/du = \lim_{\rho\to\infty} (\rho^{\frac{1}{2}}\psi_{,u}), \tag{3.6}$$

[19] For a discussion of allowable coordinate transformations in the Einstein–Rosen case see Ref. 12. The results for ρ and t can be trivially generalized.

[20] A. Jeffrey and T. Taniuti, *Nonlinear Wave Propagation* (Academic Press Inc., New York, 1964), define a wave as a disturbance propagating itself into a state in which all field quantities are constant in time, and then restrict themselves to propagation into states constant in space as well as time. This seems too restrictive in two ways: it restricts the concept of wave to pure incoming or outgoing radiation, and it does not allow the wave to propagate into regions of static field; so we have modified their definition as indicated in the text.

[21] One's first thought might be to define outgoing radiation in terms of the asymptotic behavior of the solution at infinity: for example, in terms of the power series expansion in powers of $1/r$ for spherical waves. But the example given in the next section shows that, at least for fixed u_0, it may be possible to expand certain incoming solutions (on the above definition) into the corresponding cylindrical power series in $1/\rho^{\frac{1}{2}n}$).

$$dc_2/du = \lim_{\rho\to\infty} (\rho^{\frac{1}{2}}\tilde{\chi}_{,u}). \tag{3.7}$$

Clearly, for these functions to exist, $\psi_{,u}$ and $\tilde{\chi}_{,u}$ must behave as $O(\rho^{\frac{1}{2}})$ as $\rho \to \infty$.[22] But this is characteristic of the behavior of outgoing cylindrical waves; in the next section we show that any outgoing wave solution to the standard equations satisfies this criterion. If we restrict ourselves to the Einstein–Rosen case for simplicity, the most general outgoing wave solution can be built up out of a superposition of Hankel functions of zero-order times exp (*ikl*) terms, which asymptotically behaves like $\sum_{n=0}^{\infty} f_n(u)/\rho^{(n+\frac{1}{2})}$.[23] Thus, it is certainly true that a news function exists for all outgoing solutions to the field equations. More interesting is the question: does a news function exist for non-outgoing solutions? If we take an incoming solution which is asymptotically of the form $f(v)/\rho^{\frac{1}{2}}$, it is clearly possible that if $f(v)$ vanishes for v greater than some finite value, then this solution may have a news function. This is not as trivial as in the case of spherical waves, because of the existence of tails on two-dimensional wave pulses,[24] but it can be shown to be possible for certain incoming pulses by the examination of the asymptotic form of the tail in such cases. Therefore, finite pulses of incoming radiation cannot be excluded by a news function requirement. More unexpectedly, we find that certain infinite incoming wave trains also have finite news functions—namely zero news—if they fall off fast enough as v approaches infinity. For example, take our solution $\tilde{\chi} = g(v)/\rho^{\frac{1}{2}}$, clearly of the expected asymptotic form. If we let $g(v) = \sum_{n=1}^{\infty} a_n/v^n$ for v greater than some fixed v_0, then $\lim_{\rho\to\infty} \tilde{\chi} = 0$ on $u_0 = \text{const}$ and $dc_2/du = 0$, so this describes a solution meeting requirements analogous to those in the asymptotically spherical case and representing an infinite incoming wave train. Of course, this solution is of no physical interest; but it does serve to show that it cannot be automatically assumed that all infinite incoming waves are eliminated by the Bondi–Sachs definition of asymptotically flat spaces. It may turn out that examples such as this are peculiar to the cylindrical case. But should there prove to be analogs in the asymptotically spherical case, they might shed some light on the problem of the physical significance of the class of time-dependent solutions with vanishing news function.[25]

[22] We use the expressions "order of" or "*O*" loosely, and assume that all asymptotic series needed can be differentiated as often as necessary. More careful mathematical treatment might be possible and useful in the cylindrical case, but we ignore these problems here.

[23] A careful discussion of the asymptotic behavior of Hankel functions of zero order is found in A. Erdelyi, *Asymptotic Expansions* (Dover Publications, Inc., New York, 1956). A useful mathematical discussion of solutions of the scalar wave equation in two dimensions is found in B. Baker and E. Copson, *The Mathematical Theory of Huyghens' Principle* (Clarendon Press, Oxford, England, 1950), 2nd ed.

[24] H. Lamb [*Hydrodynamics* (Dover Publications, Inc., New York, 1945), 6th ed., p. 299] discusses the tails of cylindrical waves.

E. Mass/Unit Length

Now we define the cylindrical analog of the mass aspect, whose integral over the two-sphere at infinity is interpreted as the total enclosed mass in the asymptotically spherical case.[2] In our case it can only be a mass per unit length that we can expect to define, since the field is independent of the z coordinate; and because of the cylindrical symmetry the mass aspect will be a function of the null variable u only, and so will coincide with the mass per unit length. It is the supplementary condition, Eq. (A11), which gives the relation between the mass/unit length and the news functions. If we define $dM(u)/du = \frac{1}{2} \lim_{\rho\to\infty} \gamma_{,u}$, then for those solutions for which the asymptotic behavior of ψ and $\tilde{\chi}$ is of the form $f(u)/\rho^{\frac{1}{2}}$ as $\rho \to \infty$, taking the limit of the supplementary condition yields

$$dM/du = -[(dc_1/du)^2 + (dc_2/du)^2], \tag{3.8}$$

so that the mass/unit length always decreases if there is any news. Looking at the hypersurface equation, Eq. (A7), we see that it can be integrated to give

$$\gamma(\rho, u) = \int_{\infty}^{\rho} F\, d\rho + \int_{u_0}^{u} (dM/du)\, du + M_0, \tag{3.9}$$

where F is a function of the values of ψ and $\tilde{\chi}$ on the initial hypersurface $u = u_0$, and M_0, is a constant of integration representing the mass per unit length on the initial hypersurface. Thus, to specify a unique solution to the field equations we need to give the values of ψ and $\tilde{\chi}$ on the initial null hypersurface, the news functions on the timelike tube at infinity, and the initial mass/unit length.

One might wonder why no functions analogous to N, defined on the two-sphere at infinity and needed for the full specification of the initial data in the spherical case, occurs here. The presence of N, which seems to be connected with the angular momentum and dipole moment of a rotating source, depends on the occurrence of cross terms in the metric between u and the coordinates of the space-

[25] R. Penrose [Proc. Roy. Soc. (London) **A284**, 159 (1965)] has also suggested that the time-dependent solutions associated with no news are connected with the presence of incoming radiation.

like two-surfaces on a null hypersurface—z and ϕ in our case. Since the metric (2.1) does not involve such cross terms, N vanishes here.

If ψ contains a static term $a \ln \rho + b$, γ will contain a term $a^2 \ln \rho + c$, so that we cannot define M as $\lim_{\rho\to\infty} \gamma$. We can remove this infinite contribution from the static cylinder by defining

$$M = \lim_{\rho\to\infty} (\gamma - a^2 \ln \rho). \tag{3.10}$$

Since it is only dM/du which is defined by the supplementary condition, this does not affect the energy balance equation.

If an initially static solution $\psi = a \ln \rho + b$, $\gamma = a^2 \ln \rho + c$ emits a pulse of radiation, it never becomes fully static again because of the well-known tail property of two-dimensional wave propagation. However, as $u \to \infty$, the solution becomes asymptotically static again, as Marder[11] has shown. He further shows that, while b and c change values during this process, a remains unchanged. Thus, the renormalization of the mass/unit length suggested here is independent of the surface on which it is carried out, and makes M a measure of what we might call the disposable gravitational mass, the amount available for radiation. Some further comments on the possible connection of this phenomenon with the recently discovered ten exactly conserved quantities in the asymptotically spherical case are found in the next section.

It is of some interest to note that this result agrees with the definition of mass/unit length of a cylindrical gravitational field arrived at by Thorne,[12] on an entirely different basis. He defines a C-energy flux vector, uses it to define the energy per unit length enclosed within a cylinder of radius ρ, and finds it to be proportional to $\gamma(\rho, t)$. As $\rho \to \infty$ this result passes over into the mass/unit length defined above, if we renormalize Thorne's definition as suggested above.

IV. ASYMPTOTIC FORM OF THE RIEMANN TENSOR

A. Petrov Type

The physical components of the Riemann tensor with respect to the orthonormal tetrad and null tetrad of Sec. III are given in the Appendix. Note that Ψ_1 and Ψ_3 are zero for this entire class of metrics. This enables us to make a simple application of the technique outlined by Janis and Newman[4] for finding the Petrov type from the Ψ's. This consists essentially in carrying out the four null rotations about n^μ which take l^μ into a principal null vector. Since Ψ_0 vanishes for a principal null vector, and since the effect of such a null rotation is to change Ψ_0 into

$$\Psi_0' = \Psi_0 + 4b\Psi_1 + 6b^2\Psi_2 + 4b^3\Psi_3 + b^4\Psi_4 = 0, \tag{4.1}$$

where b is the parameter specifying the null rotation, we need merely look at the multiplicity of the roots of Eq. (4.1) to find out about the possible coincidences of principal null vectors that characterize the Petrov type in the Penrose formulation.[26] Since $\Psi_1 = \Psi_3 = 0$ in our case, we can only have four distinct roots or two double roots, and the metrics must be of type I or D. The static solution $\psi = a \ln \rho$ is of type D; all others examined have proved to be of type I, including the linearized general solution; but no proof that other exceptional cases of type D cannot occur has been found.

B. Peeling Theorem

For an examination of the asymptotic behavior of the Riemann tensor it proves much more convenient to use the function $\tilde{\chi} = e^{2\psi}\chi/2\rho$ to characterize the second degree of freedom. While the field equations and Riemann tensor become more complicated, ψ and $\tilde{\chi}$ have a common asymptotic behavior. We note that if we linearize the standard equations in terms of ψ and $\tilde{\chi}$, they become identical to the equations for A_z and A_ϕ, respectively, in the corresponding cylindrically symmetric problem for the Maxwell field. To study the asymptotic form of the outgoing wave solutions of the standard equations, Eqs. (3.4) and (3.5), we look at what happens on the null hypersurface $u = u_0$ separating the static and wave regions. On this surface, the field in the exterior region must be that of a static cylinder $\psi = a \ln \rho$, $\tilde{\chi} = 0$. If we assume the solution to be analytic in u, successive differentiation of Eqs. (3.4) and (3.5) enables us to study the buildup of the wave into the region of disturbance $u > u_0$. We find that $(\psi_{,u})_{u=u_0}$ and $(\tilde{\chi}_{,u})_{u=u_0}$ can only be of the form $a(u)/\rho^{\frac{1}{2}}$ and $b(u)/\rho^{\frac{1}{2}}$, respectively. The second derivatives already introduce interactions between the two degrees of freedom due to the coupled nonlinear nature of the equations. We can verify, for example, that if $\tilde{\chi}$ is zero, the ψ wave propagates obeying a linear equation, each differentiation with respect to u bringing in successive powers of $[\rho^{(n+\frac{1}{2})}]^{-1}$. If $\tilde{\chi}$ is not zero, however, the interaction between the two degrees of freedom generates a ψ wave, if one was not present initially (unless $a = \frac{1}{4}$, in accord with our result that this is the only value of a for which a pure $\tilde{\chi}$ wave

[26] R. Penrose, Ann. Phys. 10, 179 (1960).

is possible), and in any case introduces terms of order $1/\rho^n$ into the structure of ψ and $\bar{\chi}$. So instead of the expansions into powers of $1/r$ that characterize the asymptotically flat solutions in the asymptotically spherical case, we find here that a power series expansion of the form $\sum_{n=0}^{\infty} a_n(u)/\rho^{\frac{1}{2}n}$ characterizes the solutions to the standard equations in the cylindrical case, for outgoing radiation. We shall therefore examine the structure of the Riemann tensor on the assumption that ψ_{rad} and $\bar{\chi}$ are of $O(1/\rho^{\frac{1}{2}})$ as $\rho \to \infty$. In this case we find

$$\Psi_4 = e^{2(\psi-\gamma)}O(1/\rho^{\frac{1}{2}}), \qquad \Psi_2 = e^{2(\psi-\gamma)}O(1/\rho^{\frac{3}{2}}),$$
$$\Psi_0 = e^{2(\psi-\gamma)}[O(1/\rho^{5/2}) + a(a-1)(2a-1)/\rho^2], \quad (4.2)$$

so that there is a "peeling theorem" in the cylindrical case, just as in the spherical one. The ratio of the parallax to the luminosity distance $e^{2(\psi-\gamma)}$ occurs as a factor in each of our terms, and will only approach unity for $a = 0$ or 1.

For static solutions, $2(\psi - \gamma)$ has its maximum value when $a = \frac{1}{2}$; for this value $e^{2(\psi-\gamma)} = \rho^{\frac{1}{2}}$. So the static part of $e^{2(\psi-\gamma)}$ is always of order less than $\rho^{\frac{1}{2}}$, except when $a = \frac{1}{2}$ (the shear-free case). Since the highest power occurring in the other factors in the Riemann tensor is of order $1/\rho^{\frac{3}{2}}$, this means that the Riemann tensor always vanishes asymptotically, except for $a = \frac{1}{2}$. At any rate, to order $e^{2(\psi-\gamma)}/\rho^{\frac{1}{2}}$, only $\Psi_4 \neq 0$, and the metric is of type N. Terms of order $e^{2(\psi-\gamma)}/\rho$ also occur in Ψ_4, so that to order $e^{2(\psi-\gamma)}/\rho$ the Riemann tensor is still of type N. To order $e^{2(\psi-\gamma)}\rho^{\frac{1}{2}}$, Ψ_2 and Ψ_4 do not vanish; there are then two distinct and one double root, and the metric is of type II. Terms of order $e^{2(\psi-\gamma}\rho^2$ occur in Ψ_2 as well; but whether they occur in Ψ_0 depends on the value of a. If $a = 0$ or 1 (flat space) or $a = \frac{1}{2}$ (shear-free case) they do not; in this case the metric is still of type II to order $e^{2(\psi-\gamma)}\rho^2$ in the general outgoing radiation case. If a has any other value (and this includes all positive static mass cases) terms of order $e^{2(\psi-\gamma)}/\rho^2$ occur in Ψ_0, and the metric is then of type I in the general outgoing radiation case.

In the purely static case ($\psi = a \ln \rho + b$), the metric is either flat or of type D. So we may write

$$e^{2(\gamma-\psi)}R_{\mu\nu\kappa\lambda} = \frac{{}_0N_{\mu\nu\kappa\lambda}}{\rho^{\frac{1}{2}}} + \frac{{}_0N'_{\mu\nu\kappa\lambda}}{\rho} + \frac{{}_0II_{\mu\nu\kappa\lambda}}{\rho^{\frac{3}{2}}} + \frac{{}_0I(II, D)_{\mu\nu\kappa\lambda}}{\rho^2} + \frac{{}_0I_{\mu\nu\kappa\lambda}}{\rho^{5/2}}, \quad (4.3)$$

where it is clear from the preceding discussion when the $1/\rho^2$ term is of type II or D.

One might ask: If we reverse our reasoning and assume that $e^{2(\psi-\gamma)}\Psi_0$ is of order $1/\rho^{5/2}$ in its radiative part, do we thereby eliminate infinite incoming radiation trains from the class of admissible matrics? Our previous incoming wave example, $\bar{\chi} = \sum_{n=1}^{\infty} a_n/v^n$ for $v > v_0$, shows that we do not. For it gives contributions of order $1/\rho^{5/2}$ to $e^{2(\gamma-\psi)}\Psi_0$, of order $1/\rho^{\frac{3}{2}}$ to $e^{2(\gamma-\psi)}\Psi_2$, and of order $1/\rho^{\frac{1}{2}}$ to $e^{2(\gamma-\psi)}\Psi_4$. This raises the question of whether the Newman–Penrose condition[2] that Ψ_0 be of $O(1/r^5)$ in the asymptotically spherical case really eliminates all infinite incoming radiation which falls off fast enough as $v \to \infty$. The remarks at the end of the last section on the Bondi–Sachs definition of asymptotic flatness apply here as well.

C. Exactly Conserved Quantity

The expression $e^{2(\gamma-\psi)}\Psi_0$ always leads off with the term $[a(a-1)(2a-1)]/\rho^2$ for any outgoing radiation solution, and for the class of incoming solutions just discussed, which we have seen meet the asymptotic conditions. This suggests that this term may be the cylindrical analogue of the ten exactly conserved quantities discovered by Newman and Penrose[27] in the asymptotically spherical case. There is a difference in the order in which these terms enter Ψ_0. In the spherical case the ten conserved quantities are associated with the term of $O(1/r^6)$ in Ψ_0, just after the leading term of $O(1/r^5)$. In the cylindrical case the $1/\rho^2$ term is the leading term in $e^{2(\gamma-\Psi)}\Psi_0$ (if it doesn't vanish). Indeed, in the cylindrical case we may define the conserved quantity as $\lim_{\rho\to\infty} \rho^2 e^{2(\psi-\gamma)}\Psi_0$ for this reason. But this may just be another instance of the strong effect of a on the geometry in the cylindrical case—it already enters the shear to $O(1/\rho)$. Marder[11] has noted that there are two constants associated with a static cylinder, as mentioned in the last section; one of them changes when the cylinder radiates, the other (our a) does not. At the time, this seemed a peculiarity of the cylindrical case. With the discovery of the ten exactly conserved gravitational quantities, this may actually be an additional analogy between the two cases.

V. CONCLUSIONS

We have seen how a news function analysis of the cylindrical radiation metrics described by Eq. (2.1) may be carried out. The class of solutions that can be included in such an analysis in this case, where we can pick an invariant set of forward and backward null hypersurfaces and thus give an independent definition of the concepts of outgoing and incom-

[27] E. T. Newman and R. Penrose, Phys. Rev. Letters **15**, 231 (1965).

ing radiative solutions, has been shown to be wider than the class of outgoing radiative solutions (by time reversal, all our comments could be applied to incoming solutions, of course). It includes mixtures of outgoing and incoming pulses; and perhaps more surprisingly, certain infinite incoming waves. Whether this can happen in the physically more important case of asymptotically spherical solutions remains an open question.

A peeling theorem for the cylindrical case has been established, allowing the distinction between three zones: a far zone of type N, an intermediate zone of type II, and a near zone of type I in general.

The mass/unit length has been defined, and shown to always decrease if there is any news. A quantity which seems to have a close analogy with the ten exactly conserved quantities in the asymptotically spherical case has been defined for the cylindrical case.

Enough qualitative similarity thus exists between the cylindrical case and the asymptotically spherical case to suggest that it is worthwhile to attempt to study certain questions in this case which are so far intractable in the spherical case.

The absence of any analog of N, the function connected with the dipole moment in the spherical initial data, suggests that it may be possible to generalize this class of solutions. A study of the analogous problem of cylindrically symmetric Maxwell fields shows ψ and $\tilde{\chi}$ to be closely analogous to A_z and A_ϕ. If the Maxwell tensor $F_{\mu\nu}$ is projected onto the null tetrad analogous to ours in flat space, three complex Φ's are obtained, Φ_0, Φ_1, Φ_2, as shown by Janis and Newman.[4] If only cylindrically symmetric A_z and A_ϕ are used, Φ_1 vanishes, just as Ψ_1 and Ψ_3 do for the metric Eq. (2.1). A difference between the two cases is that in the electromagnetic case, the vanishing of Φ_1 means that the field of a static charged cylinder cannot be represented by A_z and A_ϕ alone; while in the gravitational case no such restriction arises. However, there are additional static magnetic solutions, in the electromagnetic case, representing the field of a rotating charged cylinder and a cylindrically symmetric current in the direction of the axis of the cylinder, which require ϕ and A_ρ for their description. Such fields are presumed to have analogs in the gravitational case, which cannot be described by metrics of the type Eq. (2.1). In particular, van Stockum[28] has given an exterior stationary metric for a rotating cylinder which is not of this type. We hope to examine the question of generalizing the cylindrical radiation metrics to allow for these possibilities.

Another important question is the relation between the exterior field and the sources. While some work has been done on the relation between cylindrically symmetric radiation fields and interior solutions that might generate them,[11] much remains to be learned about just what types of source behavior are correlated with the two degrees of freedom of the radiation field—whether only explicitly time-dependent equations of state can be used to generate gravitational radiation, etc.

A study of the physical role of a in the exterior field and its relation to interior continuations of the metric, started by Marder,[11] may throw further light on the ten exactly conserved quantities in the spherical case, if the analogy suggested above is valid. We hope to look into some of these questions, too.

We have defined a class of outgoing and a class of incoming wave solutions in the cylindrical case by demanding that there be a region of time-independent fields separated from the time-dependent wave region by a forward-pointing null cone $u = u_0$, or a backward-pointing null cone $v = v_0$, respectively. Possibly these definitions are too restrictive, but they certainly represent nothing but incoming or outgoing radiation, so let us call these solutions *pure* outgoing or *pure* incoming radiation solutions, respectively. This immediately suggests the generalization of these concepts to the asymptotically spherical case, where we again demand the existence of *some* null cone (forward or backward) separating regions of stationary field from nonstationary field. We then have one invariant null cone in each solution of these two classes. Thus supertranslations,[29] and indeed all translations are ruled out by this invariance property. On the other hand, homogeneous Lorentz transformations merely take this null cone into itself. Thus, these two classes of solutions are invariant under the homogeneous Lorentz group. The absence of translations and supertranslations should make it possible to define angular momentum and multipole structure uniquely and invariantly for this type of solution. We are investigating these questions.

ACKNOWLEDGMENTS

The author is indebted to Professor E. T. Newman, Professor A. Janis, and Professor P. Havas for many useful comments. William J. Sarill made an in-

[28] W. van Stockum, Proc. Roy. Soc. (Edinburgh) **57**, 135 (1937).

[29] R. K. Sachs, Phys. Rev. **128**, 2851 (1962).

dependent calculation of the Ψ's and calculated the spin coefficients listed in the Appendix. Dr. J. Ehlers made available to me a calculation of the Riemann tensor for this metric done with Dr. M. Trümper, as well as other helpful comments[30]; thanks are hereby expressed to both.

APPENDIX

We list here some quantities that may be useful in working with the metric Eq. (2.1).

(a) Christoffel symbols in null coordinates:

$$\Gamma^u_{uu} = 2(\gamma_{,u} - \psi_{,u}) - (\gamma_{,\rho} - \psi_{,\rho}), \qquad \Gamma^\rho_{zz} = e^{4\psi-2\gamma}(\psi_{,u} - \psi_{,\rho}),$$

$$\Gamma^u_{zz} = e^{4\psi-2\gamma}\psi_{,\rho}, \qquad \Gamma^\rho_{z\phi} = e^{4\psi-2\gamma}[\tfrac{1}{2}(\chi_{,u} - \chi_{,\rho}) + \chi(\psi_{,u} - \psi_{,\rho})],$$

$$\Gamma^u_{z\phi} = e^{4\psi-2\gamma}(\tfrac{1}{2}\chi_{,\rho} + \psi_{,\rho}\chi), \qquad \Gamma^\rho_{\phi\phi} = e^{4\psi-2\gamma}[\chi(\chi_{,u} - \chi_{,\rho}) + \chi^2(\psi_{,u} - \psi_{,\rho})] + e^{-2\gamma}[(\psi_{,\rho} - \psi_{,u})\rho^2 - \rho],$$

$$\Gamma^u_{\phi\phi} = e^{4\psi-2\gamma}\chi(\chi_{,\rho} + \chi\psi_{,\rho}) + e^{-2\gamma}\rho(1 - \rho\psi_{,\rho}), \qquad \Gamma^z_{uz} = \psi_{,u} - \tfrac{1}{2}(\chi\chi_{,u}e^{4\psi}/\rho^2),$$

$$\Gamma^\rho_{uu} = (\gamma_{,\rho} - \psi_{,\rho}) - (\gamma_{,u} - \psi_{,u}), \qquad \Gamma^z_{u\phi} = \psi_{,u} + \psi_{,u}\chi + (e^{4\psi}/\rho^2)[\chi^2(\psi_{,u} - \chi_{,u}) - \chi^3\psi_{,u}].$$

$$\Gamma^\rho_{u\rho} = \gamma_{,\rho} - \psi_{,\rho}, \tag{A1}$$

(b) Spin coefficients for null tetrad $\bar{l}^\mu$, $\bar{n}^\mu$, m^μ[31]:

$$\rho = -\frac{e^{2(\psi-\gamma)}}{2\rho}, \qquad \mu = -1/4\rho,$$

$$\epsilon = -(ie^{4\psi-2\gamma}/4\rho)\chi_{,\rho}, \qquad \gamma = \tfrac{1}{2}(\gamma - \psi)_{,\rho} - (\gamma - \psi)_{,u} + (ie^{2\psi}/4\rho)\left(\frac{\chi_{,\rho}}{2} - \chi_{,u}\right),$$

$$\sigma = (e^{2(\psi-\gamma)}/2\rho)(1 - 2\rho\psi_{,\rho} - ie^{2\psi}\chi_{,\rho}), \qquad \lambda = \psi_{,u} - \frac{\psi_{,\rho}}{2} + 1/4\rho + \frac{ie^{2\psi}}{2\rho}\left(\frac{\chi_{,\rho}}{2} - \chi_{,u}\right), \tag{A2}$$

$$\alpha = \beta = \kappa = \nu = \pi = \tau = 0.$$

(c) Physical components of the Riemann tensor with respect to the orthonormal tetrad:

$$R_{0101} = e^{2(\psi-\gamma)}[(\gamma_{,tt} - \gamma_{,\rho\rho}) - (\psi_{,tt} - \psi_{,\rho\rho})],$$

$$R_{0202} = e^{2(\psi-\gamma)}[\psi_{,tt} + 2\psi^2_{,t} + \psi^2_{,\rho} - \gamma_{,t}\psi_{,t} - \gamma_{,\rho}\psi_{,\rho} - \tfrac{1}{4}\chi^2_{,t}e^{4\psi}/\rho^2],$$

$$R_{0303} = e^{2(\psi-\gamma)}\left[-\psi_{,tt} - \psi^2_{,\rho} + \gamma_{,t}\psi_t + \gamma_{,\rho}\psi_{,\rho} + \frac{\psi_{,\rho} - \gamma_{,\rho}}{\rho} + \tfrac{3}{4}\chi^2_{,t}e^{4\psi}/\rho^2\right],$$

$$R_{1212} = e^{2(\psi-\gamma)}[\psi_{,\rho\rho} + 2\psi^2_{,\rho} + \psi^2_{,t} - \gamma_{,t}\psi_{,t} - \gamma_{,\rho}\psi_{,\rho} - \tfrac{1}{4}\chi^2_{,\rho}e^{4\psi}/\rho^2],$$

$$R_{1313} = e^{2(\psi-\gamma)}\left[-\psi_{,\rho\rho} - \psi^2_{,t} + \gamma_{,t}\psi_{,t} + \gamma_{,\rho}\psi_{,\rho} - \frac{\psi_{,\rho} + \gamma_{,\rho}}{\rho} + \tfrac{3}{4}\chi^2_{,\rho}e^{4\psi}/\rho^2\right],$$

$$R_{2323} = e^{2(\psi-\gamma)}\{\psi^2_{,t} - \psi^2_{,\rho} + \psi_{,\rho}/\rho + \tfrac{1}{4}[(\chi^2_{,t} - \chi^2_{,\rho})/\rho^2]e^{4\psi}\},$$

$$R_{1023} = e^{4\psi-2\gamma}[\psi_{,\rho}\chi_{,t}/\rho - \chi_{,\rho}\psi_{,t}/\rho - \chi_{,t}/2\rho^2],$$

$$R_{1230} = e^{4\psi-2\gamma}\left[-\frac{\chi_{,\rho t}}{2\rho} - \frac{\chi_{,\rho}\psi_{,t}}{\rho} + \frac{\chi_{,\rho}\gamma_{,t}}{2\rho} - \frac{2\chi_{,t}\psi_{,\rho}}{\rho} + \frac{\chi_{,t}\gamma_{,\rho}}{2\rho} + \frac{\chi_{,t}}{2\rho^2}\right],$$

$$R_{1302} = e^{4\psi-2\gamma}\left[\frac{\chi_{,\rho t}}{2\rho} + \frac{2\chi_{,\rho}\psi_{,t}}{\rho} - \frac{\chi_{,\rho}\gamma_{,t}}{2\rho} + \frac{\chi_{,t}\psi_{,\rho}}{\rho} - \frac{\chi_{,t}\gamma_{,\rho}}{2\rho}\right],$$

$$R_{1213} = e^{4\psi-2\gamma}\left[\frac{\chi_{,\rho\rho}}{2\rho} - \frac{\chi_{,\rho}}{2\rho^2} + \frac{5}{2}\frac{\chi_{,\rho}\psi_{,\rho}}{\rho} - \frac{\chi_{,\rho}\gamma_{,\rho}}{2\rho} + \frac{\chi_{,t}\psi_{,t}}{2\rho} - \frac{\chi_{,t}\gamma_{,t}}{2\rho}\right],$$

[30] Dr. J. Ehlers (personal communication).
[31] Calculated by William J. Sarill.

$$R_{0203} = e^{4\psi-2\gamma}\left[\frac{\chi_{,tt}}{2\rho} + \frac{\chi_{,\rho}\psi_{,\rho}}{2\rho} - \frac{\chi_{,\rho}\gamma_{,\rho}}{2\rho} + \frac{5}{2}\frac{\chi_{,t}\psi_{,t}}{\rho} - \frac{\chi_{,t}\gamma_{,t}}{2\rho}\right],$$

$$R_{2021} = e^{2(\psi-\gamma)}[\psi_{,\rho t} + 3\psi_{,t}\psi_{,\rho} - \psi_{,t}\gamma_{,\rho} - \psi_{,\rho}\gamma_{,t} - \tfrac{1}{4}\chi_{,\rho}\chi_{,t}e^{4\psi}/\rho^2],$$

$$R_{3031} = e^{2(\psi-\gamma)}\left[-\psi_{,\rho t} - \psi_{,t}\psi_{,\rho} + \psi_{,t}\gamma_{,\rho} + \psi_{,\rho}\gamma_{,t} - \frac{\gamma_{,t}}{\rho} + \tfrac{3}{4}\chi_{,\rho}\chi_{,t}e^{4\psi}/\rho^2\right]. \quad \text{(A3)}$$

(d) Complex components of Riemann tensor with respect to the null tetrad[31]:

$$\Psi_0 = -C_{\mu\nu\rho\sigma}l^\mu m^\nu l^\rho m^\sigma$$

$$= e^{2(\psi-\gamma)}\Big\{[\gamma_{,\rho}/\rho + 2\gamma_{,\rho}\psi_{,\rho} - 4\psi_{,\rho}^2 - \psi_{,\rho\rho}] + \frac{ie^{2\psi}}{\rho}[\gamma_{,\rho}\chi_{,\rho} + \chi_{,\rho}/2\rho - 3\psi_{,\rho}\chi_{,\rho} - \chi_{,\rho\rho}/2]\Big\}, \quad \text{(A4)}$$

$$\Psi_2 = -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu l^\rho m^\sigma$$

$$= e^{2(\psi-\gamma)}\Big\{[\tfrac{1}{2}(\gamma-\psi)_{,\rho\rho} - (\gamma-\psi)_{,\rho u}] + (ie^{2\psi}/2\rho)\left[\psi_{,u}\chi_{,\rho} + \frac{\chi_{,u}}{2\rho} - \psi_{,\rho}\chi_{,u}\right]\Big\}, \quad \text{(A5)}$$

$$\Psi_4 = -C_{\mu\nu\rho\sigma}\bar{m}^\mu n^\nu \bar{m}^\rho n^\sigma$$

$$= e^{2(\psi-\gamma)}\Big\{\Big[\psi_{,\rho u} - \psi_{,uu} - \psi_{,\rho\rho}/4 + \left(\psi_{,\rho} - 2\psi_{,u} - \frac{1}{2\rho}\right)\left(\frac{\gamma_{,\rho}}{2} - \gamma_{,u}\right) - \tfrac{1}{2}(\psi_{,\rho} - 2\psi_{,u})^2 + (e^{4\psi}/2\rho^2)\left(\frac{\chi_{,\rho}}{2} - \chi_{,u}\right)^2\Big] + (ie^{2\psi}/2\rho)[(\tfrac{1}{2}\chi_{,\rho} - \chi_{,u}) \times (3\psi_{,\rho} - 6\psi_{,u} - \gamma_{,\rho} + 2\gamma_{,u} - 1/2\rho) - \chi_{,\rho u} + \chi_{,uu} + \chi_{,\rho\rho}/4]\Big\}. \quad \text{(A6)}$$

(e) Field equations in null coordinates:

(1) hypersurface equation

$$R_{\mu\nu}\bar{l}^\mu\bar{l}^\nu = e^{4(\psi-\gamma)}[2\psi_{,\rho}^2 - 2\gamma_{,\rho}/\rho + (e^{4\psi}/2\rho^2)\chi_{,\rho}^2]; \quad \text{(A7)}$$

(2) standard equations

$$\text{Re}\,(R_{\mu\nu}m^\mu m^\nu) = e^{2(\psi-\gamma)}[-2\psi_{,\rho u} + \psi_{,\rho\rho} + (\psi_{,\rho} - \psi_{,u})/\rho + (e^{4\psi}/2\rho^2)(2\chi_{,\rho}\chi_{,u} - \chi_{,\rho}^2)], \quad \text{(A8)}$$

$$\text{Im}\,(R_{\mu\nu}m^\mu m^\nu) = (e^{4\psi-2\gamma}/2\rho)[-2\chi_{,\rho u} + \chi_{,\rho\rho} - (\chi_{,\rho} - \chi_{,u})/\rho + 4(\chi_{,\rho}\psi_{,\rho} - \chi_{,\rho}\psi_{,u} - \chi_{,u}\psi_{,\rho})]; \quad \text{(A9)}$$

(3) trivial equation

$$R_{\mu\nu}\bar{l}^\mu\bar{n}^\nu = e^{2(\psi-\gamma)}[(2\gamma_{,\rho u} - \gamma_{,\rho\rho}) - (2\psi_{,\rho u} - \psi_{,\rho\rho}) + \psi_{,\rho}(2\psi_{,u} - \psi_{,\rho}) + (\psi_{,\rho} - \psi_{,u})/\rho + (e^{4\psi}/4\rho^2)(2\chi_{,\rho}\chi_{,u} - \chi_{,\rho}^2)]; \quad \text{(A10)}$$

(4) supplementary condition

$$R_{\mu\nu}\bar{n}^\mu\bar{n}^\nu = 2\psi_{,u}^2 + \psi_{,\rho}^2/2 - 2\psi_{,\rho}\psi_{,u} + (1/2\rho)(2\gamma_{,u} - \gamma_{,\rho}) + (e^{4\psi}/2\rho^2)(\chi_{,u}^2 + \tfrac{1}{4}\chi_{,\rho}^2 - \chi_{,\rho}\chi_{,u}). \quad \text{(A11)}$$

In order that the two degrees of freedom enter into the standard equations in such a way as to show the same asymptotic behavior as $\rho \to \infty$, it is necessary to replace χ by $\tilde{\chi} = e^{2\psi}\chi/2\rho$. The expressions for the Riemann tensor, etc., become must longer in this case, but the two degrees of freedom enter symmetrically with respect to their asymptotic behavior. We give only the standard equations in terms of ψ and $\tilde{\chi}$:

$$-(2\psi_{,\rho u} - \psi_{,\rho\rho}) + (\psi_{,\rho} - \psi_{,u})/\rho + 4(\tilde{\chi}_{,u} - 2\psi_{,u}\tilde{\chi})(\tilde{\chi}_{,\rho} - 2\psi_{,\rho}\tilde{\chi} + \tilde{\chi}/\rho) - 2(\tilde{\chi}_{,\rho} - 2\psi_{,\rho}\tilde{\chi} + \tilde{\chi}/\rho)^2 = 0; \quad \text{(A12)}$$

$$-(2\tilde{\chi}_{,\rho u} - \tilde{\chi}_{,\rho\rho}) + (\tilde{\chi}_{,\rho} - \tilde{\chi}_{,u})/\rho - \tilde{\chi}/\rho^2 + 2\tilde{\chi}[(2\psi_{,\rho u} - \psi_{,\rho\rho}) + (\psi_{,\rho} - \psi_{,u})/\rho - 2\psi_{,\rho}(2\psi_{,u} + \psi_{,\rho})] = 0. \quad \text{(A13)}$$

Using Eq. (A12) to eliminate $(2\psi_{,u\rho} - \psi_{,\rho\rho})$ from Eq. (A13), we get Eqs. (3.4) and (3.5) of Sec. III, with

$$f = 2\tilde{\chi}_{,\rho} - 4\tilde{\chi}\psi_{,\rho} + 2\tilde{\chi}/\rho,$$

$$g = -4\tilde{\chi}_{,\rho}\tilde{\chi} + 8\psi_{,\rho}\tilde{\chi}^2 - 4\tilde{\chi}^2/\rho,$$

$$h = (\rho^{\frac{1}{2}}/2)\psi_{,\rho\rho} + \psi_{,\rho}/2\rho^{\frac{1}{2}} + 4\rho^{\frac{1}{2}}\tilde{\chi}_{,\rho}\tilde{\chi}\psi_{,\rho} - 4\rho^{\frac{1}{2}}\psi_{,\rho}^2\tilde{\chi}^2 - \rho^{\frac{1}{2}}\tilde{\chi}_{,\rho}^2 - 2\tilde{\chi}_{,\rho}\tilde{\chi}/\rho^{\frac{1}{2}} + 4\psi_{,\rho}\tilde{\chi}^2/\rho^{\frac{1}{2}} - \tilde{\chi}^2/\rho^{\frac{3}{2}},$$

$$u = -4\psi_{,\rho}\tilde{\chi} - 2\tilde{\chi}/\rho - 8\tilde{\chi}^2\tilde{\chi}_{,\rho} + 16\tilde{\chi}^3\psi_{,\rho} - 8\tilde{\chi}^3/\rho,$$

$$v = 4\tilde{\chi}\tilde{\chi}_{,\rho} - 8\tilde{\chi}^2\psi_{,\rho} + 4\tilde{\chi}^2/\rho,$$

$$w = \rho^{\frac{1}{2}}\tilde{\chi}_{,\rho\rho}/2 + \tilde{\chi}_{,\rho}/2\rho^{\frac{1}{2}} - \tilde{\chi}/2\rho^{\frac{3}{2}} + 2\psi_{,\rho}\tilde{\chi}/\rho^{\frac{1}{2}} - 2\rho^{\frac{1}{2}}\psi_{,\rho}^2\tilde{\chi} + 8\rho^{\frac{1}{2}}\tilde{\chi}_{,\rho}\tilde{\chi}^2\psi_{,\rho} - 8\rho^{\frac{1}{2}}\psi_{,\rho}^2\tilde{\chi}^3 - 2\rho^{\frac{1}{2}}\tilde{\chi}_{,\rho}^2\tilde{\chi} - 4\tilde{\chi}_{,\rho}\tilde{\chi}^2/\rho^{\frac{1}{2}} + 8\psi_{,\rho}\tilde{\chi}^3/\rho^{\frac{1}{2}} - 2\tilde{\chi}^3/\rho^{\frac{3}{2}}. \quad \text{(A14)}$$

On the Motion of Electrons Scattered from an Infinite Slab*

C. D. TAYLOR
Sandia Corporation, Albuquerque, New Mexico
(Received 20 August 1965)

A nonlinear integral equation of the Volterra type is derived to describe the motion of the electrons egressing from an infinite slab—one formed by parallel planes of infinite area. This motion depends on (1) the flux density of electrons egressing from the slab, and (2) the distribution of the forward components of the electrons' velocities upon egressing from the slab, i.e., the velocity components along the normal to the slab. It is proved that the nonlinear integral equation can be solved by a method of successive approximations and that the solution is unique. The integral equation is solved for two special cases.

The most important application of this work is in the study of electrons scattered by the high-density gamma flux from a nuclear explosion. Therefore, the treatment presented is made directly applicable to that study.

INTRODUCTION

WHEN the high-density gamma flux from a nuclear explosion interacts with an object, a copious supply of electrons recoils from it. Not only does this result in a large static electric field buildup but also a large rf field may be generated.[1,2] Both of these phenomena depend upon the equation of motion of the recoil electrons. But the determination of this motion requires treatment of a complicated self-consistent field problem.

For convenience the electrons are considered to recoil from an infinite slab (one having parallel surfaces of infinite area) which is irradiated uniformly over one surface by unidirectionally propagating gamma rays. Also the electrons are considered to emerge into a vacuum or a medium where energy losses due to collisions are negligible. The flux density of the egressing electrons along with their velocity distribution is assumed to be known.[3] Based on this knowledge a nonlinear integral equation of the Volterra type is derived. It is proved rigorously that the integral equation can be solved by a method of successive approximations and that the solution is unique.

The motion of the egressing electrons is obtained for two special cases: (1) the flux density of the egressing electrons is considered to be approximately that which is scattered out of the slab by the flux of gamma rays from a nuclear explosion, and all the electrons are considered to emerge from the slab with the same velocity (the average velocity of the electrons scattered from the slab by gamma rays from a nuclear explosion); and (2) all the electrons are considered to egress from the slab at once with a velocity distribution which crudely approximates that of the electrons which are scattered from the slab by the gamma rays from a nuclear explosion. After obtaining the electron motions in each case the resulting electric fields are obtained.

ANALYSIS

Equation of Motion

The electric field and the equations of motion of the electrons scattered from an object depend upon the geometrical configuration of the object. Since the object is considered to be an infinite slab irradiated uniformly over one surface by unidirectionally propagating gamma rays, the flux of electrons recoiling from the slab is uniform over planes parallel to the surface of the slab (see Fig. 1). Consider the temporal variation of the incident gamma-flux density to be given by $T(\tau)$, i.e., $T(\tau)d\tau$ is the number of gamma photons per unit area striking the slab surface at times between τ and $\tau + d\tau$. The flux density of the electrons that are scattered from the

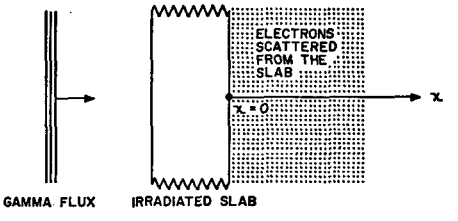

FIG. 1. The irradiated slab with the incident gamma flux and the electrons that are knocked out of the slab.

* Portions of this work are abstracted from a doctoral dissertation by C. D. Taylor, completed under the direction of Dr. R. H. Duncan at New Mexico State University, University Park, New Mexico.

This work was supported by the U. S. Atomic Energy Commission.

[1] W. J. Karzas and R. Latter, Phys. Rev. 126, 1919 (1962).

[2] C. D. Taylor, "Some Electromagnetic Effects of High Density Gamma Flux Interacting with Matter," SC-R-65-933, Sandia Corporation, Albuquerque, New Mexico (1965).

[3] These have been obtained in approximate form in Ref. 2.

slab is $T(\tau)\int_0^1 G(u)\,du$, where $T(\tau)G(u)\,du$ is defined as the number of electrons per unit area egressing from the slab with forward velocity components[4] in the range u and $u + du$, and u is the velocity component in units of the speed of light.

The number of electrons scattered from the slab per unit area with forward velocity components between u and $u + du$ and at times between τ and $\tau + d\tau$ is given by $G(u)T(\tau)\,du d\tau$. The space charge density $d^2\rho(x, t)$ with only these electrons present is

$$d^2\rho(x, t) = eG(u)T(\tau)\,du\,d\tau\,\{\delta(x) - \delta(x - \xi)\}, \quad (1)$$

where the first Dirac delta gives the residual charge density on the slab surface at $x = 0$; the second Dirac delta is the contribution made by the electrons that were scattered from the slab; and $x = \xi(u, \tau; t)$ is the position at time t of the electron that is scattered from the slab with initial velocity component along the normal to the slab u at time $t = \tau$ (see the figure). The integration of (1) yields

$$\rho(x, t) = e\int_0^t d\tau \int_0^1 du\, T(\tau)G(u) \times \{\delta(x) - \delta[x - \xi(u, \tau; t)]\}. \quad (2)$$

The electric field produced by this charge density is obtained by the use of Maxwell's equation

$$\text{div } \mathbf{D} = \rho(x', t). \quad (3)$$

Because of symmetry there is only one component of the electric field. Integrating (3) over x' from 0 to x and considering the field to be zero inside the slab—requiring it to be a good conductor—yields

$$E(x, t) = \frac{e}{\epsilon}\int_0^t d\tau\, T(\tau)\int_0^1 du\, G(u)\eta[\xi(u, \tau; t) - x], \quad (4)$$

where

$$\eta(z) = 0, \quad z < 0,$$
$$= 1, \quad z > 0.$$

The electric field is directed normal to and away from the slab.

For the electron energies of interest, relativistic corrections amount to only a few percent in the motion of the electrons; therefore, the formulation presented here will be nonrelativistic for the sake of simplicity. Starting with force equation of the electron and integrating it twice over time, the following equation is obtained:

$$\xi(u, \tau; t) = cu(t - \tau) - \frac{e}{m}\int_\tau^t dt'\,(t - t')E(\xi', t'). \quad (5)$$

Equation (4) can be combined with (5) to obtain the nonlinear integral equation of motion

$$\xi(u, \tau; t) = \xi_0 - A\int_\tau^t dt'\, K[\xi(u, \tau; t'), t'], \quad (6)$$

where

$$A = e^2/m\epsilon, \qquad \xi_0 = cu(t - \tau),$$

and

$$K[\xi(u, \tau; t'), t'] = (t - t')\int_0^{t'} d\tau' \times \int_0^1 du'\, G(u')T(\tau')\eta[\xi(u', \tau'; t') - \xi(u, \tau; t')], \quad (7)$$

which is a nonlinear integral equation of the Volterra type.

Existence of Solution

In many cases it is impossible to obtain a closed-form solution to a nonlinear integral equation. And, even though the nonlinear equation represents a physical process, it may not possess a unique solution.[5] Evidently it is essential to show that the solution of (6) is unique.

It is readily seen that $\xi_0 = cu(t - \tau)$ is integrable and bounded in the interval $\tau \leq t \leq b$ and that ξ_0 satisfies the Lipschitz condition in the interval $\tau \leq t \leq b$, where b is some finite positive number greater than τ. The kernel, $K[\xi(u, \tau; t'), t']$, is proportional to the number of electrons further from the slab than $x = \xi(u, \tau; t')$ and thus is continuous and bounded in the domain $t \geq \tau$, $t' \leq b$, and $|\xi| < c_1$, where c_1 is some finite positive number. Suppose that $c_2K[\xi(u, \tau; t'), t']$ gives the total number of electrons further from the slab than $\xi(u, \tau; t')$. Then

$$|K[\xi, t'] - K(\xi', t')| < M\,|\xi - \xi'|/c_2, \quad (8)$$

where $M = 1 +$ the maximum density of electrons occurring outside the slab. From (8) it is now established that the kernel satisfies the Lipschitz condition. Therefore, according to the theorem by Davis[6] on nonlinear integral equations of the Volterra type, (6) possesses a unique solution that may be obtained by the method of Picard, a method of successive

[4] The forward component of an electron's velocity is the component along the normal to the slab.

[5] E. E. O'Donnell, unpublished Notes, (January 1964).

[6] H. T. Davis, *Introduction to Nonlinear Differential and Integral Equations* (Dover Publications, New York, 1962), p. 415.

approximations which takes ξ_0 as the zeroth approximation.

Solutions of the Equation of Motion

The solution (6) may become a tedious process depending upon the functional forms of $G(u)$ and $T(\tau)$. Two special cases are solved illustrating the method of solution. First consider

$$G_1(u) = G_0\delta(u - u_0), \tag{9}$$

$$T_1(\tau) = \begin{cases} \alpha e^{\alpha\tau} & \tau \leq \tau_0, \\ 0, & \tau > \tau_0. \end{cases} \tag{10}$$

This time dependence roughly corresponds to the temporal variation of the production of gamma rays by a nuclear explosion. The distribution of forward-velocity components, $G_1(u)$, requires that every electron egress from the slab with the same forward-velocity component. The first-order approximation to the equation of motion is

$$\xi_1(u, \tau; t) = \begin{cases} cu_0(t-\tau) - \frac{1}{2}AG_0e^{\alpha\tau}(t-\tau)^2, & \tau \leq \tau_0, \\ cu_0(t-\tau) - \frac{1}{2}AG_0e^{\alpha\tau_0}(t-\tau)^2, & \tau > \tau_0. \end{cases} \tag{11}$$

The second-order approximation for all t cannot be obtained without numerical computations. However, it can be easily demonstrated that the first-order approximation is the exact solution for times $t \leq t_0$, where

$$t_0 = (1/\alpha) \ln (2cu_0\alpha/AG_0) \tag{12}$$

is the emission time of the electrons that are attracted back to the slab first. The electric field produced by the scattered electrons for $t \leq t_0$ is

$$E(x, t) = (e/\epsilon)G_0\eta(cu_0t - x)\{\exp \tau_1(x, t) - 1\}, \tag{13}$$

where τ_1 is the τ root of the function $\xi_1(u_0, \tau, t) - x$. If there is more than one root, τ_1 is taken as the root that is less than t_0.

Consider a second special case:

$$T_2(\tau) = T_0\delta(\tau), \tag{14}$$

$$G_2(u) = \begin{cases} 2G_0u/u_m^2, & 0 \leq u \leq u_m, \\ 0, & u > u_m, \end{cases} \tag{15}$$

where $T_0 = e^{\alpha\tau_0} - 1$ and u_m is the maximum occurring forward velocity component. This time dependence of the incident flux is a rough approximation to the temporal production of gamma rays by a nuclear explosion, and the distribution of forward components of the egressing electron velocities is a crude approximation to the actual case. The first-order approximation to the solution turns out to be the exact solution,

$$\xi(u, t) = \text{cut} - \tfrac{1}{2}AT_0G_0(1 - u^2/u_m^2)t^2. \tag{16}$$

The electric field produced by the scattered electrons is

$$E(x, t) = +(e/\epsilon)T_0G_0[1 - (u_1/u_m)^2]\eta(u_m - u_1), \tag{17}$$

where

$$u_1(x, t) = \{-cu_m^2 + [c^2u_m^4 + AG_0T_0u_m^2(2x + AG_0T_0t^2)]^{\frac{1}{2}}\}/AG_0T_0t.$$

CONCLUDING REMARKS

It has been possible to treat rigorously the problem of obtaining the motion of electrons egressing from an infinite slab. A knowledge of this motion allows the determination of the resulting electromagnetic field. This is of interest particularly in the study of the effects produced by the high-density gamma flux from a nuclear explosion.

Both special cases treated have definite merits in predicting the behavior of the electrons that are scattered from a slab by the gamma flux from a nuclear explosion. The first case represents the collective behavior of the electrons and is more important in determining the coherent rf radiation that may be produced. This is the case treated by Karzas and Latter.[1] The second case is more easily treated analytically. By following the development presented, it is possible to treat the general case as accurately as desired.

ACKNOWLEDGMENTS

The author is most grateful to his research advisor, Dr. R. H. Duncan, who proffered many useful suggestions. Also, he thanks Dr. E. E. O'Donnell for valuable counseling.

A Note on the Quantum Theory of Electrical Conductivity*

J. A. TJON†
Department of Physics, University of California, San Diego, La Jolla, California
(Received 3 January 1966)

A formal method is given for calculating the electrical conductivity tensor in ascending powers of the interaction for a system of electrons interacting with a random distribution of scattering centers.

IN the past few years there has been considerable interest in the calculation of transport coefficients for many particle systems. In particular, the electrical conductivity tensor for a system of electrons interacting with a random distribution of scattering centers has been studied in detail, and explicit calculation has been made up to zeroth order in the interaction.[1,2] In the present note we should like to show how these results can be obtained in a systematic way utilizing a method which has been developed previously in a somewhat different context.[3] It may also serve as another example where this technique can be applied successfully.

Our starting point is the Kubo formula in the one-electron approximation (we refer for notation to Ref. 2).

$$\sigma_{\mu\nu} = -\lim_{p\to+0} \mathrm{Tr}_e \left\{ \frac{\partial f}{\partial H} \tfrac{1}{2}[\hat{j}_\mu(p) j_\nu + j_\mu \hat{j}_\nu(p)] \right\} \tag{1}$$

with

$$\hat{j}_\mu(p) = \int_0^\infty dt\, e^{-pt} e^{i(H_0+\lambda V)t} j_\mu e^{-i(H_0+\lambda V)t}. \tag{2}$$

In order to compute Eq. (1) in ascending powers of λ we have to make a perturbation expansion of the operator $\hat{j}_\mu(p)$. We thereby use the momentum representation $|\mathbf{k}\rangle$ in which the operators H_0 and j_μ are diagonal. The technique described in Ref. 3 can now be applied to calculate the diagonal matrix elements of the operator $\hat{j}_\mu(p)$. As is well known, the ordinary perturbation expansion for these elements suffers from the difficulty that each term in the expansion diverges in the limit of $p \to +0$. What the method of Ref. 3 essentially does is to remove these divergences by a certain resummation procedure. The result is given formally by

$$\{\hat{j}_\mu(p)\}_d = \left[p + \sum_{n=2}^{\infty} \lambda^n \hat{L}_d^{(n)}(p) \right]^{-1} j_\mu, \tag{3}$$

where $\hat{L}_d^{(n)}(p)$ are linear operators defined in the Hilbert space $\mathfrak{L}$ of the linear operators which are diagonal in the representation $|\mathbf{k}\rangle$. They are given by

$$\hat{L}_d^{(n)}(p)a = (-i)^n \int_0^\infty d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-2}} d\tau_{n-1} e^{-p\tau_1} \times [a; V(\tau_1), V(\tau_2), \cdots, V(\tau_{n-1}), V]_{n;sd} \tag{4}$$

with

$$V(\tau) = e^{iH_0\tau} V e^{-iH_0\tau},$$

and where $a \in \mathfrak{L}$. In Eq. (4) $[a; \cdots]_{n;sd}$ designates the simple diagonal part of the nth commutator which is defined according to

$$[a; V, \cdots, V]_{n;sd} = [\cdots [a, V]_{nd}, V]_{nd}, \cdots V]_{nd}, V]_d,$$

i.e., of each commutator we have to retain only its nondiagonal contribution, except the last one, of which we should keep the diagonal part. It should be mentioned that, in general, the simple diagonal part is not the same as the so-called irreducible part defined by Van Hove,[4] but it also contains reducible contributions. Furthermore, we note that the series in Eq. (3) starts with $n = 2$ because we have assumed that $\langle \mathbf{k}| V |\mathbf{k}\rangle = 0$. We now make use of the well-known expansion for two noncommuting operators A and B,

$$[A + B]^{-1} = A^{-1} - A^{-1}BA^{-1} + \cdots.$$

With

$$A \equiv p + \lambda^2 \hat{L}_d^{(2)}(p) \quad \text{and} \quad B \equiv \sum_{n=3}^{\infty} \lambda^n \hat{L}_d^{(n)}(p)$$

we obtain

$$\{\hat{j}_\mu(p)\}_d = [p + \lambda^2 \hat{L}_d^{(2)}(p)]^{-1} j_\mu - [p + \lambda^2 \hat{L}_d^{(2)}(p)]^{-1} \times \left[\sum_{n=3}^{\infty} \lambda^n \hat{L}_d^{(n)}(p) \right] [p + \lambda^2 \hat{L}_d^{(2)}(p)]^{-1} j_\mu + \cdots. \tag{5}$$

* This work was supported in part by the U. S. Atomic Energy Commission.
† On leave from the University of Nymegen, The Netherlands.
[1] G. V. Chester and A. Tellung, Proc. Phys. Soc. (London) **73**, 745 (1959); E. Verboven, Physica **26**, 1091 (1960); A. Janner, Helv. Phys. Acta **36**, 857 (1963).
[2] P. Berger, J. M. J. Van Leeuwen, and E. Verboven, Physica **29**, 1409 (1963).
[3] J. A. Tjon, Phys. Rev. **143**, 259 (1966).
[4] L. Van Hove, Physica **23**, 441 (1957).

The reason for using this expansion is that for $p \to +0$ the quantity $[p + \lambda^2 \hat{L}_d^{(2)}(p)]^{-1} a$ exists, and it is of the order λ^{-2}. In particular, for the case of spherical symmetric scattering centers one has the simple result

$$\lim_{p \to +0} \langle \mathbf{k} | [p + \lambda^2 \hat{L}_d^{(2)}(p)]^{-1} a | \mathbf{k} \rangle = \lambda^{-2} \tau(E) a(\mathbf{k}). \tag{6}$$

In a similar way the nondiagonal part of $\hat{j}_\mu(p)$ can be treated. As a result we find

$$\{\hat{j}_\mu(p)\}_{nd} = \sum_{n=1}^{\infty} \lambda^n \hat{L}_{nd}^{(n)}(p) \{\hat{j}_\mu(p)\}_d \tag{7}$$

with

$$\hat{L}_{nd}^{(n)}(p) a = (-i)^n \int_0^\infty d\tau_1 \int_0^{\tau_1} d\tau_2 \cdots \int_0^{\tau_{n-1}} d\tau_n$$

$$e^{-p\tau_1} \times [\cdots [a, V(\tau_1)]_{nd}, V(\tau_2)]_{nd} \cdots, V(\tau_n)]_{nd}.$$

Finally, we also need the perturbation expansion of the operator $\partial f / \partial H$ which, in fact, has been given by Chester and Tellung.[1] It is of the following form:

$$\frac{\partial f}{\partial H} = \sum_{n=0}^{\infty} \lambda^{2n} f_d^{(2n)} + \sum_{n=0}^{\infty} \lambda^{2n+1} f_{nd}^{(2n+1)}, \tag{8}$$

where $f_d^{(2n)}$ and $f_{nd}^{(2n+1)}$ are diagonal and nondiagonal, respectively, in the representation $|\mathbf{k}\rangle$. The series expansion in λ of the electrical conductivity tensor

$$\sigma_{\mu\nu} = \sum_{n=-2}^{\infty} \lambda^n \sigma_{\mu\nu}^{(n)} \tag{9}$$

can be obtained on substituting Eqs. (5), (7), and (8) into Eq. (1) and collecting all the terms with the same power in λ. For the first three terms of Eq. (9), we find

$$\sigma_{\mu\nu}^{(-2)} = -\tfrac{1}{2}[(j_\nu f_d^{(0)}, \hat{L}_d^{(2)-1} j_\mu) + (f_d^{(0)} j_\mu, \hat{L}_d^{(2)-1} j_\nu)], \tag{10a}$$

$$\sigma_{\mu\nu}^{(-1)} = \tfrac{1}{2}[(j_\nu f_d^{(0)}, \hat{L}_d^{(2)-1} \hat{L}_d^{(3)} \hat{L}_d^{(2)-1} j_\mu) + (f_d^{(0)} j_\mu, \hat{L}_d^{(2)-1} \hat{L}_d^{(3)} \hat{L}_d^{(2)-1} j_\nu)], \tag{10b}$$

$$\begin{aligned}\sigma_{\mu\nu}^{(0)} = {} & -\tfrac{1}{2}[(j_\nu f_{nd}^{(1)}, \hat{L}_{nd}^{(1)} \hat{L}_d^{(2)-1} j_\nu) + (f_{nd}^{(1)} j_\mu, \hat{L}_{nd}^{(1)} \hat{L}_d^{(2)-1} j_\nu)] \\ & - \tfrac{1}{2}[(j_\nu f_d^{(0)}, \{-\hat{L}_d^{(2)-1} \hat{L}_d^{(4)} \hat{L}_d^{(2)-1} \\ & + \hat{L}_d^{(2)-1} \hat{L}_d^{(3)} \hat{L}_d^{(2)-1} \hat{L}_d^{(3)} \hat{L}_d^{(2)-1}\} j_\mu) \\ & + (j_\nu f_d^{(2)}, \hat{L}_d^{(2)-1} j_\mu) \\ & + \text{similar terms with } \mu \rightleftarrows \nu].\end{aligned} \tag{10c}$$

We have used here the notation $(a, b) \equiv \mathrm{Tr}.\, ab$. Furthermore, the variable p in Eqs. (10a)–(10c) should be taken to be $+0$. It can now easily be shown that Eqs. (10a)–(10c) are equivalent to the final results of Ref. 2 by writing these expressions out in the representation $|\mathbf{k}\rangle$.

A Closed Form for Elsasser Integrals*

SEYMOUR J. WOLFSON†
Wayne State University, Detroit, Michigan
(Received 22 December 1965)

An Elsasser integral is defined as the definite integral of a particular product of three spherical harmonics, namely,

$$\int_0^{2\pi}\int_0^{\pi} Y_{n_1}^{m_1*}\left[\frac{\partial Y_{n_2}^{m_2}}{\partial\theta}\frac{\partial Y_{n_3}^{m_3}}{\partial\varphi}-\frac{\partial Y_{n_2}^{m_2}}{\partial\varphi}\frac{\partial Y_{n_3}^{m_3}}{\partial\theta}\right]d\theta\,d\varphi.$$

This integral occurs in connection with the "dynamo theory" of the magnetic field of the earth and thus it is desirable to have a value of the integral involving only the indices of the spherical harmonics. Such a closed form is developed and some properties of the integral are discussed.

1. INTRODUCTION

THE type of integral herein discussed is one involving the product of three spherical harmonics and their derivatives in the form

$$L_{n_1n_2n_3}^{m_1m_2m_3}=\int_0^{2\pi}\int_0^{\pi} Y_{n_1}^{m_1*}\times\left\{\frac{\partial Y_{n_2}^{m_2}}{\partial\theta}\frac{\partial Y_{n_3}^{m_3}}{\partial\varphi}-\frac{\partial Y_{n_2}^{m_2}}{\partial\varphi}\frac{\partial Y_{n_3}^{m_3}}{\partial\theta}\right\}d\theta\,d\varphi, \quad (1)$$

where

$$Y_n^m(\theta,\varphi)=P_n^m(\theta)\,e^{im\varphi}/(2\pi)^{\frac{1}{2}} \quad (2)$$

and $P_n^m(\theta)$ are normalized associated Legendre polynomials.[1]

It is desired to evaluate (1) in a closed form depending only on some function of the indices.

Such integrals have thus far occurred in physics in connection with the dynamo theory of the magnetic field of the earth.[2-4] Since Elsasser[2] was the first to make use of such integrals, the author often refers to (1) as an "Elsasser" integral. The evaluation of these integrals has been considered by several authors[2,5,6] and tables of values are available for some n and m,[5] but no closed form solution has been given.

Infeld and Hull[6] outline a method whereby one can obtain a closed form for (1), but when their method is followed it appears that it does not give the correct result. The difficulty seems to lie in the result obtained by using Eq. (9.3.4) of their article in connection with (1) above. All of the steps up to that point are valid for (1), however, the result (9.3.4) requires that the sum of the upper indices be equal to n_1 which is not the case for the Elsasser integrals at that point of the derivation.

The method described below reduces the Elsasser integral to a sum of integrals for which a closed form solution is known.

2. CLOSED FORM OF THE INTEGRAL

Substituting (2) in (1), the φ integration is quite easily done and leads to the result

$$L_{n_1n_2n_3}^{m_1m_2m_3}=\frac{i}{(2\pi)^{\frac{1}{2}}}\int_0^{\pi} P_{n_1}^{m_1}\times\left\{m_3P_{n_3}^{m_3}\frac{dP_{n_2}^{m_2}}{d\theta}-m_2P_{n_2}^{m_2}\frac{dP_{n_3}^{m_3}}{d\theta}\right\}d\theta \quad (3)$$

with the restriction that $m_1=m_2+m_3$; otherwise, the integral vanishes. Using the recurrence relation

$$\{m\cot\theta-d/d\theta\}P_n^m=[(n-m)(n+m+1)]^{\frac{1}{2}}P_n^{m+1}$$

to replace the derivatives in (3) leads to the result

$$L_{n_1n_2n_3}^{m_1m_2m_3}=\frac{i}{(2\pi)^{\frac{1}{2}}}\times\int_0^{\pi} P_{n_1}^{m_1}\{m_2[(n_3-m_3)(n_3+m_3+1)]^{\frac{1}{2}}P_{n_3}^{m_3}P_{n_2}^{m_2+1}-m_3[(n_2-m_2)(n_2+m_2+1)]^{\frac{1}{2}}P_{n_2}^{m_2}P_{n_3}^{m_3+1}\}\,d\theta. \quad (4)$$

* This work was sponsored by the Air Force Cambridge Research Laboratories Office of Aero Space Research under contract No. AF 19(628)-2497 and is part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Wayne State University (1965).

† Present address: Roland F. Beers, Inc., Alexandria, Virginia.

[1] For their definition see, E. Merzbacher, *Quantum Mechanics* (John Wiley & Sons, Inc., New York, 1961) p. 180.

[2] W. M. Elsasser, Phys. Rev. **69**, 106 (1946).

[3] E. C. Bullard and H. Gellman, Phil. Trans. **A247**, 213 (1954).

[4] S. J. Wolfson, unpublished Ph. D. dissertation, Wayne State University (1965).

[5] J. R. Bird, unpublished M. Sc. thesis, The University of Toronto (1949).

[6] L. Infeld and T. E. Hull, Rev. Mod. Phys. **23**, 21 (1951).

Since the first term in (4) can be obtained from the second by interchanging n_2 with n_3 and m_2 with m_3, one need only consider the integral

$$\int_0^\pi P_{n_1}^{m_1} P_{n_2}^{m_2+1} P_{n_3}^{m_3}\, d\theta. \tag{5}$$

Making use of the recurrence relation

$$\sin\theta\, P_{n-1}^m = \left[\frac{(n+m)(n+m+1)}{(2n-1)(2n+1)}\right]^{\frac{1}{2}} P_n^{m+1} - \left[\frac{(n-m-2)(n-m-1)}{(2n-1)(2n-3)}\right]^{\frac{1}{2}} P_{n-2}^{m+1}$$

we can, by repeated application, show that

$$P_n^{m+1} = \left[(2n+1)\frac{(n-m-1)!}{(n+m+1)!}\right]^{\frac{1}{2}} \sin\theta \times \sum_{j=0}\left[(2n-1-4j)\frac{(n+m-1-2j)!}{(n-m-1-2j)!}\right]^{\frac{1}{2}} P_{n-1-2j}^m, \tag{6}$$

where the summation over the index j continues up to the largest integer J such that

$$J \leq (n-m-1)/2$$

and we use the convention that $0! = 1$.

Thus, using (6) to replace $P_{n_3}^{m_3+1}$ in (5), we have

$$\int_0^\pi P_{n_1}^{m_1} P_{n_2}^{m_2+1} P_{n_3}^{m_3}\, d\theta = \left[(2n_2+1)\frac{(n_2-m_2-1)!}{(n_2+m_2+1)!}\right]^{\frac{1}{2}} \times \sum_{j=0}\left[(2n_2-1-4j)\frac{(n_2+m_2-1-2j)!}{(n_2-m_2-1-2j)!}\right]^{\frac{1}{2}} \times \int_0^\pi P_{n_1}^{m_1} P_{n_2}^{m_2} P_{n_3-1-2j}^{m_3} \sin\theta\, d\theta. \tag{7}$$

The integrals occurring in the summation of (7) are of the type considered by Gaunt[7] (with differently normalized functions) and others.[6]

If we define

$$K_{n_1 n_2 n_3}^{m_1 m_2 m_3} = \int_0^{2\pi}\int_0^\pi Y_{n_1}^{m_1 *} Y_{n_2}^{m_2} Y_{n_3}^{m_3} \sin\theta\, d\theta\, d\varphi, \tag{8}$$

then the integrals in (7) can be written as

$$\int_0^\pi P_{n_1}^{m_1} P_{n_2}^{m_2} P_{n_3-1-2j}^{m_3} \sin\theta\, d\theta = (2\pi)^{\frac{1}{2}} K_{n_1,n_2,n_3-1-2j}^{m_1 m_2 m_3}, \tag{9}$$

where the K's [Eq. (8)] are known in closed form.

If we introduce (9) into (7), we obtain a closed form for (5). We then interchange the indices n_2 with n_3 and m_2 with m_3 to obtain the first term in (4). Combining these results in (4) leads to the formula

$$L_{n_1 n_2 n_3}^{m_1 m_2 m_3} = i\Bigg\{m_2\left[(2n_3+1)\frac{(n_3-m_3)!}{(n_3+m_3)!}\right]^{\frac{1}{2}} \times \sum_{j=0}^{J_3}\left[(2n_3-1-4j)\frac{(n_3+m_3-1-2j)!}{(n_3-m_3-1-2j)!}\right]^{\frac{1}{2}} \times K_{n_1,n_2,n_3-1-2j}^{m_1,m_2,m_3} - m_3\left[(2n_2+1)\frac{(n_2-m_2)!}{(n_2+m_2)!}\right]^{\frac{1}{2}} \times \sum_{j=0}^{J_2}\left[(2n_2-1-4j)\frac{(n_2+m_2-1-2j)!}{(n_2-m_2-1-2j)!}\right]^{\frac{1}{2}} \times K_{n_1,n_2-1-2j,n_3}^{m_1,m_2,m_3}\Bigg\}, \tag{10}$$

where J_i is the largest integer such that $J_i \leq (n_i - m_i - 1)/2$ and the summation over j is to be taken only if consistent with the factorial notation (i.e., terms containing negative factorials are to be omitted).

If the result containing only the indices is desired, one can substitute in (10) the following formula[8] for the K's:

$$K_{n_1 n_2 n_3}^{m_1 m_2 m_3} = \frac{1}{(2\pi)^{\frac{1}{2}}}\frac{(n_3+n_2-n_1-1)!![(2n_1+1)(2n_2+1)(2n_3+1)]^{\frac{1}{2}}}{(n_3+n_1-n_2)!!(n_2+n_1-n_3)!!(n_1+n_2+n_3+1)!!}(-1)^{m_3+\frac{1}{2}(n_3-n_3+n_1)} \times \left[\frac{(n_1+m_1)!(n_1-m_1)!(n_2-m_2)!(n_3-m_3)!}{2(n_2+m_2)!(n_3+m_3)!}\right]^{\frac{1}{2}} \times \sum_t (-1)^t \frac{(n_3+m_3+t)!(n_2+n_1-m_3-t)!}{t!(n_1-m_1-t)!(n_3-m_3-t)!(n_2-n_1+m_3+t)!}, \tag{11}$$

where the summation over t is taken from $t = 0$ up to $t = n_1 - m_1$ or until one of the factorials becomes negative.[9]

[7] J. A. Gaunt, Phil. Trans. Roy. Soc. (London) **A228**, 151 (1929).

[8] The result of Infeld and Hull is cited here and has been modified by the factor $1/(2\pi)^{\frac{1}{2}}$ to account for the φ dependence of the Spherical Harmonics.

[9] The following convention is followed:

$$n!! = n(n-2)\ (n-4)\ldots 1 (2 \text{ if } n \text{ is even}).\quad 0!! = (-1)!! = 1.$$

However, substitution of (11) into (10) only leads to a rather involved expression which cannot be simplified.

To compute a large number of the Elsasser integrals, it would appear that one should first calculate the K's from (11) and then use these in the computation in (10). Such calculations were carried out by the author for n and m values up to 5 and are available in decimal form to six figures.

3. SELECTION RULES

Both the K- and L-type integrals will vanish unless the indices satisfy certain "selection rules." These rules have been worked out by several authors[2,5-7] and will only be cited here for convenience.

For the nonvanishing of the K's the indices must satisfy

$$m_1 = m_2 + m_3, \tag{12a}$$

$$|n_2 - n_1| \leq n_2 + n_3 \text{ (sometimes referred to as the triangle condition)}, \tag{12b}$$

$$n_1 + n_2 + n_3 = \text{even integer}. \tag{12c}$$

Similar selection rules for the Elsasser integrals are

$$m_1 = m_2 + m_3, \tag{13a}$$

$$|n_2 - n_3| + 1 \leq n_1 \leq n_2 + n_3 - 1, \tag{13b}$$

$$n_1 + n_2 + n_3 = \text{odd integer}. \tag{13c}$$

The selection rules must be satisfied in order that (1) be different from zero. However, even if these rules *are* satisfied, it is possible that due to the oscillations of the integrand of (1) the integral vanishes. Such cases have been found for the K-type integrals[5] (8). For the Elsasser integrals with n, m values up to 5, the only one found is $L_{5.4.2}^{3.2.1} = 0$, which vanishes upon integration.

4. SYMMETRIES

The calculation which resulted in Eq. (10) holds only for m_2 and m_3 nonnegative. From (1) it is easy to establish by partial integration the relations

$$L_{n_1n_2n_3}^{m_1m_2m_3} = -L_{n_1n_2n_3}^{m_1m_2m_3}, \tag{14a}$$

$$L_{n_1n_2n_3}^{m_1,-m_2,m_3} = (-1)^{m_2} L_{n_3n_2n_1}^{m_3m_2m_1}, \tag{14b}$$

$$L_{n_1n_2n_3}^{m_1,m_2,-m_3} = (-1)^{m_3} L_{n_2n_1n_3}^{m_2m_1m_3}. \tag{14c}$$

From (14), together with (10), one can obtain all the L's for both positive and negative m values. From (14a) it is also easily seen that for $n_2 = n_3$, $m_2 = m_3$, the value of L is identically zero.

Erratum: Notes on the Evaluation of Some Fermi Integrals

M. L. GLASSER
[J. Math. Phys. 5, 1150 (1964)]
(Received 24 January 1966)

The expressions following Eq. (12) should read

(1) k an integer:

$$F_k(\gamma\zeta) = \frac{\pi k!}{\gamma^{k+1}}\left\{\frac{1}{(k+1)!}\frac{d^{k+1}}{dz^{k+1}}\left[z\csc(\pi z)e^{z\gamma\zeta}\right] + \frac{(-1)^{k+1}}{\pi}\Phi(k+1, -e^{\gamma\zeta})\right\}$$

exactly.

(2) k not an integer:

$$F_k(\gamma\zeta) \sim \frac{\pi\Gamma(k+1)}{\gamma^{k+1}}\left\{\frac{(\gamma\zeta)^{k+1}}{\pi\Gamma(k+2)} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(2^{2n-1}-1)\pi^{2n}B_n}{(2n)!\,\Gamma(k-2n+2)}(\gamma\zeta)^{k-2n+1} + \frac{\cos\pi(k+1)}{\pi}\Phi(k+1, -e^{-\gamma\zeta})\right\}$$

for $\gamma\zeta$ large.

In Appendix B, the second sentence of paragraph two should be replaced by: If $\mu \neq 0$, $B_\sigma(\lambda, \mu)$ is simply the residue at the origin; $B_\sigma(\lambda, 0)$ is the sum of the residues of the integrand at $\xi = 0, -1, -2, \cdots$.

I am grateful to Dr. Diethard Schiller, University of Timişoara, Rumania, for bringing these points to my attention.

Erratum: On Polynomial Systems in a Banach Ring

JOHN G. TAYLOR
[J. Math. Phys. 6, 1148 (1965)]
(Received 25 January 1966)

In Sec. 6 directly after Eq. (15), delete the sentence "This is because . . . associative 3 ring" and replace by: This is so, even though the triple product is nonassociative. This is because the nonassociativity does not prevent iteration of (15), which is proved to converge by use of the identity $x^3 - y^3 = x^2(x-y) + (x-y)xy + y(x-y)y$.

I would like to thank Professor H. Araki for drawing the error to my attention.

NOTICE

After 1 September 1966, all manuscripts submitted to the *Journal of Mathematical Physics* should be addressed as follows:

Journal of Mathematical Physics
Dr. Elliott W. Montroll, Editor
Physics Department
University of Rochester
Rochester, New York 14627